Efficient Numerical Solution of Parabolic Optimization Problems by Finite Element Methods

Roland Becker, Dominik Meidner, Boris Vexler

Efficient Numerical Solution of Parabolic Optimization Problems by Finite Element Methods

Roland Becker[†], Dominik Meidner[‡], and Boris Vexler[¶]

We present an approach for efficient numerical solution of optimization problems governed by parabolic partial differential equations. The main ingredients are: space-time finite element discretization, second order optimization algorithms and storage reduction techniques. We discuss the combination of these components for the solution of large scale optimization problems.

Keywords: optimal control, parabolic equations, finite elements, storage reduction

AMS Subject Classification:

1 Introduction

In this paper, we discuss efficient numerical methods for solving 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 are based on the finite element method. This allows a natural translation of the optimality conditions from the continuous to the discrete level. For this type of discretizations, we present a systematic approach for precise computation of the derivatives required in optimization algorithms. The evaluation of these derivatives is based on the solutions of appropriate adjoint (dual) and sensitivity (tangent) equations.

The solution of the underlying state equation is typically required in the whole time interval for the computation of these additional solutions. If all data are stored, the storage grows linearly with respect to the number of time

[†]Laboratoire de Mathématiques Appliquées, Université de Pau et des Pays de l'Adour BP 1155, 64013 PAU Cedex, France

[‡]Institut für Angewandte Mathematik, Ruprecht-Karls-Universität Heidelberg, INF 294, 69120 Heidelberg, Germany. This work has been partially supported by the German Research Foundation (DFG) through the *Internationales Graduiertenkolleg 710* 'Complex Processes: Modeling, Simulation, and Optimization'

[¶]Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenberger Straße 69, 4040 Linz, Austria

intervals in the time-discretization. This makes the optimization procedure prohibitive for fine discretizations. We suggest an algorithm, which allows to reduce the required storage. We analyze the complexity of this algorithm and prove that the required storage grows only *logarithmic* with respect to the number of time intervals. Such results are well-known for gradient evaluations in the context of automatic differentiation, see Griewank [12, 13] and Griewank and Walther [14]. However, to the authors' knowledge, the analysis of the required numerical effort for the whole optimization algorithm is new. The presented approach is an extension of windowing strategies introduced in Berggren, Glowinski, and Lions [5].

The main contribution of this paper is the combination of the exact computation of the derivatives based on the space-time finite element discretization with the storage reduction techniques.

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

$$\begin{aligned} \partial_t u + A(q,u) &= f \\ u(0) &= u_0(q). \end{aligned} \tag{1}$$

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 (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).

For covering additional constraints on the control variable, one may seek q in an admissible set describing, e.g., box constraints on q. For clarity of presentation, we consider here the case of no additional control contraints. However, the algorithms discussed in the sequel, can be used as an interior loop within a primal-dual active set strategy, see, e.g., Bergounioux, Ito, and Kunisch [6] and Kunisch and Rösch [16].

The paper is organized as follows: In the next section we describe an abstract optimization problem, with a parabolic state equation written in a weak form and discuss optimality conditions. Then, the problem is reformulated as an

unconstrained (reduced) optimization problem and the expressions for the required derivatives are provided. After that, we describe Newton-type methods for solution of the problem on the continuous level. In Section 3, we discuss the space and time discretizations. The space discretization is done by conforming finite elements and for the time discretization we use two approaches: discontinuous Galerkin (dG) and continuous Galerkin (cG) methods, see, e.g., Eriksson, Johnson, and Thomée [9]. For both, we provide techniques for precise evaluation of the derivatives in the corresponding discrete problems. This allows for simple translation of the optimization algorithm described in Section 2 from the continuous to the discrete level. Section 4 is devoted to the storage reduction techniques. Here, we present and analyze an algorithm, which we call *Multi-Level Windowing*, allowing for drastically reduction of the required storage by the computation of adjoint solutions. This algorithm is then specified for the computation of the derivatives required in the optimization loop. In the last section we present numerical results illustrating our approach.

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^* a Gelfand triple: $V \hookrightarrow H \hookrightarrow V^*$. The duality pairing between the Hilbert space V and its dual V^* is denoted by $\langle\cdot,\cdot\rangle_{V^*\times V}$ and the scalar product in H by $(\cdot,\cdot)_H$.

Remark 2.1 By the definition of the Gelfand triple, the space H is dense in V^*. Therefore, every functional $v^* \in V^*$ can be uniformly approximated by scalar products in H. That is, we can regard the continuous continuation of $(\cdot,\cdot)_H$ onto $V \times V^*$ as new representation formula for the functionals in V^*.

Let, moreover, $I = (0,T)$ be a time interval and the space X be defined as

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

It is well known, that the space X is continuously embedded in $C(\bar{I},H)$, see, e.g., Dautray and Lions [8].

After these preliminaries, we pose the *state equation* in a weak form using the form $a(\cdot,\cdot)(\cdot)$ defined on $Q\times V\times V$, which is assumed to be twice continuously differentiable and linear in the third argument. The *state variable* $u \in X$ is

determined by

$$\begin{aligned} \int_0^T \{(\partial_t u, \varphi)_H + a(q,u)(\varphi)\}\, dt = \int_0^T (f,\varphi)_H\, dt \quad \forall \varphi \in X, \\ u(0) = u_0(q), \end{aligned} \tag{3}$$

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 twice continuously differentiable mapping describing parameter-dependent initial conditions. Note, that the scalar products $(\partial_t u, \varphi)_H$ and $(f,\varphi)_H$ have to be understood according to Remark 2.1. For brevity of notation, we omit the arguments t and x of time-dependent functions whenever possible.

The cost functional $J \colon Q \times X \to \mathbb{R}$ is defined using two twice continuously differentiable functionals $I \colon V \to \mathbb{R}$ and $K \colon H \to \mathbb{R}$ by:

$$J(q,u) = \int_0^T I(u)\, dt + K(u(T)) + \frac{\alpha}{2}\|q - \bar{q}\|_Q^2, \tag{4}$$

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

The corresponding optimization problem is formulated as follows:

$$\text{Minimize } J(q,u) \text{ subject to (3)}, \ (q,u) \in Q \times X. \tag{5}$$

The question of existence and uniqueness of solutions to such optimization problems is discussed in, e.g., Lions [17], Fursikov [11], and Litvinov [18]. Throughout the paper, we assume problem (5) to admit a (locally) unique solution.

Furthermore, under a regularity assumption on $a'_u(q,u)$ at the solution of (5), the implicit function theorem ensures the existence of an open subset $Q_0 \subset Q$ containing the solution of the optimization problem under consideration, and of a twice continuously differentiable solution operator $S \colon Q_0 \to X$ of the state equation (3). Thus, for all $q \in Q_0$ we have

$$\begin{aligned} \int_0^T \{(\partial_t S(q), \varphi)_H + a(q, S(q))(\varphi)\}\, dt = \int_0^T (f,\varphi)_H\, dt \quad \forall \varphi \in X, \\ S(q)(0) = u_0(q). \end{aligned} \tag{6}$$

Using this solution operator we introduce the *reduced cost functional* $j \colon Q_0 \to \mathbb{R}$, defined by $j(q) = J(q, S(q))$. This definition allows to reformulate

problem (5) as an unconstrained optimization problem:

$$\text{Minimize } j(q), \quad q \in Q_0. \tag{7}$$

If q is an optimal solution of the unconstrained problem above, the first and second order *necessary* optimality condition are fulfilled:

$$\begin{aligned} j'(q)(\tau q) &= 0, && \forall \tau q \in Q, \\ j''(q)(\tau q, \tau q) &\geq 0, && \forall \tau q \in Q. \end{aligned}$$

For the unconstrained optimization problem (7), a second order *sufficient* optimality condition is given by the positive definiteness of the second derivatives $j''(q)$.

To express the first and second derivatives of the reduced cost functional j, we introduce the *Lagrangian* $\mathcal{L}\colon Q \times X \times X \times H \to \mathbb{R}$, defined as

$$\begin{aligned} \mathcal{L}(q, u, z, \tilde{z}) &= J(q, u) \\ &\quad + \int_0^T \{(f - \partial_t u, z)_H - a(q, u)(z)\}\, dt - (u(0) - u_0(q), \tilde{z})_H. \end{aligned} \tag{8}$$

With the help of the Lagrangian, we now present three auxiliary equations, which we will use in the sequel to give expressions of the derivatives of the reduced functional. Each equation will thereby be given in two formulations, first in terms of the Lagrangian and then using the concrete form of the optimization problem under consideration.

- *Dual Equation*: For given $q \in Q_0$ and $u = S(q) \in X$, find $(z, \tilde{z}) \in X \times H$ such that

$$\mathcal{L}'_u(q, u, z, \tilde{z})(\varphi) = 0, \quad \forall \varphi \in X. \tag{9}$$

- *Tangent Equation*: For given $q \in Q_0$, $u = S(q) \in X$, and a given direction $\delta q \in Q$, find $\delta u \in X$ such that

$$\begin{aligned} &\mathcal{L}''_{qz}(q, u, z, \tilde{z})(\delta q, \varphi) + \mathcal{L}''_{uz}(q, u, z, \tilde{z})(\delta u, \varphi) + \mathcal{L}''_{q\tilde{z}}(q, u, z, \tilde{z})(\delta q, \psi) \\ &\qquad + \mathcal{L}''_{u\tilde{z}}(q, u, z, \tilde{z})(\delta u, \psi) = 0, \quad \forall (\varphi, \psi) \in X \times H. \end{aligned} \tag{10}$$

- *Dual for Hessian Equation*: For given $q \in Q_0$, $u = S(q) \in X$, $(z, \tilde{z}) \in X \times H$ the corresponding solution of the dual equation (9), and $\delta u \in X$ the solution of the tangent equation (10) for the given direction δq, find $(\delta z, \delta \tilde{z}) \in X \times H$

such that

$$\mathcal{L}''_{qu}(q,u,z,\tilde{z})(\delta q,\varphi) + \mathcal{L}''_{uu}(q,u,z,\tilde{z})(\delta u,\varphi) + \mathcal{L}''_{zu}(q,u,z,\tilde{z})(\delta z,\varphi) + \mathcal{L}''_{\tilde{z}u}(q,u,z,\tilde{z})(\delta\tilde{z},\varphi) = 0, \quad \forall\varphi \in X. \tag{11}$$

Equivalently, we may rewrite these equations more detailed in the following way:

- *Dual Equation:* For given $q \in Q_0$ and $u = S(q) \in X$, find $(z,\tilde{z}) \in X \times H$ such that

$$\begin{aligned} \int_0^T \{-(\varphi,\partial_t z)_H + a'_u(q,u)(\varphi,z)\}\,dt &= \int_0^T I'(u)(\varphi)\,dt, \quad \forall\varphi \in X, \\ z(T) &= K'(u(T)), \\ \tilde{z} &= z(0). \end{aligned} \tag{12}$$

- *Tangent Equation*: For $q \in Q_0$, $u = S(q) \in X$, and a given direction $\delta q \in Q$, find $\delta u \in X$ such that

$$\begin{aligned} \int_0^T \{(\partial_t \delta u,\varphi)_H + a'_u(q,u)(\delta u,\varphi)\}\,dt &= -\int_0^T a'_q(q,u)(\delta q,\varphi)\,dt, \quad \forall\varphi \in X, \\ \delta u(0) &= u'_0(q)(\delta q). \end{aligned} \tag{13}$$

- *Dual for Hessian Equation*: For given $q \in Q_0$, $u = S(q) \in X$, $(z,\tilde{z}) \in X \times H$ the corresponding solution of the dual equation (12), and $\delta u \in X$ the solution of the tangent equation (13) for the given direction δq, find $(\delta z, \delta\tilde{z}) \in X \times H$ such that

$$\begin{aligned} \int_0^T \{-(\varphi,\partial_t \delta z)_H + a'_u(q,u)(\varphi,\delta z)\}\,dt &= \int_0^T I''(u)(\delta u,\varphi)\,dt \\ - \int_0^T \{a''_{uu}(q,u)(\delta u,\varphi,z) + a''_{qu}(q,u)(\delta q,\varphi,z)\}\,dt, &\quad \forall\varphi \in X, \\ \delta z(T) &= K''(u(T))(\delta u(T)), \\ \delta\tilde{z} &= \delta z(0). \end{aligned} \tag{14}$$

To get the representation (13) of the *tangent* equation from (10), we only need to calculate the derivatives of the Lagrangian (8). The derivation of the representations (12) and (14) for the *dual* and the *dual for Hessian* equation requires more care. Here, we integrate by parts and separate the arising boundary terms by appropriate variation of the test functions.

In virtue of the dual equation defined above, we can now state an expression for the first derivatives of the reduced functional:

THEOREM 2.1 *Let for given $q \in Q_0$:*

(i) $u = S(q) \in X$ *be a solution of the state equation* (3).
(ii) $(z, \tilde{z}) \in X \times H$ *fulfill the dual equation* (12).

Then there holds

$$j'(q)(\tau q) = \mathcal{L}'_q(q, u, z, \tilde{z})(\tau q), \quad \forall \tau q \in Q,$$

which we may expand as

$$j'(q)(\tau q) = \alpha(q - \bar{q}, \tau q)_Q - \int_0^T a'_q(q, u)(\tau q, z)\, dt + (u'_0(q)(\tau q), \tilde{z})_H, \quad \forall \tau q \in Q. \tag{15}$$

Proof Since condition (i) ensures that u is the solution of the state equation (3), and due to both, the definition (6) of the solution operator S and the definition (8) of the Lagrangian, we obtain:

$$j(q) = \mathcal{L}(q, u, z, \tilde{z}). \tag{16}$$

By taking (total) derivative of (16) with respect to q in direction τq, we get

$$\begin{aligned} j'(q)(\tau q) = {} & \mathcal{L}'_q(q, u, z, \tilde{z})(\tau q) + \mathcal{L}'_u(q, u, z, \tilde{z})(\tau u) \\ & + \mathcal{L}'_z(q, u, z, \tilde{z})(\tau z) + \mathcal{L}'_{\tilde{z}}(q, u, z, \tilde{z})(\tau \tilde{z}), \end{aligned}$$

where $\tau u = S'(q)(\tau q)$, and $\tau z \in X$ as well as $\tau \tilde{z} \in H$ are the derivatives of z or respectively $\tilde{z}$ with respect to q in direction τq. Noting the equivalence of condition (i) with

$$\mathcal{L}'_z(q, u, z, \tilde{z})(\varphi) + \mathcal{L}'_{\tilde{z}}(q, u, z, \tilde{z})(\psi) = 0, \quad \forall (\varphi, \psi) \in X \times H,$$

and calculating the derivative of the Lagrangian (8) completes the proof. □

To use Newton's method for solving the considered optimization problems, we have to compute the second derivatives of the reduced functional. The following theorem presents two alternatives for doing that. These two versions lead to two different optimization loops, which are presented in the sequel.

THEOREM 2.2 *Let for given $q \in Q_0$ the conditions of Theorem 2.1 be fulfilled.*

(a) Moreover, let for given $\delta q \in Q$:

(i) $\delta u \in X$ *fulfill the tangent equation* (13).
(ii) $(\delta z, \delta\tilde{z}) \in X \times H$ *fulfill the dual for Hessian equation* (14).
Then there holds

$$\begin{aligned} j''(q)(\delta q, \tau q) &= \mathcal{L}''_{qq}(q,u,z,\tilde{z})(\delta q, \tau q) + \mathcal{L}''_{uq}(q,u,z,\tilde{z})(\delta u, \tau q) \\ &\quad + \mathcal{L}''_{zq}(q,u,z,\tilde{z})(\delta z, \tau q) + \mathcal{L}''_{\tilde{z}q}(q,u,z,\tilde{z})(\delta\tilde{z}, \tau q), \quad \forall \tau q \in Q, \end{aligned}$$

which we may equivalently express as

$$\begin{aligned} j''(q)(\delta q, \tau q) &= \alpha(\delta q, \tau q)_Q \\ &\quad - \int_0^T \{a''_{qq}(q,u)(\delta q, \tau q, z) + a''_{uq}(q,u)(\delta u, \tau q, z) + a'_q(q,u)(\tau q, \delta z)\}\, dt \\ &\quad + (u'_0(q)(\tau q), \delta\tilde{z})_H + (u''_0(q)(\delta q, \tau q), \tilde{z})_H, \quad \forall \tau q \in Q. \end{aligned} \tag{17}$$

(b) Moreover, let for given $\delta q, \tau q \in Q$*:*
(i) $\delta u \in X$ *fulfill the tangent equation* (13) *for the given direction* δq.
(ii) $\tau u \in X$ *fulfill the tangent equation* (13) *for the given direction* τq.
Then there holds

$$\begin{aligned} j''(q)(\delta q, \tau q) &= \mathcal{L}''_{qq}(q,u,z,\tilde{z})(\delta q, \tau q) + \mathcal{L}''_{uq}(q,u,z,\tilde{z})(\delta u, \tau q) \\ &\quad + \mathcal{L}''_{qu}(q,u,z,\tilde{z})(\delta q, \tau u) + \mathcal{L}''_{uu}(q,u,z,\tilde{z})(\delta u, \tau u), \end{aligned}$$

which we may equivalently express as

$$\begin{aligned} j''(q)(\delta q, \tau q) &= \alpha(\delta q, \tau q)_Q + \int_0^T I''_{uu}(u)(\delta u, \tau u)\, dt \\ &\quad - \int_0^T \{a''_{qq}(q,u)(\delta q, \tau q, z) + a''_{uq}(q,u)(\delta u, \tau q, z) + a''_{qu}(q,u)(\delta q, \tau u, z) \\ &\quad + a''_{uu}(q,u)(\delta u, \tau u, z)\}\, dt + K''_{uu}(u)(\delta u, \tau u). \end{aligned} \tag{18}$$

Proof Due to condition (i) of Theorem 2.1, we obtain as before

$$\begin{aligned} j'(q)(\delta q) &= \mathcal{L}'_q(q,u,z,\tilde{z})(\delta q) + \mathcal{L}'_u(q,u,z,\tilde{z})(\delta u) \\ &\quad + \mathcal{L}'_z(q,u,z,\tilde{z})(\delta z) + \mathcal{L}'_{\tilde{z}}(q,u,z,\tilde{z})(\delta\tilde{z}), \end{aligned}$$

and taking (total) derivatives with respect to q in direction τq yields

$$
\begin{aligned}
j''(q)(\delta q, \tau q) = \\
&\mathcal{L}'_{qq}()(\delta q, \tau q) + \mathcal{L}'_{qu}()(\delta q, \tau u) + \mathcal{L}'_{qz}()(\delta q, \tau z) + \mathcal{L}'_{q\tilde{z}}()(\delta q, \tau \tilde{z}) \\
+ &\mathcal{L}'_{uq}()(\delta u, \tau q) + \mathcal{L}'_{uu}()(\delta u, \tau u) + \mathcal{L}'_{uz}()(\delta u, \tau z) + \mathcal{L}'_{u\tilde{z}}()(\delta u, \tau \tilde{z}) \\
+ &\mathcal{L}'_{zq}()(\delta z, \tau q) + \mathcal{L}'_{zu}()(\delta z, \tau u) \\
+ &\mathcal{L}'_{\tilde{z}q}()(\delta \tilde{z}, \tau q) + \mathcal{L}'_{\tilde{z}u}()(\delta \tilde{z}, \tau u) \\
&+ \quad \mathcal{L}'_{u}()(\delta^2 u) \quad + \quad \mathcal{L}'_{z}()(\delta^2 z) \quad + \quad \mathcal{L}'_{\tilde{z}}()(\delta^2 \tilde{z}).
\end{aligned}
$$

(For abbreviation we have omitted the content of the first parenthesis of the Lagrangian.) In addition to the notations in the proof of Theorem 2.1, we have defined $\delta^2 u = S''(q)(\delta q, \tau q)$, and $\delta^2 z \in X$ as well as $\delta^2 \tilde{z} \in H$ as the second derivatives of z or respectively $\tilde{z}$ in the directions δq and τq.

We complete the proof applying the stated conditions to this expression. □

In the sequel, we present two variants of the Newton based optimization loop on the continuous level. The difference between these variants consists in the way of computing the update. Newton-type methods are used for solving optimization problem governed by time-dependent partial differential equations, see, e.g., Hinze and Kunisch [15] and Tröltzsch [20].

From here on, we consider finite dimensional control space Q with a basis:

$$\{ \tau q_i \mid i = 1, \ldots, \dim Q \}. \tag{19}$$

Both, Algorithm 2.1 and Algorithm 2.3, describe an usual Newton-type method for the unconstrained optimization problem (7), which requires the solution of the following linear system in each iteration:

$$\nabla^2 j(q)\delta q = -\nabla j(q), \tag{20}$$

where the gradient $\nabla j(q)$ and the Hessian $\nabla^2 j(q)$ are defined as usual by the identifications:

$$(\nabla j(q), \tau q)_Q = j'(q)(\tau q), \quad \forall \tau q \in Q,$$

$$(\tau q, \nabla^2 j(q)\delta q)_Q = j''(q)(\delta q, \tau q), \quad \forall \delta q, \tau q \in Q.$$

In both algorithms, the required gradient $\nabla j(q)$ is computed using representation (15) from Theorem 2.1. However, the algorithms differ in the way how

they solve the linear system (20) to obtain a correction δq for the current control q. Algorithm 2.1 treats the computation of this system using the conjugate gradients method. It basically necessitates products of the Hessian with given vectors and does not need the entire Hessian.

Algorithm 2.1 Optimization Loop *without* building up the Hessian:

1: Choose initial $q^0 \in Q_0$ and set $n = 0$.
2: **repeat**
3: Compute $u^n \in X$, i.e. solve the state equation (3).
4: Compute $(z^n, \tilde{z}^n) \in X \times H$, i.e. solve the dual equation (12).
5: Build up the gradient $\nabla j(q^n)$. To compute its i-th component $(\nabla j(q^n))_i$, evaluate the right hand side of representation (15) for $\tau q = \tau q_i$.
6: Solve

$$\nabla^2 j(q^n)\delta q = -\nabla j(q^n)$$

by use of the method of conjugate gradients.(For the computation of the required matrix-vector products, apply the procedure described in Algorithm 2.2.)
7: Set $q^{n+1} = q^n + \delta q$.
8: Increment n.
9: **until** $\|\nabla j(q^n)\| < \mathrm{TOL}$

The computation of the required matrix-vector products can be done with the representation given in Theorem 2.2(a) and is described in Algorithm 2.2. We note that in order to obtain the product of the Hessian with a given vector, we have to solve one tangent equation and one dual for Hessian equation. This has to be done in each step of the method of conjugate gradients.

Algorithm 2.2 Computation of the matrix-vector product $\nabla^2 j(q^n)\delta q$:

Require: u^n, z^n, and $\tilde{z}^n$ are already computed for the given q^n
1: Compute $\delta u^n \in X$, i.e. solve the tangent equation (13).
2: Compute $(\delta z^n, \delta \tilde{z}^n) \in X \times H$, i.e. solve the dual for Hessian equation (14).
3: Build up the product $\nabla^2 j(q^n)\delta q$. To compute its i-th component $(\nabla^2 j(q^n)\delta q)_i$, evaluate the right hand side of representation (17) for $\tau q = \tau q_i$.

In contrast to Algorithm 2.1, Algorithm 2.3 builds up the whole Hessian. Consequently we may use every linear solver to the linear system (20). To compute the Hessian, we use the representation of the second derivatives of the reduced functional given in Theorem 2.2(b). Thus, in each Newton step we have to solve the tangent equation for each basis vector in (19).

Algorithm 2.3 Optimization Loop *with* building up the Hessian:

1: Choose initial $q^0 \in Q_0$ and set $n = 0$.
2: **repeat**
3: Compute $u^n \in X$, i.e. solve the state equation (3).
4: Compute $\{ \tau u_i^n \mid i = 1, \dots, \dim Q \} \subset X$ for the chosen basis of Q, i.e. solve the tangent equation (13) for each of the basis vectors τq_i in (19).
5: Compute $z^n \in X$, i.e. solve the dual equation (12).
6: Build up the gradient $\nabla j(q^n)$. To compute its i-th component $(\nabla j(q^n))_i$, evaluate the right hand side of representation (15) for $\tau q = \tau q_i$.
7: Build up the Hessian $\nabla^2 j(q^n)$. To compute its ij-th entry $(\nabla^2 j(q^n))_{ij}$, evaluate the right hand side of representation (18) for $\delta q = \tau q_j$ $\tau q = \tau q_i$, $\delta u = \tau u_j$, and $\tau u = \tau u_i$.
8: Compute δq as the solution of

$$\nabla^2 j(q^n)\delta q = -\nabla j(q^n)$$

by use of an arbitrary linear solver.
9: Set $q^{n+1} = q^n + \delta q$.
10: Increment n.
11: **until** $\|\nabla j(q^n)\| < \mathrm{TOL}$

We now compare the efficiency of the two presented algorithms. For one step of Newton's method, Algorithm 2.1 requires the solution of two linear problems (tangent equation and dual for Hessian equation) for each step of the CG-iteration, whereas for Algorithm 2.3 it is necessary to solve $\dim Q$ tangent equations. Thus, if we have to perform n_{CG} steps of the method of conjugate gradients per Newton step, we should favor Algorithm 2.3, if

$$\frac{\dim Q}{2} \leq n_{\mathrm{CG}}. \tag{21}$$

In Section 4, we will discuss a comparison of these two algorithms in the context of windowing.

3 Discretization

In this section, we discuss the discretization of the optimization problem (5). To this end, we use finite element method in time and space to discretize the state equation. This allows us to give a natural computable representation of the discrete gradient and Hessian. The use of exact discrete derivatives is important for the convergence of the optimization algorithms.

We discuss the corresponding (discrete) formulation of the auxiliary problems (dual, tangent and dual for Hessian) introduced in Section 2. The first

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 the time discretization. We also present the form of the required auxiliary equations for one concrete realization of the cG and the dG discretization respectively.

3.1 *Time Discretization*

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

3.1.1 ***Discontinuous Galerkin (dG) Methods.*** We introduce for $r \in \mathbb{N}_0$ the discontinuous trial and test space

$$X_k^r = \left\{ v_k \in L^2(I, V) \;\middle|\; v_k\big|_{I_m} \in P_r(I_m, V),\ m = 1, \ldots, M,\ v_k(0) \in H \right\}. \tag{22}$$

Here, we denote $P_r(I_m, V)$ the space of polynomial of degree r defined on I_m with values in V. Additionally, we will use the following notations for functions $v_k \in X_k^r$:

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

The dG discretization of the state equation (3) now reads: Find $u \in X_k^r$ such that

$$\begin{aligned} \sum_{m=1}^{M} \int_{I_m} \{(\partial_t u_k, \varphi)_H + a(q, u_k)(\varphi)\}\, dt + \sum_{m=1}^{M} ([u_k]_{m-1}, \varphi_{m-1}^+)_H \\ = \sum_{m=1}^{M} \int_{I_m} (f, \varphi)_H\, dt, \quad \forall \varphi \in X_k^r, \\ u_{k,0}^- = u_0(q). \end{aligned} \tag{23}$$

For the analysis of the discontinuous finite element time discretization we refer to Estep and Larsson [10] and Eriksson, Johnson, and Thomée [9].

The corresponding semi-discrete optimization problem is given by:

$$\text{Minimize } J(q, u_k) \text{ subject to (23)}, \ (q, u_k) \in Q \times X_k^r, \tag{24}$$

with the cost functional J from (4).

Similar to the continuous case, we introduce a semi-discrete solution operator $S_k \colon Q_{k,0} \to X_k^r$ such that $S_k(q)$ fulfills for $q \in Q_{k,0}$ the semi-discrete state equation (23). As in Section 2, we define the semi-discrete reduced cost functional $j_k \colon Q_{k,0} \to \mathbb{R}$ as

$$j_k(q) = J(q, S_k(q)),$$

and reformulate the optimization problem (24) as unconstrained problem:

$$\text{Minimize } j_k(q), \quad q \in Q_{k,0}.$$

To derive a representation of the derivatives of j_k, we define the semi-discrete Lagrangian $\mathcal{L}_k \colon Q \times X_k^r \times X_k^r \times H \to \mathbb{R}$, similar to the continuous case, as

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

With these preliminaries, we obtain similar expressions for the three auxiliary equations in terms of the semi-discrete Lagrangian as stated in the section before. However, the derivation of the explicite representations for the auxiliary

equations requires some care due to the special form of the Lagrangian $\mathcal{L}_k$ for the dG discretization:

- *Dual Equation for dG*: For given $q \in Q_{k,0}$ and $u_k = S_k(q) \in X_k^r$, find $(z_k, \tilde{z}_k) \in X_k^r \times H$ such that

$$\begin{aligned} \sum_{m=1}^{M} \int_{I_m} \{-(\varphi, \partial_t z_k)_H + a_u'(q,u_k)(\varphi, z_k)\}\, dt - \sum_{m=1}^{M-1} (\varphi_m^-, [z_k]_m)_H \\ + (\varphi_M^-, z_{k,M}^-)_H = \sum_{m=1}^{M} \int_{I_m} I'(u_k)(\varphi)\, dt + K'(u_{k,M}^-)(\varphi_M^-), \quad \forall \varphi \in X_k^r, \\ \tilde{z}_k = z_{k,0}^+. \end{aligned} \tag{25}$$

- *Tangent Equation for dG*: For $q \in Q_{k,0}$, $u_k = S_k(q) \in X_k^r$, and a given direction $\delta q \in Q$, find $\delta u_k \in X_k^r$ such that

$$\begin{aligned} \sum_{m=1}^{M} \int_{I_m} \{(\partial_t \delta u_k, \varphi)_H + a_u'(q,u_k)(\delta u_k, \varphi)\}\, dt + \sum_{m=1}^{M} ([\delta u_k]_{m-1}, \varphi_{m-1}^+)_H \\ = -\sum_{m=1}^{M} \int_{I_m} a_q'(q,u_k)(\delta q, \varphi)\, dt, \quad \forall \varphi \in X_k^r, \\ \delta u_{k,0}^- = u_0'(q)(\delta q). \end{aligned} \tag{26}$$

- *Dual for Hessian Equation for dG*: For given $q \in Q_{k,0}$, $u_k = S_k(q) \in X_k^r$, $(z_k, \tilde{z}_k) \in X_k^r \times H$ the corresponding solution of the dual equation (25), and $\delta u_k \in X_k^r$ the solution of the tangent equation (26) for the given direction δq, find $(\delta z_k, \delta \tilde{z}_k) \in X_k^r \times H$ such that

$$\begin{aligned} \sum_{m=1}^{M} \int_{I_m} \{-(\varphi, \partial_t \delta z_k)_H + a_u'(q,u_k)(\varphi, \delta z_k)\}\, dt - \sum_{m=1}^{M-1} (\varphi_m^-, [\delta z_k]_m)_H \\ + (\varphi_M^-, \delta z_{k,M}^-)_H = -\sum_{m=1}^{M} \int_{I_m} \{a_{uu}''(q,u_k)(\delta u_k, \varphi, z_k) \\ + a_{qu}''(q,u_k)(\delta q, \varphi, z_k)\}\, dt + \sum_{m=1}^{M} \int_{I_M} I''(u_k)(\delta u_k, \varphi)\, dt \\ + K''(u_{k,M}^-)(\delta u_{k,M}^-, \varphi_M^-), \quad \forall \varphi \in X_k^r, \\ \delta \tilde{z}_k = \delta z_{k,0}^+. \end{aligned} \tag{27}$$

As on the continuous level, the tangent equation can be obtained directly by calculating the derivatives of the Lagrangian, and for the dual equations,

we additionally integrate by parts. But, since the test functions are piecewise polynomials, we can not separate the terms containing φ_M^- as we did it for the boundary terms in the continuous formulation before. However, because the support of φ_0 is just the point 0, separation of the equation to determine $\tilde{z}_k$ or $\delta\tilde{z}_k$ is still possible.

Now, the representations from Theorem 2.1 and Theorem 2.2 can be translated to the semi-discrete level: We have

$$\begin{aligned} j_k'(q)(\tau q) &= \alpha(q-\bar{q},\tau q)_Q \\ &\quad - \sum_{m=1}^{M} \int_{I_m} a_q'(q,u_k)(\tau q, z_k)\,dt + (u_0'(q)(\tau q),\tilde{z}_k)_H, \quad \forall \tau q \in Q, \end{aligned} \tag{28}$$

and, depending on whether we use version (a) or (b) of Theorem 2.2,

$$\begin{aligned} j_k''(q)(\delta q,\tau q) &= \alpha(\delta q,\tau q)_Q \\ -\sum_{m=1}^{M} \int_{I_m} &\{a_{qq}''(q,u_k)(\delta q,\tau q,z_k) + a_{uq}''(q,u_k)(\delta u_k,\tau q,z_k) + a_q'(q,u_k)(\tau q,\delta z_k)\}\,dt \\ &+ (u_0'(q)(\tau q),\delta\tilde{z}_k)_H + (u_0''(q)(\delta q,\tau q),\tilde{z}_k)_H, \quad \forall \tau q \in Q, \end{aligned} \tag{29}$$

or

$$\begin{aligned} j_k''(q)(\delta q,\tau q) &= \alpha(\delta q,\tau q)_Q + \sum_{m=1}^{M} \int_{I_m} I_{uu}''(u_k)(\delta u_k,\tau u_k)\,dt \\ -\sum_{m=1}^{M} \int_{I_m} &\{a_{qq}''(q,u_k)(\delta q,\tau q,z_k) + a_{uq}''(q,u_K)(\delta u_k,\tau q,z_k) + a_{qu}''(q,u_k)(\delta q,\tau u_k,z_k) \\ &+ a_{uu}''(q,u_k)(\delta u_k,\tau u_k,z_k)\}\,dt + K_{uu}''(u_k)(\delta u_k,\tau u_k). \end{aligned} \tag{30}$$

3.1.2 ***Continuous Galerkin (cG) Methods.*** In this subsection, we discuss the time discretization by Galerkin methods with continuous trial functions and discontinuous test functions, the so called cG methods. For the test space, we use the space X_k^r defined in (22), and additionally, we introduce a trial space given by:

$$Y_k^s = \left\{ v_k \in C(\bar{I},V) \;\middle|\; v_k\big|_{I_m} \in P_s(I_m,V),\ m=1,\dots,M \right\}.$$

To simplify the notation, we will use in this subsection the same symbols for the Lagrangian and the several solutions as in the subsection above for the dG

discretization.

In virtue of these two spaces, we state the semi-discrete state equation in the cG context: Find $u_k \in Y_k^s$, such that

$$\begin{aligned} \int_0^T \{(\partial_t u_k, \varphi)_H + a(q, u_k)(\varphi)\}\, dt = \int_0^T (f, \varphi)_H\, dt, \quad \forall \varphi \in X_k^r \\ u_k(0) = u_0(q). \end{aligned} \tag{31}$$

Similarly to the previous subsection, we define the semi-discrete optimization problem

$$\text{Minimize } J(q, u_k) \text{ subject to (31)}, \ (q, u_k) \in Q \times Y_k^s, \tag{32}$$

and the Lagrangian $\mathcal{L}_k \colon Q \times Y_k^s \times X_k^r \times H \to \mathbb{R}$ as

$$\begin{aligned} \mathcal{L}_k(q, u_k, z_k, \tilde{z}_k) &= J(q, u_k) \\ &\quad + \int_0^T \{(f - \partial_t u_k, z_k)_H - a(q, u_k)(z_k)\}\, dt - (u_k(0) - u_0(q), \tilde{z}_k)_H. \end{aligned}$$

Now, we can recall the process described in the previous subsection for the dG discretization to obtain the solution operator $S_k \colon Q_{k,0} \to Y_k^s$, the reduced functional j_k, and the unconstrained optimization problem.

For the cG discretization, the three auxiliary equations read as follows:

- *Dual Equation for cG*: For given $q \in Q_{k,0}$ and $u_k = S_k(q) \in Y_k^s$, find $(z_k, \tilde{z}_k) \in X_k^r \times H$ such that

$$\begin{aligned} \sum_{m=1}^{M} \int_{I_m} \{-(\varphi, \partial_t z_k)_H + a_u'(q, u_k)(\varphi, z_k)\}\, dt - \sum_{m=1}^{M-1} (\varphi(t_m), [z_k]_m)_H \\ + (\varphi(T), z_{k,M}^-)_H = \sum_{m=1}^{M} \int_{I_m} I'(u_k)(\varphi)\, dt + K'(u_k(T))(\varphi(T)) \\ + (\varphi(0), z_{k,0}^+ - \tilde{z}_k)_H, \quad \forall \varphi \in Y_k^s. \end{aligned} \tag{33}$$

- *Tangent Equation for cG*: For $q \in Q_{k,0}$, $u_k = S_k(q) \in Y_k^s$, and a given

direction $\delta q \in Q$, find $\delta u_k \in Y_k^s$ such that

$$\begin{aligned}\int_0^T \{(\partial_t \delta u_k, \varphi)_H + a_u'(q,u_k)(\delta u_k, \varphi)\}\, dt = \\ -\int_0^T a_q'(q,u_k)(\delta q, \varphi)\, dt, \quad \forall \varphi \in X_k^r, \\ \delta u_k(0) = u_0'(q)(\delta q).\end{aligned} \tag{34}$$

- *Dual for Hessian Equation for cG*: For given $q \in Q_{k,0}$, $u_k = S_k(q) \in Y_k^s$, $(z_k, \tilde{z}_k) \in X_k^r \times H$ the corresponding solution of the dual equation (33), and $\delta u_k \in Y_k^s$ the solution of the tangent equation (34) for the given direction δq, find $(\delta z_k, \delta \tilde{z}_k) \in X_k^r \times H$ such that

$$\begin{aligned}\sum_{m=1}^{M} \int_{I_m} \{-(\varphi, \partial_t \delta z_k)_H + a_u'(q,u_k)(\varphi, \delta z_k)\}\, dt - \sum_{m=1}^{M-1} (\varphi(t_m), [\delta z_k]_m)_H \\ + (\varphi(T), \delta z_{k,M}^-)_H = -\sum_{m=1}^{M} \int_{I_m} \{a_{uu}''(q,u_k)(\delta u_k, \varphi, z_k) \\ + a_{qu}''(q,u_k)(\delta q, \varphi, z_k)\}\, dt + \sum_{m=1}^{M} \int_{I_M} I''(u_k)(\delta u_k, \varphi)\, dt \\ + K''(u_k)(\delta u_k(T), \varphi(T)) + (\varphi(0), z_{k,0}^+ - \tilde{z}_k)_H, \quad \forall \varphi \in Y_k^s.\end{aligned} \tag{35}$$

The derivation of the tangent equation (34) is straightforward and similar to the continuous case. However, the dual equation (33) and the dual for Hessian equation (35) contain jump terms such as $[z_k]_m$ or $[\delta z_k]_m$ due to the interval-wise integration by parts. As for the case of dG semi-discretization described in the previous subsection, the initial conditions can not be separated as in the continuous case, cf. (12) and (14). In contrast to the dG semi-discretization, we also can not separate the conditions to determine $\tilde{z}_k$ or $\delta \tilde{z}_k$ here, since for the cG methods the test functions of the dual equations are continuous, see the discussion for a concrete realization of the cG method in the next section.

Again, Theorem 2.1 and Theorem 2.2 are translated to the semi-discrete level by replacing the equations (12), (13) and (14) by the semi-discrete equations (33), (34) and (35). The representations of the derivatives of j_k for the cG discretization have then the same form as in the dG case. Therefore, one should use formulas (28), (29) and (30), where $u_k, \delta u_k, z_k, \delta z_k, \tilde{z}_k$ and $\delta \tilde{z}_k$ are now determined by (31), (33), (34), and (35).

3.2 *Space-time Discretization*

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., Ciarlet [7]. A mesh consists of cells K, which constitute a non-overlapping cover of the computational domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$. The corresponding mesh is denoted by $\mathcal{T}_h = \{K\}$, where we define the 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 finite element space $V_h \subset V$ in standard way:

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

Here, $\widetilde{Q}_l(K)$ consists of shape functions obtained via (bi-)linear transformations of functions in $Q_l(\widehat{K})$ defined on the reference cell $\widehat{K} = (0, 1)^2$.

Now, the time-discretized schemes developed in the two previous subsections can be transfered to the full discretized level. For doing this, we use the spaces

$$X_{hk}^r = \left\{ v_{hk} \in L^2(I, V_h) \;\middle|\; v_{hk}\big|_{I_m} \in P_r(I_m, V_h),\ m = 1, \ldots, M,\ v_{hk}(0) \in V_h \right\}$$

and

$$Y_{hk}^s = \left\{ v_{hk} \in C(\bar{I}, V_h) \;\middle|\; v_{hk}\big|_{I_m} \in P_s(I_m, V_h),\ m = 1, \ldots, M \right\}$$

instead of X_k^r and Y_k^s.

Remark 3.1 Often, by solving problems with complex dynamical behavior, it is desirable to use time-dependent meshes $\mathcal{T}_{h_m}$. Then, h_m describes the mesh used in time interval I_m and we can use the same definition of $X_{h_m k}^r$ as before, because of the discontinuity in time. Consequently, dG discretization is directly applicable to time-dependent space discretizations. The definition of $Y_{h_m k}^s$ requires more care due to the request for continuity in time. An approach overcoming this difficulty can be found in Becker [1].

In the sequel, we present one concrete time-stepping scheme for the dG and the cG discretization combined with the finite element space discretization. These schemes correspond to the implicit Euler scheme and the Crank-Nicolson scheme, respectively.

To obtain the standard implicit Euler scheme as a special case of dG discretization, we choose $r = 0$ and approximate the integrals arising by the box rule. Furthermore, we define $U_m = u_{hk}\big|_{I_m}$, $\Delta U_m = \delta u_{hk}\big|_{I_m}$, $Z_m = z_{hk}\big|_{I_m}$, and $\Delta Z_m = z_{hk}\big|_{I_m}$ for $i = 1, \ldots, M$, and $U_0 = u_{hk,0}^-$, $\Delta U_0 = \delta u_{hk,0}^-$, $Z_0 = \tilde{z}_{hk}$, and

$\Delta Z_0 = \delta \tilde{z}_{hk}$. With this, we obtain the following schemes for the dG-discretized state and auxiliary equations, which should be fulfilled for all $\varphi \in V_h$:

- *State Equation for dG*:
  - $m = 0$:

$$(U_0, \varphi)_H = (u_0(q), \varphi)_H$$

  - $m = 1, \ldots, M$:

$$(U_m, \varphi)_H + k_m a(q, U_m)(\varphi) = (U_{m-1}, \varphi)_H + k_m (f(t_m), \varphi)_H$$

- *Dual Equation for dG*:
  - $m = M$:

$$(\varphi, Z_M)_H + k_M a'_u(q, U_M)(\varphi, Z_M) = K'(U_M)(\varphi) + k_M I'(U_M)(\varphi)$$

  - $m = M-1, \ldots, 1$:

$$(\varphi, Z_m)_H + k_m a'_u(q, U_m)(\varphi, Z_m) = (\varphi, Z_{m+1})_H + k_m I'(U_m)(\varphi)$$

  - $m = 0$:

$$(\varphi, Z_0)_H = (\varphi, Z_1)_H$$

- *Tangent Equation for dG*:
  - $m = 0$:

$$(\Delta U_0, \varphi)_H = (u'_0(q)(\delta q), \varphi)_H$$

  - $m = 1, \ldots, M$:

$$(\Delta U_m, \varphi)_H + k_m a'_u(q, U_m)(\Delta U_m, \varphi) = (\Delta U_{m-1}, \varphi)_H - k_m a'_q(q, U_m)(\delta q, \varphi)$$

- *Dual for Hessian Equation for dG*:
  - $m = M$:

$$\begin{aligned}(\varphi, \Delta Z_M)_H + k_M a'_u(q, U_M)(\varphi, \Delta Z_M) = \\ K''(U_M)(\Delta U_M, \varphi) + k_M I''(U_M)(\Delta U_M, \varphi) \\ - k_M \Big\{ a''_{uu}(q, U_M)(\Delta U_M, \varphi, Z_M) + a''_{qu}(q, U_M)(\delta q, \varphi, Z_M) \Big\}\end{aligned}$$

- $m = M-1, \dots, 1$:

$$
\begin{aligned}
(\varphi, \Delta Z_m)_H + k_m a'_u(q, U_m)(\varphi, \Delta Z_m) = \\
(\varphi, \Delta Z_{m+1})_H + k_m I''(U_m)(\Delta U_m, \varphi) \\
- k_m \Big\{ a''_{uu}(q, U_m)(\Delta U_m, \varphi, Z_m) + a''_{qu}(q, U_m)(\delta q, \varphi, Z_m) \Big\}
\end{aligned}
$$

- $m = 0$:

$$
(\varphi, \Delta Z_0)_H = (\varphi, \Delta Z_1)_H
$$

Remark 3.2 The implicit Euler scheme is known to be a first order strongly A-stable method. The resulting schemes for the auxiliary equations have basically the same structure and lead consequently to first order approximation, too. However, the precise a priori error analysis for the optimization problem requires more care and depends on the given structure of the problem under consideration.

The Crank-Nicolson scheme can be obtained in the context of cG discretization by choosing $r = 0$, $s = 1$ and approximating the integrals arising by the trapezoidal rule. Using the representation of the Crank-Nicolson scheme as a cG-scheme, allows us directly to give a concrete form of the auxiliary equations leading to the exact computation of the discrete gradient and Hessian.

We set $U_m = u_{hk}(t_m)$, $\Delta U_m = \delta u_{hk}(t_m)$, $Z_m = z_{hk}\big|_{I_m}$, and $\Delta Z_m = z_{hk}\big|_{I_m}$ for $i = 1, \dots, M$, and $U_0 = u_{hk}(0)$, $\Delta U_0 = \delta u_{hk}(0)$, $Z_0 = \tilde{z}_{hk}$, and $\Delta Z_0 = \delta \tilde{z}_{hk}$. With this, we obtain the following schemes for the cG-discretized state and auxiliary equations, which should be fulfilled for all $\varphi \in V_h$:

- *State Equation for cG*:
  - $m = 0$:

$$
(U_0, \varphi)_H = (u_0(q), \varphi)_H
$$

  - $m = 1, \dots, M$:

$$
\begin{aligned}
(U_m, \varphi)_H + \frac{k_m}{2} a(q, U_m)(\varphi) = (U_{m-1}, \varphi)_H \\
- \frac{k_m}{2} a(q, U_{m-1})(\varphi) + \frac{k_m}{2} \Big\{ (f(t_{m-1}), \varphi)_H + (f(t_m), \varphi)_H \Big\}
\end{aligned}
$$

- *Dual Equation for cG*:

- $m = M$:

$$(\varphi, Z_M)_H + \frac{k_M}{2} a'_u(q, U_M)(\varphi, Z_M) = K'(U_M)(\varphi) + \frac{k_M}{2} I'(U_M)(\varphi)$$

- $m = M-1, \ldots, 1$:

$$\begin{aligned}(\varphi, Z_m)_H + \frac{k_m}{2} a'_u(q, U_m)(\varphi, Z_m) &= (\varphi, Z_{m+1})_H \\ &\quad - \frac{k_{m+1}}{2} a'_u(q, U_m)(\varphi, Z_{m+1}) + \frac{k_m + k_{m+1}}{2} I'(U_m)(\varphi)\end{aligned}$$

- $m = 0$:

$$(\varphi, Z_0)_H = (\varphi, Z_1)_H - \frac{k_1}{2} a'_u(q, U_0)(\varphi, Z_1) + \frac{k_1}{2} I'(U_0)(\varphi)$$

- *Tangent Equation for cG*:
  - $m = 0$:

$$(\Delta U_0, \varphi)_H = (u'_0(q)(\delta q), \varphi)_H$$

  - $m = 1, \ldots, M$:

$$\begin{aligned}(\Delta U_m, \varphi)_H + \frac{k_m}{2} a'_u(q, U_m)(\Delta U_m, \varphi) &= \\ (\Delta U_{m-1}, \varphi)_H - \frac{k_m}{2} a'_u(q, U_{m-1})(\Delta U_{m-1}, \varphi) &\\ - \frac{k_m}{2} \Big\{ a'_q(q, U_{m-1})(\delta q, \varphi) + a'_q(q, U_m)(\delta q, \varphi) \Big\} &\end{aligned}$$

- *Dual for Hessian Equation for cG*:
  - $m = M$:

$$\begin{aligned}(\varphi, \Delta Z_M)_H + \frac{k_M}{2} a'_u(q, U_M)(\varphi, \Delta Z_M) &= \\ K''(U_M)(\Delta U_M, \varphi) + \frac{k_M}{2} I''(U_M)(\Delta U_M, \varphi) &\\ - \frac{k_M}{2} \Big\{ a''_{uu}(q, U_M)(\Delta U_M, \varphi, Z_M) + a''_{qu}(q, U_M)(\delta q, \varphi, Z_M) \Big\} &\end{aligned}$$

- $m = M-1, \dots, 1$:

$$
\begin{aligned}
(\varphi, \Delta Z_m)_H + \frac{k_m}{2} a'_u(q, U_m)(\varphi, \Delta Z_m) &= (\varphi, \Delta Z_{m+1})_H \\
&- \frac{k_{m+1}}{2} a'_u(q, U_m)(\varphi, \Delta Z_{m+1}) + \frac{k_m + k_{m+1}}{2} I''(U_m)(\Delta U_m, \varphi) \\
&- \frac{k_m}{2} \Big\{ a''_{uu}(q, U_m)(\Delta U_m, \varphi, Z_m) + a''_{qu}(q, U_m)(\delta q, \varphi, Z_m) \Big\} \\
&- \frac{k_{m+1}}{2} \Big\{ a''_{uu}(q, U_m)(\Delta U_m, \varphi, Z_{m+1}) + a''_{qu}(q, U_m)(\delta q, \varphi, Z_{m+1}) \Big\}
\end{aligned}
$$

- $m = 0$:

$$
\begin{aligned}
(\varphi, \Delta Z_0)_H = (\varphi, \Delta Z_1)_H &- \frac{k_1}{2} a'_u(q, U_0)(\varphi, Z_1) + \frac{k_1}{2} I''(U_0)(\Delta U_0, \varphi) \\
&- \frac{k_1}{2} \Big\{ a''_{uu}(q, U_0)(\Delta U_0, \varphi, Z_1) + a''_{qu}(q, U_0)(\delta q, \varphi, Z_1) \Big\}
\end{aligned}
$$

The resulting Crank-Nicolson scheme is known to be of second order. However, in contrast to the implicit Euler scheme, this method does not possess the strong A-stability property. The structure of the time-steps for the dual and the dual for Hessian equations is quite unusual. In the first and in the last steps, "half-steps" occur, and in the other steps, terms containing the sizes of two neighboring time intervals k_m and k_{m+1} appear. This complicates the a priori error analysis for the dual scheme, which can be found in Becker [1].

4 Windowing

When computing the gradient of the reduced cost functional as described in the algorithms in Section 2, we need to have access to the solution of the state equation at all points in space and time while computing the dual equation. Similarly, we need the solution of the state, tangent, and dual equations to solve the dual for Hessian equation when computing matrix-vector products with the Hessian of the reduced functional. For large problems, especially in three dimensions, storing all the necessary data might be impossible. However, there are techniques to reduce the storage requirements drastically, known as checkpointing techniques.

In this section, we present an approach, which relies on ideas from Berggren, Glowinski, and Lions [5]. In the sequel, we extend these ideas to obtain two concrete algorithms and present an extension to apply the algorithms to the

whole optimization loops showed in Section 2. Due to its structure, we call this approach *Multi-Level Windowing*.

4.1 *The Abstract Algorithm*

First, we consider the following abstract setting: Let two time stepping schemes be given:

$$\begin{aligned} \mathrm{x}_{m-1} &\mapsto \mathrm{x}_m, && \text{for } m = 1, \dots, M, \\ (\mathrm{y}_{m+1}, \mathrm{x}_m) &\mapsto \mathrm{y}_m, && \text{for } m = M-1, \dots, 0, \end{aligned}$$

together with a given initial value x_0 and the mapping $\mathrm{x}_M \mapsto \mathrm{y}_M$. All time stepping schemes given for dG and cG discretization in the previous section are concrete realizations of these abstract schemes.

Additionally, we assume that the solutions x_m as well as y_m require for all $m = 0, \dots, M$ the same amount of storage. However, if this is not the case, the windowing technique presented in the sequel can be applied to clusters of time steps similar in size instead of single time steps. Such clustering is, e.g., important by using dynamical meshes, since in this case, the amount of storage for a solution x_m depends on the current mesh.

The trivial approach to perform the forward and backwards iterations is to compute and store the whole forward solution $\{\mathrm{x}_m\}_{m=0}^{M}$, and use these values to compute the backwards solution $\{\mathrm{y}_m\}_{m=0}^{M}$. The required amount of storage S_0 in terms of the size of one forward solution x_m to do this is $S_0 = M + 1$. The number of forward steps W_0 necessary to compute the whole backwards solution is $W_0 = M$.

The aim of the following windowing algorithms is to reduce the needed storage by performing some additional forward steps. To introduce the windowing, we additionally assume that we can factorize the number of given time steps M as $M = PQ$ with positive integers P and Q. With this, we can separate the set of time points $\{0, \dots, M\}$ in P slices each containing $Q - 1$ time steps and $P + 1$ sets containing one element as

$$\begin{aligned} \{0, \dots, M\} = \{0\} \cup \{1, \dots, Q-1\} \cup \{Q\} \cup \cdots \\ \cdots \cup \{(P-1)Q\} \cup \{(P-1)Q+1, \dots, PQ-1\} \cup \{PQ\}. \end{aligned}$$

The algorithm now works as follows: First, we compute the forward solution x_m for $m = 1, \dots, M$ and store the $P + 1$ samples $\{\mathrm{x}_{Ql}\}_{l=0}^{P}$. Additionally, we store the $Q - 1$ values of x in the last slice. Now, we have the necessary information on x to compute y_m for $m = M, \dots, (P-1)Q + 1$. Thus, the

values of x in the last slice are not longer needed. We can replace them with the values of x in the next-last slice, which we can directly compute using the time stepping scheme since we stored the value $\mathrm{x}_{(P-2)Q}$ in the first run. Thereby, we can compute y_m for $m = (P-1)Q, \ldots, (P-2)Q+1$. This can now be done iteratively till we have computed y in the first slice and finally obtain the value y_0. This so called *One-Level Windowing* is presented on detail in Algorithm 4.1.

Algorithm 4.1 OneLevelWindowing(P, Q, M):

Require: $M = PQ$.

Store x_0.
Take x_0 as initial value for x.
for $m = 1$ **to** $(P-1)Q$ **do**
  Compute x_m.
  if m is a multiple of Q **then**
    Store x_m.
  end if
end for
for $n = (P-1)Q$ **downto** 0 **step** Q **do**
  Take x_n as initial value for x.
  for $m = n+1$ **to** $n+Q-1$ **do**
    Compute x_m.
    Store x_m.
  end for
  if $n = M - Q$ **then**
    Compute x_M.
    Store x_M.
  end if
  for $m = n+Q$ **downto** $n+1$ **do**
    Compute y_m in virtue of x_m.
    Delete x_m from memory.
  end for
  if $n = 0$ **then**
    Compute y_0.
    Delete x_0 from memory.
  end if
end for

During the Execution of Algorithm 4.1, the needed amount of memory is not exceeding $(P+1)+(Q-1)$ forward solutions. Each of the y_m's is computed exactly once, so we need M solving steps to obtain the whole solution y. To compute the necessary values of x_m, we have to solve $M + (P-1)(Q-1)$ forward steps, since we have to compute each of the values of x in the first

$P-1$ slices. We summarize:

$$S_1(P,Q) = P + Q, \qquad W_1(P,Q) = 2M - P - Q + 1,$$

where again S_1 denotes the required amount of memory in terms of the size of one forward solution and W_1 the number of time steps to provide the forward solution x needed to compute the whole backwards solution y.

Here, the subscript 1 suggests that we can extend this approach to factorizations of M in $L+1$ factors for $L \in \mathbb{N}$. This extension can be obtained via the following inductive argumentation: Assuming $M = M_0 M_1 \cdots M_L$ with positive integers M_l, we can apply the algorithm described above to the factorization $M = PQ$ with $P = M_0$ and $Q = M_1 M_2 \cdots M_L$, and then recursively to each of the P slices. This so called *Multi-Level Windowing* is described in Algorithm 4.2. It has to be started with the call MultiLevelWindowing$(0, 0, L, M_0, M_1, \ldots, M_L, M)$. Of course, there holds by construction

$$\text{OneLevelWindowing}(P, Q, M) = \text{MultiLevelWindowing}(0, 0, 1, P, Q, M).$$

Algorithm 4.2 MultiLevelWindowing$(s, l, L, M_0, M_1, \ldots, M_L, M)$:

Require: $M = M_0 M_1 \cdots M_L$.
1: Set $P = M_l$ and $Q = M_{l+1} \cdots M_L$.
2: **if** $l = 0$ **and** $s = 0$ **then**
3: Store x_0.
4: **end if**
5: Take x_s as initial value for x.
6: **for** $m = 1$ **to** $(P-1)Q$ **do**
7: Compute x_{s+m}.
8: **if** m is a multiple of Q **then**
9: Store x_{s+m}.
10: **end if**
11: **end for**
12: **for** $n = (P-1)Q$ **downto** 0 **step** Q **do**
13: **if** $l + 1 < L$ **then**
14: Call MultiLevelWindowing$(s+n, l+1, L, M_0, M_1, \ldots, M_L, M)$.
15: **else**
16: Take x_{s+n} as initial value for x.
17: **for** $m = n+1$ **to** $n+Q-1$ **do**
18: Compute x_{s+m}.
19: Store x_{s+m}.
20: **end for**

```
      if s + n = M − Q then
        Compute x_M.
        Store x_M.
      end if
      for m = n + Q downto n + 1 do
        Compute y_{s+m} in virtue of x_{s+m}.
        Delete x_{s+m} from memory.
      end for
      if s + n = 0 then
        Compute y_0.
        Delete x_0 from memory.
      end if
    end if
  end for
```

Remark 4.1 The presented approach can be extended to cases where a suitable factorization $M = M_0 M_1 \cdots M_L$ does not exist. We then consider a representation of M as $M = (M_0 - 1)Q_0 + R_0$ with positive integers M_0, Q_0 and R_0 with $Q_0 \leq R_0 < 2Q_0$ and apply this idea recursively to the generated subintervals of length Q_0 or R_0. This can easily be done, since by construction, the reminder interval of length R_0 has at least the same length as the regular subintervals.

In the following theorem, we calculate the necessary amount of storage and the number of needed forward steps to perform the Multi-Level Windowing described in Algorithm 4.2 for a given factorization $M = M_0 M_1 \cdots M_L$ of length $L + 1$:

THEOREM 4.1 *For given $L \in \mathbb{N}_0$ and a factorization of the number of time steps M as $M = M_0 M_1 \cdots M_L$ with $M_l \in \mathbb{N}$, the required amount of memory in the Multi-Level Windowing to perform all backwards solution steps is*

$$S_L(M_0, M_1, \ldots, M_L) = \sum_{l=0}^{L} (M_l - 1) + 2.$$

To achieve this storage reduction, the number of performed forward steps enhances to

$$W_L(M_0, M_1, \ldots, M_L) = (L+1)M - \sum_{l=0}^{L} \frac{M}{M_l} + 1.$$

Proof We prove the theorem by mathematical induction:

- $L = 0$: Here we use the trivial approach where the entire forward solution

x is saved. As considered in the beginning of this subsection, we then have $S_0(M) = M + 1$ and $W_0(M) = M$.

- $L - 1 \rightsquigarrow L$: We consider the factorization $M = M_0 M_1 \cdots M_{L-2}(M_{L-1}M_L)$ of length L additionally to the given one of length $L+1$. Then we obtain in the same way as for the One-Level Windowing, where we reduce the storage mainly from $PQ - 1$ to $(P-1) + (Q-1)$,

$$\begin{aligned} &S_L(M_0, M_1, \ldots, M_{L-1}, M_L) \\ &= S_{L-1}(M_0, M_1, \ldots, M_{L-1}M_L) - (M_{L-1}M_L - 1) + (M_{L-1} - 1) + (M_L - 1). \end{aligned}$$

In virtue of the induction hypothesis for S_{L-1}, it follows

$$\begin{aligned} S_L(M_0, M_1, \ldots, M_{L-1}, M_L) &= \sum_{l=0}^{L-2}(M_l - 1) + (M_{L-1} - 1) + (M_L - 1) + 2 \\ &= \sum_{l=0}^{L}(M_l - 1) + 2. \end{aligned}$$

Now, we prove the assertion for W_L. For this, we justify the equality

$$\begin{aligned} &W_L(M_0, M_1, \ldots, M_{L-1}, M_L) \\ &\quad = W_{L-1}(M_0, M_1, \ldots, M_{L-1}M_L) + \frac{M}{M_{L-1}M_L}(M_{L-1} - 1)(M_L - 1). \end{aligned}$$

This follows directly from the fact that we divide each of the $\frac{M}{M_{L-1}M_L}$ slices $\{s+1, \ldots, s + M_{L-1}M_L - 1\}$ of length $M_{L-1}M_L - 1$ as

$$\begin{aligned} &\{s+1, \ldots, s + M_{L-1}M_L - 1\} = \{s+1, \ldots, s + M_L - 1\} \cup \{s + M_L\} \cup \cdots \\ &\cdots \cup \{s + (M_{L-1} - 1)M_L\} \cup \{s + (M_{L-1} - 1)M_L + 1, \ldots, s + M_{L-1}M_L - 1\}. \end{aligned}$$

Since we just need to compute the forward solution in the first $M_{L-1} - 1$ subslices when we change from the factorization of length L to the one of length $L + 1$, the additional work is

$$\frac{M}{M_{L-1}M_L}(M_{L-1} - 1)(M_L - 1)$$

as stated. Then we obtain in virtue of the induction hypothesis for W_{L-1}

$$W_L(M_0, M_1, \ldots, M_{L-1}, M_L) = LM + M - \sum_{l=0}^{L-2} \frac{M}{M_l} - \frac{M}{M_{L-1}} - \frac{M}{M_L} + 1$$
$$= (L+1)M - \sum_{l=0}^{L} \frac{M}{M_l} + 1.$$

□

If $M^{\frac{1}{L+1}} \in \mathbb{N}$, the minimum of $\widetilde{S}_L$ of all possible factorizations of length $L+1$ is

$$\widetilde{S}_L = S_L(M^{\frac{1}{L+1}}, \ldots, M^{\frac{1}{L+1}}) = (L+1)(M^{\frac{1}{L+1}} - 1) + 2.$$

The numbers of forward steps for the memory-optimal factorization then results in

$$\widetilde{W}_L = W_L(M^{\frac{1}{L+1}}, \ldots, M^{\frac{1}{L+1}}) = (L+1)(M - M^{\frac{L}{L+1}}) + 1.$$

If we choose $L \approx \log_2 M$, then we obtain for the optimal factorization from above logarithmic growth of the necessary amount of storage:

$$\widetilde{S}_L = O(\log_2 M), \qquad \widetilde{W}_L = O(M \log_2 M).$$

Remark 4.2 If we consider time stepping schemes which depend not only on the immediate but on p predecessors, i.e.

$$(\mathrm{x}_{m-p}, \mathrm{x}_{m-p+1}, \ldots, \mathrm{x}_{m-1}) \mapsto \mathrm{x}_m, \qquad \text{for } m = p, \ldots, M$$

with given initial values x_0, x_1,..., x_{p-1}, the presented windowing approach can not be used directly. One possibility to extend this concept to such cases is to save p values of x instead of one at each checkpoint. Then, during the backwards run, we will always have access to the necessary information on x to compute y.

4.2 *Application to Optimization*

In this subsection, we consider the Multi-Level Windowing, described in the previous subsection, in the context of nonstationary optimization. We give a detailed estimate of the number of steps and the amount of memory required to perform one Newton step for a given number of levels $L \in \mathbb{N}$. For

brevity, we will just write W_L and S_L instead of $W_L(M_0, M_1, \ldots, M_L)$ and $S_L(M_0, M_1, \ldots, M_L)$.

4.2.1 ***Optimization Loop* without *Building up the Hessian.*** First, we treat the variant of the optimization algorithms, which does not build up the entire Hessian of the reduced functional and is given in Algorithm 2.1. As stated in this algorithm, it is necessary to compute the value of the reduced functional and the gradient one time per Newton step. To apply the derived windowing techniques, we set $\mathrm{x} = u$, $\mathrm{y} = z$ and note, that Algorithm 4.2 can easily be extended to compute the necessary terms for evaluating the functional and the gradient during the forward or backwards computation, respectively. Thus, the total number of times steps needed to do this, is $W^{\mathrm{grad}} = W_L + M$. The required amount of memory is $S^{\mathrm{grad}} = S_L$.

Additionally to the gradient, we need to compute one matrix-vector product of the Hessian times a given vector in each of the n_{CG} steps of the conjugate gradient method. This is done as described in Algorithm 2.2. For avoiding the storage of u or z in all time steps, we have to compute u, δu, z, and δz again in every CG step. Consequently, we set here $\mathrm{x} = (u, \delta u)$ and $\mathrm{y} = (z, \delta z)$. We obtain $W^{\mathrm{hess}} = 2(W_L + M)$ and $S^{\mathrm{hess}} = 2S_L$.

In total we achieve

$$\begin{aligned} W^{(1)} &= W^{\mathrm{grad}} + n_{\mathrm{CG}} W^{\mathrm{hess}} = (1 + 2n_{\mathrm{CG}})(W_L + M), \\ S^{(1)} &= \max(S^{\mathrm{grad}}, S^{\mathrm{hess}}) = 2S_L. \end{aligned}$$

Remark 4.3 The windowing algorithm 4.2 can be modified to reduce the necessary forward steps under acceptance of increasing the needed amount of storage as follows: We do not to delete u while computing z at the points where u is saved before starting the computation of z. Additionally, we save z at these checkpoints. These saved values of u and z can be used to reduce the necessary number of forward steps to provide the values of u and δu for computing one matrix-vector product with the Hessian. Of course, when saving additional samples of u and z, the needed amount of storage increases. For one Newton step we obtain the total work $\widetilde{W}^{(1)}$ and storage $\widetilde{S}^{(1)}$ as

$$\widetilde{W}^{(1)} = W^{(1)} - 2n_{\mathrm{CG}} \min(S_L, M) \quad \text{and} \quad \widetilde{S}^{(1)} = S^{(1)} + 2S_L - M_0 - 2.$$

This modified algorithm includes the case of not using windowing for $L = 0$, while the original algorithm also for $L = 0$ deletes u during the computation of z.

4.2.2 ***Optimization Loop* with *Building up the Hessian.*** For using Algorithm 2.3, it is necessary to compute u, δu_i $(i = 1, \dots, \dim Q)$, and z. Again, the evaluation of the reduced functional is done during the first forward computation, and the evaluation of the gradient and the Hessian is done during the computation of z. So, we set $\mathrm{x} = (u, \delta u_1, \delta u_2, \dots, \delta u_{\dim Q})$ and $\mathrm{y} = z$. The required number of steps and the needed amount of memory are

$$W^{(2)} = (1 + \dim Q)W_L + M \quad \text{and} \quad S^{(2)} = (1 + \dim Q)S_L.$$

Remark 4.4 If we apply globalization techniques as line search to one of the presented optimization algorithms, we have to compute the solution of the state equation and the value of the cost functional several times without computing the gradient or the Hessian. The direct approach for doing this, is to compute the state solution, evaluate it and delete it afterwards. This might be not optimal, since for the following computation of the gradient (and the Hessian) via windowing, the needful preparations are not done. So, the better way of doing this is to run Algorithm 4.2 until line 23, and break so after completing the forward solution. If after that, the value of the gradient is needed, it is possible to restart directly on line 25 with the computation of the backwards solutions. If we consider the version presented in the actual subsection with building up the Hessian, we have to compute the tangent solutions in an extra forward run in which we can also use the saved values of the state solution.

4.2.3 ***Comparison of the Two Variants of the Optimization Algorithm.*** For $\dim Q \geq 1$, we obtain directly $S^{(2)} \geq S^{(1)}$. The relation between $W^{(1)}$ and $W^{(2)}$ depends on the factorization of M. A simple calculation leads to the following condition:

$$W^{(2)} \leq W^{(1)} \quad \Longleftrightarrow \quad \frac{\dim Q}{2} \leq n_{\mathrm{CG}} \left(1 + \frac{M}{W_L}\right).$$

If we choose L such that $W_L \approx M \log_2 M$, we can express the condition above just in terms of M as

$$W^{(2)} \lesssim W^{(1)} \quad \Longleftrightarrow \quad \frac{\dim Q}{2} \lesssim n_{\mathrm{CG}} \left(1 + \frac{1}{\log_2 M}\right).$$

This means, that even thought the required memory for the second algorithm with building up the Hessian is greater, this algorithm needs only then fewer steps than the first one, if the necessary numbers of CG steps performed in each Newton step is greater than half of the dimension of Q times a factor depending logarithmic on the number of time steps M.

5 Numerical Results

In this last section, we present some illustrative numerical examples. Throughout, the spatial discretization is done with piecewise bilinear/trilinear finite elements on quadrilateral or hexahedral cells in two respectively three dimensions. The resulting nonlinear state equations are solved by Newton's method, whereas the linear sub-problems are treated by a multigrid method. For time discretization, we consider the variants of the cG and dG methods which we have presented in Section 3. Throughout this section, we only present results using the variant of the optimization loop building up the entire Hessian, described in Algorithm 2.3 since the results of the variant without building up the Hessian are mainly the same.

All computations are done based on the software packages RoDoBo [4] and Gascoigne [2]. To depict the computed solutions, the visualization software VisuSimple [3] was used.

We consider the following two example problems on the space-time domain $\Omega \times (0,T)$ with $T = 1$.

- *Example 1*: In the first example we, discuss an optimal control problem with terminal observation, where the control variable enters the initial condition of the (nonlinear) state equation. We choose $\Omega = (0,1)^3 \subset \mathbb{R}^3$ and pose the state equation as

$$\begin{aligned} \partial_t u - \nu \Delta u + u^2 &= 0 \quad \text{in } \Omega \times (0,T), \\ \partial_n u &= 0 \quad \text{on } \partial\Omega \times (0,T), \\ u(0,\cdot) &= g_0 + \sum_{i=1}^{8} g_i q_i \quad \text{on } \Omega, \end{aligned} \tag{36}$$

  where $\nu = 0.1$, $g_0 = (1 - 2\|x - \bar{x}_0\|)^{30}$ with $\bar{x}_0 = (0.5, 0.5, 0.5)^T$ and $g_i = (1 - 0.5\|x - \bar{x}_i\|)^{30}$ with $\bar{x}_i \in \{0.2, 0.8\}^3$ for $i = 1, \ldots, 8$ are given.

  For an additionally given reference solution

$$\bar{u}_T(x) = \frac{3 + x_1 + x_2 + x_3}{6}, \quad x = (x_1, x_2, x_3)^T,$$

  the optimization problem now reads as:

$$\text{Minimize } \frac{1}{2}\int_\Omega (u(T,\cdot) - \bar{u}_T)^2\, dx + \frac{\alpha}{2}\|q\|_Q \text{ subject to (36), } (q,u) \in Q \times X,$$

  where $Q = \mathbb{R}^8$ and X is chosen in virtue of (2) with $V = H^1(\Omega)$ and

$H = L^2(\Omega)$. The regularization parameter α is set to 10^{-4}.

- *Example 2*: In the second example, we choose $\Omega = (0,1)^2 \subset \mathbb{R}^2$ and consider a parameter estimation problem with the state equation given by

$$\begin{aligned} \partial_t u - \nu\Delta u + q_1\partial_1 u + q_2\partial_2 u &= 2 + \sin(10\pi t) \quad \text{in } \Omega \times (0,T), \\ u &= 0 \quad \text{on } \partial\Omega \times (0,T), \\ u(0,\cdot) &= 0 \quad \text{on } \Omega, \end{aligned} \tag{37}$$

where we again set $\nu = 0.1$.

We assume to be given measurements $\bar{u}_{T,1}, \dots, \bar{u}_{T,5} \in \mathbb{R}$ of the point values $u(T, p_i)$ for five different measurement points $p_i \in \Omega$. The unknown parameters $(q_1, q_2) \in Q = \mathbb{R}^2$ are estimated using a least squares approach resulting in the following optimization problem:

$$\text{Minimize } \frac{1}{2}\sum_{i=1}^{5}(u(T,p_i) - \bar{u}_{T,i})^2 \text{ subject to (37), } (q,u) \in Q \times X.$$

The consideration of point measurements does not fulfill the assumption on the cost functional in (4), 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 Rannacher and Vexler [19].

5.1 *Validation of the Compututation of Derivatives*

To verify the computation of the gradient ∇j_{hk} and the Hessian $\nabla^2 j_{hk}$ of the reduced cost functional, we consider the first and second difference quotients

$$\begin{aligned} \frac{j_{hk}(q + \varepsilon\delta q) - j_{hk}(q - \varepsilon\delta q)}{2\varepsilon} &= (\nabla j_{hk}, \delta q) + e_1, \\ \frac{j_{hk}(q + \varepsilon\delta q) - 2j_{hk}(q) + j_{hk}(q - \varepsilon\delta q)}{\varepsilon^2} &= (\delta q, \nabla^2 j_{hk}\delta q) + e_2. \end{aligned}$$

We obtain using standard convergence and stability analysis the concrete form of the errors e_1 and e_2 as

$$e_1 \approx c_1\varepsilon^2\nabla^3 j_{hk}(\xi_1) + c_2\varepsilon^{-1}, \qquad e_2 \approx c_3\varepsilon^2\nabla^4 j_{hk}(\xi_2) + c_4\varepsilon^{-2},$$

where $\xi_1, \xi_2 \in (q - \varepsilon\delta q, q + \varepsilon\delta q)$ is an intermediate point and the constants c_i do not depend on ε.

The Tables 1 and 2 show the errors between the values of the derivatives computed by use of the difference quotients above and by use of the approach presented in the Sections 2 and 3, for the considered examples. The values of these errors and the orders of convergence of the reduction of these errors for $\varepsilon \to 0$ are given in the Tables 1 and 2. Note, that the values of the derivatives computed via the approach based on the ideas presented in Section 2 do not depend on ε.

The content of these tables does not considerably depend on the discretization parameters h and k, so we have the exact discrete derivatives also on coarse meshes or when using large time steps.

Table 1. Convergence of the difference quotients for the gradient and the Hessian of the reduced cost functional for Example 1 with $q = (0, \ldots, 0)^T$ and $\delta q = (1, \ldots, 1)^T$

	Discontinuous Galerkin				Continuous Galerkin			
	Gradient		Hessian		Gradient		Hessian	
ε	e_1	Conv.	e_2	Conv.	e_1	Conv.	e_2	Conv.
1.0e-00	8.56e-01	—	6.72e-01	—	7.96e-01	—	5.97e-01	—
1.0e-01	5.37e-03	2.20	4.32e-03	2.19	5.28e-03	2.17	4.08e-03	2.16
1.0e-02	5.35e-05	2.00	4.27e-05	2.00	5.26e-05	2.00	4.05e-05	2.00
1.0e-03	5.34e-07	2.00	3.28e-05	0.11	5.26e-07	2.00	3.27e-05	0.09
1.0e-04	5.30e-09	2.00	8.49e-05	-0.41	5.41e-09	1.98	8.47e-05	-0.41
1.0e-05	2.91e-10	1.25	9.16e-05	-0.03	3.24e-10	1.22	7.25e-05	0.06

Table 2. Convergence of the difference quotients for the gradient and the Hessian of the reduced cost functional for Example 2 with $q = (6, 6)^T$ and $\delta q = (1, 1)^T$

	Discontinuous Galerkin				Continuous Galerkin			
	Gradient		Hessian		Gradient		Hessian	
ε	e_1	Conv.	e_2	Conv.	e_1	Conv.	e_2	Conv.
1.0e-00	1.44e-01	—	8.09e-02	—	2.80e-01	—	1.33e-01	—
1.0e-01	1.36e-03	2.02	7.76e-04	2.01	2.59e-03	2.03	1.27e-03	2.02
1.0e-02	1.36e-05	2.00	7.75e-06	2.00	2.59e-05	2.00	1.27e-05	2.00
1.0e-03	1.36e-07	1.99	4.32e-07	1.25	2.59e-07	1.99	3.97e-07	1.50
1.0e-04	2.86e-09	1.67	5.01e-05	-2.06	2.83e-09	1.96	5.56e-06	-1.14
1.0e-05	5.94e-08	-1.31	2.18e-02	-2.63	9.95e-08	-1.54	2.00e-02	-3.55

5.2 *Optimization*

In this subsection, we apply the two optimization algorithms described in Section 2 to the two considered optimization problems. For both examples, we present the results for the two time discretization schemes presented in Section 3.

In Table 3 and Table 4, we show the progression of the norm of the gradient of the reduced functional $\|\nabla j_{hk}\|_2$ and the reduction of the cost functional j_{hk} during the Newton iteration for Example 1 and Example 2, respectively.

The computations for Example 1 were done on a mesh consisting of 4096 hexahedral cells with diameter $h = 0.0625$. The time interval $(0, 1)$ is split into 100 slices of size $k = 0.01$.

Table 3. Results of the optimization loop with dG and cG discretization for Example 1 starting with initial guess $q_0 = (0, \ldots, 0)^T$

	Discontinuous Galerkin			Continuous Galerkin		
Step	n_{CG}	$\|\nabla j_{hk}\|_2$	j_{hk}	n_{CG}	$\|\nabla j_{hk}\|_2$	j_{hk}
0	—	1.21e-01	2.76e-01	—	1.21e-01	2.76e-01
1	2	4.99e-02	1.34e-01	2	4.98e-02	1.34e-01
2	2	2.00e-02	6.28e-02	2	1.99e-02	6.33e-02
3	3	7.61e-03	2.94e-02	3	7.62e-03	3.00e-02
4	3	2.55e-03	1.64e-02	3	2.57e-03	1.70e-02
5	3	6.03e-04	1.32e-02	3	6.21e-04	1.37e-02
6	3	5.72e-05	1.29e-02	3	6.18e-05	1.34e-02
7	3	6.37e-07	1.29e-02	3	7.62e-07	1.34e-02
8	3	1.75e-10	1.29e-02	3	1.21e-10	1.34e-02

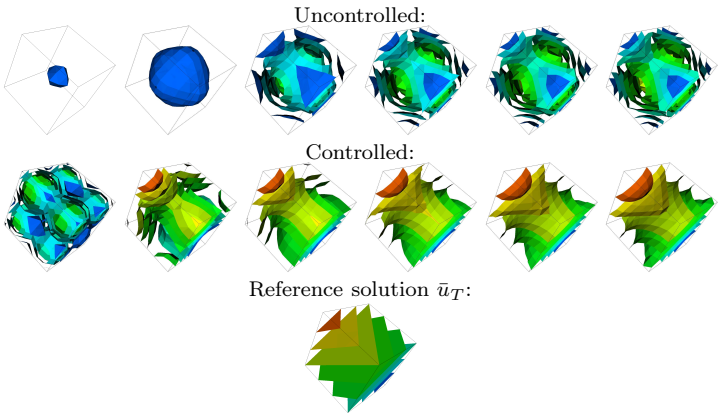

Figure 1. Solution of example problem 1 for time $t = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ before and after optimization

For Example 2, we chose a quadrilateral mesh with mesh size $h = 0.03125$ consisting of 1024 cells. The size of the time steps was set as $k = 0.005$ corresponding to 200 time steps. In Table 4, we additionaly show the value of the estimated parameters during the optimization run. The values of "measurements" are taken from a solution of the state equation on a fine mesh consisting of 65536 cells with 5000 time steps for the "exact" values of parameters chosen as $q_{\mathrm{exact}} = (7, 9)^T$.

Table 4. Results of the optimization loop with dG and cG discretization for Example 2

	Discontinuous Galerkin				Continuous Galerkin			
Step	n_{CG}	$\|\nabla j_{hk}\|_2$	j_{hk}	q	n_{CG}	$\|\nabla j_{hk}\|_2$	j_{hk}	q
0	—	1.54e-02	1.73e-03	$(6.00, 6.00)^T$	—	1.25e-02	1.23e-03	$(6.00, 6.00)^T$
1	2	5.37e-04	4.53e-04	$(5.97, 7.72)^T$	2	4.35e-04	3.07e-03	$(6.06, 7.43)^T$
2	2	1.65e-04	7.85e-05	$(6.80, 8.52)^T$	2	1.29e-04	4.75e-05	$(6.48, 8.37)^T$
3	2	3.44e-05	5.56e-06	$(7.18, 9.19)^T$	2	2.48e-05	2.35e-06	$(6.87, 8.84)^T$
4	2	2.54e-06	9.20e-07	$(7.35, 9.39)^T$	2	1.47e-06	9.29e-09	$(6.99, 8.98)^T$
5	2	1.66e-08	8.91e-07	$(7.36, 9.41)^T$	2	6.10e-09	2.04e-10	$(6.99, 8.99)^T$
6	2	7.35e-13	8.91e-07	$(7.36, 9.41)^T$	1	5.89e-11	2.04e-10	$(6.99, 8.99)^T$

We note that due to condition (21), for Example 1 the variant of the optimization algorithm, which only uses matrix-vector products of the Hessian is the more efficient one, whereas for Example 2 one should use the variant which builds up the entire Hessian.

5.3 *Windowing*

This subsection is devoted to the practical verification of the presented Multi-Level Windowing. For this, we consider Example 1 with dG time discretization on a grid consisting of 32768 cells performing 500 time steps. Table 5 demonstrates the reduction of the storage requirement described in Section 4. We can achieve a storage reduction about the factor 30 for both variants of the optimization loop. Thereby total number of steps only growths about the factor 3.2 for the algorithm with, and 4.0 for the algorithm without building up the entire Hessian.

Table 5. Reduction of the storage requirement due to Windowing in Example 1 with dG discretization and 32768 cells in each time step

	With Hessian		Without Hessian	
Factorization	Memory in MB	Time Steps	Memory in MB	Time Steps
500	1236	45000	274	35000
$5 \cdot 100$	259	80640	58	87948
$10 \cdot 50$	148	84690	32	90783
$2 \cdot 2 \cdot 5 \cdot 25$	78	120582	17	118503
$5 \cdot 10 \cdot 10$	59	114174	13	113463
$4 \cdot 5 \cdot 5 \cdot 5$	41	136512	9	130788
$2 \cdot 2 \cdot 5 \cdot 5 \cdot 5$	39	146646	9	138663

We remark that although the factorization $2 \cdot 2 \cdot 5 \cdot 25$ consists of more factors than the factorization $5 \cdot 10 \cdot 10$, both, the storage requirement and the total number of time steps are greater for first factorization than for the second one. The reason for this is the imbalance of the size of the different factors in $2 \cdot 2 \cdot 5 \cdot 25$. As showed in Section 4, in the optimal factorization are all factors

the same. So, it is evident, that a factorization as $5 \cdot 10 \cdot 10$ is more efficient than one where the size factors varies very much.

Table 5 also proves the asserted dependence of the condition when to use which variant of the optimization loop on the considered factorization on M. For the factorizations $5 \cdot 100$ and $10 \cdot 50$ the variant with building up the Hessian needs less forward steps than the other variant without building up the Hessian. However, for the remaining factorizations the situation is the opposite way around.

References

[1] Becker, R., 2001. *Adaptive Finite Elements for Optimal Control Problems*. Habilitationsschrift, Institut für Angewandte Mathematik, Universität Heidelberg.

[2] Becker, R., Braack, M., Meidner, D., Richter, T., Schmich, M., and Vexler, B., 2005. The finite element toolkit GASCOINGE. URL `http://www.gascoigne.uni-hd.de`.

[3] Becker, R., Dunne, T., and Meidner, D., 2005. VISUSIMPLE: An interactive VTK-based visualization and graphics/mpeg-generation program. URL `http://www.visusimple.uni-hd.de`.

[4] Becker, R., Meidner, D., and Vexler, B., 2005. RODOBO: A C++ library for optimization with stationary and nonstationary PDEs based on GASCOIGNE [2]. URL `http://www.rodobo.uni-hd.de`.

[5] Berggren, M., Glowinski, R., and Lions, J.-L., 1996. A computational approach to controllability issues for flow-related models. (I): Pointwise control of the viscous burgers equation. *Int. J. Comput. Fluid Dyn.*, **7**(3), 237–253.

[6] Bergounioux, M., Ito, K., and Kunisch, K., 1999. Primal-dual strategy for constrained optimal control problems. *SIAM J. Control Optim.*, **37**(4), 1176–1194.

[7] Ciarlet, P. G., 2002. *The Finite Element Method for Elliptic Problems*, volume 40 of *Classics Appl. Math.* SIAM, Philadelphia.

[8] Dautray, R. and Lions, J.-L., 1992. *Mathematical Analysis and Numerical Methods for Science and Technology: Evolution Problems I*, volume 5. Springer-Verlag, Berlin.

[9] Eriksson, K., Johnson, C., and Thomée, V., 1985. Time discretization of parabolic problems by the discontinuous Galerkin method. *RAIRO Modelisation Math. Anal. Numer.*, **19**, 611–643.

[10] Estep, D. and Larsson, S., 1993. The discontinuous Galerkin method for semilinear parabolic problems. *RAIRO Modelisation Math. Anal. Numer.*, **27**(1), 35–54.

[11] Fursikov, A. V., 1999. *Optimal Control of Distributed Systems: Theory and Applications*, volume 187 of *Transl. Math. Monogr.* AMS, Providence.

[12] Griewank, A., 1992. Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation. *Optim. Methods Softw.*, **1**(1), 35–54.

[13] Griewank, A., 2000. *Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation*, volume 19 of *Frontiers Appl. Math.* SIAM, Philadelphia.

[14] Griewank, A. and Walther, A., 2000. Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation. *ACM Trans. Math. Software*, **26**(1), 19–45.

[15] Hinze, M. and Kunisch, K., 2001. Second order methods for optimal control of time-dependent fluid flow. *SIAM J. Control Optim.*, **40**(3), 925–946.

[16] Kunisch, K. and Rösch, A., 2002. Primal-dual active set strategy for a general class of constrained optimal control problems. *SIAM J. Optim.*, **13**(2), 321–334.

[17] Lions, J.-L., 1971. *Optimal Control of Systems Governed by Partial Differential Equations*, volume 170 of *Grundlehren Math. Wiss.* Springer-Verlag, Berlin.

[18] Litvinov, W. G., 2000. *Optimization in Elliptic Problems With Applications to Mechanics of Deformable Bodies and Fluid Mechanics*, volume 119 of *Oper. Theory Adv. Appl.* Birkhäuser Verlag, Basel.

[19] Rannacher, R. and Vexler, B., 2004. A priori error estimates for the finite element discretization of elliptic parameter identification problems with pointwise measurements. *SIAM J. Control Optim.* To appear.

[20] Tröltzsch, F., 1999. On the Lagrange-Newton-SQP method for the optimal control of semilinear parabolic equations. *SIAM J. Control Optim.*, **38**(1), 294–312.