SIAM J. Sci. Comput.
Vol. 29, No. 1, pp. 22–48

NUMERICAL SENSITIVITY ANALYSIS FOR THE QUANTITY OF INTEREST IN PDE-CONSTRAINED OPTIMIZATION*

ROLAND GRIESSE† AND BORIS VEXLER†

Abstract. In this paper, we consider the efficient computation of derivatives of a functional (the quantity of interest) which depends on the solution of a PDE-constrained optimization problem with inequality constraints and which may be different from the cost functional. The optimization problem is subject to perturbations in the data. We derive conditions under with the quantity of interest possesses first and second order derivatives with respect to the perturbation parameters. An algorithm for the efficient evaluation of these derivatives is developed, with considerable savings over a direct approach, especially in the case of high-dimensional parameter spaces. The computational cost is shown to be small compared to that of the overall optimization algorithm. Numerical experiments involving a parameter identification problem for Navier–Stokes flow and an optimal control problem for a reaction-diffusion system are presented which demonstrate the efficiency of the method.

Key words. sensitivity analysis, PDE-constrained optimization, quantity of interest

AMS subject classifications. 49J20, 49K40, 90C31

DOI. 10.1137/050637273

1. Introduction.

In this paper we consider PDE-constrained optimization problems with inequality constraints. The optimization problems are formulated in a general setting including optimal control as well as parameter identification problems. The problems are subject to perturbation in the data. We suppose we are given a quantity of interest (output functional), which depends on both the state and the control variables and which may be different from the cost functional used during the optimization.

The quantity of interest is shown to possess first and, under tighter assumptions, second order derivatives with respect to the perturbation parameters. In the presence of control constraints, strict complementarity and compactness of certain derivatives of the state equation are assumed; for second order derivatives, stability of the active set is required in addition. The precise conditions are given in section 3. The main contribution of this paper is to devise an efficient algorithm to evaluate these sensitivity derivatives which offers considerable savings over a direct approach, especially in the case of high-dimensional parameter spaces. We show that the derivatives of the quantity of interest can be computed with little additional numerical effort in comparison to the corresponding derivatives of the cost functional. Moreover, the computational cost for the evaluation of the gradient of the quantity of interest is independent of the dimension of the parameter space and low compared to that of the overall optimization algorithm. The cost to evaluate the Hessian grows linearly with the dimension of the parameter space. We refer to Table 3.1 for details.

The parametric derivatives of the quantity of interest offer a significant amount of additional information on top of an optimal solution. The derivative information can be used to assess the stability of an optimal solution, or to compute a Taylor

*Received by the editors August 1, 2005; accepted for publication (in revised form) June 27, 2006; published electronically January 8, 2007.
http://www.siam.org/journals/sisc/29-1/63727.html
†Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Linz, Austria (roland.griesse@oeaw.ac.at, boris.vexler@oeaw.ac.at).

expansion which allows the fast prediction of the perturbed value of the quantity of interest in a neighborhood of a reference parameter.

We note that a quantity of interest different from the cost functional is often natural. For instance, an optimization problem in fluid flow may aim at minimizing the drag of a given body, e.g., by adjusting the boundary conditions. The quantity of interest, however, may be the lift coefficient of the optimal configuration. We also mention the applicability of our results to bilevel optimization problems where the outer variable is the "perturbation" parameter and the outer objective is the output functional, whose derivatives are needed to employ efficient optimization algorithms.

The necessity to compute higher order derivatives may impose possible limitations to the applicability of the methods presented in this paper. Second order derivatives of the cost functional and the PDE constraint are required to evaluate the gradient of the quantity of interest, and third order derivatives are required to evaluate the Hessian.

Let us put our work into perspective. The existence of first and second order sensitivity derivatives of the *objective function* (cost functional) in optimal control of PDEs with control constraints has been proved in [7, 18]. Moreover, [8] addresses the numerical computation of these derivatives. Recently, the computation of the gradient of the quantity of interest in the *absence of inequality constraints* has been discussed in [3].

Problem setting. We consider the PDE-constrained optimization problem in the following abstract form: The state variable u in an appropriate Hilbert space $\mathcal{V}$ with scalar product $(\cdot,\cdot)_{\mathcal{V}}$ is determined by a PDE (*state equation*) in weak form:

$$a(u,q,p)(\phi) = f(\phi) \quad \forall \phi \in \mathcal{V}, \tag{1.1}$$

where q denotes the control, or, more generally, design variable, in the Hilbert space $\mathcal{Q} = L^2(\omega)$ with the standard scalar product $(\cdot,\cdot)$. Typically, ω is a subset of the computational domain Ω or a subset of its boundary $\partial\Omega$. In case of finite-dimensional controls we set $\mathcal{Q} = \mathbb{R}^n$ and identify this space with $L^2(\omega)$, where $\omega = \{1, 2, \dots, n\}$ to keep the notation consistent. The parameter p from a normed linear space $\mathcal{P}$ describes the perturbations of the data.

For fixed $p \in \mathcal{P}$, the semilinear form $a(\cdot,\cdot,p)(\cdot)$ is defined on the Hilbert space $\mathcal{V} \times \mathcal{Q} \times \mathcal{V}$. Semilinear forms are written with two parentheses; the first one refers to the nonlinear arguments, whereas the second one embraces all linear arguments. The partial derivatives of the semilinear form $a(\cdot,\cdot,p)(\cdot)$ are denoted by $a'_u(\cdot,\cdot,p)(\cdot,\cdot)$, $a'_q(\cdot,\cdot,p)(\cdot,\cdot)$, etc. The linear functional $f \in \mathcal{V}'$ represents the right-hand side of the state equation, where $\mathcal{V}'$ denotes the dual space of $\mathcal{V}$. For the cost functional (objective functional) we assume the form

$$J(u,p) + \frac{\alpha}{2}\|q - \overline{q}\|^2_{\mathcal{Q}}, \tag{1.2}$$

which is typical in PDE-constrained optimization problems. Here, $\alpha > 0$ is a regularization parameter and $\overline{q} \in \mathcal{Q}$ is a reference control. The functional $J : \mathcal{V} \times \mathcal{P} \to \mathbb{R}$ is also subject to perturbation. It is possible to extend our analysis to more general cost functionals than (1.2). In particular, only notational changes are necessary if J contains linear terms in q, and if α and $\overline{q}$ also depend on the perturbation parameter. However, full generality of the cost functional comes at the expense of additional assumptions which would unnecessarily complicate the discussion.

In order to cover additional control constraints we introduce a nonempty closed convex subset $\mathcal{Q}_{ad} \subset \mathcal{Q}$ by

$$\mathcal{Q}_{ad} = \{q \in \mathcal{Q} \mid b_-(x) \le q(x) \le b_+(x) \text{ a.e. on } \omega\},$$

with bounds $b_- \le b_+ \in \mathcal{Q}$. In the case of finite-dimensional controls the inequality $b_- \le q \le b_+$ is meant to hold componentwise.

The problem under consideration is to

$$(\mathbf{OP}(p)) \qquad \begin{array}{c} \text{minimize (1.2) over } \mathcal{Q}_{ad} \times \mathcal{V} \\ \text{subject to the state equation (1.1)} \end{array}$$

for fixed $p \in \mathcal{P}$. We assume that in a neighborhood of a reference parameter p_0, there exist functions $u = U(p)$ and $q = Q(p)$, which map the perturbation parameter p to a local solution (u, q) of the problem $(\mathbf{OP}(p))$. In section 3, we give sufficient conditions ensuring the existence and differentiability of these functions. Our results complement previous findings in [7, 10, 18].

The quantity of interest is denoted by a functional

$$I : \mathcal{V} \times \mathcal{Q} \times \mathcal{P} \to \mathbb{R}. \tag{1.3}$$

This gives rise to the definition of the reduced quantity of interest $i : \mathcal{P} \to \mathbb{R}$,

$$i(p) = I(U(p), Q(p), p). \tag{1.4}$$

Likewise, we denote by $j : \mathcal{P} \to \mathbb{R}$ the reduced cost functional

$$j(p) = J(U(p), p) + \frac{\alpha}{2}\|Q(p) - \overline{q}\|^2_{\mathcal{Q}}. \tag{1.5}$$

As stated above, the main contribution of this paper is to devise an efficient algorithm to evaluate the first and second derivatives of the reduced quantity of interest $i(p)$.

The outline of the paper is as follows: In the next section we specify the first order necessary optimality conditions for the problem under consideration. We recall a primal-dual active set method for its solution. The core step of this method is described in some detail since it is also used for the problems arising during the sensitivity computation. In section 3 we use duality arguments for the efficient evaluation of the first and second order sensitivities of the quantity of interest with respect to perturbation parameters. Throughout, we compare the standard sensitivity analysis for the reduced cost functional $j(p)$ with our analysis for the reduced quantity of interest $i(p)$. In the last section we discuss two numerical examples illustrating our approach. The first example deals with a parameter identification problem for a channel flow described by the incompressible Navier–Stokes equations. In the second example we consider the optimal control of time-dependent three-species reaction-diffusion equations under control constraints.

2. Optimization algorithm. In this section we recall the first order necessary conditions for the problem $(\mathbf{OP}(p))$ and describe the optimization algorithm with active set strategy which we use in our numerical examples. In particular, we specify the Newton step taking into account the active sets since the sensitivity problems arising in section 3 are solved by the same technique.

Throughout the paper we make the following assumption.

ASSUMPTION 2.1.
1. *Let* $a(\cdot,\cdot,\cdot)(\cdot)$ *be three times continuously differentiable with respect to* (u,q,p).
2. *Let* $J(\cdot,\cdot)$ *be three times continuously differentiable with respect to* (u,p).
3. *Let* $I(\cdot,\cdot,\cdot)$ *be twice continuously differentiable with respect to* (u,q,p).

In order to establish the optimality system, we introduce the Lagrangian $\mathcal{L}$: $\mathcal{V}\times\mathcal{Q}\times\mathcal{V}\times\mathcal{P}\to\mathbb{R}$ as follows:

$$\mathcal{L}(u,q,z,p) = J(u,p) + \frac{\alpha}{2}\|q-\overline{q}\|^2_{\mathcal{Q}} + f(z) - a(u,q,p)(z), \tag{2.1}$$

where $z\in\mathcal{V}$ denotes the adjoint state. The first order necessary conditions for the problem $(\mathbf{OP}(p))$ read

$$\mathcal{L}'_u(u,q,z,p)(\delta u) = 0 \quad \forall\delta u\in\mathcal{V}, \tag{2.2}$$

$$\mathcal{L}'_q(u,q,z,p)(\delta q - q) \geq 0 \quad \forall\delta q\in\mathcal{Q}_{ad}, \tag{2.3}$$

$$\mathcal{L}'_z(u,q,z,p)(\delta z) = 0 \quad \forall\delta z\in\mathcal{V}. \tag{2.4}$$

They can be explicitly rewritten as follows:

$$J'_u(u,p)(\delta u) - a'_u(u,q,p)(\delta u, z) = 0 \quad \forall\delta u\in\mathcal{V}, \tag{2.5}$$

$$\alpha(q-\overline{q},\delta q - q) - a'_q(u,q,p)(\delta q - q, z) \geq 0 \quad \forall\delta q\in\mathcal{Q}_{ad}, \tag{2.6}$$

$$f(\delta z) - a(u,q,p)(\delta z) = 0 \quad \forall\delta z\in\mathcal{V}. \tag{2.7}$$

For given u,q,z,p, we introduce an additional Lagrange multiplier $\mu\in L^2(\omega)$ by the following identification:

$$\begin{aligned}(\mu,\delta q) &:= -\mathcal{L}'_q(u,q,z,p)(\delta q)\\ &= -\alpha(q-\overline{q},\delta q) + a'_q(u,q,p)(\delta q, z) \quad \forall\delta q\in L^2(\omega).\end{aligned}$$

The variational inequality (2.6) is known to be equivalent to the following pointwise conditions a.e. on ω:

$$q(x) = b_-(x) \Rightarrow \mu \leq 0, \tag{2.8}$$

$$q(x) = b_+(x) \Rightarrow \mu \geq 0, \tag{2.9}$$

$$b_-(x) < q(x) < b_+(x) \Rightarrow \mu = 0. \tag{2.10}$$

In addition to the necessary conditions above, in the following lemma we recall second order sufficient optimality conditions.

LEMMA 2.2 (sufficient optimality conditions). *Let* $x=(u,q,z)$ *satisfy the first order necessary conditions* (2.2)–(2.4) *of* $(\mathbf{OP}(p))$. *Moreover, let* $a'_u(u,q,p)\colon\mathcal{V}\to\mathcal{V}'$ *be surjective. If there exists* $\rho>0$ *such that*

$$\begin{pmatrix}\delta u & \delta q\end{pmatrix}\begin{bmatrix}\mathcal{L}''_{uu}(x,p) & \mathcal{L}''_{uq}(x,p)\\ \mathcal{L}''_{qu}(x,p) & \mathcal{L}''_{qq}(x,p)\end{bmatrix}\begin{pmatrix}\delta u\\ \delta q\end{pmatrix} \geq \rho\left(\|\delta u\|^2_{\mathcal{V}} + \|\delta q\|^2_{\mathcal{Q}}\right)$$

holds for all $(\delta u,\delta q)$ *satisfying the linear (tangent) PDE*

$$a'_u(u,q,p)(\delta u,\varphi) + a'_q(u,q,p)(\delta q,\varphi) = 0 \quad \forall\varphi\in\mathcal{V},$$

then (u,q) *is a strict local optimal solution of* $(\mathbf{OP}(p))$.

For the proof we refer the reader to [19].

For the solution of the first order necessary conditions (2.5)–(2.7) for fixed $p \in \mathcal{P}$, we employ a nonlinear primal-dual active set strategy; see [4, 12, 15, 16]. In the following we sketch the corresponding algorithm on the continuous level:

Nonlinear primal-dual active set strategy

1. Choose initial guess u^0, q^0, z^0, μ^0 and $c > 0$ and set $n = 1$
2. While not converged
3. Determine the active sets $\mathcal{A}_+^n$ and $\mathcal{A}_-^n$
$$\mathcal{A}_-^n = \{x \in \omega \mid q^{n-1} + \mu^{n-1}/c - b_- \le 0\},$$
$$\mathcal{A}_+^n = \{x \in \omega \mid q^{n-1} + \mu^{n-1}/c - b_+ \ge 0\}$$
4. Solve the equality-constrained optimization problem
$$\text{Minimize} \quad J(u^n, p) + \frac{\alpha}{2}\|q^n - \overline{q}\|_{\mathcal{Q}}^2 \text{ over } \mathcal{V} \times \mathcal{Q}$$
subject to (1.1) and to
$$q^n(x) = b_-(x) \text{ on } \mathcal{A}_-^n, \qquad q^n(x) = b_+(x) \text{ on } \mathcal{A}_+^n$$
with adjoint variable z^n
5. Set $\mu^n = -\alpha(q^n - \overline{q}) + a_q'(u^n, q^n, p)(\cdot, z^n)$
6. Set $n = n + 1$ and go to 2.

Remark 2.3.
1. The initial guess for the Lagrange multiplier μ^0 can be taken according to step 5. Another possibility is choosing $\mu^0 = 0$ and $q^0 \in \mathcal{Q}_{ad}$, which leads to solving the optimization problem (step 4) without control constraints in the first iteration.
2. The convergence in step 2 can be determined conveniently from agreement of the active sets in two consecutive iterations. The problem is then solved only on the inactive set $\mathcal{I}^n = \omega \setminus (A_-^n \cup A_+^n)$, e.g., by applying Newton's method to the corresponding first order necessary conditions.

Later on, the above algorithm is applied on the discrete level. The concrete discretization schemes are described in section 4 for each individual example.

Clearly, the main step in the primal-dual algorithm is the solution of the equality-constrained nonlinear optimization problem in step 4. We shall describe the Lagrange Newton SQP method for its solution in some detail since exactly the same procedure may be used to solve the sensitivity problems in section 3, which are the main focus of our paper.

For given active and inactive sets $\mathcal{A} = \mathcal{A}_+ \cup \mathcal{A}_-$ and $\mathcal{I} = \omega \setminus \mathcal{A}$, let us define the "restriction" operator $R_{\mathcal{I}} : L^2(\omega) \to L^2(\omega)$ by

$$R_{\mathcal{I}}(q) = q \cdot \chi_{\mathcal{I}},$$

where $\chi_{\mathcal{I}}$ is a characteristic function of the set $\mathcal{I}$. Similarly, the operators $R_{\mathcal{A}}$, $R_{\mathcal{A}_+}$, and $R_{\mathcal{A}_-}$ are defined. Note that $R_{\mathcal{I}}$, etc., are obviously self-adjoint.

The first order necessary conditions for the purely equality-constrained problem in step 4 are (compare (2.2)–(2.4), respectively, (2.5)–(2.7)):

$$\mathcal{L}'_u(u,q,z,p)(\delta u) = 0 \quad \forall \delta u \in \mathcal{V}, \tag{2.11}$$

$$\mathcal{L}'_q(u,q,z,p)(\delta q) = 0 \quad \forall \delta q \in L^2(\mathcal{I}^n), \tag{2.12}$$

$$q - b_- = 0 \quad \text{on } A^n_-, \tag{2.13}$$

$$q - b_+ = 0 \quad \text{on } A^n_+, \tag{2.14}$$

$$\mathcal{L}'_z(u,q,z,p)(\delta z) = 0 \quad \forall \delta z \in \mathcal{V}, \tag{2.15}$$

with the inactive set $\mathcal{I}^n = \omega \backslash (\mathcal{A}^n_- \cup \mathcal{A}^n_+)$. Using the restriction operators, (2.12)–(2.14) can be reformulated as

$$\mathcal{L}'_q(u,q,z,p)(R_{\mathcal{I}^n}\delta q) + (q - b_-, R_{\mathcal{A}^n_-}\delta q) + (q - b_+, R_{\mathcal{A}^n_+}\delta q) = 0 \quad \forall \delta q \in \mathcal{Q}.$$

The Lagrange Newton SQP method is defined as Newton's method, applied to (2.11)–(2.15). To this end, we define B as the Hessian operator of the Lagrangian $\mathcal{L}$, i.e.,

$$B(x,p) = \begin{bmatrix} \mathcal{L}''_{uu}(x,p)(\cdot,\cdot) & \mathcal{L}''_{uq}(x,p)(\cdot,\cdot) & \mathcal{L}''_{uz}(x,p)(\cdot,\cdot) \\ \mathcal{L}''_{qu}(x,p)(\cdot,\cdot) & \mathcal{L}''_{qq}(x,p)(\cdot,\cdot) & \mathcal{L}''_{qz}(x,p)(\cdot,\cdot) \\ \mathcal{L}''_{zu}(x,p)(\cdot,\cdot) & \mathcal{L}''_{zq}(x,p)(\cdot,\cdot) & 0 \end{bmatrix}. \tag{2.16}$$

To shorten the notation, we abbreviate $x = (u,q,z)$ and $\mathcal{X} = \mathcal{V} \times \mathcal{Q} \times \mathcal{V}$. Note that $B(x,p)$ is a bilinear operator on the space $\mathcal{X}$. By "multiplication" of B with an element $\delta x \in \mathcal{X}$ from the left, we mean the insertion of the components of δx into the first argument. Similarly we define the "multiplication" of B with an element $\delta x \in \mathcal{X}$ from the right as insertion of the components of δx into the second argument. When only one element is inserted, B is interpreted as a linear operator $B : \mathcal{X} \to \mathcal{X}'$. In what follows, we shall omit the $(\cdot,\cdot)$ notation if no ambiguity arises.

In the absence of control constraints, the Newton update $(\Delta u, \Delta q, \Delta z)$ for (2.11)–(2.15) at the current iterate (u_k, q_k, z_k) is given by the solution of

$$B(x_k,p) \begin{pmatrix} \Delta u \\ \Delta q \\ \Delta z \end{pmatrix} = - \begin{pmatrix} \mathcal{L}'_u(x_k,p) \\ \mathcal{L}'_q(x_k,p) \\ \mathcal{L}'_z(x_k,p) \end{pmatrix}. \tag{2.17}$$

With nonempty active sets $\mathcal{A}^n_-$ and $\mathcal{A}^n_+$, however, (2.17) is replaced by

$$\widetilde{B}(x_k,p) \begin{pmatrix} \Delta u \\ \Delta q \\ \Delta z \end{pmatrix} = - \begin{pmatrix} \mathcal{L}'_u(x_k,p) \\ R_{\mathcal{I}^n}\mathcal{L}'_q(x_k,p) + R_{\mathcal{A}^n_-}(q_k - b_-) + R_{\mathcal{A}^n_+}(q_k - b_+) \\ \mathcal{L}'_z(x_k,p) \end{pmatrix}, \tag{2.18}$$

where

$$\widetilde{B}(x_k,p) = \begin{pmatrix} id & & \\ & R_{\mathcal{I}^n} & \\ & & id \end{pmatrix} B(x_k,p) \begin{pmatrix} id & & \\ & R_{\mathcal{I}^n} & \\ & & id \end{pmatrix} + \begin{pmatrix} 0 & & \\ & R_{\mathcal{A}^n} & \\ & & 0 \end{pmatrix}. \tag{2.19}$$

In other words, $\widetilde{B}$ is obtained from B by replacing those components in the derivatives with respect to the control q by the identity which belongs to the active set. In our practical realization, we reduce the system (2.18) to the control space $L^2(\omega)$ using Schur complement techniques; see, e.g., [17]. The reduced system is solved iteratively using the conjugate gradient method, where each step requires the evaluation of a matrix-vector product for the reduced Hessian, which in turn requires the solution of one tangent and one dual problem; see, e.g., [13] or [2] for a detailed description of this procedure in the context of space-time finite element discretization of the problem. In fact, the reduced system needs to be solved only on the currently inactive part $L^2(\mathcal{I}^n)$ of the control space since on the active sets, the update Δq satisfies the trivial relation $R_{\mathcal{A}^n_\pm}(\Delta q) = R_{\mathcal{A}^n_\pm}(b_\pm - q_{k-1})$.

The Newton step is completed by applying the update

$$(u_{k+1}, q_{k+1}, z_{k+1}) = (u_k, q_k, z_k) + (\Delta u, \Delta q, \Delta z).$$

3. Sensitivity analysis. In this section we analyze the behavior of local optimal solutions for $(\mathbf{OP}(p))$ under perturbations of the parameter p. We derive formulas for the first and second order derivatives of the reduced quantity of interest and develop an efficient method for their evaluation.

To set the stage, we outline the main ideas in section 3.1 by means of a finite-dimensional optimization problem, without partitioning the optimization variables into states and controls, and in the absence of control constraints. To facilitate the discussion of the infinite-dimensional case, we treat the case of no control constraints in section 3.2 and turn to problems with these constraints in section 3.3. Throughout, we compare the standard sensitivity analysis for the reduced cost functional $j(p)$ (1.5) with our analysis for the reduced quantity of interest $i(p)$ (1.4). The main results can be found in Theorem 3.6 for the unconstrained case and in Theorems 3.18 and 3.21 for the case with control constraints. An algorithm at the end of section 3 summarizes the necessary steps to evaluate the various sensitivity quantities.

3.1. Outline of ideas. Let us consider the nonlinear finite-dimensional equality-constrained optimization problem

$$\text{Minimize } J(x,p) \text{ s.t. } g(x,p) = 0, \tag{3.1}$$

where $x \in \mathbb{R}^n$ denotes the optimization variable, $p \in \mathbb{R}^d$ is the perturbation parameter, and $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^m$ collects a number of equality constraints. The Lagrangian of (3.1) is $\mathcal{L}(x,p) = J(x,p) - z^\top g(x,p)$, and under standard constraint qualifications, a local minimizer x_0 of (3.1) at the reference parameter p_0 has an associated Lagrange multiplier $z_0 \in \mathbb{R}^m$ such that

$$\begin{aligned} \mathcal{L}'_x(x_0, z_0, p_0) &= J'_x(x_0,p_0) - z_0^\top g'_x(x_0,p_0) = 0, \\ \mathcal{L}'_z(x_0, z_0, p_0) &= g(x_0,p_0) = 0 \end{aligned} \tag{3.2}$$

hold. If we assume second order sufficient conditions to hold in addition, then the implicit function theorem yields the local existence of functions $X(p)$ and $Z(p)$ which satisfy (3.2) with p instead of p_0, and $X(p_0) = x_0$ and $Z(p_0) = z_0$ hold. Moreover, (3.2) can be differentiated totally with respect to the parameter and we obtain

$$\begin{pmatrix} \mathcal{L}''_{xx}(x_0,z_0,p_0) & g'_x(x_0,p_0)^\top \\ g'_x(x_0,p_0) & 0 \end{pmatrix} \begin{pmatrix} X'(p_0)\,\delta p \\ Z'(p_0)\,\delta p \end{pmatrix} = - \begin{pmatrix} \mathcal{L}''_{xp}(x_0,z_0,p_0)\,\delta p \\ g'_p(x_0,p_0)\,\delta p \end{pmatrix}. \tag{3.3}$$

The solution of (3.3) is a directional derivative of $X(p)$ (and $Z(p)$) at $p = p_0$, and we note that it is equivalent to the solution of a linear-quadratic optimization problem. Hence the evaluation of the full Jacobian $X'(p_0)$ requires $d = \dim \mathcal{P}$ solves of (3.3) with different δp. In our context of large-scale problems, iterative solvers need to be used and the numerical effort to evaluate the full Jacobian scales linearly with the number of right-hand sides, i.e., with the dimension of the parameter space $d = \dim \mathcal{P}$.

We adapt the definition of the reduced cost functional and the reduced quantity of interest to our current setting, $j(p) = J(X(p), p)$ and $i(p) = I(X(p), p)$. Since we wish to compare the effort to compute the first and second order derivatives of both, we begin by recalling the following result.

LEMMA 3.1. *Under the conditions above, the reduced cost functional is twice differentiable and*

$$\begin{aligned} j'(p_0)\,\delta p &= \mathcal{L}'_p(x_0, z_0, p_0)\,\delta p, \\ \delta p^\top j''(p_0)\,\widehat{\delta p} &= \delta p^\top \big[\mathcal{L}''_{px}(x_0, z_0, p_0) X'(p_0)\,\widehat{\delta p} + \mathcal{L}''_{pz}(x_0, z_0, p_0) Z'(p_0)\,\widehat{\delta p} \\ &\qquad + \mathcal{L}''_{pp}(x_0, z_0, p_0)\,\widehat{\delta p}\big]. \end{aligned}$$

Proof. We have $j(p) = \mathcal{L}(X(p), Z(p), p)$ and hence by the chain rule $j'(p_0) = \mathcal{L}'_x(x_0, z_0, p_0) X'(p_0) + \mathcal{L}'_z(x_0, z_0, p_0) Z'(p_0) + \mathcal{L}'_p(x_0, z_0, p_0)$, where the first two terms vanish in view of (3.2). Differentiating again totally with respect to p yields the expression for the second derivative. ∎

Lemma 3.1 shows that the evaluation of the gradient of $j(\cdot)$ does not require any linear solves of the sensitivity system (3.3), while the evaluation of the Hessian requires $d = \dim \mathcal{P}$ such solves. The corresponding results for the infinite-dimensional case can be found below in Propositions 3.5 and 3.16 for the unconstrained and control constrained cases.

We will show now that the derivatives of the reduced quantity of interest $i(\cdot)$ can be evaluated efficiently, requiring just one additional system solve. This is a significant improvement over a direct approach; see Table 3.1.

From a first look at

$$i'(p_0)\,\delta p = I'_x(x_0, p_0) X'(p_0)\,\delta p + I'_p(x_0, p_0)\,\delta p$$

it seems that the evaluation of the gradient $i'(p_0)$ requires $d = \dim \mathcal{P}$ solves of the system (3.3). This is referred to as the direct approach in Table 3.1. However, using (3.3), we may rewrite this as

$$i'(p_0)\,\delta p = \big[-\big(I'_x(x_0, p_0), 0\big)\, B_0^{-1}\big] \begin{pmatrix} \mathcal{L}''_{xp}(x_0, z_0, p_0)\,\delta p \\ g'_p(x_0, p_0)\,\delta p \end{pmatrix} + I'_p(x_0, p_0)\,\delta p,$$

where B_0 is the matrix on the left-hand side of (3.3). Realizing that $I'_x(x_0, p_0)$ has just one row, evaluating the term in square brackets amounts to only one linear system solve. We define the dual quantities (v, y) by

$$B_0^\top \begin{pmatrix} v \\ y \end{pmatrix} = -\begin{pmatrix} I'_x(x_0, p_0) \\ 0 \end{pmatrix}$$

and finally obtain

$$i'(p_0)\,\delta p = v^\top \mathcal{L}''_{xp}(x_0, z_0, p_0)\,\delta p + y^\top \mathcal{L}''_{zp}(x_0, z_0, p_0)\,\delta p + I'_p(x_0, p_0)\,\delta p. \tag{3.4}$$

We refer to this as a dual approach. In our context, B_0 is symmetric and hence the computation of the dual quantities requires just one solve of (3.3) with a modified right-hand side; see again Table 3.1.

For the second derivative, we differentiate (3.4) totally with respect to p. From the chain rule we infer that the sensitivities $X'(p_0)$ and $Z'(p_0)$ now come into play. In addition, v and y need to be differentiated with respect to p, but again a duality technique can be used in order to avoid computing these extra terms. Hence the extra computational cost to evaluate the Hessian of $i(\cdot)$ amounts to $d = \dim \mathcal{P}$ solves for the evaluation of the sensitivity matrices $X'(p_0)$ and $Z'(p_0)$; see Table 3.1. Details can be found in the proofs of Theorem 3.6 for the unconstrained case and Theorems 3.18 and 3.21 for the case with control constraints.

3.2. The case of no control constraints. Throughout this and the following section, we denote by $p_0 \in \mathcal{P}$ a given reference parameter and by $x_0 = (u_0, q_0, z_0)$ a solution to the corresponding first order optimality system (2.11)–(2.15). Moreover, we make the following regularity assumption which we require throughout.

ASSUMPTION 3.2. *Let the derivative $a'_u(u_0, q_0, p_0) : \mathcal{V} \to \mathcal{V}'$ be both surjective and injective, so that it possesses a continuous inverse.*

In the case of no control constraints, i.e., $\mathcal{Q}_{ad} = \mathcal{Q}$, the first order necessary conditions (2.11)–(2.15) simplify to

$$\mathcal{L}'_u(u, q, z, p)(\delta u) = 0 \quad \forall \delta u \in \mathcal{V}, \tag{3.5}$$

$$\mathcal{L}'_q(u, q, z, p)(\delta q) = 0 \quad \forall \delta q \in \mathcal{Q}, \tag{3.6}$$

$$\mathcal{L}'_z(u, q, z, p)(\delta z) = 0 \quad \forall \delta z \in \mathcal{V}. \tag{3.7}$$

The analysis in this subsection is based on the classical implicit function theorem. We denote by $B_0 = B(x_0, p_0)$ the previously defined Hessian operator at the given reference solution. For the results in this section we require that B_0 is boundedly invertible. This property follows from the second order sufficient conditions; see, for instance, [14].

LEMMA 3.3. *Let the second order sufficient conditions set forth in Lemma* 2.2 *hold at x_0 for $\mathbf{OP}(p_0)$. Then B_0 is boundedly invertible.*

The following lemma is a direct application of the implicit function theorem (see [5]) to the first order optimality system (3.5)–(3.7).

LEMMA 3.4. *Let B_0 be boundedly invertible. Then there exist neighborhoods $\mathcal{N}(p_0) \subset \mathcal{P}$ of p_0 and $\mathcal{N}(x_0) \subset \mathcal{X}$ of x_0 and a continuously differentiable function $(U, Q, Z) : \mathcal{N}(p_0) \to \mathcal{N}(x_0)$ with the following properties:*

(a) *For every $p \in \mathcal{N}(p_0)$, $(U(p), Q(p), Z(p))$ is the unique solution to the system (3.5)–(3.7) in the neighborhood $\mathcal{N}(x_0)$.*

(b) *$(U(p_0), Q(p_0), Z(p_0)) = (u_0, q_0, z_0)$ holds.*

(c) *The derivative of (U, Q, Z) at p_0 in the direction $\delta p \in \mathcal{P}$ is given by the unique solution of*

$$B_0 \begin{pmatrix} U'(p_0)(\delta p) \\ Q'(p_0)(\delta p) \\ Z'(p_0)(\delta p) \end{pmatrix} = - \begin{pmatrix} \mathcal{L}''_{up}(x_0, p_0)(\cdot, \delta p) \\ \mathcal{L}''_{qp}(x_0, p_0)(\cdot, \delta p) \\ \mathcal{L}''_{zp}(x_0, p_0)(\cdot, \delta p) \end{pmatrix}. \tag{3.8}$$

In the following proposition we recall the first and second order sensitivity derivatives of the *cost functional* $j(p)$; compare this to [18].

PROPOSITION 3.5. *Let B_0 be boundedly invertible. Then the reduced cost functional $j(p) = J(U(p), p) + \frac{\alpha}{2} \|Q(p) - \overline{q}\|^2_{\mathcal{Q}}$ is twice continuously differentiable in $\mathcal{N}(p_0)$.*

The first order derivative at p_0 in the direction $\delta p \in \mathcal{P}$ is given by

$$j'(p_0)(\delta p) = \mathcal{L}'_p(x_0, p_0)(\delta p). \tag{3.9}$$

For the second order derivative in the directions of δp and $\widehat{\delta p}$, we have

$$\begin{aligned} j''(p_0)(\delta p, \widehat{\delta p}) &= \mathcal{L}''_{up}(x_0, p_0)(U'(p)(\delta p), \widehat{\delta p}) + \mathcal{L}''_{qp}(x_0, p_0)(Q'(p)(\delta p), \widehat{\delta p}) \\ &\quad + \mathcal{L}''_{zp}(x_0, p_0)(Z'(p)(\delta p), \widehat{\delta p}) + \mathcal{L}''_{pp}(x_0, p_0)(\delta p, \widehat{\delta p}). \end{aligned} \tag{3.10}$$

Proof. Since $(U(p), Q(p))$ satisfies the state equation, we have

$$j(p) = \mathcal{L}(U(p), Q(p), Z(p), p)$$

for all $p \in \mathcal{N}(p_0)$. By the chain rule, the derivative of $j(p)$ reads

$$\begin{aligned} j'(p_0)(\delta p) = \mathcal{L}'_u(x_0, p_0)(U'(p_0)(\delta p)) + \mathcal{L}'_q(x_0, p_0)(Q'(p_0)(\delta p)) + \mathcal{L}'_z(x_0, p_0)(Z'(p_0)(\delta p)) \\ + \mathcal{L}'_p(x_0, p_0)(\delta p). \end{aligned}$$

The three terms in the first line vanish in view of the optimality system (3.5)–(3.7). Differentiating (3.9) again totally with respect to p in the direction of $\widehat{\delta p}$ yields (3.10), which completes the proof. ∎

The previous proposition allows us to evaluate the first order derivative of the reduced cost functional *without* computing the sensitivity derivatives of the state, control, and adjoint variables. That is, the effort to evaluate $j'(p_0)$ is negligible compared to the effort required to solve the optimization problem. In order to obtain second order derivative $j''(p_0)$, however, the sensitivity derivatives have to be computed according to formula (3.8). This corresponds to the solution of one additional linear-quadratic optimization problem per perturbation direction δp, whose optimality system is given by (3.8).

We now turn to our main result in the absence of control constraints. In the following theorem, we show that the first and second order derivatives of the *quantity of interest* can be evaluated at practically the same effort as those of the cost functional. To this end, we use a duality technique (see section 3.1) and formulate the following dual problem for the dual variables $v \in \mathcal{V}$, $r \in \mathcal{Q}$, and $y \in \mathcal{V}$:

$$B_0 \begin{pmatrix} v \\ r \\ y \end{pmatrix} = - \begin{pmatrix} I'_u(q_0, u_0, p_0) \\ I'_q(q_0, u_0, p_0) \\ 0 \end{pmatrix}. \tag{3.11}$$

We remark that this dual problem involves the same operator matrix B_0 as the sensitivity problem (3.8) since B_0 is self-adjoint.

THEOREM 3.6. *Let B_0 be boundedly invertible. Then the reduced quantity of interest $i(p)$ defined in* (1.4) *is twice continuously differentiable in $\mathcal{N}(p_0)$. The first order derivative at p_0 in the direction $\delta p \in \mathcal{P}$ is given by*

$$\begin{aligned} i'(p_0)(\delta p) = \mathcal{L}''_{up}(x_0, p_0)(v, \delta p) + \mathcal{L}''_{qp}(x_0, p_0)(r, \delta p) + \mathcal{L}''_{zp}(x_0, p_0)(y, \delta p) \\ + I'_p(u_0, q_0, p_0)(\delta p). \end{aligned} \tag{3.12}$$

For the second order derivative in the directions of δp and $\widehat{\delta p}$, we have

$$\begin{aligned} i''(p_0)(\delta p, \widehat{\delta p}) &= \langle v, \eta \rangle_{\mathcal{V} \times \mathcal{V}'} + \langle r, \kappa \rangle_{\mathcal{Q} \times \mathcal{Q}'} + \langle y, \sigma \rangle_{\mathcal{V} \times \mathcal{V}'} \\ &+ \begin{pmatrix} U'(p_0)(\delta p) \\ Q'(p_0)(\delta p) \\ \delta p \end{pmatrix}^\top \begin{pmatrix} I''_{uu}(q_0, u_0, p_0) & I''_{uq}(q_0, u_0, p_0) & I''_{up}(q_0, u_0, p_0) \\ I''_{qu}(q_0, u_0, p_0) & I''_{qq}(q_0, u_0, p_0) & I''_{qp}(q_0, u_0, p_0) \\ I''_{pu}(q_0, u_0, p_0) & I''_{pq}(q_0, u_0, p_0) & I''_{pp}(q_0, u_0, p_0) \end{pmatrix} \begin{pmatrix} U'(p_0)(\widehat{\delta p}) \\ Q'(p_0)(\widehat{\delta p}) \\ \widehat{\delta p} \end{pmatrix}. \end{aligned} \tag{3.13}$$

Here, $(\eta, \kappa, \sigma) \in \mathcal{V}' \times \mathcal{Q}' \times \mathcal{V}'$ *is given by*

$$\begin{pmatrix} \eta \\ \kappa \\ \sigma \end{pmatrix} = \begin{pmatrix} \mathcal{L}'''_{upp}()(\cdot, \delta p, \widehat{\delta p}) \\ \mathcal{L}'''_{qpp}()(\cdot, \delta p, \widehat{\delta p}) \\ \mathcal{L}'''_{zpp}()(\cdot, \delta p, \widehat{\delta p}) \end{pmatrix}$$
$$+ \begin{pmatrix} \mathcal{L}'''_{upu}()(\cdot, \delta p, U'(p_0)(\widehat{\delta p})) + \mathcal{L}'''_{upq}()(\cdot, \delta p, Q'(p_0)(\widehat{\delta p})) + \mathcal{L}'''_{upz}()(\cdot, \delta p, Z'(p_0)(\widehat{\delta p})) \\ \mathcal{L}'''_{qpu}()(\cdot, \delta p, U'(p_0)(\widehat{\delta p})) + \mathcal{L}'''_{qpq}()(\cdot, \delta p, Q'(p_0)(\widehat{\delta p})) + \mathcal{L}'''_{qpz}()(\cdot, \delta p, Z'(p_0)(\widehat{\delta p})) \\ \mathcal{L}'''_{zpu}()(\cdot, \delta p, U'(p_0)(\widehat{\delta p})) + \mathcal{L}'''_{zpq}()(\cdot, \delta p, Q'(p_0)(\widehat{\delta p})) \end{pmatrix}$$
$$+ \Big(B'_u()(U'(p_0)(\widehat{\delta p})) + B'_q()(Q'(p_0)(\widehat{\delta p})) + B'_z()(Z'(p_0)(\widehat{\delta p})) + B'_p()(\widehat{\delta p}) \Big) \begin{pmatrix} U'(p_0)(\delta p) \\ Q'(p_0)(\delta p) \\ Z'(p_0)(\delta p) \end{pmatrix}.$$

Remark 3.7.

(a) In the definition of (η, κ, σ) we have abbreviated the evaluation at the point (x_0, p_0) by $()$.

(b) The bracket $\langle \cdot, \cdot \rangle_{\mathcal{V} \times \mathcal{V}'}$ in (3.13) denotes the duality pairing between $\mathcal{V}$ and its dual space $\mathcal{V}'$. For instance, the evaluation of $\langle v, \eta \rangle_{\mathcal{V} \times \mathcal{V}'}$ amounts to plugging in v instead of $\cdot$ in the definition of η. A similar notation is used for the control space $\mathcal{Q}$.

(c) It is tedious but straightforward to check that (3.13) coincides with (3.10) if the quantity of interest is chosen equal to the cost functional. In this case, it follows from (3.11) that the dual quantities v and r vanish and $y = z_0$ holds.

Proof of Theorem 3.6. From the definition of the reduced quantity of interest (1.4), we infer that

$$i'(p_0)(\delta p) = I'_u(u_0, q_0, p_0)(U'(p_0)(\delta p)) + I'_q(u_0, q_0, p_0)(Q'(p_0)(\delta p)) + I'_p(u_0, q_0, p_0)(\delta p) \tag{3.14}$$

holds. In virtue of (3.8) and (3.11), the sum of the first two terms equals

$$-\begin{pmatrix} I'_u(u_0, q_0, p_0) \\ I'_q(u_0, q_0, p_0) \\ 0 \end{pmatrix}^\top B_0^{-1} \begin{pmatrix} \mathcal{L}''_{up}(x_0, p_0)(\cdot, \delta p) \\ \mathcal{L}''_{qp}(x_0, p_0)(\cdot, \delta p) \\ \mathcal{L}''_{zp}(x_0, p_0)(\cdot, \delta p) \end{pmatrix} = \begin{pmatrix} v \\ r \\ y \end{pmatrix}^\top \begin{pmatrix} \mathcal{L}''_{up}(x_0, p_0)(\cdot, \delta p) \\ \mathcal{L}''_{qp}(x_0, p_0)(\cdot, \delta p) \\ \mathcal{L}''_{zp}(x_0, p_0)(\cdot, \delta p) \end{pmatrix},$$

which implies (3.12). In order to obtain the second derivative, we differentiate (3.14) totally with respect to p in the direction of $\widehat{\delta p}$. This yields

$$\begin{aligned} &i''(p_0)(\delta p, \widehat{\delta p}) \\ &= \begin{pmatrix} U'(p_0)(\delta p) \\ Q'(p_0)(\delta p) \\ \delta p \end{pmatrix}^\top \begin{pmatrix} I''_{uu}(q_0, u_0, p_0) & I''_{uq}(q_0, u_0, p_0) & I''_{up}(q_0, u_0, p_0) \\ I''_{qu}(q_0, u_0, p_0) & I''_{qq}(q_0, u_0, p_0) & I''_{qp}(q_0, u_0, p_0) \\ I''_{pu}(q_0, u_0, p_0) & I''_{pq}(q_0, u_0, p_0) & I''_{pp}(q_0, u_0, p_0) \end{pmatrix} \begin{pmatrix} U'(p_0)(\widehat{\delta p}) \\ Q'(p_0)(\widehat{\delta p}) \\ \widehat{\delta p} \end{pmatrix} \\ &\qquad + \begin{pmatrix} I'_u(u_0, q_0, p_0) \\ I'_q(u_0, q_0, p_0) \\ 0 \end{pmatrix}^\top \begin{pmatrix} U''(p_0)(\delta p, \widehat{\delta p}) \\ Q''(p_0)(\delta p, \widehat{\delta p}) \\ Z''(p_0)(\delta p, \widehat{\delta p}) \end{pmatrix}. \end{aligned} \tag{3.15}$$

From differentiating (3.8) totally with respect to p in the direction of $\widehat{\delta p}$, we obtain

$$B_0 \begin{pmatrix} U''(p_0)(\delta p, \widehat{\delta p}) \\ Q''(p_0)(\delta p, \widehat{\delta p}) \\ Z''(p_0)(\delta p, \widehat{\delta p}) \end{pmatrix} = - \begin{pmatrix} \eta \\ \kappa \\ \sigma \end{pmatrix}. \tag{3.16}$$

From here, (3.13) follows. □

The main statement of the previous theorem is that the first and second order derivatives of the reduced quantity of interest can be evaluated at the additional expense of just one dual problem (3.11), compared to the evaluation of the reduced cost functional's derivatives. More precisely, computing the gradient of $i(p)$ at p_0 requires only the solution of (3.11). In addition, in order to compute the Hessian of $i(p)$ at p_0, the sensitivity quantities $U'(p_0)$, $Q'(p_0)$, and $Z'(p_0)$ need to be evaluated in the directions of a collection of basis vectors of the parameter space $\mathcal{P}$. That is, $\dim \mathcal{P}$ sensitivity problems (3.8) need to be solved. These are exactly the same problems which have to be solved for the computation of the Hessian of the *reduced cost functional*; see Table 3.1. Note that in the combined effort $1 + \dim \mathcal{P}$, "1" refers to the same dual problem (3.11) that has already been solved during the computation of the gradient of $i(p)$. In case that the space $\mathcal{P}$ is infinite-dimensional, it needs to be discretized first. Finally, in order to evaluate the second order Taylor expansion for a given direction δp,

$$i(p_0 + \delta p) \approx i(p_0) + i'(p_0)(\delta p) + \frac{1}{2} i''(p_0)(\delta p, \delta p),$$

the same dual problem (3.11) and one sensitivity problem (3.8) in the direction of δp are needed; see Table 3.1.

TABLE 3.1
Number of linear-quadratic problems to be solved to evaluate the derivatives of $j(p)$ and $i(p)$.

	Reduced cost functional $j(p)$	Reduced quantity of interest $i(p)$	
		Dual approach	Direct approach
Gradient	0	1	$\dim \mathcal{P}$
Hessian	$\dim \mathcal{P}$	$1 + \dim \mathcal{P}$	$(\dim \mathcal{P})\ (\dim \mathcal{P} + 1)/2$

Note that the sensitivity and dual problems (3.8) and (3.11), respectively, are solved by the technique described in section 2. The solution of such problems amounts to the computation of one additional QP step (2.17), with a different right-hand side. Therefore, the numerical effort to compute, e.g., the second order Taylor expansion for a given direction is typically low compared to the solution of the nonlinear optimization problem $\mathbf{OP}(p_0)$.

3.3. The control-constrained case. The analysis is based on the notion of strong regularity for the problem $\mathbf{OP}(p)$. Strong regularity extends the previous assumption of bounded invertibility of B_0 used throughout section 3.2.

Below, we make use of $\mu_0 \in \mathcal{Q}$ given by the following identification:

$$(\mu_0, \delta q) = -\mathcal{L}'_q(x_0, p_0)(\delta q) \quad \forall \delta q \in \mathcal{Q}. \tag{3.17}$$

This quantity acts as a Lagrange multiplier for the control constraint $q \in \mathcal{Q}_{ad}$. For the definition of strong regularity we introduce the following linearized optimality system which depends on $\varepsilon = (\varepsilon^u, \varepsilon^q, \varepsilon^z) \in \mathcal{V} \times \mathcal{Q} \times \mathcal{V}$:

$(\mathbf{LOS}(\varepsilon))$

$$\begin{aligned}\mathcal{L}''_{uu}(x_0,p_0)(\delta u, u-u_0) + \mathcal{L}''_{uq}(x_0,p_0)(\delta u, q-q_0) \\ + \mathcal{L}''_{uz}(x_0,p_0)(\delta u, z-z_0) + \mathcal{L}'_u(x_0,p_0)(\delta u) \\ + (\varepsilon^u, \delta u)_{\mathcal{V}} = 0 \quad \forall \delta u \in \mathcal{V},\end{aligned} \tag{3.18}$$

$$\begin{aligned}\mathcal{L}''_{uq}(x_0,p_0)(u-u_0, \delta q - q) + \mathcal{L}''_{qq}(x_0,p_0)(\delta q - q, q-q_0) \\ + \mathcal{L}''_{qz}(x_0,p_0)(\delta q - q, z-z_0) \\ + \mathcal{L}'_q(x_0,p_0)(\delta q - q) + (\varepsilon^q, \delta q - q) \geq 0 \quad \forall \delta q \in \mathcal{Q}_{ad},\end{aligned} \tag{3.19}$$

$$\begin{aligned}\mathcal{L}''_{zu}(x_0,p_0)(\delta z, u-u_0) + \mathcal{L}''_{zq}(x_0,p_0)(\delta z, q-q_0) \\ + \mathcal{L}'_z(x_0,p_0)(\delta z) + (\varepsilon^z, \delta z)_{\mathcal{V}} = 0 \quad \forall \delta z \in \mathcal{V}.\end{aligned} \tag{3.20}$$

In what follows, we refer to (3.18)–(3.20) as $(\mathbf{LOS}(\varepsilon))$.

DEFINITION 3.8 (strong regularity). *Let $p_0 \in \mathcal{P}$ be a given reference parameter and let $x_0 = (u_0, q_0, z_0)$ be a solution to the corresponding first order optimality system* (2.5)–(2.7)*. If there exist neighborhoods $\mathcal{N}(0) \subset \mathcal{X} = \mathcal{V} \times \mathcal{Q} \times \mathcal{V}$ of 0 and $\mathcal{N}(x_0) \subset \mathcal{X}$ of x_0 such that*

(a) *for every $\varepsilon \in \mathcal{N}(0)$, there exists a solution $(u^\varepsilon, q^\varepsilon, z^\varepsilon)$ to the linearized optimality system* (3.18)–(3.20)*,*

(b) *$(u^\varepsilon, q^\varepsilon, z^\varepsilon)$ is the unique solution of* (3.18)–(3.20) *in $\mathcal{N}(x_0)$, and*

(c) *$(u^\varepsilon, q^\varepsilon, z^\varepsilon)$ depends Lipschitz-continuously on ε, i.e., there exists $L > 0$ such that*

$$\|u^{\varepsilon_1} - u^{\varepsilon_2}\|_{\mathcal{V}} + \|q^{\varepsilon_1} - q^{\varepsilon_2}\|_{\mathcal{Q}} + \|z^{\varepsilon_1} - z^{\varepsilon_2}\|_{\mathcal{V}} \leq L\, \|\varepsilon_1 - \varepsilon_2\|_{\mathcal{X}} \tag{3.21}$$

holds for all $\varepsilon_1, \varepsilon_2 \in \mathcal{N}(0)$,

then the first order optimality system (2.5)–(2.7) *is called* strongly regular *at x_0.*

Note that (u_0, q_0, z_0) solves (3.18)–(3.20) for $\varepsilon = 0$. It is not difficult to see that in the case of no control constraints, i.e., $\mathcal{Q} = \mathcal{Q}_{ad}$, strong regularity is nothing else than bounded invertibility of B_0 which we had to assume in section 3.2. In the following lemma we show that strong regularity holds under suitable second order sufficient optimality conditions, in analogy to Lemma 3.3. The proof can be carried out using the techniques presented in [21].

LEMMA 3.9. *Let the second order sufficient optimality conditions set forth in Lemma 2.2 hold at x_0 for $\mathbf{OP}(p_0)$. Then for any $\varepsilon \in \mathcal{X}$,* (3.18)–(3.20) *has a unique solution $(u^\varepsilon, q^\varepsilon, z^\varepsilon)$ and the map*

$$\mathcal{X} \ni \varepsilon \mapsto (u^\varepsilon, q^\varepsilon, z^\varepsilon) \in \mathcal{X} \tag{3.22}$$

is Lipschitz continuous. That is, the optimality system is strongly regular at x_0.

In the next step, we proceed to prove that the solution $(u^\varepsilon, q^\varepsilon, z^\varepsilon)$ of the linearized optimality system (3.18)–(3.20) is directionally differentiable with respect to the perturbation ε. To this end, we need the following assumption.

ASSUMPTION 3.10. *At the reference point (u_0, q_0, z_0), let the following linear operators be compact:*

1. $\mathcal{V} \ni u \mapsto a''_{qu}(u_0, q_0, p_0)(\cdot, u, z_0) \in \mathcal{Q}'$,
2. $\mathcal{Q} \ni q \mapsto a''_{qq}(u_0, q_0, p_0)(\cdot, q, z_0) \in \mathcal{Q}'$,
3. $\mathcal{V} \ni z \mapsto a'_q(u_0, q_0, p_0)(\cdot, z) \in \mathcal{Q}'$.

Remark 3.11. The previous assumption is satisfied for the following important classes of PDE-constrained optimization problems on bounded domains $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$:

(a) If ($\mathbf{OP}(p)$) is a distributed optimal control problem for a semilinear elliptic PDE, e.g.,

$$-\Delta u = f(u) + q \text{ on } \Omega$$

with $\mathcal{V} = H_0^1(\Omega)$ and $\mathcal{Q} = L^2(\Omega)$, then $a''_{qu} = a''_{qq} = 0$ and a'_q is the compact injection of $\mathcal{V}$ into $\mathcal{Q}$.

(b) In the case of Neumann boundary control on $\partial\Omega$, e.g.,

$$-\Delta u = f(u) \text{ on } \Omega \quad \text{ and } \quad \frac{\partial}{\partial n} u = q \text{ on } \partial\Omega,$$

we have $\mathcal{V} = H^1(\Omega)$ and $\mathcal{Q} = L^2(\partial\Omega)$. Again, $a''_{qu} = a''_{qq} = 0$ and a'_q is the compact Dirichlet trace operator from $\mathcal{V}$ to $\mathcal{Q}$.

(c) For bilinear control problems, e.g.,

$$-\Delta u = qu + f \text{ on } \Omega$$

with $\mathcal{V} = H_0^1(\Omega)$, $\mathcal{Q} = L^2(\Omega)$, and an appropriate admissible set $\mathcal{Q}_{ad}$, we have $a''_{qq} = 0$. Moreover, the operators $u \mapsto a''_{qu}(u_0, q_0, p_0)(\cdot, u, z_0) = (uz_0, \cdot)$ and $z \mapsto a'_q(u_0, q_0, z_0) = (u_0 z, \cdot)$ are compact from $\mathcal{V}$ to $\mathcal{Q}'$ since the pointwise product of two functions in $\mathcal{V}$ embeds compactly into $\mathcal{Q}$.

(d) For parabolic equations such as

$$u_t = \Delta u + f(u) + q$$

with solutions in $\mathcal{V} = \{u \in L^2(0, T; H_0^1(\Omega)) : u_t \in L^2(0, T; H^{-1}(\Omega))\}$ we have $a''_{qu} = a''_{qq} = 0$ and a'_q is the compact injection of $\mathcal{V}$ into $\mathcal{Q} = L^2(\Omega \times (0, T))$.

(e) Finally, Assumption 3.10 is always satisfied if the space $\mathcal{Q}$ is finite-dimensional. This includes all cases of parameter identification problems without any additional restrictions on the coupling between the parameters q and the state variable u. For instance, the Arrhenius law leads to reaction-diffusion equations of the form

$$-\Delta u = f(u) + e^{qu} \text{ on } \Omega$$

with unknown Arrhenius parameter $q \in \mathbb{R}$.

For the following theorem, we introduce the admissible set $\widehat{\mathcal{Q}}_{ad}$, defined as

$$\widehat{\mathcal{Q}}_{ad} = \{\hat{q} \in \mathcal{Q} : \widehat{b}_-(x) \le \hat{q}(x) \le \widehat{b}_+(x) \text{ a.e. on } \omega\}$$

with bounds

$$\widehat{b}_-(x) = \begin{cases} 0 & \text{if } \mu_0(x) \neq 0 \text{ or } q_0(x) = b_-(x), \\ -\infty & \text{else,} \end{cases}$$

$$\widehat{b}_+(x) = \begin{cases} 0 & \text{if } \mu_0(x) \neq 0 \text{ or } q_0(x) = b_+(x), \\ +\infty & \text{else.} \end{cases}$$

THEOREM 3.12. *Let the second order sufficient optimality conditions set forth in Lemma* 2.2 *hold at* x_0 *for* $\mathbf{OP}(p_0)$ *in addition to Assumption* 3.10. *Then the map*

(3.22) *is directionally differentiable at* $\varepsilon = 0$ *in every direction* $\delta\varepsilon = (\delta\varepsilon^u, \delta\varepsilon^q, \delta\varepsilon^z) \in \mathcal{X}$. *The directional derivative is given by the unique solution* $(\hat{u}, \hat{q})$ *and adjoint variable* $\hat{z}$ *of the following linear-quadratic optimal control problem, termed* $\mathbf{DQP}(\delta\varepsilon)$*:*

$(\mathbf{DQP}(\delta\varepsilon))$

$$\text{Minimize} \quad \frac{1}{2} \begin{pmatrix} \hat{u} & \hat{q} \end{pmatrix} \begin{pmatrix} \mathcal{L}''_{uu}(x_0, p_0) & \mathcal{L}''_{uq}(x_0, p_0) \\ \mathcal{L}''_{qu}(x_0, p_0) & \mathcal{L}''_{qq}(x_0, p_0) \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{q} \end{pmatrix} + (\hat{u}, \delta\varepsilon^u)_{\mathcal{V}} + (\hat{q}, \delta\varepsilon^q)$$

subject to $\hat{q} \in \widehat{\mathcal{Q}}_{ad}$ *and*

$$a'_u(u_0, q_0, p_0)(\hat{u}, \phi) + a'_q(u_0, q_0, p_0)(\hat{q}, \phi) + (\delta\varepsilon^z, \phi) = 0 \quad \forall \phi \in \mathcal{V}.$$

The first order optimality conditions for this problem read

$$\begin{aligned} &\mathcal{L}''_{uu}(x_0, p_0)(\delta u, \hat{u}) + \mathcal{L}''_{uq}(x_0, p_0)(\delta u, \hat{q}) \\ &\quad + \mathcal{L}''_{uz}(x_0, p_0)(\delta u, \hat{z}) + (\delta\varepsilon^u, \delta u) = 0 \quad \forall \delta u \in \mathcal{V}, \end{aligned} \tag{3.23}$$

$$\begin{aligned} &\mathcal{L}''_{uq}(x_0, p_0)(\hat{u}, \delta q - \hat{q}) + \mathcal{L}''_{qq}(x_0, p_0)(\delta q - \hat{q}, \hat{q}) \\ &\quad + \mathcal{L}''_{qz}(x_0, p_0)(\delta q - \hat{q}, \hat{z}) + (\delta\varepsilon^q, \delta q - \hat{q}) \geq 0 \quad \forall \delta q \in \hat{\mathcal{Q}}_{ad}, \end{aligned} \tag{3.24}$$

$$\begin{aligned} &\mathcal{L}''_{zu}(x_0, p_0)(\delta z, \hat{u}) + \mathcal{L}''_{zq}(x_0, p_0)(\delta z, \hat{q}) \\ &\quad + (\delta\varepsilon^z, \delta z) = 0 \quad \forall \delta z \in \mathcal{V}. \end{aligned} \tag{3.25}$$

Proof. Let $\delta\varepsilon = (\delta\varepsilon^u, \delta\varepsilon^q, \delta\varepsilon^z) \in \mathcal{X}$ be given and let $\{\tau_n\} \subset \mathbb{R}^+$ denote a sequence converging to zero. We denote by $(u_n, q_n, z_n) \in \mathcal{X}$ the unique solution of $\mathbf{LOS}(\varepsilon_n)$, where $\varepsilon_n = \tau_n \delta\varepsilon$. Note that (u_0, q_0, z_0) is the unique solution of $\mathbf{LOS}(0)$ and that $(u_n, q_n, z_n) \to (u_0, q_0, z_0)$ strongly in $\mathcal{X}$. From Lemma 3.9 we infer that

$$\left\| \frac{u_n - u_0}{\tau_n} \right\|_{\mathcal{V}} + \left\| \frac{q_n - q_0}{\tau_n} \right\|_{\mathcal{Q}} + \left\| \frac{z_n - z_0}{\tau_n} \right\|_{\mathcal{V}} \leq L \, \|\delta\varepsilon\|_{\mathcal{X}}.$$

This implies that a subsequence (still denoted by index n) of the difference quotients converges weakly to some limit element $(\hat{u}, \hat{q}, \hat{z}) \in \mathcal{X}$. The proof proceeds with the construction of the *pointwise* limit $\widetilde{q}$ of $(q_n - q_0)/\tau_n$, which is later shown to coincide with $\hat{q}$. It is well known that the variational inequality (3.19) in $\mathbf{LOS}(\varepsilon_n)$ can be equivalently rewritten as

$$q_n(x) = \Pi_{[b_-(x), b_+(x)]}\big(d_n(x)\big) \text{ a.e. on } \omega, \tag{3.26}$$

where $\Pi_{[b_-(x), b_+(x)]}$ is the projection onto the interval $[b_-(x), b_+(x)]$ and

$$\begin{aligned} d_n = \bar{q} + \frac{1}{\alpha} \Big(&a''_{qu}(u_0, q_0, p_0)(\cdot, u_n - u_0, z_0) + a''_{qq}(u_0, q_0, p_0)(\cdot, q_n - q_0, z_0) \\ &+ a'_q(u_0, q_0, p_0)(\cdot, z_n) - \varepsilon_n^q \Big) \in \mathcal{Q}. \end{aligned} \tag{3.27}$$

The linear operators in (3.27) are understood as their Riesz representations in $\mathcal{Q}$. Similarly, we have $q_0(x) = \Pi_{[b_-(x), b_+(x)]}\big(d_0(x)\big)$ a.e. on ω, where

$$d_0 = \bar{q} + \frac{1}{\alpha} a'_q(u_0, q_0, p_0)(\cdot, z_0) \in \mathcal{Q}. \tag{3.28}$$

Note that $d_n \to d_0$ strongly in $\mathcal{Q}$ since the Fréchet derivatives in (3.27) are bounded linear operators. From the compactness properties in Assumption 3.10 we infer that

$$\frac{d_n - d_0}{\tau_n} \to \hat{d} \text{ strongly in } \mathcal{Q}, \quad \text{where}$$

$$\hat{d} = \frac{1}{\alpha} \Big(a''_{qu}(u_0, q_0, p_0)(\cdot, \hat{u}, z_0) + a''_{qq}(u_0, q_0, p_0)(\cdot, \hat{q}, z_0) + a'_q(u_0, q_0, p_0)(\cdot, \hat{z}) - \delta\varepsilon^q \Big).$$

By taking another subsequence, we obtain that $d_n \to d_0$ and $(d_n - d_0)/\tau_n \to \hat{d}$ hold also pointwise a.e. on ω. The construction of the pointwise limit

$$\widetilde{q}(x) = \lim_{n\to\infty} \frac{q_n(x) - q_0(x)}{\tau_n}$$

uses the following partition of ω into five disjoint subsets:

$$\omega = \omega^I \cup \omega_0^+ \cup (\omega^+ \setminus \omega_0^+) \cup \omega_0^- \cup (\omega^- \setminus \omega_0^-), \tag{3.29}$$

where

(3.30a)	$\omega^I = \{x \in \omega : b_-(x) < q_0(x) < b_+(x)\}$	(inactive),
(3.30b)	$\omega_0^+ = \{x \in \omega : \mu_0(x) > 0\}$	(upper strongly active),
(3.30c)	$\omega^+ = \{x \in \omega : q_0(x) = b_+(x)\}$	(upper active),
(3.30d)	$\omega_0^- = \{x \in \omega : \mu_0(x) < 0\}$	(lower strongly active),
(3.30e)	$\omega^- = \{x \in \omega : q_0(x) = b_-(x)\}$	(lower active).

The Lagrange multiplier μ_0 belonging to the constraint $q_0 \in \mathcal{Q}_{ad}$ defined in (3.17) allows the following representation:

$$\mu_0 = \alpha(d_0 - q_0). \tag{3.31}$$

Note that the five sets in (3.29) are guaranteed to be disjoint if $b_-(x) < b_+(x)$ holds a.e. on ω. However, one can easily check that $\widetilde{q}$ is well defined also in the case that the bounds coincide on all or part of ω. We now distinguish five cases according to the sets in (3.29).

Case 1. For almost every x in the inactive subset ω^I, we have $q_0(x) = d_0(x)$ and $q_n(x) = d_n(x)$ for all sufficiently large n. Therefore,

$$\widetilde{q}(x) = \lim_{n\to\infty} \frac{q_n(x) - q_0(x)}{\tau_n} = \hat{d}(x).$$

Case 2. For almost every $x \in \omega_0^+$, $\mu_0(x) > 0$ implies $d_0(x) > q_0(x)$ by (3.31). Therefore, $q_0(x) = b_+(x)$ and $d_n(x) > q_0(x)$ for sufficiently large n. Hence $q_n = b_+(x)$ for these n and

$$\widetilde{q}(x) = \lim_{n\to\infty} \frac{q_n(x) - q_0(x)}{\tau_n} = 0.$$

Case 3. For almost every $x \in \omega^+ \setminus \omega_0^+$, we have $q_0(x) = b_+(x) = d_0(x)$.

(a) If $\hat{d}(x) > 0$, then $d_n(x) > b_+(x)$ for sufficiently large n. Therefore, $q_n(x) = b_+(x)$ for these n and hence $\widetilde{q}(x) = 0$.

(b) If $\hat{d}(x) = 0$, then $(q_n(x) - q_0(x))/\tau_n = \min\{0, d_n(x) - b_+(x)\}/\tau^n$ for sufficiently large n; hence $\widetilde{q}(x) = 0$.

(c) If $\hat{d}(x) < 0$, then $d_n(x) < b_+(x)$ and hence $q_n(x) = d_n(x)$ for sufficiently large n. Therefore, $\widetilde{q}(x) = \hat{d}(x)$ holds.

Case 3 can be summarized as

$$\widetilde{q}(x) = \lim_{n\to\infty} \frac{q_n(x) - q_0(x)}{\tau_n} = \min\{0, \hat{d}(x)\}.$$

Case 4. For almost every $x \in \omega_0^-$, we obtain, similarly to Case 2,

$$\widetilde{q}(x) = \lim_{n\to\infty} \frac{q_n(x) - q_0(x)}{\tau_n} = 0.$$

Case 5. For almost every $x \in \omega^- \setminus \omega_0^-$, we obtain, similarly to Case 3,

$$\widetilde{q}(x) = \lim_{n\to\infty} \frac{q_n(x) - q_0(x)}{\tau_n} = \max\{0, \hat{d}(x)\}.$$

Summarizing all previous cases, we have shown that

$$\widetilde{q}(x) = \Pi_{[\hat{b}_-(x), \hat{b}_+(x)]}(\hat{d}(x)). \tag{3.32}$$

We proceed by showing that

$$\frac{q_n - q_0}{\tau_n} \to \widetilde{q} \text{ strongly in } \mathcal{Q} = L^2(\omega). \tag{3.33}$$

From the Lipschitz continuity of the projection Π, it follows that

$$\begin{aligned}\left\| \frac{q_n - q_0}{\tau_n} - \widetilde{q} \right\|_{\mathcal{Q}} &= \left\| \frac{1}{\tau_n} \left(\Pi_{\mathcal{Q}_{ad}}(d_n) - \Pi_{\mathcal{Q}_{ad}}(d_0) \right) - \Pi_{\widehat{\mathcal{Q}}_{ad}}(\hat{d}) \right\|_{\mathcal{Q}} \\ &\le \left\| \frac{d_n - d_0}{\tau_n} \right\|_{\mathcal{Q}} + \|\hat{d}\|_{\mathcal{Q}} \to 2\|\hat{d}\|_{\mathcal{Q}}.\end{aligned}$$

From Lebesgue's dominated convergence theorem, (3.33) follows. Consequently, we have $\widetilde{q} = \hat{q}$. The projection formula (3.32) is equivalent to the variational inequality (3.24). Using (3.18) and (3.20) for (u_n, q_n, z_n) and for (u_0, q_0, z_0), we infer that the weak limit $(\hat{u}, \hat{q}, \hat{z})$ satisfies (3.23) and (3.25). It is readily checked that (3.23)–(3.25) are the first order necessary conditions for ($\mathbf{DQP}(\delta\varepsilon)$). In view of the second order sufficient optimality conditions (Lemma 2.2), ($\mathbf{DQP}(\delta\varepsilon)$) is strictly convex and thus it has a unique solution. In view of Assumption 3.2 and (3.25), we obtain

$$\left\| \frac{u_n - u_0}{\tau_n} - \hat{u} \right\|_{\mathcal{V}} \le C \left\| \frac{q_n - q_0}{\tau_n} - \hat{q} \right\|_{\mathcal{Q}},$$

where C is independent of n. Hence $\hat{u}$ is also the strong limit of the difference quotient in $\mathcal{V}$. The same argument holds for $\hat{z}$. Our whole argument remains valid if in the beginning, we start with an arbitrary subsequence of $\{\tau_n\}$. Since the limit $(\hat{u}, \hat{q}, \hat{z})$ is always the same, the convergence extends to the whole sequence. ∎

From the previous theorem we derive the following important corollary. The proof follows from a direct application of the implicit function theorem for generalized equations; see [6, Theorem 2.4].

COROLLARY 3.13. *Under the conditions of the previous theorem, there exist neighborhoods $\mathcal{N}(p_0) \subset \mathcal{P}$ of p_0 and $\mathcal{N}(x_0) \subset \mathcal{X}$ of x_0 and a directionally differentiable function $(U, Q, Z) : \mathcal{N}(p_0) \to \mathcal{N}(x_0)$ with the following properties:*

(a) *For every $p \in \mathcal{N}(p_0)$, $(U(p), Q(p), Z(p))$ is the unique solution to the system (2.5)–(2.7) in the neighborhood $\mathcal{N}(x_0)$.*

(b) *$(U(p_0), Q(p_0), Z(p_0)) = (u_0, q_0, z_0)$ holds.*

(c) *The directional derivative of* (U, Q, Z) *at* p_0 *in the direction* $\delta p \in \mathcal{P}$ *is given by the derivative of* $\varepsilon \mapsto (u^\varepsilon, q^\varepsilon, z^\varepsilon)$ *at* $\varepsilon = 0$ *in the direction*

$$\delta\varepsilon = \begin{pmatrix} \mathcal{L}''_{up}(x_0, p_0)(\cdot, \delta p) \\ \mathcal{L}''_{qp}(x_0, p_0)(\cdot, \delta p) \\ \mathcal{L}''_{zp}(x_0, p_0)(\cdot, \delta p) \end{pmatrix}, \tag{3.34}$$

i.e., by the solution and adjoint $(\hat{u}, \hat{q}, \hat{z})$ *of* $\mathbf{DQP}(\delta\varepsilon)$.

We remark that computing the sensitivity derivative of (U, Q, Z) for a given direction δp amounts to solving the linear-quadratic optimal control problem $\mathbf{DQP}(\delta\varepsilon)$ for $\delta\varepsilon$ given by (3.34). Note that this problem, like the original one $\mathbf{OP}(p_0)$, is subject to pointwise inequality constraints for the control variable. Due to the structure of the admissible set $\widehat{\mathcal{Q}}_{ad}$, the directional derivative of (U, Q, Z) is in general not a linear function of the direction δp, but only positively homogeneous. Note, however, that if the admissible set $\widehat{\mathcal{Q}}_{ad}$ is a linear space (which follows from a condition known as strict complementarity; see below), then the directional derivative becomes a linear function of the direction (i.e., it is the Gateaux differential).

DEFINITION 3.14 (strict complementarity). *Strict complementarity is said to hold at* (x_0, p_0) *if*

$$\Big\{ x \in \omega \,:\, q_0(x) \in \{b_-(x), b_+(x)\} \text{ and } \mu_0(x) = 0 \Big\}$$

is a set of measure zero.

A consequence of the strict complementarity condition is that the sensitivity derivatives are characterized by a linear system of equations set forth in the following lemma. We recall that $\widetilde{B}$ was defined in (2.19) and that R_I denotes the multiplication of a function in $L^2(\omega)$ with the characteristic function of the inactive set $\omega^I = \{x \in \omega : b_-(x) < q_0(x) < b_+(x)\}$; see section 2.

LEMMA 3.15. *Under the conditions of Theorem* 3.12 *and if strict complementarity holds at* (x_0, p_0)*, then the directional derivative of* (U, Q, Z) *is characterized by the following* linear *system of equations:*

$$\widetilde{B}(x_0, p_0) \begin{pmatrix} U'(p_0)(\delta p) \\ Q'(p_0)(\delta p) \\ Z'(p_0)(\delta p) \end{pmatrix} = - \begin{pmatrix} \mathcal{L}''_{up}(x_0, p_0)(\cdot, \delta p) \\ R_I \mathcal{L}''_{qp}(x_0, p_0)(\cdot, \delta p) \\ \mathcal{L}''_{zp}(x_0, p_0)(\cdot, \delta p) \end{pmatrix}. \tag{3.35}$$

Moreover, the operator $\widetilde{B}(x_0, p_0) : \mathcal{X} \to \mathcal{X}'$ *is boundedly invertible.*

Proof. By virtue of the strict complementarity property, the admissible set $\widehat{\mathcal{Q}}_{ad}$ defined in Theorem 3.12 becomes

$$\widehat{\mathcal{Q}}_{ad} = \Big\{ \hat{q} \in \mathcal{Q} \,:\, \hat{q}(x) = 0, \text{ where } q_0(x) \in \{b_-(x), b_+(x)\} \Big\}.$$

Consequently, the variational inequality (3.24) simplifies to the following equation for $Q'(p_0)(\delta p) \in \widehat{\mathcal{Q}}_{ad}$:

$$\begin{aligned} &\mathcal{L}''_{qu}(x_0, p_0)(\delta q, U'(p_0)(\delta p)) + \mathcal{L}''_{qq}(x_0, p_0)(\delta q, Q'(p_0)(\delta p)) \\ &\qquad + \mathcal{L}''_{qz}(x_0, p_0)(\delta q, Z'(p_0)(\delta p)) = -\mathcal{L}''_{qp}(x_0, p_0)(\delta q, \delta p) \quad \forall \delta q \in \widehat{\mathcal{Q}}_{ad}, \end{aligned}$$

which is equivalent to the middle equation in (3.35). The first and third equation in (3.35) coincide with (3.23) and (3.25), which proves the first claim. From Theorem 3.12 we conclude that $\widetilde{B}(x_0, p_0)$ is bijective. Since it a continuous linear operator from $\mathcal{X} \to \mathcal{X}'$, so is its inverse. ∎

We are now in the position to recall the first and second order sensitivity derivatives of the reduced cost functional $j(p)$; see again [18]. Note that we do not make use of strict complementarity in the following proposition.

PROPOSITION 3.16. *Under the conditions of Theorem* 3.12*, the reduced cost functional*

$$j(p) = J(U(p), p) + \frac{\alpha}{2} \|Q(p) - \overline{q}\|^2_{\mathcal{Q}}$$

is continuously differentiable in $\mathcal{N}(p_0)$*. The derivative at* p_0 *in the direction* $\delta p \in \mathcal{P}$ *is given by*

$$j'(p)(\delta p) = \mathcal{L}'_p(x_0, p_0)(\delta p). \tag{3.36}$$

Additionally, the second order directional derivatives of the reduced cost function j *exist and are given by the following formula:*

$$\begin{aligned} j''(p_0)(\delta p, \widehat{\delta p}) = \mathcal{L}''_{up}(x_0, p_0)(U'(p_0)(\delta p), \widehat{\delta p}) + \mathcal{L}''_{qp}(x_0, p_0)(Q'(p_0)(\delta p), \widehat{\delta p}) \\ + \mathcal{L}''_{zp}(x_0, p_0)(Z'(p_0)(\delta p), \widehat{\delta p}) + \mathcal{L}''_{pp}(x_0, p_0)(\delta p, \widehat{\delta p}). \end{aligned} \tag{3.37}$$

Proof. As in the unconstrained case there holds

$$\begin{aligned} j'(p_0)(\delta p) = \mathcal{L}'_u(x_0, p_0)(U'(p_0)(\delta p)) + \mathcal{L}'_q(x_0, p_0)(Q'(p_0)(\delta p)) \\ + \mathcal{L}'_z(x_0, p_0)(Z'(p_0)(\delta p)) + \mathcal{L}'_p(x_0, p_0)(\delta p), \end{aligned}$$

and the terms $\mathcal{L}'_u$ and $\mathcal{L}'_z$ vanish. Moreover,

$$\mathcal{L}'_q(x_0, p_0)(Q'(p_0)(\delta p)) = -(\mu_0, Q'(p_0)(\delta p)) = 0$$

since $Q'(p_0)(\delta p)$ is zero on the strongly active set and μ_0 vanishes on its complement. The formula for the second order derivative follows as in Proposition 3.5 by total directional differentiation of the first order formula. ∎

Remark 3.17. We note that the expressions for the first and second order derivatives in Proposition 3.16 are the same as in the unconstrained case; see Proposition 3.5.

We now turn to our main result in the control-constrained case, concerning the differentiability and efficient evaluation of the sensitivity derivatives for the reduced quantity of interest (1.4). We recall that in the unconstrained case, we have made use of a duality argument for the efficient computation of the first and second order derivatives; see section 3.2. However, in the presence of control constraints, this technique seems to be applicable only in the case of strict complementarity since otherwise, the derivatives $(U'(p_0)(\delta p), \xi'(p_0)(\delta p), Z'(p_0)(\delta p))$ do not depend linearly on the direction δp. In analogy to (3.11) and (3.35), we define the dual quantities $(\widetilde{v}, \widetilde{r}, \widetilde{y}) \in \mathcal{X}$ by

$$\widetilde{B}(x_0, p_0) \begin{pmatrix} \widetilde{v} \\ \widetilde{r} \\ \widetilde{y} \end{pmatrix} = - \begin{pmatrix} I'_u(q_0, u_0, p_0) \\ R_I I'_q(q_0, u_0, p_0) \\ 0 \end{pmatrix}. \tag{3.38}$$

THEOREM 3.18. *Under the conditions of Theorem* 3.12*, the reduced quantity of interest* $i(p)$ *is directionally differentiable at the reference parameter* p_0*. If, in addition, strict complementarity holds at* (x_0, p_0)*, then the first order directional derivative* *at p_0 in the direction $\delta p \in \mathcal{P}$ is given by*

$$(3.39)\quad i'(p_0)(\delta p) = \mathcal{L}''_{up}(x_0,p_0)(\widetilde{v},\delta p) + \mathcal{L}''_{qp}(x_0,p_0)(R_I\widetilde{r},\delta p) + \mathcal{L}''_{zp}(x_0,p_0)(\widetilde{y},\delta p) + I'_p(u_0,q_0,p_0)(\delta p).$$

Proof. The proof is carried out similarly to the proof of Theorem 3.6 using Lemma 3.15. □

Our next goal is to consider second order derivatives of the reduced quantity of interest. In order to apply the approach used in the unconstrained case, we rely on the existence of second order directional derivatives of (U,Q,Z) at p_0. However, these second order derivatives do not exist without further assumptions, as seen from the following simple consideration: Suppose that near a given reference parameter $p_0 = 0$, the local optimal control is given by $Q(p)(x) = \max\{0, x+p\} \in L^2(\omega)$ for $x \in \omega = (-1,1)$ and $p \in \mathbb{R}$. (An appropriate optimal control problem $(\mathbf{OP}(p))$ can be easily constructed.) Then $Q'(p)(x) = H(x+p)$ (the Heaviside function), which is not directionally differentiable with respect to p and values in $L^2(\omega)$. Note that the point $x = -p$ of discontinuity marks the boundary between the active and inactive sets of $(\mathbf{OP}(p))$. Hence we conclude that the reason for the nonexistence of the second order directional derivatives of Q lies in the change of the active set with p.

The preceding argument leads to the following assumption.

ASSUMPTION 3.19. *There exists a neighborhood $\mathcal{N}(p_0) \subset \mathcal{P}$ of the reference parameter p_0 such that for every $p \in \mathcal{N}(p_0)$, strict complementarity holds at the solution $(U(p),Q(p),Z(p))$, and the active sets coincide with those of (u_0,q_0,z_0).*

Remark 3.20. The previous assumption seems difficult to satisfy in the general case. However, if the control variable is finite-dimensional and strict complementarity is assumed at the reference solution (u_0,q_0,z_0), then Assumption 3.19 is satisfied since the Lagrange multiplier $\mu(p) = -\mathcal{L}'_q(U(p),Q(p),Z(p),p)$ is continuous with respect to p and has values in $\mathbb{R}^n$.

We now proceed to our main result concerning second order derivatives of the reduced quantity of interest. In the theorem below, we use again () to denote evaluation at the point (x_0,p_0).

THEOREM 3.21. *Under the conditions of Theorem* 3.12 *and Assumption* 3.19, *the reduced quantity of interest $i(p)$ is twice directionally differentiable at p_0. The second order directional derivatives in the directions of δp and $\widehat{\delta p}$ are given by*

$$(3.40)\quad i''(p_0)(\delta p,\widehat{\delta p}) = \langle \widetilde{v},\eta\rangle_{\mathcal{V}\times\mathcal{V}'} + \langle \widetilde{r},\kappa\rangle_{\mathcal{Q}\times\mathcal{Q}'} + \langle \widetilde{y},\sigma\rangle_{\mathcal{V}\times\mathcal{V}'}$$
$$+ \begin{pmatrix} U'(p_0)(\delta p) \\ Q'(p_0)(\delta p) \\ \delta p \end{pmatrix}^{\top} \begin{pmatrix} I''_{uu}(q_0,u_0,p_0) & I''_{uq}(q_0,u_0,p_0) & I''_{up}(q_0,u_0,p_0) \\ I''_{qu}(q_0,u_0,p_0) & I''_{qq}(q_0,u_0,p_0) & I''_{qp}(q_0,u_0,p_0) \\ I''_{pu}(q_0,u_0,p_0) & I''_{pq}(q_0,u_0,p_0) & I''_{pp}(q_0,u_0,p_0) \end{pmatrix} \begin{pmatrix} U'(p_0)(\widehat{\delta p}) \\ Q'(p_0)(\widehat{\delta p}) \\ \widehat{\delta p} \end{pmatrix}.$$

Here, $(\eta,\kappa,\sigma) \in \mathcal{V}' \times \mathcal{Q}' \times \mathcal{V}'$ is given, as in the unconstrained case, by

$$(3.41)\quad \begin{pmatrix} \eta \\ \kappa \\ \sigma \end{pmatrix} = \begin{pmatrix} \mathcal{L}'''_{upp}()(\cdot,\delta p,\widehat{\delta p}) \\ \mathcal{L}'''_{qpp}()(\cdot,\delta p,\widehat{\delta p}) \\ \mathcal{L}'''_{zpp}()(\cdot,\delta p,\widehat{\delta p}) \end{pmatrix}$$

Evaluation of sensitivity derivatives

1. Evaluate $j'(p_0)\,\delta p$ according to (3.36)
2. Compute the sensitivities $U'(p_0)\,\delta p$, $Q'(p_0)\,\delta p$, and $Z'(p_0)\,\delta p$ from (3.35)
3. Evaluate $j''(p_0)(\delta p, \widehat{\delta p})$ according to (3.37)
4. Compute the dual quantities $(\widetilde{v}, \widetilde{r}, \widetilde{y})$ from (3.38)
5. Evaluate $i'(p_0)\,\delta p$ according to (3.39)
6. Compute the sensitivities $U'(p_0)\,\widehat{\delta p}$, $Q'(p_0)\,\widehat{\delta p}$, and $Z'(p_0)\,\widehat{\delta p}$ from (3.35)
7. Compute the auxiliary quantities (η, κ, σ) from (**??**)
8. Evaluate $i''(p_0)(\delta p, \widehat{\delta p})$ according to (3.40).

ALGORITHM 3.1. *Evaluating the first and second order derivatives of the reduced cost function j and the reduced quantity of interest i.*

$$
+ \begin{pmatrix} \mathcal{L}'''_{upu}()(\cdot, \delta p, U'(p_0)(\widehat{\delta p})) + \mathcal{L}'''_{upq}()(\cdot, \delta p, Q'(p_0)(\widehat{\delta p})) + \mathcal{L}'''_{upz}()(\cdot, \delta p, Z'(p_0)(\widehat{\delta p})) \\ \mathcal{L}'''_{qpu}()(\cdot, \delta p, U'(p_0)(\widehat{\delta p})) + \mathcal{L}'''_{qpq}()(\cdot, \delta p, Q'(p_0)(\widehat{\delta p})) + \mathcal{L}'''_{qpz}()(\cdot, \delta p, Z'(p_0)(\widehat{\delta p})) \\ \mathcal{L}'''_{zpu}()(\cdot, \delta p, U'(p_0)(\widehat{\delta p})) + \mathcal{L}'''_{zpq}()(\cdot, \delta p, Q'(p_0)(\widehat{\delta p})) \end{pmatrix}
$$

$$
+ \Big(\widetilde{B}'_u()(U'(p_0)(\widehat{\delta p})) + \widetilde{B}'_q()(Q'(p_0)(\widehat{\delta p})) + \widetilde{B}'_z()(Z'(p_0)(\widehat{\delta p})) + \widetilde{B}'_p()(\widehat{\delta p}) \Big) \begin{pmatrix} U'(p_0)(\delta p) \\ Q'(p_0)(\delta p) \\ Z'(p_0)(\delta p) \end{pmatrix}.
$$

Proof. The proof uses the same argument as the proof of Theorem 3.6. Note that in view of Assumption 3.19, $\widetilde{B}(U(p), Q(p), Z(p), p)$ is totally directionally differentiable with respect to p at p_0. In the direction $\widehat{\delta p}$, the derivative is

$$
\widetilde{B}'_u()(U'(p_0)(\widehat{\delta p})) + \widetilde{B}'_q()(Q'(p_0)(\widehat{\delta p})) + \widetilde{B}'_z()(Z'(p_0)(\widehat{\delta p})) + \widetilde{B}'_p()(\widehat{\delta p}).
$$

Due to the constant active sets, these partial derivatives have the following form:

$$
\widetilde{B}'_u() = \begin{pmatrix} id & & \\ & R_I & \\ & & id \end{pmatrix} B'_u(x_0, p_0) \begin{pmatrix} id & & \\ & R_I & \\ & & id \end{pmatrix}.
$$

In view of the bounded invertibility of $\widetilde{B}(x_0, p_0)$ (see Lemma 3.15), the second order partial derivatives of (U, Q, Z) at p_0 exist by the implicit function theorem. They satisfy the analogue of (3.16). ∎

We conclude this section by stating Algorithm 3.1 which collects the necessary steps to evaluate the first and second order sensitivity derivatives $j'(p_0)\,\delta p$ and $j''(p_0)(\delta p, \widehat{\delta p})$ as well as $i'(p_0)\,\delta p$ and $i''(p_0)(\delta p, \widehat{\delta p})$ for given $\delta p, \widehat{\delta p} \in \mathcal{P}$. We suppose that the original optimization problem ($\mathbf{OP}(p)$) has been solved, e.g., by the primal-dual active set approach in section 2, for the nominal parameter p_0. We denote by

$\mathcal{A}_\pm$ and $\mathcal{I}$ the active and inactive sets belonging to the nominal solution (u_0, q_0) and adjoint state z_0. For the definition of $\widetilde{B}(x_0, p_0)$ appearing in (3.35) and (3.38), we refer the reader to (2.19).

4. Numerical examples. In this section we illustrate our approach using two examples from different areas. The first example is concerned with a parameter identification problem for the stationary Navier–Stokes system. No inequality constraints are present in this problem, and first and second order derivatives of the quantity of interest are obtained. In the second example, we consider a control-constrained optimal control problem for an instationary reaction-diffusion system subject to an infinite-dimensional parameter, which demonstrates the full potential of our approach.

4.1. Example 1. In this section we illustrate our approach using as an example a parameter identification flow problem without inequality constraints. We consider the configuration sketched in Figure 4.1.

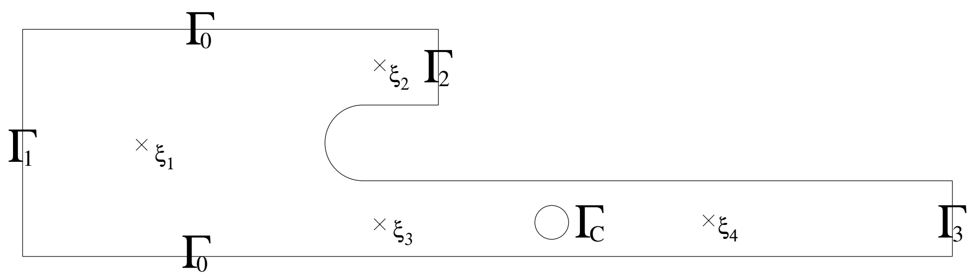

FIG. 4.1. *Configuration of the system of pipes with measurement points.*

The (stationary) flow in this system of pipes around the cylinder Γ_C is described by incompressible Navier–Stokes equations, with unknown viscosity q:

$$\begin{array}{rcll} -q\Delta v + v\cdot\nabla v + \nabla p &=& f & \text{in } \Omega,\\ \nabla\cdot v &=& 0 & \text{in } \Omega,\\ v &=& 0 & \text{on } \Gamma_0\cup\Gamma_C,\\ v &=& v_{in} & \text{on } \Gamma_1,\\ q\frac{\partial v}{\partial n} - pn &=& \pi n & \text{on } \Gamma_2,\\ q\frac{\partial v}{\partial n} - pn &=& 0 & \text{on } \Gamma_3. \end{array} \tag{4.1}$$

Here, the state variable $u = (v, p)$ consists of the velocity $v = (v^1, v^2) \in H^1(\Omega)^2$ and the pressure $p \in L^2(\Omega)$. The inflow Dirichlet boundary condition on Γ_1 is given by a parabolic inflow v_{in}. The outflow boundary conditions of the Neumann type are prescribed on Γ_2 and Γ_3 involving the perturbation parameter $\pi \in P = \mathbb{R}$. (Unlike in previous sections, we denote the perturbation parameter by π to avoid the confusion with the pressure p.) Physically, the perturbation parameter π describes the pressure difference between Γ_2 and Γ_3; see [11] for a detailed discussion of this type of outflow boundary condition. The reference parameter chosen is $\pi_0 = 0.029$.

The aim is to estimate the unknown viscosity $q \in Q = \mathbb{R}$ using the measurements of the velocity in four given points; see Figure 4.1. By the least squares approach, this results in the following parameter identification problem:

$$\text{Minimize} \sum_{i=1}^{4}\sum_{j=1}^{2}(v^j(\xi_i) - \bar{v}_i^j)^2 + \alpha q^2, \quad \text{subject to (4.1).}$$

Here, $\bar{v}_i^j$ are the measured values of the components of the velocity at the point ξ_i and α is a regularization parameter. For a priori error analysis for finite element discretization of parameter identification problems with pointwise measurements we refer the reader to [20].

The sensitivity analysis of previous sections allows us to study the dependence on the perturbation parameter π. To illustrate this, we define two functionals describing the possible quantities of interest:

$$I_1(u,q) = q, \qquad I_2(u,q) = c_d(u),$$

where $c_d(u)$ is the drag coefficient on the cylinder Γ_C defined as

$$c_d(u) = c_0 \int_{\Gamma_C} n \cdot \sigma \cdot d \, ds, \tag{4.2}$$

with a chosen direction $d = (1,0)$, a given constant c_0, and the stress tensor σ given by

$$\sigma = \frac{\nu}{2}(\nabla v + (\nabla v)^T) - pI.$$

For the discretization of the state equation we use conforming finite elements on a shape-regular quadrilateral mesh $\mathcal{T}_h$. The trial and test spaces consist of cellwise bilinear shape-functions for both pressure and velocities. We add further terms to the finite element formulation in order to obtain a stable formulation with respect to both the pressure-velocity coupling and convection dominated flow. This type of stabilization technique is based on local projections of the pressure first introduced in [1]. The resulting parameter identification problem is solved by Newton's method on the parameter space as described in [3] which is known to be mesh-independent. The nonlinear state equation is likewise solved by Newton's method, whereas the linear subproblems are computed using a standard multigrid algorithm. With these ingredients, the total numerical cost for the solution of this parameter identification problem on a given mesh behaves like $\mathcal{O}(N)$, where N is the number of degrees of freedom (dofs) for the state equation.

For the reduced quantities of interest $i_1(\pi)$ and $i_2(\pi)$ we compute the first and second derivatives using the representations from Theorem 3.6. In Tables 4.1 and 4.2 we collect the values of these derivatives for a sequence of uniformly refined meshes.

In order to verify the computed sensitivity derivatives, we make a comparison with the derivatives computed by the second order difference quotients. To this end we choose $\varepsilon = 10^{-4}$ and compute

$$di_l = \frac{i_l(\pi_0+\varepsilon) - i_l(\pi_0-\varepsilon)}{2\,\varepsilon}, \quad ddi_l = \frac{i_l(\pi_0+\varepsilon) - 2i_l(\pi_0) + i_l(\pi_0-\varepsilon)}{\varepsilon^2}$$

by solving the optimization problem additionally for $\pi = \pi_0 - \varepsilon$ and $\pi = \pi_0 + \varepsilon$. The results are shown in Table 4.3.

Remark 4.1. The relative errors in Table 4.3 are of the order of the estimated finite difference truncation error. We therefore consider the correctness of our method to have been verified to within the accuracy of this test. The same holds for Example 2 and Table 4.4 below.

TABLE 4.1
The values of $i_1(\pi)$ and its derivatives on a sequence of uniformly refined meshes.

Cells	dofs	$i_1(\pi)$	$i_1'(\pi)$	$i_1''(\pi)$
60	270	1.0176e–2	–3.9712e–1	1.4065e–1
240	900	1.0086e–2	–3.9386e–1	–3.2022e–1
960	3240	1.0013e–2	–3.9613e–1	–8.5278e–1
3840	12240	1.0003e–2	–3.9940e–1	–1.0168e–0
15360	47520	1.0000e–2	–4.0030e–1	–1.0601e–0

TABLE 4.2
The values of $i_2(\pi)$ and its derivatives on a sequence of uniformly refined meshes.

Cells	dofs	$i_2(\pi)$	$i_2'(\pi)$	$i_2''(\pi)$
60	270	3.9511e–1	–13.4846	9.89988
240	900	3.9106e–1	–13.8759	–4.09824
960	3240	3.9293e–1	–13.8151	16.5239
3840	12240	3.9242e–1	–13.7357	19.3916
15360	47520	3.9235e–1	–13.7144	19.9385

TABLE 4.3
Comparison of the computed derivatives of i_l $(l = 1, 2)$ with difference quotients, on the finest grid.

l	i_l'	di_l	$\frac{\mathbf{di_l - i_l'}}{\mathbf{i_l'}}$	i_l''	ddi_l	$\frac{\mathbf{ddi_l - i_l''}}{\mathbf{i_l''}}$
1	–0.399403	–0.399404	**2.5e–6**	–1.01676	–1.01678	**2.0e–5**
2	–13.73574	–13.73573	**–7.3e–7**	19.3916	19.3917	**5.2e–6**

4.2. Example 2. The second example concerns a control-constrained optimal control problem for an instationary reaction-diffusion model in three spatial dimensions. As the problem setup was described in detail in [9], we will be brief here. The reaction-diffusion state equation is given by

$$(c_1)_t = D_1 \Delta c_1 - k_1 c_1 c_2 \quad \text{in } \Omega \times (0, T), \tag{4.3a}$$

$$(c_2)_t = D_2 \Delta c_2 - k_2 c_1 c_2 \quad \text{in } \Omega \times (0, T), \tag{4.3b}$$

where c_i denotes the concentration of the ith substance, and hence $u = (c_1, c_2)$ is the state variable. Ω is a domain in $\mathbb{R}^3$, in this case an annular cylinder (Figure 4.2), and T is the given final time. The control q enters through the inhomogeneous boundary conditions

$$D_1 \frac{\partial c_1}{\partial n} = 0 \quad \text{in } \partial\Omega \times (0, T), \tag{4.4a}$$

$$D_2 \frac{\partial c_2}{\partial n} = q(t)\, \alpha(t, x) \quad \text{in } \partial\Omega_c \times (0, T), \tag{4.4b}$$

$$D_2 \frac{\partial c_2}{\partial n} = 0 \quad \text{in } (\partial\Omega \setminus \partial\Omega_c) \times (0, T), \tag{4.4c}$$

and α is a given shape function on the boundary, modeling a revolving nozzle on the control surface $\partial\Omega_c$, the upper annulus. Initial conditions

$$c_1(0, x) = c_{10}(x) \quad \text{in } \Omega, \tag{4.5a}$$

$$c_2(0, x) = c_{20}(x) \quad \text{in } \Omega \tag{4.5b}$$

are also given. The objective to be minimized is

$$J(c_1, c_2, q) = \frac{1}{2} \int_\Omega \alpha_1 \left| c_1(T, \cdot) - c_{1T} \right|^2 + \alpha_2 \left| c_2(T, \cdot) - c_{2T} \right|^2 \mathrm{d}x + \frac{\gamma}{2} \int_0^T |q - q_d|^2 \, \mathrm{d}t + \frac{1}{\varepsilon} \max \left\{ 0, \int_0^T q(t) \, \mathrm{d}t - q_c \right\}^3 ;$$

i.e., it contains contributions from deviation of the concentrations at the given terminal time T from the desired ones c_{iT}, plus control cost and a term stemming from a penalization of excessive total control action. We consider here the particular setup described in [9, Example 1], where substance c_1 is to be driven to zero at time T (i.e., we have $\alpha_1 = 1$ and $\alpha_2 = 0$) from given uniform initial state $c_{10} \equiv 1$. This problem features a number of parameters, and differentiability of optimal solutions with respect to these parameters was proved in [10]; hence, we may apply the results of section 3. The nominal as well as the sensitivity and dual problems were solved using a primal-dual active set strategy; see [9, 15]. The nominal control is depicted in Figure 4.2. One clearly sees that the upper and lower bounds with values 5 and 1, respectively, are active in the beginning and end of the time interval. All computations were carried out using piecewise linear finite elements on a tetrahedral grid with roughly 3300 vertices, 13200 tetrahedra, and 100 time steps.

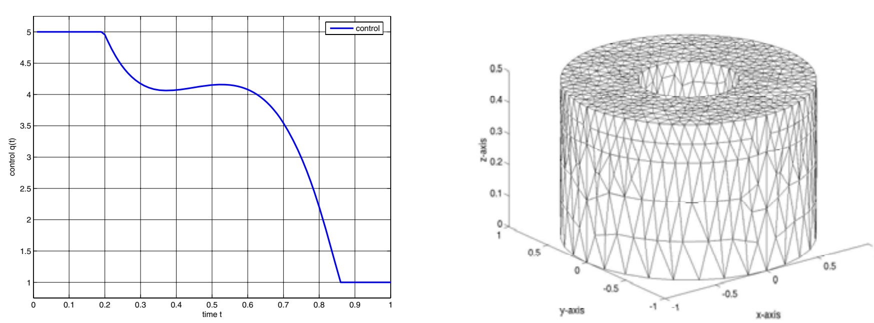

Fig. 4.2. *Optimal (unperturbed) control q (left) and computational domain (right).*

Since the control variable is infinite-dimensional and control constraints are active in the solution, the active sets will in general change even under arbitrarily small perturbations; hence second order derivatives of the reduced quantity of interest $i(p)$ may not exist (see the discussion before Asumption 3.19).

We choose as quantity of interest the total amount of control action

$$I(u, q) = \int_0^T q(t) \, \mathrm{d}t.$$

In contrast to the previous example, we consider now an infinite-dimensional parameter $p = c_{10}$, the initial value of the first substance. After discretization on the given spatial grid, the parameter space has a dimension $\dim P \approx 3300$. A look at Table 3.1 now reveals the potential of our method: The direct evaluation of the derivative $i'(p_0)$ would have required the solution of 3300 auxiliary linear-quadratic problems, an unbearable effort. By our dual approach, however, we need to solve only one additional such problem (3.38) for the dual quantities. The derivative $i'(p_0)$ is shown

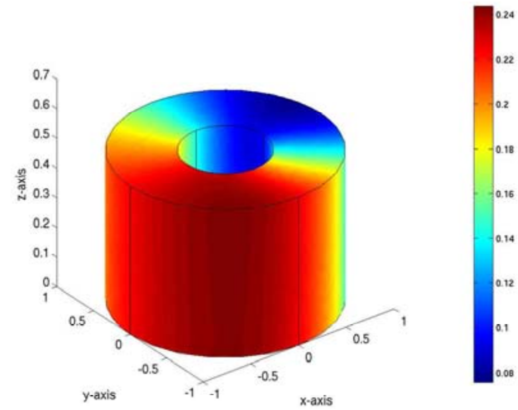

FIG. 4.3. *Gradient of the quantity of interest.*

in Figure 4.3 as a distributed function on Ω. In the unperturbed setup, the desired terminal state $c_1(T)$ is everywhere above the desired state $c_{1T} \equiv 0$. By increasing the value of the initial state c_{10}, the desired terminal state is even more difficult to reach, which leads to an increased control effort and thus an increased value of the quantity of interest. This is reflected by the sign of the function in Figure 4.3, which is everywhere positive. Moreover, one can identify the region of Ω where perturbations in the initial state have the greatest impact on the value of the quantity of interest.

In order to check the derivative, we use again a comparison with a difference quotient in the given direction of $\delta p \equiv 1$. Table 4.4 shows the analogue of Table 4.3 with $\varepsilon = 10^{-2}$ for this example.

TABLE 4.4
Comparison of the computed derivatives of i with difference quotients.

i'	di	$\frac{\mathbf{di-i'}}{\mathbf{i'}}$
0.222770	0.222463	**−1.4e–3**

5. Conclusion. In this paper, we considered PDE-constrained optimization problems with inequality constraints, which depend on a perturbation parameter p. The differentiability of optimal solutions with respect to this parameter is shown in Theorem 3.12. This result complements previous findings in [7, 18] and makes precise the compactness assumptions needed for the proof.

We obtained sensitivity results for a quantity of interest which depends on the optimal solution and is different from the cost functional. The main contribution of this paper is to devise an efficient algorithm to evaluate these sensitivity derivatives. Using a duality technique, we showed that the numerical cost of evaluating the gradient or the Hessian of the quantity of interest is only marginally higher than the evaluation of the gradient or the Hessian of the cost functional. The small additional effort is spent for the solution of one additional linear-quadratic optimization problem for a suitable dual quantity. A comparison with a direct approach for the evaluation of the gradient and the Hessian revealed the tremendous savings of the dual approach especially in the case of a high-dimenensional parameter space. Two numerical examples confirmed the correctness of our derivative formulae and illustrated the applicability of our results.

REFERENCES

[1] R. Becker and M. Braack, *A finite element pressure gradient stabilization for the Stokes equations based on local projections*, Calcolo, 38 (2001), pp. 173–199.

[2] R. Becker, D. Meidner, and B. Vexler, *Efficient numerical solution of parabolic optimization problems by finite element methods*, Optim. Methods Softw., in revision.

[3] R. Becker and B. Vexler, *Mesh refinement and numerical sensitivity analysis for parameter calibration of partial differential equations*, J. Comput. Phys., 206 (2005), pp. 95–110.

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

[5] J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, New York, 1969.

[6] A. Dontchev, *Implicit function theorems for generalized equations*, Math. Programming, 70 (1995), pp. 91–106.

[7] R. Griesse, *Parametric sensitivity analysis in optimal control of a reaction-diffusion system—part* I: *Solution differentiability*, Numer. Funct. Anal. Optim., 25 (2004), pp. 93–117.

[8] R. Griesse, *Parametric sensitivity analysis in optimal control of a reaction-diffusion system—part* II: *Practical methods and examples*, Optim. Methods Softw., 19 (2004), pp. 217–242.

[9] R. Griesse and S. Volkwein, *A primal-dual active set strategy for optimal boundary control of a nonlinear reaction-diffusion system*, SIAM J. Control Optim., 44 (2005), pp. 467–494.

[10] R. Griesse and S. Volkwein, *Parametric sensitivity analysis for optimal boundary control of a* 3*D reaction-diffusion system*, in Large-Scale Nonlinear Optimization, G. D. Pillo and M. Roma, eds., Nonconvex Optim. Appl. 83, Springer-Verlag, New York, 2006, pp. 127–149.

[11] J. Heywood, R. Rannacher, and S. Turek, *Artificial boundaries and flux and pressure conditions for the incompressible Navier–Stokes equations*, Internat. J. Numer. Methods Fluids, 22 (1996), pp. 325–352.

[12] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888.

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

[14] K. Ito and K. Kunisch, *Augmented Lagrangian-SQP-methods in Hilbert spaces and application to control in the coefficients problem*, SIAM J. Optim., 6 (1996), pp. 96–125.

[15] K. Ito and K. Kunisch, *The primal-dual active set method for nonlinear optimal control problems with bilateral constraints*, SIAM J. Control Optim., 43 (2004), pp. 357–376.

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

[17] F.-S. Kupfer, *An infinite-dimensional convergence theory for reduced SQP methods in Hilbert space*, SIAM J. Optim., 6 (1996), pp. 126–163.

[18] K. Malanowski, *Sensitivity analysis for parametric optimal control of semilinear parabolic equations*, J. Convex Anal., 9 (2002), pp. 543–561.

[19] H. Maurer and J. Zowe, *First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems*, Math. Programming, 16 (1979), pp. 98–110.

[20] 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.

[21] F. Tröltzsch, *Lipschitz stability of solutions of linear-quadratic parabolic control problems with respect to perturbations*, Dyn. Contin. Discrete Impuls. Syst. Ser. A Math. Anal., 7 (2000), pp. 289–306.