SENSITIVITY ANALYSIS AND THE ADJOINT UPDATE STRATEGY FOR OPTIMAL CONTROL PROBLEMS WITH MIXED CONTROL-STATE CONSTRAINTS

ROLAND GRIESSE AND DANIEL WACHSMUTH

Abstract. In this article, an optimal control problem subject to a semilinear elliptic equation and mixed control-state constraints is investigated. The problem data depends on certain parameters. Under an assumption of seperation of the active sets and a second-order sufficient optimality condition, Bouligand-differentiability (B-differentiability) of the solutions with respect to the parameter is established. Furthermore, an adjoint update strategy is proposed which yields a better approximation of the optimal controls and multipliers than the classical Taylor expansion, with remainder terms vanishing in L^∞.

1 Introduction

In this paper we consider an optimal control problem for a semilinear elliptic equation, subject to mixed control-state constraints. Some of the problem data depends on a parameter θ. The problem under consideration is

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|y - y_d(\theta)\|^2_{L^2(\Omega)} + \frac{\gamma(\theta)}{2}\|u - u_d(\theta)\|^2_{L^2(\Omega)} \\ &\text{s.t.} \quad \begin{cases} -\Delta y + d(y,\theta) = u + f(\theta) & \text{in } \Omega \\ y = 0 & \text{on } \Gamma \end{cases} \qquad (\mathbf{P}(\theta)) \\ &\text{and} \quad u + b_i(\theta)\, y - c_i(\theta) \ge 0 \quad \text{in } \Omega, \quad i = 1,\ldots,s. \end{aligned}$$

The assumptions on the problem data are made precise below. The parameter θ varies in a normed linear space Θ. We are concerned here with the sensitivity analysis of local optimal solutions of $(\mathbf{P}(\theta))$ with respect to perturbations of θ near a reference value θ_0. To our knowledge, this is the first publication dealing with the sensitivity analysis of optimal control problems involving partial differential equations and mixed control-state constraints.

Suppose that $u[\theta_0]$ is a local optimal reference solution for $(\mathbf{P}(\theta_0))$ satisfying second-order sufficient conditions and a separation assumption for the sets where the inequality constraints are active. Our main results are:

(1) In a neighborhood of $u[\theta_0]$, the solutions of neighboring problems $u[\theta]$ are locally unique and B-differentiable w.r.t. θ (see Theorems 4.2 and 4.3).

(2) We analyze two different update fomulas in order to approximate $u[\theta]$, using $u[\theta_0]$ and derivative information. Our so-called adjoint update scheme permits an L^∞ estimate for the remainder, while the classical Taylor formula does not (see Theorem 5.2).

Date: November 30, 2007.

2000 *Mathematics Subject Classification.* 49K20; 49K40; 46G05; 46T20.

Key words and phrases. Optimal control, control and state constraints, Bouligand derivative, Taylor expansion.

Let us put our work into perspective. Optimal control problems with pointwise constraints are subject of ongoing research. Necessary as well as sufficient optimality conditions are being investigated. Mixed control and state constraints as in ($\mathbf{P}(\theta)$) appear as Lavrentiev-type regularizations for problems with pure state constraints, but they are also interesting in their own right. Recently, the existence of regular Lagrange multipliers associated to mixed constraints was proved by Rösch and Tröltzsch [16]. Regular Lagrange multipliers allow the formulation of necessary optimality conditions in function spaces, and these are the basis for many optimization algorithms, for instance the semi-smooth Newton's method, see Hintermüller, Tröltzsch, and Yousept [5] in the context of problems with mixed constraints.

Problem ($\mathbf{P}(\theta)$) under consideration is in general non-convex. In this case, an essential assumption to find local minima are second-order sufficient conditions. We will employ such a condition involving strongly active sets, as presented by Meyer and Tröltzsch [15]. Problems with more than one mixed inequality constraint feature the additional difficulty that associated Lagrange multipliers need not to be unique. This may occur when several constraints are simultaneously active on a nontrivial subset of the domain Ω. This circumstance certainly has an impact on the stability properties of solutions of ($\mathbf{P}(\theta)$) and thus on numerical methods. In our analysis we thus assume that the active sets are separated, which in turn implies the uniqueness of multipliers, see Alt et al. [1].

In a series of articles, Malanowski [7, 8, 10, 11] proved stability and sensitivity properties of optimal control problems with *ordinary* differential equations in the presence of mixed control-state and pure state constraints. A Lipschitz stability result for optimal control problems with partial differential equations and mixed constraints was derived in [1]. Fréchet differentiability of the solution of Lavrientiev regularized problems with respect to the regularization parameter was analyzed in [6] under a strict complementarity assumption.

In a previous paper [4], we showed the B-differentiability of solutions for a simpler problem setting which accomodates linear-quadratic optimal control problems with pure control constraints. We also analyzed several update schemes which allow to approximately recover the optimal control $u[\theta]$ for parameters θ near a reference value θ_0. It was shown that the Taylor expansion

$$u[\theta] \approx u[\theta_0] + u'[\theta_0](\theta - \theta_0)$$

is not an optimal update scheme because it does not allow an error estimate in L^∞. This was remedied by the use of a so-called adjoint update scheme involving a certain adjoint quantity.

In this paper, we extend the results of [4] in two directions. First of all, we allow here a semilinear state equation. This resolves the unsatisfactory situation that arises for linear-quadratic optimal control problems, where the compuation of $u'[\theta_0](\theta - \theta_0)$ during the evaluation of the update formula is no less expensive than the solution of problem ($\mathbf{P}(\theta)$) itself with the perturbed parameter θ. Secondly, we consider here several mixed control-state constraints rather than one single pure control constraint, which in addition may depend on the parameter θ.

The outline of the paper is as follows. In the remainder of this section, we fix most of our standing assumptions. In Section 2, we state the system of necessary optimality conditions for ($\mathbf{P}(\theta)$). We also reformulate it in terms of a generalized equation, whose linearization is analyzed in Section 3. In particular, we prove that solutions to the linearized optimality system depend Lipschitz continuously and B-differentiably on the parameter. This result is used in Section 4 to prove the same properties about the nonlinear optimality system for $(\mathbf{P}(\theta))$, in virtue of an implicit function theorem for generalized equations. In Section 5, we devise two different update strategies to recover $u[\theta]$ from $u[\theta_0]$ and derivative information. We prove that the so-called adjoint scheme allows error estimates in L^∞ while the classical Taylor expansion does not. These findings are confirmed by numerical experiments in Section 6.

For convenience, we recall the notation of Bouligand or B-differentiability, which is also called a directional Fréchet derivative, see [2]. Recall that an operator $A : X \to Y$ is said to be positively homogeneous if $A(\lambda x) = \lambda A(x)$ holds for all $\lambda \geq 0$ and all $x \in X$.

Definition 1.1. *A function $f : X \to Y$ between normed linear spaces X and Y is said to be B-differentiable at $x_0 \in X$ if there exists $\varepsilon > 0$ and a positively homogeneous operator $f'(x_0) : X \to Y$ such that*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + r(x_0; x - x_0)$$

holds for all $x \in X$, where the remainder satisfies $\|r(x_0; x - x_0)\|_Y / \|x - x_0\|_X \to 0$ as $\|x - x_0\|_X \to 0$.

Clearly, if the B-derivative $f'(x_0)$ is a linear and continuous operator, then the B-derivative becomes the Fréchet derivative.

Throughout the paper, we denote by $B_\rho^X(x)$ or simply $B_\rho(x)$ the open ball of radius ρ, centered at $x \in X$. Moreover, $(\cdot,\cdot)_\Omega$ denotes the scalar product in $L^2(\Omega)$.

Standing Assumptions. Throughout, $\Omega \subset \mathbb{R}^N$, $N \in \{2, 3\}$, denotes a bounded domain with $C^{1,1}$ boundary Γ. The parameter θ varies in a some normed linear space Θ. Moreover, suppose that $B_\rho(\theta_0)$ is an open ball around some fixed reference value θ_0.

- **(A1)** The desired state $y_d : B_\rho(\theta_0) \to L^2(\Omega)$ is Lipschitz and B-differentiable at θ_0.
- **(A2)** The reference control $u_d : B_\rho(\theta_0) \to L^\infty(\Omega)$ is Lipschitz and B-differentiable at θ_0.
- **(A3)** The control cost parameter $\gamma : B_\rho(\theta_0) \to \mathbb{R}^+$ is Lipschitz and B-differentiable at θ_0.
- **(A4)** The right hand side $f(\cdot) : B_\rho(\theta_0) \to L^2(\Omega)$ is B-differentiable w.r.t. θ.
- **(A5)** The coefficients in the inequalities $b_i, c_i : B_\rho(\theta_0) \to L^\infty(S_i)$ are Lipschitz and B-differentiable for $i = 1, \ldots, s$.
- **(A6)** The nonlinearity $d(y, \theta)$ may depend on the spatial variable ξ.
  - (a) For any fixed $\theta \in B_\rho(\theta_0)$, $d(y, \theta)$ is measurable w.r.t. ξ and of class C^2 w.r.t. y, for almost all $\xi \in \Omega$.
  - (b) $d(\cdot, \theta)$ is assumed monotone w.r.t. y, i.e., $d_y(y, \theta) \geq 0$ holds for all $y \in \mathbb{R}$ and all $\theta \in B_\rho(\theta_0)$, a.e. in Ω.
  - (c) Moreover, $d(\cdot, \theta)$, $d_y(\cdot, \theta)$ and $d_{yy}(\cdot, \theta)$ are assumed to be a locally bounded and locally Lipschitz continuous function w.r.t. y, uniformly in $B_\rho(\theta_0)$, i.e., the following conditions hold: there exists $K > 0$ such that

    $$|d(0, \theta)| + |d_y(0, \theta)| + |d_{yy}(0, \theta)| \leq K \quad \text{a.e. in } \Omega \text{ and for all } \theta \in B_\rho(\theta_0),$$

    and for any $M > 0$, there exists $L(M) > 0$ such that

    $$\begin{aligned} |d(y_1, \theta) - d(y_2, \theta)| &\leq L(M)\,|y_1 - y_2| \quad \text{a.e. in } \Omega \text{ and for all } \theta \in B_\rho(\theta_0) \\ |d_y(y_1, \theta) - d_y(y_2, \theta)| &\leq L(M)\,|y_1 - y_2| \quad \text{a.e. in } \Omega \text{ and for all } \theta \in B_\rho(\theta_0) \\ |d_{yy}(y_1, \theta) - d_{yy}(y_2, \theta)| &\leq L(M)\,|y_1 - y_2| \quad \text{a.e. in } \Omega \text{ and for all } \theta \in B_\rho(\theta_0) \end{aligned}$$

hold for all $y_i \in \mathbb{R}$ satisfying $|y_i| \le M$.

(d) $d(y,\cdot) : B_\rho(\theta_0) \to \mathbb{R}$ is locally Lipschitz w.r.t. θ, uniformly on bounded sets of $\mathbb{R}$, i.e., for any $M > 0$, there exists $L(M) > 0$ such that

$$|d(y,\theta_1) - d(y,\theta_2)| \le L(M)\,|\theta_1 - \theta_2| \quad \text{a.e. in } \Omega$$

holds for all $y \in \mathbb{R}$ satisfying $|y| \le M$. This condition also holds for d_y and d_{yy}.

(e) For all $y \in L^\infty(\Omega)$, $d(y,\cdot) : B_\rho(\theta_0) \to L^\infty(\Omega)$ is B-differentiable at θ_0. The same holds for $d_y(y,\cdot)$.

Further assumptions will be stated as needed.

Preliminary results. Let us collect briefly some known results for solvability of the non-linear state equation. For future reference, we define the state space

$$Y = H^2(\Omega) \cap H^1_0(\Omega),$$

endowed with the norm $\|y\|_Y = \|y\|_{H^2(\Omega)}$. We note that $Y \hookrightarrow C_0(\Omega)$, the space of continuous functions on Ω with zero boundary values. It is well-known that under the assumptions above, the state equation admits a unique solution $y \in Y$ for every $u \in L^2(\Omega)$ and $\theta \in B_\rho(\theta_0)$.

In the context of optimality conditions, we will make use of the unique solvability of the linearized equation

$$\begin{aligned} -\Delta y + d_y(y_0,\theta_0)\,y &= g \quad \text{in } \Omega \\ y &= 0 \quad \text{on } \Gamma. \end{aligned} \tag{1.1}$$

Thanks to the Assumption **(A6)** (a)–(d) on d it holds

Lemma 1.2. *Let $y_0 \in C(\bar\Omega)$ be given. Then for every $g \in L^2(\Omega)$, the linearized system (1.1) admits a unique solution $y \in Y$. The mapping $g \mapsto y$ is linear and continuous.*

2 Optimality Conditions and Generalized Equation

We assume that problem $(\mathbf{P}(\theta_0))$ admits a local solution. This reference solution is denoted in the sequel by $(y_0, u_0) \in Y \times L^2(\Omega)$. For the existence of Lagrange multipliers, we employ the following linearized Slater condition:

(A7) There exist $\hat u \in L^\infty(\Omega)$ and $\beta > 0$ such that

$$\hat u + b_i(\theta_0)\,\hat y - c_i(\theta_0) \ge \beta \quad \text{in } \Omega, \quad i = 1,\dots,s$$

holds, where $\hat y$ is the solution of the linearized system (1.1) to the right hand side $\hat u$.

The Lagrange multipliers and adjoint state associated to (y_0, u_0) need not be unique without further assumptions, see [1, Remark 2.6 and Proposition 3.5]. Therefore, we follow the approach taken there and introduce the so-called security sets at the reference solution.

(A8) Suppose that there exists $\sigma > 0$ such that the sets

$$S_i^\sigma := \{x \in \Omega : 0 \le u_0 + b_i(\theta_0)\,y_0 - c_i(\theta_0) \le \sigma\}, \quad i = 1,\dots,s$$

are pairwise disjoint.

Note that **(A8)** implies that the active sets, i.e., the subsets of Ω where the inequality constraints are active, are separated with a security margin σ from each other.

However, it turns out that this separation assumption alone is not sufficient to show the uniqueness of multipliers. Therefore, we work in the sequel with the following regularity assumption.

(A9) We assume that the system

$$\begin{aligned} -\Delta y + d_y(y_0,\theta_0)\, y + \sum_{i=1}^{s} \chi_{S_i^\sigma} b_i(\theta_0)\, y &= g \qquad \text{in } \Omega \\ y &= 0 \qquad \text{on } \Gamma \end{aligned} \tag{2.1}$$

admits a unique solution $y \in Y$ for all right-hand sides $g \in L^2(\Omega)$, with continuous mapping $g \mapsto y$.

It should be noted that the regularity assumption is automatically fullfilled for $b \geq 0$. For the case of a parabolic state equation, the assumption $b \in L^\infty(\Omega)$ is already sufficient. For further discussions of the regularity assumption we refer to [14, 15].

Assuming **(A7)**, **(A8)**, and **(A9)** from now on, we obtain the existence, uniqueness and regularity of Lagrange multipliers associated to the mixed constraints.

Proposition 2.1. *There exist unique Lagrange multipliers $\mu_{0,i} \in L^\infty(\Omega)$, $i = 1,\dots,s$, and a unique adjoint state $p_0 \in Y$ such that $(y_0, u_0, p_0, \mu_{0,1}, \dots, \mu_{0,s})$ satisfies the following optimality conditions, where all data is evaluated at θ_0:*

$$-\Delta p + d_y(y,\theta_0)\, p = -(y - y_d) + \textstyle\sum_{i=1}^{s} b_i\, \mu_i \qquad \text{in } \Omega, \quad p = 0 \quad \text{on } \Gamma \tag{2.2a}$$

$$\gamma\,(u - u_d) - p - \textstyle\sum_{i=1}^{s} \mu_i = 0 \qquad \text{in } \Omega \tag{2.2b}$$

$$-\Delta y + d(y,\theta_0) = u + f \qquad \text{in } \Omega, \quad y = 0 \quad \text{on } \Gamma \tag{2.2c}$$

$$0 \leq \mu_i \quad \perp \quad u + b_i\, y - c_i \geq 0, \quad i = 1,\dots,s \qquad \text{in } \Omega. \tag{2.2d}$$

The complementarity system (2.2d) *can be equivalently stated as*

$$\sum_{i=1}^{s} \mu_i = \max\big\{0, \gamma \min_{i=1,\dots,s} \big(c_i - b_i\, y - u_d\big) - p\big\} \qquad \text{in } \Omega. \tag{2.2e}$$

Proof. Using the assumption **(A7)**, the existence of multipliers $\mu_{0,i} \in L^1(\Omega)$ and adjoint state p fulfilling the optimality system follows directly from [16, Theorem 3.3]. There, also the equivalent formulation (2.2e) is derived [16, Lemma 5.3]. The active sets associated to the mixed constraints are separated due to **(A8)**. Hence, one can apply a bootstrapping argument to (2.2e) similar to [16, Theorem 5.4] to obtain the claimed regularity. Multiplying (2.2b) with $\chi_{S_i^\sigma} b_i$ gives $\sum_{i=1}^{s} b_i\, \mu_i = (\sum_{i=1}^{s} \chi_{S_i^\sigma} b_i)(\gamma\,(u - u_d) - p)$. Substituting this expression into the adjoint equation (2.2a) yields

$$-\Delta p + d_y(y_0,\theta_0)\, p + \sum_{i=1}^{s} \chi_{S_i^\sigma} b_i\, p = g \quad \text{in } \Omega, \qquad p = 0 \quad \text{on } \Gamma,$$

with the uniquely determined right hand side $g = -(y - y_d) + (\sum_{i=1}^{s} \chi_{S_i^\sigma} b_i)\, \gamma\,(u - u_d)$. This equation is uniquely solvable by **(A9)**, hence the adjoint state p is unique, which in turn gives uniqueness of the multipliers by **(A8)** and (2.2b). ☐

Based on assumption **(A8)**, we now introduce the relaxed optimal control problem

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|y - y_d(\theta)\|^2_{L^2(\Omega)} + \frac{\gamma(\theta)}{2}\|u - u_d(\theta)\|^2_{L^2(\Omega)} \\ &\text{s.t.} \quad \begin{cases} -\Delta y + d(y,\theta) = u + f(\theta) & \text{in } \Omega \\ y = 0 & \text{on } \Gamma \end{cases} \\ &\text{and} \quad u + b_i(\theta)\, y - c_i(\theta) \geq 0 \quad \text{in } S_i^\sigma, \quad i = 1,\ldots,s, \end{aligned} \qquad (\mathbf{P}^{aux}(\theta))$$

where the inequality constraints are restricted to the sets S_i^σ and are thus separated by definition. We will prove later in Theorem 4.3 that $(\mathbf{P}(\theta))$ and $(\mathbf{P}^{aux}(\theta))$ have identical solutions in the vicinity of θ_0. This is a consequence of the Lipschitz stability result w.r.t. $L^\infty(\Omega)$ for the state y and the control u, see Theorem 4.2 *(ii)*. The advantage of $(\mathbf{P}^{aux}(\theta))$ is that it possesses unique Lagrange multipliers and a unique adjoint state not only for $\theta = \theta_0$. In this and the following sections, we consider only $(\mathbf{P}^{aux}(\theta))$. The transition back to the original problem will be made in Section 4, see Theorem 4.3.

For simplicity, we will write the mixed-constraints in $(\mathbf{P}^{aux}(\theta))$ as one single constraint on the union of all sets S_i^σ. Let us define

$$S = \bigcup_{i=1}^s S_i^\sigma, \quad b(\theta) = \sum_{i=1}^s b_i(\theta)\chi_{S_i^\sigma}, \quad \text{and} \quad c(\theta) = \sum_{i=1}^s c_i(\theta)\chi_{S_i^\sigma}.$$

Then the condition $u + b(\theta)y - c(\theta) \geq 0$ on S is equivalent to the inequality constraints of $(\mathbf{P}^{aux}(\theta))$.

The optimality system to $(\mathbf{P}^{aux}(\theta))$ is very similar to (2.2). It can be proven as in Proposition 2.1 that for every solution (y,u) of $(\mathbf{P}^{aux}(\theta))$ there exist (p,μ) fulfilling

$$-\Delta p + d_y(y,\theta)\, p = -(y - y_d) + b\,\mu \qquad \text{in } \Omega, \quad p = 0 \quad \text{on } \Gamma \tag{2.3a}$$
$$\gamma\,(u - u_d) - p - \mu = 0 \qquad \text{in } \Omega \tag{2.3b}$$
$$-\Delta y + d(y,\theta) = u + f \qquad \text{in } \Omega, \quad y = 0 \quad \text{on } \Gamma \tag{2.3c}$$
$$0 \leq \mu \quad \perp \quad u + b\,y - c \geq 0 \qquad \text{in } S, \tag{2.3d}$$

or equivalently

$$\mu = \max\big\{0, \chi_S\big[\gamma(c - b\,y - u_d) - p\big]\big\} \qquad \text{in } \Omega, \tag{2.3e}$$

where all data is evaluated at the parameter θ.

We can recover the multipliers μ_i associated to the individual mixed constraints of $(\mathbf{P}^{aux}(\theta))$ by setting $\mu_i = \chi_{S_i^\sigma}\mu$. Then the condition (2.3d) is equivalent to

$$0 \leq \mu_i \quad \perp \quad u + b_i(\theta_0)\, y - c_i(\theta_0) \geq 0 \quad \text{in } S_i^\sigma, \quad i = 1,\ldots,s.$$

It should be noted that the tuple (y_0, u_0, p_0, μ_0) is also a solution of the optimality system (2.3) of $(\mathbf{P}^{aux}(\theta))$ if one defines $\mu_0 = \sum_{i=1}^s \mu_{0,i}$.

Now, let us reformulate the optimality system (2.3) in terms of a generalized equation,

$$0 \in F(w,\theta) + \mathcal{N}(w). \tag{2.4}$$

Here and in the sequel we use the abbreviation

$$w = (y, u, p, \mu).$$

We define

$$F(w,\theta) = \begin{pmatrix} -\Delta p + d_y(y,\theta)\, p + (y - y_d(\theta)) - b(\theta)\,\mu \\ \gamma(\theta)\,(u - u_d(\theta)) - p - \mu \\ -\Delta y + d(y,\theta) - u - f(\theta) \\ \chi_S(u + b(\theta)\, y - c(\theta)) \end{pmatrix} \tag{2.5}$$

and note that F maps $W \times \Theta$ into Z, where

$$\begin{aligned} W &= Y \times L^\infty(\Omega) \times Y \times L^\infty(\Omega) \\ Z &= L^2(\Omega) \times L^\infty(\Omega) \times L^2(\Omega) \times L^\infty(\Omega). \end{aligned}$$

The set-valued part of the generalized equation (2.4) is given by

$$\mathcal{N}(w) = \{0\} \times \{0\} \times \{0\} \times N(\mu),$$

where

$$N(\mu) := \begin{cases} \{z \in L^\infty(\Omega) : (z, \mu - \nu)_\Omega \geq 0 \quad \forall \nu \in K\} & \text{if } \mu \in K, \\ \emptyset & \text{if } \mu \notin K \end{cases}$$

and K is the cone of nonnegative functions $K := \{\mu \in L^\infty(\Omega) : \mu \geq 0 \quad \text{a.e. in } \Omega\}$.

Lemma 2.2. *The generalized equation (2.4) is equivalent to the optimality system (2.3) for* $(\mathbf{P}^{aux}(\theta))$*. More precisely, a tuple* $(y,u,p,\mu) \in Y \times L^\infty(\Omega) \times Y \times L^\infty(\Omega)$ *satisfies (2.3) for a given parameter* θ *if and only if it satisfies (2.4) for* θ*.*

We note that by our assumptions and Proposition 2.1, $(y_0,u_0,p_0,\mu_0) \in W$ satisfies the generalized equation (2.4) for θ_0. We will use an implicit function theorem for generalized equations, see [3], in order to study the dependence of the solution of (2.4) on θ.

Lemma 2.3. $F : W \times B_\rho(\theta_0) \to Z$ *is partially Fréchet differentiable w.r.t.* w*. Its derivative in the direction* δw *is given by*

$$F'(w,\theta)\,\delta w = \begin{pmatrix} -\Delta\delta p + d_y(y,\theta)\,\delta p + d_{yy}(y,\theta)\,\delta y\, p + \delta y - b(\theta)\,\delta\mu \\ \gamma(\theta)\,\delta u - \delta p - \delta\mu \\ -\Delta\delta y + d_y(y,\theta)\,\delta y - \delta u \\ \chi_S(\delta u + b(\theta)\,\delta y) \end{pmatrix} \tag{2.6}$$

Proof. We need to consider only those terms in F which depend nonlinearly on w, i.e., the terms $d(y,\theta)$ and $d_y(y,\theta)\,p$. Owing to assumption **(A6)**, they are Fréchet differentiable w.r.t. (y,p) by standard results for Nemyckii operators. □

We now consider the linearized equation

$$\delta \in F(w_0,\theta_0) + F'(w_0,\theta_0)(w - w_0) + \mathcal{N}(w). \tag{2.7}$$

In our situation, (2.7) reads

$$\delta \in \begin{pmatrix} -\Delta p + d_y(y_0,\theta_0)\, p + d_{yy}(y_0,\theta_0)(y - y_0)\, p_0 + y - y_d(\theta_0) - b(\theta_0)\,\mu \\ \gamma(\theta_0)\,(u - u_d(\theta_0)) - p - \mu \\ -\Delta y + d(y_0,\theta_0) + d_y(y_0,\theta_0)(y - y_0) - u - f(\theta_0) \\ \chi_S(u + b(\theta_0)\, y - c(\theta_0)) \end{pmatrix} + \mathcal{N}(w). \tag{2.8}$$

It will be useful to reformulate the fourth row, which contains the set-valued part $N(w)$, also in terms of a nonsmooth equation.

Lemma 2.4. *The linearized generalized equation (2.8) is equivalent to*

$$-\Delta p + d_y(y_0,\theta_0)\,p + d_{yy}(y_0,\theta_0)(y-y_0)\,p_0 + y - y_d(\theta_0) - b(\theta_0)\,\mu = \delta_1 \tag{2.9a}$$

$$\gamma(\theta_0)(u-u_d(\theta_0)) - p - \mu = \delta_2 \tag{2.9b}$$

$$-\Delta y + d(y_0,\theta_0) + d_y(y_0,\theta_0)(y-y_0) - u - f(\theta_0) = \delta_3 \tag{2.9c}$$

$$\mu = \max\big\{0, \chi_S\big[\gamma(\theta_0)(c(\theta_0) - b(\theta_0)\,y - u_d(\theta_0) + \delta_4) - p - \delta_2\big]\big\} \tag{2.9d}$$

in Ω*, with boundary conditions* $y = p = 0$ *on* Γ*.*

Proof. The first three rows of (2.8) are unchanged. A simple calculation shows that the last row in (2.8) is equivalent to the complementarity system

$$0 \le \mu \quad \perp \quad u + b(\theta_0)\,y - c(\theta_0) - \delta_4 \ge 0 \qquad \text{in } S.$$

In turn, this complementarity system is equivalent to

$$\mu = \max\big\{0, \chi_S\big[\gamma(\theta_0)(c(\theta_0) - b(\theta_0)\,y - u + \delta_4) + \mu\big]\big\} \tag{2.10}$$

on the set S. On the complement $\Omega \setminus S$, (2.10) simply reads $\mu = 0$. Using the second row of (2.8), we obtain the equivalence of (2.10) and (2.9d). ☐

3 Properties of the Linearized Optimality System

In this section, we discuss further properties of the linearized optimality system (2.8). In particular, we prove its so-called *strong regularity* to prepare the application of the Implicit Function Theorem, see Proposition 3.3.

Lemma 3.1. *The linearized generalized equation (2.8) is equivalent to the optimality system of the following auxiliary linear-quadratic optimal control problem.*

$$\begin{aligned}
\textit{Minimize}\quad & \frac{1}{2}\|y - y_d(\theta_0)\|^2_{L^2(\Omega)} + \frac{\gamma(\theta_0)}{2}\|u - u_d(\theta)\|^2_{L^2(\Omega)} \\
& + \frac{1}{2}(d_{yy}(y_0,\theta_0)(y-y_0)^2, p_0)_\Omega - (\delta_1, y)_\Omega - (\delta_2, u)_\Omega \\
\textit{s.t.}\quad & \begin{cases} -\Delta y + d(y_0,\theta_0) + d_y(y_0,\theta_0)(y-y_0) = u + f(\theta_0) + \delta_3 & \textit{in } \Omega \\ y = 0 & \textit{on } \Gamma \end{cases} \\
\textit{and}\quad & u + b(\theta_0)\,y - c(\theta_0) \ge \chi_S \delta_4 \quad \textit{in } S.
\end{aligned} \tag{$\mathbf{AQP}(\delta)$}$$

In the investigations of this linear-quadratic problem the regularity assumption **(A9)** will be very useful. It implies the unique solvability of the system

$$\begin{aligned} -\Delta y + d_y(y_0,\theta_0)\,y + b(\theta_0)\,y &= g \qquad \text{in } \Omega \\ y &= 0 \qquad \text{on } \Gamma. \end{aligned} \tag{3.1}$$

We will benefit from this observation in the following way: Using the substitution $v := u + by$, we can transform the problem ($\mathbf{AQP}(\delta)$) into a control constrained problem in the new control v. To this end, the equation $v = u + y(u)$ has to be solvable w.r.t. u for every v, which leads directly to the system (3.1).

The stability and differentiability analysis of the linearized generalized equation (2.7) will rely heavily on a second-order sufficient condition, which is needed to prove the solvability and stability of ($\mathbf{AQP}(\delta)$).

We define for $\tau > 0$ the set of strongly active mixed constraints

$$A_\tau = \{x \in S : \mu_0 > \tau\}.$$

The critical cone for the second-order condition is then given as

$$C_\tau := \big\{(y,u) \in Y \times L^2(\Omega) : (y,u) \text{ solves (1.1) with right hand side } u, \\ \text{and it satisfies } u + b(\theta_0)\, y = 0 \text{ on } A_\tau \big\}.$$

We assume that the nominal solution $w_0 = (y_0, u_0, p_0, \mu_0)$ of (2.3) satisfies the following coercivity assumption:

(A10) There exist $\tau > 0$ and $\alpha > 0$ such that

$$L''(y_0, u_0, p_0, \mu_0)(y,u)^2 \geq \alpha \, \|u\|^2_{L^2(\Omega)} \quad \forall (y,u) \in C_\tau. \tag{3.2}$$

It was proven by Meyer and Tröltzsch in [15] for a slightly different setting that this condition is indeed sufficient for local optimality of the nominal solution. As a by-product, we also obtain that the nominal solution is locally optimal for $(\mathbf{P}^{aux}(\theta))$.

Corollary 3.2. *The nominal solution* (y_0, u_0) *of* $(\mathbf{P}(\theta_0))$ *is locally optimal for the nonlinear auxiliary problem* $(\mathbf{P}^{aux}(\theta_0))$*.*

Moreover, assumption **(A10)** is also sufficient to prove the existence of solutions of $(\mathbf{AQP}(\delta))$ for small δ and their Lipschitz continuity with respect to δ.

Proposition 3.3 (Strong Regularity)**.** *There exist constants* $\rho_z > 0$ *and* $\rho_\mu > 0$ *such that for every* $\delta \in Z$ *with* $\delta \in B^Z_{\rho_z}(0)$*,* $(\mathbf{AQP}(\delta))$ *has a solution, adjoint state and Lagrange multiplier* $(y_\delta, u_\delta, p_\delta, \mu_\delta) \in W$ *satisfying (2.8), which is unique in* $Y \times L^\infty(\Omega) \times Y \times B^{L^\infty}_{\rho_\mu}(\mu_0)$*. Moreover, there exists* $L > 0$ *such that*

$$\|y_\delta - y_{\delta'}\|_Y + \|u_\delta - u_{\delta'}\|_{L^\infty(\Omega)} + \|p_\delta - p_{\delta'}\|_Y + \|\mu_\delta - \mu_{\delta'}\|_{L^\infty(\Omega)} \leq L \, \|\delta - \delta'\|_Z$$

holds for all $\delta, \delta' \in B^Z_{\rho_z}(0)$*.*

Proof. Let us note first that the coercivity assumption **(A10)** is not enough to ensure solvability of $(\mathbf{AQP}(\delta))$. This is the same situation as for optimal control problems with control constraints, see the discussion in [12]. To overcome this difficulty let us consider another auxiliary problem, which is similar to $(\mathbf{AQP}(\delta))$ but has a modified mixed constraint:

$$\begin{aligned} u + b(\theta_0)\, y - c(\theta_0) &= \chi_S \delta_4 && \text{in } A_\tau, \\ u + b(\theta_0)\, y - c(\theta_0) &\geq \chi_S \delta_4 && \text{in } S \setminus A_\tau, \quad i = 1, \ldots, s. \end{aligned} \tag{$\widetilde{\mathbf{AQP}}(\delta)$}$$

Now, we apply the transformation $v := u + by$. Due to assumption **(A9)**, this substitution is reversible, i.e. for given v there exists u with associated state $y(u)$ such that $v = u + by(u)$ holds. Then problem $(\widetilde{\mathbf{AQP}}(\delta))$ is transformed equivalently to an optimization problem with box constraints on the new control v. Thanks to **(A10)**, this new auxiliary problem is strictly convex and admits for every $\delta \in Z$ a unique solution $(\tilde{y}_\delta, \tilde{u}_\delta, \tilde{p}_\delta, \tilde{\mu}_\delta)$. The mapping $\delta \mapsto (\tilde{y}_\delta, \tilde{u}_\delta, \tilde{p}_\delta, \tilde{\mu}_\delta)$ is Lipschitz continuous: Following the lines of the proofs of Proposition 3.2 and Lemma 3.6 of [1] one can show that there is a constant $L > 0$ such that

$$\|\tilde{y}_\delta - \tilde{y}_{\delta'}\|_Y + \|\tilde{u}_\delta - \tilde{u}_{\delta'}\|_{L^\infty(\Omega)} + \|\tilde{p}_\delta - \tilde{p}_{\delta'}\|_Y + \|\tilde{\mu}_\delta - \tilde{\mu}_{\delta'}\|_{L^\infty(\Omega)} \leq L \, \|\delta - \delta'\|_Z$$

holds for all $\delta, \delta' \in Z$. In particular, we have $\|\tilde{\mu}_\delta - \mu_0\|_{L^\infty(\Omega)} \leq L \, \|\delta\|_Z$, since the nominal solution is also a solution of problem $(\widetilde{\mathbf{AQP}}(\delta))$ with $\delta = 0$. Hence there exists a radius $\rho_z > 0$ such that $\tilde{\mu}_\delta(x) > \tau/2$ holds on A_τ for solutions associated to perturbations δ with $\|\delta\|_Z < \rho_z$. This in turn implies that the mixed constraint remains active on A_τ and the complementary condition to $(\mathbf{AQP}(\delta))$ is satisfied. With [15, Theorem 5] it follows that $(\tilde{y}_\delta, \tilde{u}_\delta, \tilde{p}_\delta, \tilde{\mu}_\delta)$ is a local solution of $(\mathbf{AQP}(\delta))$.

It remains to show that this solution is unique if μ is restricted to a small ball in $L^\infty(\Omega)$ around μ_0. Let $(\tilde{y}, \tilde{u}, \tilde{p}, \tilde{\mu})$ be another local solution of $(\mathbf{AQP}(\delta))$ with associated Lagrange multiplier and adjoint state such that $\|\tilde\mu_\delta - \tilde\mu\|_{L^\infty(\Omega)} \le \rho_\mu$. We can choose $\rho_\mu \le \tau/4$, and it follows that $\tilde\mu > \tau/4$ holds on A_τ. Hence, the constraint is still strongly active and $\tilde u + b\,\tilde y - c = 0$ is fulfilled on A_τ. Consequently $(\tilde y, \tilde u)$ is admissible and locally optimal for the uniquely solvable problem $(\widetilde{\mathbf{AQP}}(\delta))$. Hence it coincides with $(\tilde y_\delta, \tilde u_\delta)$. Since the Lagrange multipliers and adjoint state are also unique $(\tilde\mu, \tilde p)$ coincides with $(\tilde\mu_\delta, \tilde p_\delta)$. □

Remark 3.4. *In the previous theorem, the perturbation δ_2 was taken from $L^\infty(\Omega)$. If one allows perturbations δ in*

$$Z_p = L^2(\Omega) \times L^p(\Omega) \times L^2(\Omega) \times L^\infty(\Omega),$$

for $p \in [1,\infty)$, the existence of a solution of $(\mathbf{AQP}(\delta))$ cannot be verified by the technique above. However, if solutions $(y_\delta, u_\delta, p_\delta, \mu_\delta)$ and $(y_{\delta'}, u_{\delta'}, p_{\delta'}, \mu_{\delta'})$ are known to exist, the weaker Lipschitz estimate

$$\|y_\delta - y_{\delta'}\|_Y + \|u_\delta - u_{\delta'}\|_{L^p(\Omega)} + \|p_\delta - p_{\delta'}\|_Y + \|\mu_\delta - \mu_{\delta'}\|_{L^p(\Omega)} \le L_p\,\|\delta - \delta'\|_{Z_p}$$

can be shown. This is needed below in the proof of Theorem 3.5.

For the discussion of differentiability of the map $\delta \mapsto w_\delta$, the following sets will play a role. They denote the subsets on which the constraint for the reference solution is inactive, weakly active, or strongly active, respectively.

$$\begin{aligned} S^- &= \{x \in S : u_0 + b(\theta_0)\,y_0 - c(\theta_0) > 0\} \\ S^0 &= \{x \in S : u_0 + b(\theta_0)\,y_0 - c(\theta_0) = 0 \quad \text{and} \quad \mu_0 = 0\} \\ S^+ &= \{x \in S : \mu_0 > 0\}. \end{aligned}$$

Associated to these sets is the projection Π_{I_0} onto the following cone of functions:

$$I_0 = \big\{z \in L^2(\Omega) : z = 0 \text{ in } S^- \cup (\Omega \setminus S), \quad z \ge 0 \text{ in } S^0\big\}. \tag{3.3}$$

For the following theorem, we also need the space

$$W_p = Y \times L^p(\Omega) \times Y \times L^p(\Omega), \quad p \in [1,\infty).$$

Theorem 3.5. *The map $Z \ni \delta \mapsto (y_\delta, u_\delta, p_\delta, \mu_\delta)$ is B-differentiable at $\delta = 0$ with values in W_p for any $p \in [1,\infty)$. The derivative at $\delta = 0$ in the direction of $\widehat\delta \in Z$ is given by the unique solution, Lagrange multiplier, and adjoint state of the following auxiliary linear-quadratic optimal control problem.*

$$\textit{Minimize} \quad \frac{1}{2}\|y\|^2_{L^2(\Omega)} + \frac{\gamma(\theta_0)}{2}\|u\|^2_{L^2(\Omega)} + \frac{1}{2}(d_{yy}(y_0,\theta_0)\,y^2, p_0)_\Omega - (\widehat\delta_1, y)_\Omega - (\widehat\delta_2, u)_\Omega$$

$$\textit{s.t.} \quad \begin{cases} -\Delta y + d_y(y_0,\theta_0)\,y = u + \widehat\delta_3 & \textit{in } \Omega \\ \qquad\qquad\qquad\quad y = 0 & \textit{on } \Gamma \end{cases} \tag{$\mathbf{DQP}(\widehat\delta)$}$$

$$\textit{and} \quad \begin{cases} u + b(\theta_0)\,y - \widehat\delta_4 = 0 & \textit{in } S^+, \\ u + b(\theta_0)\,y - \widehat\delta_4 \ge 0 & \textit{in } S^0. \end{cases}$$

In other words, the derivative satisfies the system

$$-\Delta p + d_y(y_0,\theta_0)\,p + d_{yy}(y_0,\theta_0)\,y\,p_0 + y - b(\theta_0)\,\mu = \widehat\delta_1 \tag{3.4a}$$

$$\gamma(\theta_0)\,u - p - \mu = \widehat\delta_2 \tag{3.4b}$$

$$-\Delta y + d_y(y_0,\theta_0)\,y - u = \widehat\delta_3 \tag{3.4c}$$

$$\mu = \Pi_{I_0}\Big(\chi_S\big[\gamma(\theta_0)(-b(\theta_0)\,y - u + \widehat\delta_4) + \mu\big]\Big) \tag{3.4d}$$

or equivalently

$$\mu = \Pi_{I_0}\Big(\chi_S\big[\gamma(\theta_0)(-b(\theta_0)\,y + \widehat{\delta}_4) - p - \widehat{\delta}_2\big]\Big). \tag{3.4e}$$

in Ω*, with boundary conditions* $y = p = 0$ *on* Γ*.*

Before we prove the theorem, we remark that the computation of a directional derivative of $\delta \mapsto (y_\delta, u_\delta, p_\delta, \mu_\delta)$ amounts to the solution of a linear-quadratic optimal control problem. In general, the directional derivative does not depend linearly on the direction $\widehat{\delta}$ because of the inequality constraints present on the set S^0. However, if S^0 has measure zero, i.e., if strict complementarity holds at w_0, then I_0 becomes a linear space. In this case, the directional derivative depends linearly and continuously on $\widehat{\delta}$, and the B-derivative becomes the Fréchet derivative.

Proof of Theorem 3.5. Suppose that $\widehat{\delta} \in Z$ is the given direction, and let us fix $p \in [1, \infty)$. We proceed in two steps.

Step 1: We verify that (3.4) has a unique solution, i.e., the candidate for the derivative exists. Comparing the definitions of S^+ and A_τ, it follows that the critical cone C_τ of the sufficient second-order optimality condition **(A10)** contains more directions than are admissible in the problem $(\mathbf{DQP}(\widehat{\delta}))$. Hence, this problem is convex by **(A10)** and possesses a unique solution which, together with its unique adjoint state and adjoint multipliers, is characterized by the optimality system (3.4).

Step 2: We need to estimate the remainder term

$$\|y_{\widehat{\delta}} - y_0 - \widehat{y}\|_Y + \|u_{\widehat{\delta}} - u_0 - \widehat{u}\|_{L^p(\Omega)} + \|p_{\widehat{\delta}} - p_0 - \widehat{p}\|_Y + \|\mu_{\widehat{\delta}} - \mu_0 - \widehat{\mu}\|_{L^p(\Omega)},$$

where $(y_{\widehat{\delta}}, u_{\widehat{\delta}}, p_{\widehat{\delta}}, \mu_{\widehat{\delta}})$ denotes the unique solution of $(\mathbf{AQP}(\delta))$ for the direction $\widehat{\delta}$, and $(\widehat{y}, \widehat{u}, \widehat{p}, \widehat{\mu})$ denotes the candidate for the derivative, i.e., the unique solution of $(\mathbf{DQP}(\widehat{\delta}))$. First of all, we realize that the difference $(y_{\widehat{\delta}}, u_{\widehat{\delta}}, p_{\widehat{\delta}}, \mu_{\widehat{\delta}}) - (y_0, u_0, p_0, \mu_0)$ satisfies (3.4a)–(3.4c), and in addition

$$\begin{aligned}\mu_{\widehat{\delta}} - \mu_0 = \max\big\{0, \chi_S\big[\gamma(\theta_0)(c(\theta_0) - b(\theta_0)\,y_{\widehat{\delta}} - u_d(\theta_0) + \widehat{\delta}_4) - p_{\widehat{\delta}} - \widehat{\delta}_2\big]\big\}\\ - \max\big\{0, \chi_S\big[\gamma(\theta_0)(c(\theta_0) - b(\theta_0)\,y_0 - u_d(\theta_0) + 0) - p_0 - 0\big]\big\}.\end{aligned} \tag{3.5}$$

Next we use the B-differentiability of the projection from $L^\infty(\Omega)$ into $L^p(\Omega)$, see for instance [9, Theorem 2.1] or [4, Theorem 4.3]. We can thus write

$$\mu_{\widehat{\delta}} - \mu_0 = \Pi_{I_0}\Big(\chi_S\big[\gamma(\theta_0)(-b(\theta_0)\,(y_{\widehat{\delta}} - y_0) + \widehat{\delta}_4) - (p_{\widehat{\delta}} - p_0) - \widehat{\delta}_2\big]\Big) + r_1, \tag{3.6}$$

where the remainder term satisfies

$$\frac{\|r_1\|_{L^p(\Omega)}}{\big\|\chi_S\big[\gamma(\theta_0)(-b(\theta_0)\,(y_{\widehat{\delta}} - y_0) + \widehat{\delta}_4) - (p_{\widehat{\delta}} - p_0) - \widehat{\delta}_2\big]\big\|_{L^\infty(\Omega)}} \to 0 \tag{3.7}$$

as the denominator goes to zero. From the Lipschitz continuity result (Proposition 3.3), it follows that

$$\big\|\chi_S\big[\gamma(\theta_0)(-b(\theta_0)\,(y_{\widehat{\delta}} - y_0) + \widehat{\delta}_4) - (p_{\widehat{\delta}} - p_0) - \widehat{\delta}_2\big]\big\|_{L^\infty(\Omega)} \le L\,\|\widehat{\delta}\|_Z.$$

Together with (3.7), this implies

$$\frac{\|r_1\|_{L^p(\Omega)}}{\|\widehat{\delta}\|_Z} \to 0 \quad \text{as } \|\widehat{\delta}\|_Z \to 0. \tag{3.8}$$

We proceed by defining

$$y^r := y_{\widehat{\delta}} - y_0, \quad u^r := u_{\widehat{\delta}} - u_0, \quad p^r := p_{\widehat{\delta}} - p_0, \quad \mu^r := \mu_{\widehat{\delta}} - \mu_0 - r_1.$$

It is straightforward to verify that (y^r, u^r, p^r, μ^r) satisfies

$$-\Delta p^r + d_y(y_0,\theta_0)\, p^r + d_{yy}(y_0,\theta_0)\, y^r\, p_0 + y^r - b(\theta_0)\, \mu^r = \widehat{\delta}_1 + b(\theta_0)\, r_1 \tag{3.9a}$$

$$\gamma(\theta_0)\, u^r - p^r - \mu^r = \widehat{\delta}_2 + r_1 \tag{3.9b}$$

$$-\Delta y^r + d_y(y_0,\theta_0)\, y^r - u^r = \widehat{\delta}_3 \tag{3.9c}$$

$$\mu^r = \Pi_{I_0}\Big(\chi_S\big[\gamma(\theta_0)(-b(\theta_0)\, y^r + \widehat{\delta}_4) - p^r - \widehat{\delta}_2\big]\Big) \tag{3.9d}$$

in Ω, with boundary conditions $y^r = p^r = 0$ on Γ. Comparing with (3.4), we find that (y^r, u^r, p^r, μ^r) and $(\widehat{y}, \widehat{u}, \widehat{p}, \widehat{\mu})$ satisfy the same system of equations, except for the right hand side perturbations $b\, r_1$ and r_1 in (3.9a) and (3.9b). In other words, $(\widehat{y}, \widehat{u}, \widehat{p}, \widehat{\mu})$ and (y^r, u^r, p^r, μ^r) are solutions of ($\mathbf{DQP}(\widehat{\delta})$), once with $\widehat{\delta}$, and once with $\widehat{\delta} + (b(\theta_0)\, r_1, r_1, 0, 0)$. Due to the similarity of ($\mathbf{DQP}(\widehat{\delta})$) to ($\mathbf{AQP}(\delta)$), its solutions also enjoy the Lipschitz stability property given in Proposition 3.3 and Remark 3.4. Hence we can estimate

$$\begin{aligned}
&\|y_{\widehat{\delta}} - y_0 - \widehat{y}\|_Y + \|u_{\widehat{\delta}} - u_0 - \widehat{u}\|_{L^p(\Omega)} + \|p_{\widehat{\delta}} - p_0 - \widehat{p}\|_Y + \|\mu_{\widehat{\delta}} - \mu_0 - \widehat{\mu} - r_1\|_{L^p(\Omega)} \\
&= \|y^r - \widehat{y}\|_Y + \|u^r - \widehat{u}\|_{L^p(\Omega)} + \|p^r - \widehat{p}\|_Y + \|\mu^r - \widehat{\mu}\|_{L^p(\Omega)} \\
&\le L_p\, \|r_1\|_{L^p(\Omega)}.
\end{aligned}$$

Using the triangle inequality, we can finally estimate the remainder term:

$$\begin{aligned}
&\|y_{\widehat{\delta}} - y_0 - \widehat{y}\|_Y + \|u_{\widehat{\delta}} - u_0 - \widehat{u}\|_{L^p(\Omega)} + \|p_{\widehat{\delta}} - p_0 - \widehat{p}\|_Y + \|\mu_{\widehat{\delta}} - \mu_0 - \widehat{\mu}\|_{L^p(\Omega)} \\
&\le (L_p + 1)\, \|r_1\|_{L^p(\Omega)}.
\end{aligned}$$

By (3.8), the right hand side is of order $o\,(\|\widehat{\delta}_Z\|)$, and so this holds for the left hand side as well. Clearly the tuple $(\widehat{y}, \widehat{u}, \widehat{p}, \widehat{\mu})$ is positively homogeneous w.r.t. the direction $\widehat{\delta}$. Hence we conclude that indeed, $(\widehat{y}, \widehat{u}, \widehat{p}, \widehat{\mu})$ is the Bouligand derivative of $(y_\delta, u_\delta, p_\delta, \mu_\delta)$ at $\delta = 0$ in the direction $\widehat{\delta}$. □

4 Properties of the Nonlinear Optimality System

In this section we return our attention to the nonlinear optimality system (2.3) of problem ($\mathbf{P}^{aux}(\theta)$). We recall that (2.3) is equivalent to the generalized equation (2.4). We will apply an implicit function theorem for generalized equations given in [3] in order to prove a B-differentiability result for the solution of (2.4). We begin with an introductory lemma.

Lemma 4.1. *The function $F : W \times B_\rho(\theta_0) \to Z$ defined in (2.5) has the following properties.*

(i) For all $R > 0$, there exists $L(R) > 0$ such that

$$\|F(w,\theta_1) - F(w,\theta_2)\|_Z \le L(R)\, \|\theta_1 - \theta_2\|_\Theta$$

holds for all $w \in B^W_R(w_0)$ and $\theta_i \in B_\rho(\theta_0)$.

(ii) $F(w_0, \cdot)$ is B-differentiable at θ_0.

(iii) F is partially Fréchet differentiable w.r.t. w in a neighborhood of (w_0, θ_0), and the derivative F' is continuous w.r.t. (w,θ) at (w_0,θ_0).

Proof. To prove (i), we carry out the estimate for the first row of F only. The rest follows similarly.

$$\begin{aligned}
\|F_1(w,\theta_1) - F_1(w,\theta_2)\|_{L^2(\Omega)} \le\ & \|\big(d_y(y,\theta_1) - d_y(y,\theta_2)\big)\|_{L^2(\Omega)}\, \|p\|_{L^\infty(\Omega)} \\
& + \|y_d(\theta_1) - y_d(\theta_2)\|_{L^2(\Omega)} + \|b(\theta_1) - b(\theta_2)\|_{L^2(\Omega)}\, \|\mu\|_{L^\infty(\Omega)}.
\end{aligned}$$

The first term can be estimated by Assumption **(A6)**, part (d). Here we use the fact that $\|y\|_{L^\infty(\Omega}$ and $\|p\|_{L^\infty(\Omega}$ remain bounded for $w \in B_R^W(w_0)$. The second term can be estimated by **(A1)**, and the third term by **(A5)** and the observation that $\|\mu\|_{L^\infty(\Omega}$ remains bounded for $w \in B_R^W(w_0)$.
Assertion (ii) follows because we assumed all terms in F depending on θ to be B-differentiable w.r.t. at θ_0, see Assumptions **(A1)**–**(A6)**.
To prove (iii), we recall from Lemma 2.3 that F is partially Fréchet differentiable w.r.t. w on all of $W \times B_\rho(\theta_0)$. The derivative is given by (2.6), and it is easily seen to be continuous w.r.t. (w, θ). □

We are now in the position to show that the generalized equation (2.4) has a locally unique solution which is B-differentiable w.r.t. θ. However, B-differentiability does not hold w.r.t. the norm of W, but only w.r.t. the weaker norm of $W_p = Y \times L^p(\Omega) \times Y \times L^p(\Omega)$ for $p \in [1, \infty)$.

Theorem 4.2. *There exists $\rho_1 \in (0, \rho]$ and $R > 0$ such that the following holds.*

(i) *For every $\theta \in B_{\rho_1}(\theta_0)$, the nonlinear generalized equation (2.4) has a solution $w[\theta] = (y[\theta], u[\theta], p[\theta], \mu[\theta]) \in W$, which is unique in $B_R^W(w_0)$.*

(ii) *The map*
$$B_{\rho_1}(\theta_0) \ni \theta \mapsto w[\theta] \in W$$
is Lipschitz continuous

(iii) *and B-differentiable at θ_0 w.r.t. the norm of W_p for all $p \in [1, \infty)$. The derivative in the direction of $\delta\theta$, denoted by $w'[\theta_0]\,\delta\theta$, is given by the unique solution of (3.4) in the direction of*
$$\widehat{\delta} = -F_\theta(w_0, \theta_0)\,\delta\theta.$$

Proof. The result follows from an application of the Implicit Function Theorem [3, Theorem 2.4], which is given in Appendix A. Proposition 3.3 and Lemma 4.1 show that F satisfies the necessary requisites. As a consequence, the properties of the solution map $\theta \mapsto w[\theta]$ of (2.4) are inherited from the map $\delta \mapsto w_\delta$ of (2.7). In Proposition 3.3, we proved the Lipschitz continuity of $Z \ni \delta \mapsto w_\delta \in W$. In addition, Theorem 3.5 proves that $Z \ni \delta \mapsto w_\delta \in W_p$ is B-differentiable at $\delta = 0$. □

Theorem 4.3. *There exists $\rho_2 \in (0, \rho_1]$ such that for all $\rho \in B_{\rho_2}(\theta_0)$, $(y[\theta], u[\theta])$ is a local optimal solution of our original problem $(\mathbf{P}(\theta))$, and $p[\theta]$ together with $(\chi_{S_1^\sigma}\mu[\theta], \ldots, \chi_{S_s^\sigma}\mu[\theta])$ are the unique associated adjoint state and Lagrange multipliers.*

Proof. In a first step we show that $w[\theta]$ satisfies the necessary optimality conditions (2.2) for $(\mathbf{P}(\theta))$. Recall that $w[\theta]$ is a solution of the generalized equation (2.4), i.e., it is a solution of the optimality system (2.3) for the auxiliary problem $(\mathbf{P}^{aux}(\theta))$ by Lemma 2.2. Let us define $\mu_i[\theta] = \chi_{S_i^\sigma}\,\mu[\theta]$. Then $\mu[\theta] = \sum_{i=1}^s \mu_i[\theta]$ holds. Comparing (2.3) and (2.2), we only need to verify that $(y[\theta], u[\theta])$ is feasible for $(\mathbf{P}(\theta))$ and that (2.2d) holds. Indeed, on the set S_i^σ, $(y[\theta], u[\theta])$ is feasible w.r.t. the i-th constraint for $(\mathbf{P}(\theta))$ by definition. Due to the assumption **(A5)** on b and c, the separation assumption **(A8)** and the Lipschitz stability result (Theorem 4.2 (ii)), we have the following estimate on $\Omega \setminus S_i^\sigma$.

$$\begin{aligned}
&u[\theta] + b_i(\theta)\,y[\theta] - c_i(\theta)\\
&\quad \geq u_0 + b_i(\theta_0)\,y_0 - c_i(\theta_0) - \|u[\theta] - u_0\|_{L^\infty(\Omega)} - \|b_i(\theta) - b_i(\theta_0)\|_{L^\infty(\Omega)}\|y[\theta]\|_{L^\infty(\Omega)}\\
&\qquad - \|b_i(\theta_0)\|_{L^\infty(\Omega)}\|y[\theta] - y_0\|_{L^\infty(\Omega)} - \|c_i(\theta) - c_i(\theta_0)\|_{L^\infty(\Omega)}\\
&\quad \geq \sigma - C\,\|\theta - \theta_0\|
\end{aligned}$$

for some constant $c > 0$. Hence $(y[\theta], u[\theta])$ is feasible for the original problem $(\mathbf{P}(\theta))$ as long as $\|\theta - \theta_0\|$ is sufficiently small. Moreover, this estimate shows that the set where the i-th constraint is active remains a subset of S_i^σ, i.e.,

$$A_i[\theta] := \{x \in \Omega : u[\theta] + b_i(\theta)\, y[\theta] - c_i(\theta) = 0\} \subset S_i^\sigma, \quad i = 1, \ldots, s.$$

Hence assumption **(A8)** implies that $A_i[\theta]$ are pairwise disjoint for sufficiently small $\|\theta - \theta_0\|$, and one can show as before that the associated Lagrange multipliers and adjoint state are unique.

To verify (2.2d), we consider (2.3b) and (2.3e) on the set S_i^σ. It follows that

$$\begin{aligned}\mu_i[\theta] = \mu[\theta] &= \max\big\{0, \gamma\,(c_i(\theta) - b_i(\theta)\, y[\theta] - u_d(\theta)) - p[\theta]\big\}\\ &= \max\big\{0, \gamma\,(c_i(\theta) - b_i(\theta)\, y[\theta] - u) + \mu_i[\theta]\big\}\end{aligned}$$

holds on S_i^σ, which implies (2.2d) on S_i^σ. On $\Omega \setminus S_i^\sigma$, $\mu_i[\theta] = 0$ holds and again (2.2d) is verified because $(y[\theta], u[\theta])$ was shown above to be feasible for $(\mathbf{P}(\theta))$.

It remains to verify that $w[\theta]$ is indeed a local optimal solution. This can be done by showing that the second-order sufficient conditions are stable under small perturbations. That is, **(A10)** continues to hold with $\tau/2$, $\alpha/2$, μ_0 replaced by $\mu[\theta]$, y_0 replaced by $y[\theta]$ in (1.1) and $b(\theta_0)$ replaced by $b(\theta)$ in C_τ, given that $\|\theta - \theta_0\|$ is sufficiently small. The proof uses well established techniques and is therefore not carried out here. We refer the reader to [13, Lemma 5.2]. □

Theorem 4.3 implies that the local solutions of $(\mathbf{P}^{aux}(\theta))$ and $(\mathbf{P}(\theta))$ near (y_0, u_0) coincide and thus Theorem 4.2 is in fact a statement about local solutions of $(\mathbf{P}(\theta))$.

5 Adjoint Update Strategy and Error Estimates

In Theorems 4.2 and 4.3, we proved the B-differentiability at θ_0 of the locally unique solution map $w[\theta]$ of $(\mathbf{P}(\theta))$ and its optimality system (2.2), with respect to θ and with the target space $W_p = Y \times L^p(\Omega) \times Y \times L^p(\Omega)$ for $p \in [1, \infty)$. The B-differentiability cannot be achieved w.r.t. $L^\infty(\Omega)$ for the control u and the Lagrange multipliers μ_i. The reason is that B-differentiation of the projection $\max\{0, \cdot\}$ requires a norm gap [4, Theorem 4.3], which cannot be made up for because all the smoothing gained by solution operators of differential equations occurs *before* the projection is applied. To make this precise, we consider the nonsmooth part (2.3e) of the optimality system, which we recall here for convenience:

$$\mu = \max\big\{0, \chi_S\big[\gamma(\theta_0)(c(\theta_0) - b(\theta_0)\, y - u_d(\theta_0)) - p\big]\big\}. \tag{5.1}$$

From the adjoint equation (2.3a) we infer that

$$p = \mathcal{S}^\star\big(b(\theta_0)\, \mu - (y - y_d(\theta_0))\big), \tag{5.2}$$

where $\mathcal{S}$ denotes the solution operator of the elliptic equation

$$-\Delta z + d_y(y_0, \theta_0)\, z = f \quad \text{in } \Omega, \qquad z = 0 \quad \text{on } \Gamma.$$

(Note that for our problem $\mathcal{S}^\star = \mathcal{S}$ holds, but we write the customary $\mathcal{S}^\star$ in (5.2) nonetheless.) Substituting (5.2) into (5.1), we obtain the representation

$$\mu = \max\big\{0, g - \mathcal{G}\mu\big\} \quad \text{in } \Omega, \tag{5.3}$$

where

$$\begin{aligned}\mathcal{G}\mu &= \chi_S\, \mathcal{S}^\star(b(\theta_0)\, \mu)\\ g &= \chi_S\big[\gamma(\theta_0)(c(\theta_0) - b(\theta_0)\, y - u_d(\theta_0)) + \mathcal{S}^\star\big(y - y_d(\theta_0)\big)\big].\end{aligned}$$

Clearly, (5.3) shows that all smoothing by $\mathcal{G}$ occurs before the projection is applied, and hence we cannot achieve B-differentiability of $\mu[\theta]$ with values in $L^\infty(\Omega)$.

Following our ideas in [4], we consider an alternative or adjoint formulation in order to reverse the order of projecting and smoothing. We define the adjoint quantity

$$\phi = g - \mathcal{G}\mu, \tag{5.4}$$

so that $\mu = \max\{0, \phi\}$ holds. Plugging in, we see that ϕ satisfies

$$\phi = g - \mathcal{G}\max\{0, \phi\}. \tag{5.5}$$

In terms of our original variables, (5.4) reads

$$\phi = \chi_S\big[\gamma(\theta_0)(c(\theta_0) - b(\theta_0)\, y - u_d(\theta_0)) - p\big].$$

By replacing μ by ϕ in the optimality system (2.3), we obtain its alternative form:

$$-\Delta p + d_y(y_0, \theta_0)\, p = -(y - y_d(\theta_0)) + b(\theta_0)\max\{0, \phi\} \tag{5.6a}$$

$$\gamma(\theta_0)(u - u_d(\theta_0)) - p - \max\{0, \phi\} = 0 \tag{5.6b}$$

$$-\Delta y + d(y_0, \theta_0) = u + f(\theta_0) \tag{5.6c}$$

$$\phi = \chi_S\big[\gamma(\theta_0)(c(\theta_0) - b(\theta_0)\, y - u_d(\theta_0)) - p\big] \tag{5.6d}$$

in Ω, with boundary conditions $y = p = 0$ on Γ.

This alternative form of the optimality system enjoys a better B-differentiability property:

Theorem 5.1. *The system (5.6) is equivalent to (2.3) when μ and ϕ are computed by $\mu = \max\{0, \phi\}$ and (5.6d), respectively. The map $\Theta \ni \theta \mapsto (y[\theta], u[\theta], p[\theta], \phi[\theta])$ is B-differentiable with values in $Y \times L^p(\Omega) \times Y \times L^\infty(\Omega)$ for any $p \in [1, \infty)$. The derivative at θ_0 in the direction of $\delta\theta \in \Theta$ is given by the unique solution of the following system, where I_0 was defined in (3.3).*

$$-\Delta p + d_y(y_0, \theta_0)\, p + d_{yy}(y_0, \theta_0)\, y\, p_0 + y - b(\theta_0)\Pi_{I_0}(\phi) = \widehat{\delta}_1 \tag{5.7a}$$

$$\gamma(\theta_0)\, u - p - \Pi_{I_0}(\phi) = \widehat{\delta}_2 \tag{5.7b}$$

$$-\Delta y + d_y(y_0, \theta_0)\, y - u = \widehat{\delta}_3 \tag{5.7c}$$

$$\phi = \chi_S\big[\gamma(\theta_0)(-b(\theta_0) y + \widehat{\delta}_4) - p - \widehat{\delta}_2\big] \tag{5.7d}$$

in Ω, with boundary conditions $y = p = 0$ on Γ, where $\widehat{\delta} = -F_\theta(w_0, \theta_0)\, \delta\theta$.

Proof. The equivalence between (5.6) and (2.3) follows immediately from the definition of ϕ. From (5.6d) and due to the embedding $Y \hookrightarrow L^\infty(\Omega)$ it follows that $\phi[\theta]$ is B-differentiable with values in $L^\infty(\Omega)$. The conditions for the derivative (5.7) can be derived from (5.6) in a straightforward way. ☐

Now we consider problem $(\mathbf{P}(\theta))$ under small perturbations of the reference parameter θ_0, such that $\|\theta - \theta_0\| < \rho_2$ holds. We recall that our interest is in recovering an approximate solution $u[\theta]$ from the reference solution $u_0 = u[\theta_0]$ and derivative information. The motivation behind this approach is the fact that the evaluation of the derivative (i.e., the solution of the linear-quadratic problem $(\mathbf{DQP}(\widehat{\delta}))$) is numerically less expensive than the repeated solution of the nonlinear problem $(\mathbf{P}(\theta))$ for the perturbed parameter.

The B-differentiability of $u[\theta]$ suggests the use of the Taylor expansion

$$\mathcal{C}_u(\theta) := u_0 + u'[\theta_0](\theta - \theta_0) \tag{5.8}$$

as an update scheme for the control variable. We will develop an alternative update strategy based on the optimality condition (5.6b). We can solve that equation for $u[\theta]$ and obtain

$$u[\theta] = u_d(\theta) + \gamma(\theta)^{-1}\big(p[\theta] + \max\{0, \phi[\theta]\}\big).$$

Now, we use Taylor expansion of $p[\theta]$ and $\phi[\theta]$ on the right-hand side. The max-function however remains untouched. We thus obtain the following approximation of $u[\theta]$:

$$\mathcal{C}_u^{\mathrm{adj}}(\theta) := u_d(\theta) + \gamma(\theta)^{-1}\big(p[\theta_0] + p'[\theta_0](\theta-\theta_0) + \max\{0, \phi[\theta_0] + \phi'[\theta_0](\theta-\theta_0)\}\big). \tag{5.9}$$

In accordance with previous work [4], we call (5.9) the *adjoint update strategy.*

Theorem 5.2. *The update strategies $\mathcal{C}_u$ and $\mathcal{C}_u^{\mathrm{adj}}$ admit the following approximation properties:*

$$\frac{\|\mathcal{C}_u(\theta) - u[\theta]\|_{L^p(\Omega)}}{\|\theta-\theta_0\|} \to 0 \;\; \textit{as} \;\; \|\theta-\theta_0\| \to 0 \quad \textit{for all } p \in [1,\infty) \tag{5.10}$$

$$\frac{\|\mathcal{C}_u^{\mathrm{adj}}(\theta) - u[\theta]\|_{L^p(\Omega)}}{\|\theta-\theta_0\|} \to 0 \;\; \textit{as} \;\; \|\theta-\theta_0\| \to 0 \quad \textit{for all } p \in [1,\infty] \tag{5.11}$$

Proof. Equation (5.10) follows immediately from the B-differentiability result for $u[\cdot]$, see Theorem 4.2. To prove (5.11), we use the definition of $\mathcal{C}_u^{\mathrm{adj}}$ and the alternative form of the gradient equation (5.6b) to obtain

$$\mathcal{C}_u^{\mathrm{adj}}(\theta) - u[\theta] = \gamma(\theta)^{-1}\big(p[\theta_0] + p'[\theta_0](\theta-\theta_0) - p[\theta] + \max\{0, \phi[\theta_0] + \phi'[\theta_0](\theta-\theta_0)\} - \max\{0, \phi[\theta]\}\big)$$

First of all, we realize that $\gamma(\theta)^{-1}$ remains positive and bounded away from zero in a neighborhood of θ_0 by **(A3)**. Using the Lipschitz continuity of the $\max\{0,\cdot\}$ operation in $L^\infty(\Omega)$, we can estimate

$$\|\mathcal{C}_u^{\mathrm{adj}}(\theta) - u[\theta]\|_{L^\infty(\Omega)} \le c\,\big(\|p[\theta] - p[\theta_0] - p'[\theta_0](\theta-\theta_0)\|_{L^\infty(\Omega)} + \|\phi[\theta] - \phi[\theta_0] - \phi'[\theta_0](\theta-\theta_0)\|_{L^\infty(\Omega)}\big).$$

By Theorem 5.1, both $p[\cdot]$ and $\phi[\cdot]$ are B-differentiable at θ_0 with values in $L^\infty(\Omega)$, hence $\|\mathcal{C}_u^{\mathrm{adj}}(\theta) - u[\theta]\|_{L^\infty(\Omega)} = o\,(\|\theta-\theta_0\|)$ holds. □

Theorem 5.2 suggests that the adjoint update strategy $\mathcal{C}_u^{\mathrm{adj}}$ is better suited to approximate $u[\theta]$ than $\mathcal{C}_u$, since it allows an error estimate in $L^\infty(\Omega)$. This will be confirmed by numerical experiments in the next section.

Let us briefly discuss the computational cost of both update strategies $\mathcal{C}_u$ and $\mathcal{C}_u^{\mathrm{adj}}$. We recall from Theorem 4.2 that the evaluation of $u'[\theta_0](\theta-\theta_0)$ and simultaneously of $p'[\theta_0](\theta-\theta_0)$ requires the solution of the linear-quadratic optimal control problem ($\mathbf{DQP}(\widehat{\delta})$) with $\widehat{\delta} = -F_\theta(w_0,\theta_0)(\theta-\theta_0)$. When a semi-smooth Newton method [5, 17] is used for the solution of the corresponding optimality system (3.4), the term $\phi'[\theta_0](\theta-\theta_0)$, given by (5.7d), appears naturally in the determination of the active sets, hence it is available at no additional cost. We conclude that no additional system solves are necessary when using the alternative update strategy $\mathcal{C}_u^{\mathrm{adj}}$ instead of $\mathcal{C}_u$.

In an analogous way to (5.8) and (5.9) one can construct update rules also for the multiplier μ. Let us define

$$\mathcal{C}_\mu(\theta) := \mu_0 + \mu'[\theta_0](\theta - \theta_0), \tag{5.12}$$

$$\mathcal{C}_\mu^{\mathrm{adj}}(\theta) := \max\{0,\, \phi_0 + \phi'[\theta_0](\theta - \theta_0)\}. \tag{5.13}$$

As for the control approximations $\mathcal{C}_u$ and $\mathcal{C}_u^{\mathrm{adj}}$ one obtains

Corollary 5.3. *The update strategies $\mathcal{C}_\mu$ and $\mathcal{C}_\mu^{\mathrm{adj}}$ have the following approximation properties:*

$$\frac{\|\mathcal{C}_\mu(\theta) - \mu[\theta]\|_{L^p(\Omega)}}{\|\theta - \theta_0\|} \to 0 \;\; \textit{as} \;\; \|\theta - \theta_0\| \to 0 \quad \textit{for all } p \in [1, \infty) \tag{5.14}$$

$$\frac{\|\mathcal{C}_\mu^{\mathrm{adj}}(\theta) - \mu[\theta]\|_{L^p(\Omega)}}{\|\theta - \theta_0\|} \to 0 \;\; \textit{as} \;\; \|\theta - \theta_0\| \to 0 \quad \textit{for all } p \in [1, \infty]. \tag{5.15}$$

6 Numerical Experiments

In this section we report on some numerical experiments. The example employed is simple with respect to its structure. Its specific difficulty lies in the fact that no strict complementarity holds, hence the Bouligand derivative cannot be a linear (Fréchet) one, see the remark after the statement of Theorem 3.5. We consider

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|y\|^2_{L^2(\Omega)} + \frac{\gamma}{2}\|u\|^2_{L^2(\Omega)} \\ &\text{s.t.} \quad \begin{cases} -\Delta y + y^3 = u & \text{in } \Omega \\ \qquad\quad\; y = 0 & \text{on } \Gamma \end{cases} \\ &\text{and} \quad u + y - \theta \geq 0 \quad \text{in } \Omega. \end{aligned} \tag{$\mathbf{P}(\theta)$}$$

That is, the parameter $\theta \in L^\infty(\Omega)$ enters only in the mixed constraint. For the reference value θ_0 with $\theta_0(x_1, x_2) = \min\{\sin(3\pi x_1)\sin(3\pi x_2) - 0.5,\, 0\}$, a locally optimal solution is given by $u_0 = y_0 = p_0 = \mu_0 \equiv 0$, i.e., all quantities are identically zero. Strict complementarity does not hold since the constraint is active on a set of positive measure with a corresponding vanishing multiplier. All assumptions of the article are satisfied: The regularity assumption **(A9)** is satisfied because of $b = 1 > 0$. Also the second-order sufficient optimality condition **(A10)** holds at the reference solution:

$$L''(y_0, u_0, p_0, \mu_0)(y, u)^2 = \|y\|^2_{L^2(\Omega)} + \gamma\,\|u\|^2_{L^2(\Omega)} \geq \gamma\,\|u\|^2_{L^2(\Omega)} \quad \forall (y, u).$$

For the computations, we chose $\Omega = (0, 1)^2$ and $\gamma = 10^{-4}$. The problem was discretized by linear finite elements for all involved quantities with mesh size $h \approx 0.004$ which leads to more than $160{,}000$ degrees of freedom for each of u, y, p, μ. For the solution of the optimization problems, a semi-smooth Newton method (see e.g. [5, 17]) was applied to the non-smooth formulation (2.3) of the optimality system and similarly to $(\mathbf{DQP}(\widehat{\delta}))$ for the derivative problems.

We computed the corresponding solutions $(y[\theta_i], u[\theta_i], p[\theta_i], \mu[\theta_i], \phi[\theta_i])$ for a sequence of perturbations $\{\theta_i\}_{i=1}^n$. In addition, we also computed the first order updates $\mathcal{C}_u$, $\mathcal{C}_\mu$, as well as the adjoint ones $\mathcal{C}_u^{adj}$, $\mathcal{C}_\mu^{adj}$ by the schemes analyzed in Section 5. Each perturbation θ_i was obtained by a random perturbation of the finite element coordinates of θ_0. This allows us to verify that the estimate obtained in Theorem 5.2 are indeed uniform with respect to the parameter. The parameters have a prescribed norm, namely

$$\|\theta_i - \theta_0\|_{L^\infty(\Omega)} \leq \texttt{logspace(0,-4,}n\texttt{)} = 10^{-4\cdot\frac{i-1}{n-1}}, \quad i = 1, \ldots, n = 13.$$

In Figure 6.1 the approximation errors and error quotients are plotted. There, we used the following abbreviations for norms of errors in u, μ, p:

$$E_u := \|u[\theta] - \mathcal{C}_u(\theta)\|_{L^\infty(\Omega)} = \|u[\theta] - u[\theta_0] - u'[\theta_0](\theta - \theta_0)\|_{L^\infty(\Omega)},$$
$$E_\mu := \|\mu[\theta] - \mathcal{C}_\mu(\theta)\|_{L^\infty(\Omega)} = \|\mu[\theta] - \mu[\theta_0] - \mu'[\theta_0](\theta - \theta_0)\|_{L^\infty(\Omega)},$$
$$E_p := \|p[\theta] - p[\theta_0] - p'[\theta_0](\theta - \theta_0)\|_{L^\infty(\Omega)}.$$

The corresponding error quantities are denoted by

$$E_u^{\text{adj}} := \|u[\theta] - \mathcal{C}_u^{\text{adj}}(\theta)\|_{L^\infty(\Omega)},$$
$$E_\mu^{\text{adj}} := \|\mu[\theta] - \mathcal{C}_\mu^{\text{adj}}(\theta)\|_{L^\infty(\Omega)}.$$

In a similar manner, we denote the remainder quotients by

$$R_u := \frac{E_u}{\|\theta - \theta_0\|_{L^\infty(\Omega)}}, \quad R_u^{\text{adj}} := \frac{E_u^{\text{adj}}}{\|\theta - \theta_0\|_{L^\infty(\Omega)}},$$

and likewise for R_p and $R_\mu, R_\mu^{\text{adj}}$.

In the upper picture of Figure 6.1, we plotted the error norms. As one can see, the classical update strategies $\mathcal{C}_u$ and $\mathcal{C}_\mu$ (dashed lines) provide a converging sequence as the perturbation size tends to zero. As predicted, the remainder terms do not vanish, but instead these quotients remain almost constant (dashed lines in the lower picture).

For comparison, we also depicted the behaviour of error and remainder for the first-order update of the adjoint state (dotted lines). Since the mapping $p[\theta]$ is B-differentiable at θ_0 from Θ to $L^\infty(\Omega)$ by Theorem 4.2, we observe a vanishing remainder term for small perturbation sizes.

This behavior is then inherited by the auxiliary variable $\phi[\theta]$, which was used to formulate the adjoint update strategies $\mathcal{C}_u^{\text{adj}}$ and $\mathcal{C}_\mu^{\text{adj}}$ for control and multiplier. The related error norms and quotients are plotted in Figure 6.1 (solid lines). Clearly, the adjoint update strategies yields better approximations than the updates by the Taylor formula: They lead to smaller errors in the $L^\infty(\Omega)$ norm throughout. As predicted by Theorem 5.2 and Corollary 5.3, the remainder quotients for the adjoints updates tend to zero as $\theta \to \theta_0$. Thus, the results of Section 5 can also be observed numerically.

Since the adjoint update strategies are available without any additional computational cost, we advocate to use them to recover a perturbed from an unpertubed (nominal) solution.

7 Concluding Remarks

In this paper, we have performed a parametric sensitivity analysis for an optimal control problem with several mixed control-state inequality constraints. We point out that **(A7)**–**(A9)** are standard assumptions when proving the existence and uniqueness of Lagrange multipliers, and **(A10)** are typical and rather weak second-order sufficient conditions involving strongly active subsets. All assumptions were required to hold only at the reference solution $u[\theta_0]$. Nevertheless, this is enough to prove that a local solution map $\theta \mapsto w[\theta]$ of $(\mathbf{P}(\theta))$ exists which is Lipschitz in W and B-differentiable w.r.t. W_p for $p < \infty$.

We have devised the so-called adjoint update scheme in order to recover a perturbed solution $u[\theta]$ from a nearby reference value $u[\theta_0]$. In comparison with the classical Taylor expansion, the adjoint update strategy has better accuracy, which was proved and also observed numerically. Moreover, it carries no additional penalty

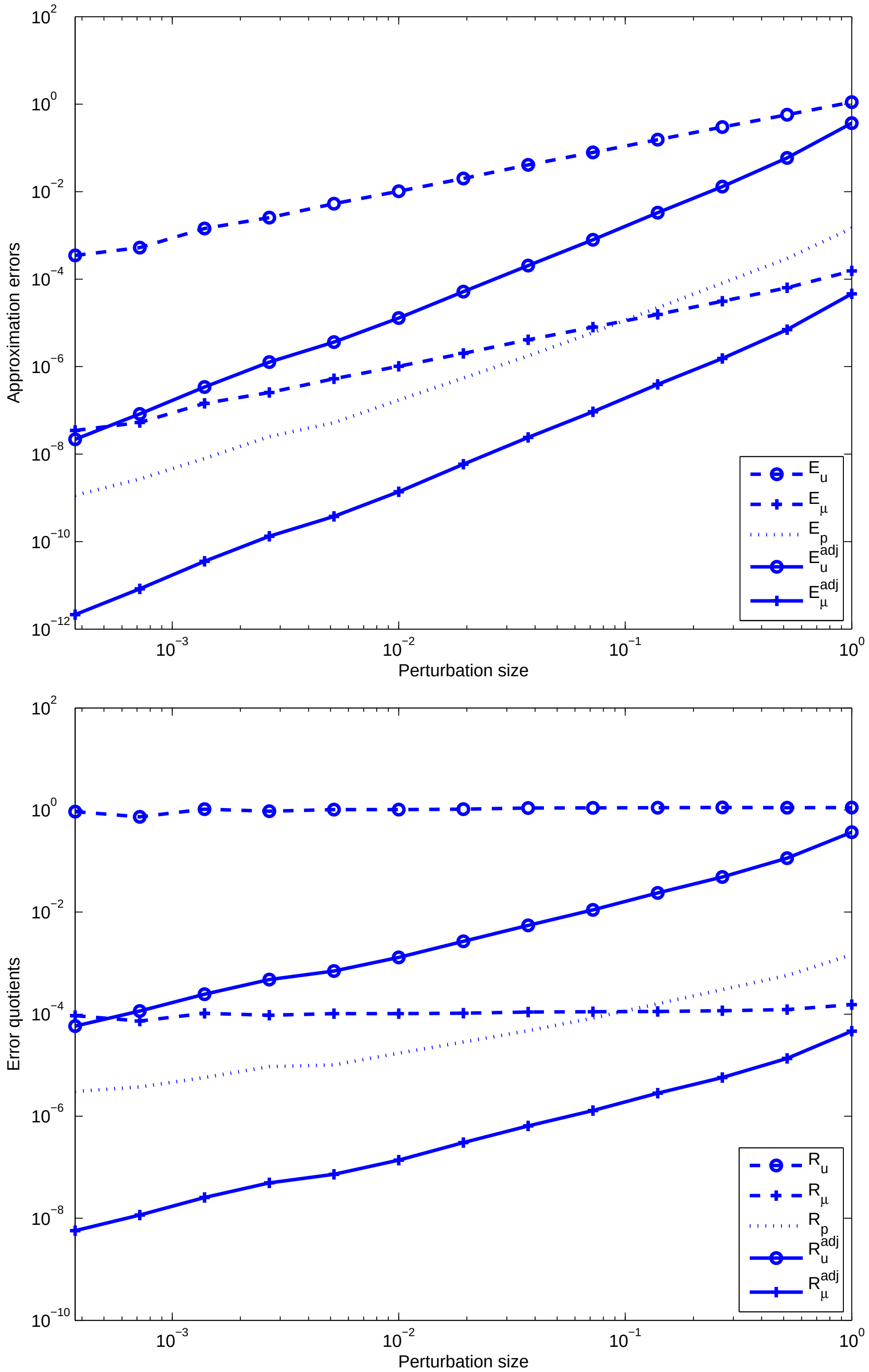

FIGURE 6.1. Approximation errors (top) and error quotients (bottom)

in terms of computation time because all involved quantities arise naturally when using semi-smooth Newton methods. We therefore advocate the use of the adjoint strategy.

We anticipate potential applications of our analysis for instance in nonlinear model predictive control (NMPC) algorithms, which repeatedly solve parametric sensitivity problems in order to update previous solutions to keep up with constantly changing parameters.

It is also possible to integrate the adjoint update into a path-following method for the solution of state-constrained problems as in [6]. There the (Fréchet-) derivative of solutions with respect to the Lavrientiev regularization parameter was employed to obtain intial guesses for the next solution step. Clearly, one can use the Bouligand derivative analyzed in the present article instead of the Fréchet derivative, if the assumption of strict complementarity is violated.

A Implicit Function Theorem

For easy reference, we cite here the Implicit Function Theorem from [3, Theorem 2.4] with adapted notation.

Theorem A.1. *Let W be a Banach space and let Θ, Z be normed linear spaces. Suppose that $F : W \times \Theta \to Z$ is a function and $\mathcal{N} : W \to Z$ is a set-valued map. Let $w_0 \in W$ be a solution to*

$$0 \in F(w, \theta) + \mathcal{N}(w) \tag{A.1}$$

for $\theta = \theta_0$, and let $\mathcal{Z}$ be a neighborhood of $0 \in Z$. Suppose that

- *(i) F is Lipschitz in θ, uniformly in w at (w_0, θ_0), and $F(w_0, \cdot)$ is B-differentiable at θ_0 with B-derivative $F_\theta(w_0, \theta_0)\,\delta\theta$ for all $\delta\theta \in \Theta$,*
- *(ii) F is partially Fréchet differentiable with respect to w in a neighborhood of (w_0, θ_0), and its partial derivative F_w is continuous in both w and θ at (w_0, θ_0),*
- *(iii) (A.1) is strongly regular at $(w_0, \theta_0$, i.e., there exists a function $\xi : \mathcal{Z} \to W$ such that $\xi(0) = w_0$, $\delta \in F(w_0, \theta_0) + F_w(w_0, \theta_0)(\xi(\delta) - w_0) + \mathcal{N}(\xi(\delta))$ for all $\delta \in \mathcal{Z}$, and ξ is Lipschitz continuous.*

Then there exist neighborhoods U of w_0 and V of θ_0 and a function

$$\theta \mapsto \Xi(\theta) = w[\theta]$$

from V to U such that $\Xi(\theta_0) = w_0$, $\Xi(\theta)$ is a solution of (A.1) *for every $\theta \in V$, and Ξ is Lipschitz continuous.*

If, in addition, $W_p \supset W$ is a normed linear space such that

- *(iv) $\xi : \mathcal{Z} \to W_p$ is B-differentiable at 0 with derivative $\xi'(0)\,\widehat{\delta}$ for all $\widehat{\delta} \in Z$,*

then $\theta \mapsto \Xi(\theta) \in W_p$ is also B-differentiable at θ_0 and its derivative is given by

$$\Xi'(\theta_0)\,\delta\theta = \xi'(0)\,(-F_\theta(w_0, \theta_0)\,\delta\theta), \tag{A.2}$$

for any $\delta\theta \in \Theta$.

References

[1] W. Alt, R. Griesse, N. Metla, and A. Rösch. Lipschitz stability for elliptic optimal control problems with mixed control-state constraints. *submitted*, 2006.

[2] F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, Berlin, 2000.

[3] A. Dontchev. Implicit function theorems for generalized equations. *Mathematical Programming*, 70:91–106, 1995.

[4] R. Griesse, T. Grund, and D. Wachsmuth. Update strategies for perturbed nonsmooth equations. *Optimization Methods and Software*, to appear.

[5] M. Hintermüller, F. Tröltzsch, and I. Yousept. Mesh-independence of semismooth Newton methods for Lavrentiev-regularized state constrained nonlinear optimal control problems. *Numerische Mathematik*, to appear.

[6] M. Hintermüller and I. Yousept. A sensitivity-based extrapolation technique for the numerical solution of state-constrained optimal control problems. *submitted*, 2007.

[7] K. Malanowski. Bouligand differentiability of solutions to parametric optimal control problems. *Numer. Funct. Anal. Optim.*, 22(7-8):973–990, 2001.

[8] K. Malanowski. Stability and sensitivity analysis for optimal control problems with control-state constraints. *Dissertationes Math. (Rozprawy Mat.)*, 394:51, 2001.

[9] K. Malanowski. Remarks on differentiability of metric projections onto cones of nonnegative functions. *Journal of Convex Analysis*, 10(1):285–294, 2003.

[10] K. Malanowski. Stability and sensitivity analysis for linear-quadratic optimal control subject to state constraints. *Optimization*, 56(4):463–478, 2007.

[11] K. Malanowski. Sufficient optimality conditions in stability analysis for state-constrained optimal control. *Appl. Math. Optim.*, 55(2):255–271, 2007.

[12] K. Malanowski and F. Tröltzsch. Lipschitz stability of solutions to parametric optimal control for parabolic equations. *ZAA*, 18:469–489, 1999.

[13] K. Malanowski and F. Tröltzsch. Lipschitz stability of solutions to parametric optimal control for parabolic equations. *Journal of Analysis and its Applications*, 18(2):469–489, 1999.

[14] C. Meyer, U. Prüfert, and F. Tröltzsch. On two numerical methods for state-constrained elliptic control problems. *Optimization Methods and Software*, 22(6):871–899, 2007.

[15] C. Meyer and F. Tröltzsch. On an elliptic optimal control problem with pointwise mixed control-state constraints. In *Recent advances in optimization*, volume 563 of *Lecture Notes in Econom. and Math. Systems*, pages 187–204. Springer, Berlin, 2006.

[16] A. Rösch and F. Tröltzsch. On the regularity of solutions and Lagrange multipliers of optimal control problems for semilinear equations with mixed pointwise control-state constraints. *SIAM Journal on Control and Optimization*, 46(3):1098–1115, 2007.

[17] A. Rösch and D. Wachsmuth. Semi-smooth Newton's method for an optimal control problem with control and mixed control-state constraints. *Matheon Preprint 415, submitted*, 2007.

Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstrasse 69, A–4040 Linz, Austria

E-mail address: roland.griesse@oeaw.ac.at

URL: http://www.ricam.oeaw.ac.at/people/page/griesse

Technische Universität Berlin, Sekretariat MA 4-5, Strasse des 17. Juni 136, D–10623 Berlin, Germany

E-mail address: wachsmut@math.tu-berlin.de

URL: http://www.math.tu-berlin.de/~wachsmut/