Optimization and Engineering (2021) 22:2039–2085
https://doi.org/10.1007/s11081-020-09547-2

RESEARCH ARTICLE

Convergence of the SQP method for quasilinear parabolic optimal control problems

Fabian Hoppe[1] · Ira Neitzel[1]

Received: 30 October 2019 / Revised: 24 July 2020 / Accepted: 24 July 2020 /
Published online: 3 September 2020

Abstract

Based on the theoretical framework recently proposed by Bonifacius and Neitzel (Math Control Relat Fields 8(1):1–34, 2018. https://doi.org/10.3934/mcrf.2018001) we discuss the sequential quadratic programming (SQP) method for the numerical solution of an optimal control problem governed by a quasilinear parabolic partial differential equation. Following well-known techniques, convergence of the method in appropriate function spaces is proven under some common technical restrictions. Particular attention is payed to how the second order sufficient conditions for the optimal control problem and the resulting L^2-local quadratic growth condition influence the notion of "locality" in the SQP method. Further, a new regularity result for the adjoint state, which is required during the convergence analysis, is proven. Numerical examples illustrate the theoretical results.

Keywords Optimal control · Quasilinear parabolic partial differential equation · Sequential quadratic programming · Convergence analysis

Mathematics Subject Classification 35K59 · 49K20 · 90C48 · 49N60 · 65K10 · 90C55 · 49M15 · 49M37

1 Overview

Optimal control problems governed by linear and semilinear parabolic partial differential equations (PDEs) have been subject to intense research for several years.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer 211504053—SFB 1060.

✉ Fabian Hoppe
hoppe@ins.uni-bonn.de

Ira Neitzel
neitzel@ins.uni-bonn.de

[1] Institut für Numerische Simulation, Universität Bonn, Endenicher Allee 19b, 53115 Bonn, Germany

Existence- and regularity of their solutions is well understood, first order necessary and second order sufficient optimality conditions have been proven, and discretization errors for different types of discretization are available, see e.g. the pioneering work of Lions (1971) concerned with linear PDEs and Hinze et al. (2009), or Tröltzsch (2010) for a recent overview covering theoretical and numerical aspects of both linear and nonlinear problems.

Recently, optimal control of quasilinear parabolic equations was addressed by Bonifacius and Neitzel (2018), Casas and Chrysafinos (2018), and Meinlschmidt et al. (2017a, b), Meinlschmidt and Rehberg (2016). The functional analytic framework for the analysis of the state equation is provided by the concept of maximal parabolic regularity of nonautonomous operators, see e.g. the work of Amann (2004, 2003, 2005), Meinlschmidt and Rehberg (2016), Haller-Dintelmann and Rehberg (2009), or further references in Bonifacius and Neitzel (2018). The highly non-trivial existence and regularity theory for solutions of the underlying PDE poses the main difficulty in the theoretical analysis of such problems. For a discussion of previous literature concerning optimal control of quasilinear PDEs see the introduction of Bonifacius and Neitzel (2018) and Casas and Chrysafinos (2018), respectively. In particular, optimal control of quasilinear elliptic equations has been considered by Casas and Tröltzsch (2009, 2011, 2012), Casas and Dhamo (2011) and Yousept (2013), de Los Reyes and Dhamo (2016), Nicaise and Tröltzsch (2017). Several physical models lead to quasilinear PDEs (e.g. temperature-dependent thermal conductivity), which motivates the analysis of this challenging class of problems from the applied point of view, see e.g. the so-called thermistor problem (Meinlschmidt et al. 2017a, b).

For the efficient numerical solution of nonlinear optimal control problems sequential quadratic programming (SQP) methods form a prominent class of state of the art algorithms: The nonlinear optimization problem is approximated by a sequence of linear quadratic subproblems that can be solved e.g. by application of the well-understood primal dual active set strategy. The analysis of such SQP methods for nonlinear optimal control problems has been addressed by several researchers, see e.g. Tröltzsch (1999), Goldberg and Tröltzsch (1998) for semilinear parabolic equations, Hintermüller and Hinze (2006), Hinze and Kunisch (2001), Wachsmuth (2007) for optimal control of time-dependent Navier–Stokes equation, Griesse et al. (2010), Griesse et al. (2008) for semilinear elliptic problems with mixed constraints, and Heinkenschloss and Tröltzsch (1999) for optimal control of a phase field equation. For an overview concerning the origins of SQP methods in the context of PDE-constrained optimization we also refer to the introduction of Goldberg and Tröltzsch (1998). As further second order methods for the solution of nonlinear optimal control problems we mention the semismooth Newton method and versions of the primal dual active set strategy, respectively, see e.g. Hinze and Kunisch (2001), Hintermüller et al. (2007), Ito and Kunisch (2004).

In the present paper, we focus on the numerical solution of quasilinear parabolic optimal control problems by the SQP method. To our best knowledge, a corresponding convergence analysis in function space has not been carried out in the existing literature. The most closely related existing publications are those by Ulbrich and Ziems (Ulbrich and Ziems 2017; Ziems 2013; Ziems and Ulbrich 2011) and chapter 8 in the thesis of Feldhordt (2017), respectively. Ulbrich and Ziems consider trust-region and trust-region SQP methods for optimal control of general nonlinear PDE.

The main difference to our result is that they include discretization in their work and prove convergence of adaptive multilevel algorithms whereas we stick to the function space setting. In return, we are able to prove locally superlinear convergence around local minima fulfilling certain second order conditions avoiding the two norm gap (Ioffe 1979; Casas and Tröltzsch 2012), whereas Ulbrich and Ziems establish global convergence to a point fulfilling first order optimality conditions, but without explicit rate. Feldhordt (2017) considers optimal control of the so-called chemotaxis system and proves convergence of the SQP method assuming a rather strong second order sufficient condition. This corresponds to our interim result in Section 6.1, whereas our main focus during the rest of the paper is on the interplay of weaker second order conditions and the notation of "locality" in the SQP method. The second order sufficient conditions we refer to in the present paper are due to Bonifacius and Neitzel (2018). For the topic of second order conditions in PDE-constrained optimization in general we refer to Goldberg and Tröltzsch (1989), Bonnans (1998), or the recent survey by Casas and Tröltzsch (2015) and the references therein.

Many of our arguments in the present paper are similar to those known from earlier publications. However, we believe that our consideration is of interest for three main reasons:

First, we demonstrate that the results on optimal control of quasilinear parabolic PDE obtained by Bonifacius and Neitzel (2018) allow to derive convergence of the SQP method. In particular, existence and regularity theory of quasilinear parabolic PDE is much more involved than the corresponding treatment of semilinear PDE. This makes the choice of the correct function spaces more complicated than in previous work on SQP methods and we believe that it is not clear a-priori that—in the end—the arguments from the existing literature apply to the present model problem as well.

Second, we show a new regularity result for the adjoint state in Sect. 7. The proof relies on maximal parabolic regularity arguments and is based on the work of Bonifacius and Neitzel (2018) and Haller-Dintelmann and Rehberg (2009). The result is crucial for our further analysis, because the improved regularity allows us to estimate the second derivative of the nonlinearity of the state equation in an appropriate way.

Finally, most proofs concerning convergence of the SQP method have been published before the introduction of a framework for second order sufficient conditions without two norm gap by Casas and Tröltzsch (2012). As shown by Bonifacius and Neitzel (2018) our model problem fits into this framework and hence it is natural to revisit convergence theory—and in particular, the question of localization of the quadratic subproblems—of the SQP method under the aspect of absence of the two norm gap: Since quadratic growth of the reduced objective functional holds L^2-locally (instead of L^∞-locally) around the optimal control, one may wonder, whether it is possible to replace L^∞-neighbourhoods from previous convergence proofs for the SQP method by L^2-neighbourhoods. The answer of this question is not straightforward, due to the fact that convergence of the SQP method is—as usual—established by showing convergence of a generalized Newton method for a certain generalized (set-valued) equation: In order to obtain a differentiable map in this generalized equation we still need to measure controls in a norm stronger than the L^2-norm. In contrast, the regularity property (analogous to the invertibility of the Hessian in the classical Newton method) relies on the L^2-coercivity property due to the second order sufficient conditions. —For our model problem, we give an answer to this question in Sect. 6.3, which is our main result.

The rest of this paper is organized as follows and keeps the main structure of previous results concerning the analysis of SQP methods, cf. in particular the work of Tröltzsch (1999), Wachsmuth (2007) and Goldberg and Tröltzsch (1998):

In Sects. 2 and 3 we briefly recall the assumptions and the model problem as well as its first order optimality conditions from Bonifacius and Neitzel (2018). The idea of the SQP method is outlined together with appropriate second order sufficient conditions. To prepare the analysis of the convergence properties of the SQP method, we provide some auxiliary results that are specifically related to our quasilinear parabolic model problem in Sect. 4. The proof of a new regularity result for the adjoint state is postponed to Sect. 7. After that, we follow the standard argument to prove convergence of the SQP method in Sects. 5 and 6: We utilize the connection to the Josephy-Newton method for a generalized equation originating from the first order optimality conditions. Convergence of this Newton method is proven in Sect. 5 and the interpretation of the iterates as the solutions of certain quadratic optimization problems is topic of Sect. 6. Assuming strong second order sufficient conditions we formulate our first main result in Sect. 6.1. The remaining two theoretical Sects. 6.2 and 6.3 of the paper are devoted to the analysis of the generalized Newton and the SQP method under weaker second order assumptions. In particular we are able to replace the L^∞-neighbourhoods in the results of Tröltzsch (1999) and Wachsmuth (2007) by L^2-neighbourhoods in our final result in Sect. 6.3. For a detailed overview of this part of the paper we refer to the introduction of Sect. 6. Finally, we give short numerical examples that illustrate our theoretical findings in Sect. 8.

Notation For a Lipschitz domain Ω and $\theta \in (0, 1]$, $k \in \mathbb{N}$, $p \in [1, \infty]$ we denote by $L^p = L^p(\Omega)$, $H^{\theta,p} = H^{\theta,p}(\Omega)$ and $W^{k,p} = W^{k,p}(\Omega)$ the usual Lebesgue-, Bessel-potential- and Sobolev-spaces, respectively. For the two latter families of spaces a subscript D denotes incorporation of previously defined homogeneous Dirichlet boundary conditions. With $H_D^{-\theta,p'}$ and $W_D^{-1,p'}$ we refer to the topological dual spaces of $H_D^{\theta,p}$ and $W_D^{1,p}$, where $\langle\cdot,\cdot\rangle$ stands for the duality pairing and—in case of Hilbert spaces—the scalar product. Norms $\|\cdot\|$ are indexed by the space they refer to. For some integrability exponent $r \in [1, \infty]$, we define the conjugate exponent r' by $1/r + 1/r' = 1$. Spaces of countinuously differentiable resp. Hölder continuous functions are denoted as usual by $\mathscr{C}^\alpha$.

The open and closed balls of radius $r > 0$ around x_0 in a Banach space X are denoted by

$$\mathbb{B}_r^X(x_0) := \{x \in X\colon \|x - x_0\|_X < r\} \quad \text{and} \quad \overline{\mathbb{B}_r^X(x_0)} := \{x \in X\colon \|x - x_0\|_X \leq r\}.$$

With $(X, Y)_{r,s}$ or $[X, Y]_r$ we refer to real or complex interpolation spaces of two normed spaces X, Y, respectively. Given $I \subset \mathbb{R}$, a Banach space X, and a function $\phi\colon I \to X$, we denote by $\mathrm{tr}_t\phi$, $t \in I$, the trace $\phi(t) \in X$, if such a pointwise evaluation is well-defined.

The notation "$\ldots \lesssim \ldots$" will be used in order to express that "$\ldots \leq C \cdot \ldots$" holds with a generic constant $C > 0$, whose dependencies are not relevant for the present context. We use double arrows "$\rightrightarrows$" to indicate set-valued maps.

2 Model problem and assumptions

2.1 The model problem

Our model problem is the same as the one in Example 2.5 of Bonifacius and Neitzel (2018), and reads as follows:

$$\left\{\begin{array}{l} \min\limits_{y,u} J(y,u) := \dfrac{1}{2}\|y - y_d\|^2_{L^2(I\times\Omega)} + \dfrac{\gamma}{2}\|u\|^2_{L^2(\Lambda)}, \\ \text{subject to} \quad u \in U_{ad} \quad \text{and} \quad \left\{\begin{aligned} \partial_t y + \mathscr{A}(y)y &= Bu \\ y(0) &= y_0. \end{aligned}\right. \end{array}\right. \tag{OCP}$$

Here, the quasilinear part $\mathscr{A}$ of the state equation is defined by

$$\mathscr{A}(y)\cdot \; := -\mathrm{div}(\xi(y)\mu\nabla\cdot\,),$$

The control operator B, Λ, and the admissible set U_{ad} will be specified in the following section.

2.2 Assumptions

We rely on the following assumptions that we repeat from Bonifacius and Neitzel (2018) with minor changes, cf. the following remark.

Assumption 2.1 $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, is a bounded domain with boundary $\partial\Omega$. $\Gamma_N \subset \partial\Omega$ is relatively open and denotes the Neumann boundary part whereas $\Gamma_D = \partial\Omega \setminus \Gamma_N$ denotes the Dirichlet boundary part equipped with homogeneous Dirichlet boundary conditions. We assume that $\Omega \cup \Gamma_N$ is Gröger regular such that every chart map in the Definition of Gröger regularity can be chosen volume preserving. The time interval $I = (0, T)$ with $T > 0$ is fixed.

For the definition of Gröger regularity we refer to the work of Gröger (1989). It has been used in a notation similiar to this work e.g. in Bonifacius and Neitzel (2018, Definition A.1). In Section 5 of Haller-Dintelmann et al. (2009) an alternative characterization can be found. The additional requirement of volume preserving chart maps is satisfied in particular by all domains with Lipschitz boundary ("strong Lipschitz domain"), but also e.g. by two crossing beams in dimension three (Haller-Dintelmann and Rehberg 2009, Remark 3.3 and section 7.3).

Assumption 2.2 The function $\xi\colon \mathbb{R} \to \mathbb{R}$ is twice differentiable with ξ'' being Lipschitz continuous on bounded subsets of $\mathbb{R}$. Let $\mu\colon\ \Omega \to \mathbb{R}^{d\times d}$, $\mu = \mu^T$, be measurable and uniformly bounded and coercive in the following sense:

$$0 < \mu_\bullet := \inf_{x\in\Omega} \inf_{z\in\mathbb{R}^d\setminus\{0\}} \frac{z^T\mu(x)z}{z^Tz}, \qquad \mu^\bullet := \sup_{x\in\Omega} \sup_{1\le i,j\le d} |\mu_{i,j}(x)| < \infty$$

We assume a coercivity condition $0 < \xi_\bullet \le \xi \le \xi^\bullet$ for ξ as well. With this we define as above

$$\langle \mathscr{A}(y)\varphi, \psi\rangle_{L^2(I,W_D^{1,2})} := \int_I \int_\Omega \xi(y)\mu\nabla\varphi\nabla\psi\, dxdt, \qquad \varphi, \psi \in L^2(I, W_D^{1,2}),$$

whenever y is a measurable function on Ω.

Assumption 2.3 We assume that there is $p \in (d, 4)$ such that

$$-\text{div}(\mu\nabla\cdot) + 1\colon\ W_D^{1,p} \to W_D^{-1,p}$$

is a topological isomorphism and fix this choice of p.

Assumption 2.4 Let $\zeta \in (0, 1)$ and $s > 2$ be fixed such that

$$\max\left\{1 - \frac{1}{p}, \frac{d}{p}\right\} < \zeta \qquad \text{and} \qquad \max\left\{\frac{2}{\zeta - d/p}, \frac{2}{1-\zeta}\right\} < s$$

holds. By $\mathscr{D}$ we denote the domain of the unbounded operator $-\text{div}(\mu\nabla\cdot) + 1$ in the Bessel potential space $H_D^{-\zeta,p}$. The desired state $y_d \in L^\infty(I, L^{p/2})$, the initial condition for the state equation $y_0 \in (H_D^{-\zeta,p}, \mathscr{D})_{1-1/s,s}$ and the regularization parameter $\gamma > 0$ are fixed.

We introduce the measure space (Λ, ρ) by $\Lambda = \{\bullet\}^m \times I$ equipped with measure ρ being the product of the counting measure on the $m-$element set $\{\bullet\}^m$ with the Lebesgue measure on I. Within the control space $U := L^s(\Lambda, \rho) = L^s(I, \mathbb{R}^m)$ the set of admissible controls is given by

$$U_{ad} := \{u \in U\colon\ u_a \le u \le u_b \qquad \rho\text{-a.e. on } \Lambda\}$$

with fixed control bounds $u_a, u_b \in L^\infty(\Lambda)$. Finally, for fixed control basis functions $b_1, \ldots, b_m \in H_D^{-\zeta,p}$ we define the bounded linear control operator by

$$B\colon\ U \to L^s(I, H_D^{-\zeta,p}), \qquad (Bu)(t) := \sum_{i=1}^m u_i(t)b_i.$$

Remark 2.5 The choice of control space and operator ("purely time-dependent controls") corresponds to Example 2.5 of Bonifacius and Neitzel (2018), where the control space is chosen as $L^\infty(\Lambda)$ instead of $L^s(\Lambda)$. We will make use of measuring controls in L^s instead of L^∞ when applying the Riesz-Thorin interpolation theorem, see the remark concluding Sect. 6.1. The reason for choosing purely time-dependent controls—apart from practical motivation, see e.g. de Los Reyes et al. (2008)—is outlined in the remark at the end of Sect. 5.1. The symmetry property $\mu = \mu^T$ as well as the slightly higher spatial integrability of the desired state y_d ($L^{p/2}$ instead of L^2) are required to derive improved regularity for the adjoint state in Sect. 7.

3 Optimality conditions and SQP method

We follow Goldberg and Tröltzsch (1998), Tröltzsch (1999), Wachsmuth (2007). From Bonifacius and Neitzel (2018), Section 4.1, recall the following notation:

$$\begin{aligned} \mathscr{A}'(y)v &:= -\operatorname{div}\left(\xi'(y)v\mu\nabla y\right), \\ \mathscr{A}''(y)[v_1, v_2] &:= -\operatorname{div}\left(\xi'(y)(v_1\mu\nabla v_2 + v_2\mu\nabla v_1) + \xi''(y)v_1v_2\mu\nabla y\right) \end{aligned}$$

for $v, v_1, v_2 \in W^{1,r}(I, W_D^{-1,p}) \cap L^r(I, W_D^{1,p})$, $r \in (1, \infty)$ and a measurable function y on $I \times \Omega$. The divergence operators have to be understood in weak form, of course.

3.1 First order necessary optimality conditions

In Bonifacius and Neitzel (2018), Lemma 4.1, the existence of a global solution to (**OCP**) is established. Further, any local solution to (**OCP**) fulfills the following system of equations, cf. Bonifacius and Neitzel (2018, Lemmas 4.6-4.8):

$$\partial_t y + \mathscr{A}(y)y = Bu, \; y(0) = y_0 \tag{SE}$$

$$\begin{gathered} -\partial_t p + \mathscr{A}(y)^* p + \mathscr{A}'(y)^* p = y - y_d, \\ p(T) = 0 \end{gathered} \tag{AE}$$

$$(\gamma u + B^* p, v - u)_{L^2(\Lambda)} \geq 0 \text{ for all } v \in U_{ad}, \tag{FON}$$

This optimality system consists of the state equation (SE), the adjoint equation (AE), and the variational inequality (FON). The underlying function spaces are introduced in the next section. For reasons of shortness we will sometimes write the state equation as

$$e(y, u) := (\partial_t y + \mathscr{A}(y)y - Bu, \quad \mathrm{tr}_0 y - y_0) = 0 \tag{1}$$

with the $\mathscr{C}^2$-map

$$e \colon (W^{1,s}(I, W_D^{-1,p}) \cap L^s(I, W_D^{1,p})) \times L^s(\Lambda) \to L^s(I, W_D^{-1,p}) \times (W_D^{-1,p}, W_D^{1,p})_{1-1/s,s}.$$

By $L_{OCP}(y,u,p) := J(y,u) - \langle p, e_1(y,u)\rangle$ we denote the Lagrangian of (**OCP**).

3.2 Generalized equation and SQP method

We reformulate the optimality system as the generalized equation

$$0 \in F(y,u,p) + N(y,u,p) \tag{GE}$$

with the maps

$$F(y,u,p) := \begin{pmatrix} \partial_t y + \mathscr{A}(y)y - Bu \\ \mathrm{tr}_0 y - y_0 \\ -\partial_t p + \mathscr{A}(y)^* p + \mathscr{A}'(y)^* p - (y - y_d) \\ \mathrm{tr}_T p \\ \gamma u + B^* p \end{pmatrix}$$
$$\text{and} \qquad N(y,u,p) := \big(\{0\}, \{0\}, \{0\}, \{0\}, N_{U_{ad}}(u)\big)^T,$$

where $N_{U_{ad}}(u)$ denotes the normal cone of the closed convex set U_{ad} at the point $u \in L^s(\Lambda)$, i.e. $N_{U_{ad}}(u) = \big\{v \in L^s(\Lambda)\colon (v, w-u)_{L^2(\Lambda)} \le 0 \text{ for all } w \in U_{ad}\big\}$. To make the definition of F and N precise, F is understood as map $F\colon X_s \to Z_s$ with

$$X_s := \Big(W^{1,s}(I, W_D^{-1,p}) \cap L^s(I, W_D^{1,p})\Big) \times L^s(\Lambda)$$
$$\times \Big(W^{1,s}(I, W_D^{-1,p'}) \cap L^s(I, W_D^{1,p'})\Big)$$

and

$$Z_s := L^s(I, W_D^{-1,p}) \times (W_D^{-1,p}, W_D^{1,p})_{1-1/s,s} \times L^s(I, W_D^{-1,p'})$$
$$\times (W_D^{-1,p'}, W_D^{1,p'})_{1-1/s,s} \times L^s(\Lambda).$$

Accordingly, N is understood as set valued map $X_s \rightrightarrows Z_s$. We equip X_s and Z_s with the canonical norms

$$\|(y,u,p)\|_{X_s} := \|y\|_{W^{1,s}(I,W_D^{-1,p}) \cap L^s(I,W_D^{1,p})} + \|u\|_{L^s(\Lambda)}$$
$$+ \|p\|_{W^{1,s}(I,W_D^{-1,p'}) \cap L^s(I,W_D^{1,p'})},$$
$$\|(f, y_0, g, p_T, r)\|_{Z_s} := \|f\|_{L^s(I,W_D^{-1,p})} + \|y_0\|_{(W_D^{-1,p},W_D^{1,p})_{1-1/s,s}} + \|g\|_{L^s(I,W_D^{-1,p'})}$$
$$+ \|p_T\|_{(W_D^{-1,p'},W_D^{1,p'})_{1-1/s,s}} + \|r\|_{L^s(\Lambda)}.$$

Having chosen these spaces, the following result holds:

Lemma 3.1 *$F\colon X_s \to Z_s$ is continuously Fréchet differentiable and $N\colon X_s \rightrightarrows Z_s$ has closed graph.*

Proof Differentiability has been used implicitely by (Bonifacius and Neitzel 2018, Lemma 4.5) where the differentiability of the control to state map is shown by the implicit function theorem. The closed graph property is standard. □

Remark 3.2 Note that we require time integrability $s \gg 1$ as in Assumption 2.4 for the y-component in order to have the embedding

$$W^{1,s}(I, W_D^{-1,p}) \cap L^s(I, W_D^{1,p}) \hookrightarrow L^\infty(I \times \Omega),$$

cf. Bonifacius and Neitzel (2018, Proposition 3.3). The latter is needed to ensure differentiability of the superposition operators associated with ξ and ξ', and hence differentiability of F. In return, this implies by the definition of F that we have to consider L^s-integrable control functions u, i.e. (GE) cannot be stated with controls measured in L^2.

Sometimes we will need the following subspaces X_∞ and Z_∞ of X_s, Z_s:

$$\begin{aligned} X_\infty &:= \left(W^{1,s}(I, W_D^{-1,p}) \cap L^s(I, W_D^{1,p})\right) \times L^\infty(\Lambda) \\ &\quad \times \left(W^{1,s}(I, W_D^{-1,p'}) \cap L^s(I, W_D^{1,p'})\right), \\ Z_\infty &:= L^s(I, W_D^{-1,p}) \times (W_D^{-1,p}, W_D^{1,p})_{1-1/s,s} \times L^s(I, W_D^{-1,p'}) \\ &\quad \times (W_D^{-1,p'}, W_D^{1,p'})_{1-1/s,s} \times L^\infty(\Lambda), \end{aligned}$$

equipped with the canonical norms similarly as above. Note that changing from X_s, Z_s to X_∞, Z_∞ means nothing more than replacing the $L^s(\Lambda)$-factors by $L^\infty(\Lambda)$-factors, i.e. considering controls in the L^∞- instead of the L^s-norm. The same result as before holds:

Lemma 3.3 *$F\colon X_\infty \to Z_\infty$ is continuously Fréchet differentiable and $N\colon X_\infty \rightrightarrows Z_\infty$ has closed graph.*

Due to Lemma 3.1 we can formulate the ansatz of the SQP method in its abstract form as the Josephy-Newton method for generalized equations, see Josephy (1979), Dontchev (1996), Alt (1990), or Hinze et al. (2009, chapter 2): Given an iterate $(y_k, u_k, p_k) \in X_s$, solve

$$0 \in F(y_k, u_k, p_k) + F'(y_k, u_k, p_k)(y - y_k, u - u_k, p - p_k) + N(y, u, p) \tag{2}$$

to obtain the new iterate $(y_{k+1}, u_{k+1}, p_{k+1}) \in X_s$. Writing down the full system of equations for (2) we find:

$$\begin{aligned} \partial_t y + \mathscr{A}(y_k)y + \mathscr{A}'(y_k)y &= Bu + \mathscr{A}'(y_k)y_k \\ \mathrm{tr}_0 y &= y_0 \end{aligned} \tag{3}$$

$$\begin{aligned} -\partial_t p + \mathscr{A}(y_k)^* p + \mathscr{A}'(y_k)^* p &= y - y_d - \mathscr{A}''(y_k)[y - y_k, \cdot]^* p_k \\ \mathrm{tr}_T p &= 0 \end{aligned} \tag{4}$$

$$0 \in \gamma u + B^* p + N_{U_{ad}}(u). \tag{5}$$

Obviously, the current u-iterate u_k has canceled out, which implies that the next iterate (y, u, p) depends on y_k and p_k but not on u_k. This is due to the structure of our model problem. Note that the first two equations (3) are equivalent to the linearized state equation

$$0 = e(y_k, u_k) + e_y(y_k, u_k)(y - y_k) + e_u(y_k, u_k)(u - u_k). \tag{6}$$

A standard computation shows that

$$\frac{1}{2} L''_{OCP}(y_k, u_k, p_k)[(y - y_k, u - u_k)]^2 + J'(y_k, u_k)(y - y_k, u - u_k) \tag{7}$$

is equal (up to addition of constants) to the expression

$$J_k(y, u) := \frac{1}{2}\|y - y_d\|^2 + \frac{\gamma}{2}\|u\|^2 - \frac{1}{2}\langle p_k, \mathscr{A}''(y_k)[y - y_k, y - y_k]\rangle, \tag{8}$$

that finally fulfills: The system of equations (3),(4),(5) is the formal optimality system of the following optimal control problem:

$$\begin{cases} \min\limits_{y,u} J_k(y, u) \\ \text{subject to} \quad u \in U_{ad} \quad \text{and equation (3).} \end{cases} \tag{\textbf{QP}}$$

This is the classical formulation of the SQP method as sequence of quadratic problems to solve. Note that these computations were completely formal in the sense that we do not know whether (**QP**) is convex or not. Hence, we cannot say whether there is a unique minimizer or whether the optimality system (3),(4),(5) is a sufficient characterization for this minimizer. This issue will be addressed in the following section utilizing the assumption of second order sufficient conditions.

3.3 Second order sufficient conditions and SQP

Depending on second order sufficient conditions (SSCs) for (**OCP**) based on those derived in Bonifacius and Neitzel (2018) we have to restrict the admissible set for (**QP**) to ensure convexity.

Assumption 3.4 From now on let $\bar{u} \in U_{ad}$ be a fixed L^2-local minimizer for (**OCP**), i.e. there is $r > 0$ such that

$$u \in U_{ad} \text{ and } \|u - \bar{u}\|_{L^2(I,\mathbb{R}^m)} < r \qquad \Longrightarrow \qquad j(u) \geq j(\bar{u}).$$

Let $\bar{y}$ and $\bar{p}$ the state and adjoint state associated with $\bar{u}$. For $\sigma \in [0, \infty]$ we define the σ-active set of $\bar{u}$ as

$$A^\sigma(\bar{u}) := \{x \in \Lambda\colon |\gamma\bar{u} + B^*\bar{p}|(x) > \sigma\}$$

and the corresponding subspace

$$C^\sigma(\bar{u}) := \{v \in L^2(\Lambda)\colon\, v = 0 \text{ on } A^\sigma(\bar{u})\}$$

of directions vanishing on $A^\sigma(\bar{u})$. We assume that the following second order sufficient condition for (**OCP**) is satisfied at $\bar{u}$: There is a fixed $\sigma \in [0, \infty]$ (whether we allow the case $\sigma = 0$ or not will be stated in our further results) such that there exists $\delta > 0$ such that

$$\begin{cases} L''_{OCP}(\bar{y}, \bar{u}, \bar{p})[(y, u)]^2 \geq \delta \|u\|^2_{L^2(\Lambda)} \\ \text{for all} \quad (y, u) \in W^{1,2}(I, W_D^{-1,p}) \cap L^2(I, W_D^{1,p}) \times L^2(\Lambda) \quad \text{s.t.} \\ \qquad u \in C^\sigma(\bar{u}), \\ \qquad e_y(\bar{y}, \bar{u})y + e_u(\bar{y}, \bar{u})u = 0. \end{cases} \tag{SSC-σ}$$

Condition (SSC-σ) is stronger than the second order sufficient condition derived by Bonifacius and Neitzel (2018, Theorem 4.14) which has smallest possible gap to the corresponding necessary condition. However we conclude from the cited result:

Theorem 3.5 *Let Assumption 3.4 hold with some* $\sigma \in [0, \infty]$. *Then there are* $\epsilon, \eta > 0$ *such that the quadratic growth condition*

$$j(u) \geq j(\bar{u}) + \eta \|u - \bar{u}\|^2_{L^2(\Lambda)}$$

holds for all $u \in U_{ad} \cap \overline{\mathbb{B}^{L^2}_\epsilon(\bar{u})}$.

We also mention the work of Casas and Chrysafinos (2018) in which second order optimality conditions analogous to those of Bonifacius and Neitzel (2018), but for a slightly different setting w.r.t. the domain, the boundary conditions and the boundedness properties of the nonlinearity, were derived. Casas and Chrysafinos deal with $C^{1,1}$-smooth domains, homogeneous Dirichlet boundary conditions and locally Lipschitz continuous coefficients for the state equation, which enables them to consider W^2-regularity of the states.

Given $\sigma \in [0, \infty]$ that will become clear from the context, we introduce the modified admissible set as

$$U^\sigma_{ad} := U_{ad} \cap (\bar{u} + C^\sigma(\bar{u})) = \{u \in U_{ad}\colon\, u = \bar{u} \text{ on } A^\sigma(\bar{u})\} \tag{9}$$

and define the corresponding restricted quadratic problem as follows:

$$\begin{cases} \min\limits_{y,u} J_k(y, u) \\ \text{subject to} \quad u \in U^\sigma_{ad} \quad \text{and Equation (3)} \end{cases} \tag{$\mathbf{QP}$-σ}$$

Using the relation of J_k to the second derivative of the Lagrangian of (**OCP**) (see (7) and (8)) it is clear that (**QP**-σ) is a linear quadratic and under Assumption 3.4

strictly coercive and therefore strictly convex optimal control problem, at least for $(y_k, u_k, p_k) = (\bar{y}, \bar{u}, \bar{p})$. This will be crucial for the convergence analysis of the SQP method.

Remark 3.6 Second order sufficient conditions related to strongly active sets turned out to be suitable assumptions for the analysis of SQP methods, see e.g. Tröltzsch (1999), Goldberg and Tröltzsch (1998), Wachsmuth (2007), which work with the same assumption as we do. That we do not work with the SSCs formulated by Bonifacius and Neitzel (2018) directly has two reasons: First, we require the coercivity condition in (SSC-σ) to hold on a vector space instead of just a cone in the proof of the L^2-stability result in Sect. 5.1. Second, in Sect. 6.2 we will make use of the fact that strongly active sets behave well under small perturbations for $\sigma > 0$.

Remark 3.7 Strongest possible second order conditions, i.e. coercivity of L''_{OCP} on the whole space $L^2(\Lambda)$ will be refered to by $\sigma = \infty$. In this case it holds $C^\infty(\bar{u}) = L^2(\Lambda)$ and $U^\infty_{ad} = U_{ad}$. See e.g. Griesse et al. (2010, 2008), Feldhordt (2017) or Heinkenschloss and Tröltzsch (1999) for such an assumption in the context of SQP methods. In Sect. 6.1 we state our main theorem for this special case.

4 Auxiliary results

Before going into the details of the convergence analysis for the SQP method we collect some auxiliary results in the following section.

4.1 Regularity of the adjoint state

For our further analysis we will heavily rely on $L^\infty(I, W^{1,p'})$-regularity of the adjoint state $\bar{p}$ associated with the optimal control $\bar{u}$, cf. the remarks in Sect. 4.3. For better readability we postpone the proof of the corresponding regularity theorem to Sect. 7 and state here only

Lemma 4.1 *It holds $\bar{p} \in L^\infty(I, W^{1,p'}_D)$.*

Proof Set $r = s$, $y = \bar{y}$, $w = \bar{y} - y_d$ and $w_T = 0$ in Theorem 7.2 of Sect. 7. Due to $y_d \in L^\infty(I, L^{p/2})$ and $L^{p/2} \hookrightarrow H^{-\zeta,p}$ all requirements are fulfilled. It follows $\bar{p} \in W^{1,s}(I, H^{-\zeta,p}_D) \cap L^s(I, \mathscr{D})$, i.e. even $\bar{p} \in L^\infty(I, W^{1,p}_D)$ by Theorem 7.1 (b). □

4.2 A property of the control operator

Recall from Assumption 2.4 the definition of the control operator that refers to the case of purely time-dependent controls. Obviously, B is continuous from $L^2(\Lambda)$ to $L^2(I, W^{-1,p}_D)$ and therefore its adjoint B^* is defined on $L^2(I, W^{1,p'}_D)$ with values in $L^2(\Lambda)$. To derive the L^∞-stability result from the L^2-stability result in Sect. 5.1, we need to perform a bootstrapping argument that requires us to know how B^* behaves restricted to a space of more regular functions.

To simplify notation, let B: $L^s(I,\mathbb{R}) \to L^s(I, H^{-\zeta,p})$ be defined by $u \mapsto u \cdot b_1$ with only a single fixed control function $b_1 \in H_D^{-\zeta,p}$. Of course, this yields

$$(B^*v)(t) = \langle b_1, v(t)\rangle_{W_D^{-1,p}, W_D^{1,p'}} \qquad \text{for every } v \in L^2(I, W_D^{1,p'}).$$

It is obvious that B maps $L^r(\Lambda)$ into $L^r(I, H_D^{-\zeta,p})$ for $r \in [2,\infty]$. To obtain $B^*v \in L^q(\Lambda)$, we have to ensure that $v \in L^q(I, H_D^{\zeta,p'})$ holds. We need the following lemma:

Lemma 4.2 *It holds*

$$(W_D^{-1,q}, W_D^{1,q})_{\theta,1} \hookrightarrow H_D^{2\theta-1,q}$$

for $0 < \theta < 1$ *and* $q \in (1,\infty)$ *as long as* $2\theta - 1 \notin \{1/q, -1/q'\}$.

Proof This is a direct consequence of Griepentrog et al. (2002, Theorem 3.5). □

Now, set $\theta := (\zeta+1)/2$. For $r \in (1,\infty)$ there are two possibilities: If $\theta < 1 - 1/r$, then it holds for $0 \le \rho < 1 - 1/r - \theta$

$$W^{1,r}(I, W_D^{-1,p'}) \cap L^r(I, W_D^{1,p'}) \hookrightarrow \mathscr{C}^\rho(I, (W_D^{-1,p'}, W_D^{1,p'})_{\theta,1}) \hookrightarrow \mathscr{C}^\rho(I, H_D^{\zeta,p'}),$$

i.e. B^* is continuous from $W^{1,r}(I, W_D^{-1,p'}) \cap L^r(I, W_D^{1,p'})$ to $L^\infty(\Lambda)$. Otherwise, if $\theta > 1 - 1/r$, we obtain $q \ge 1$ such that $1/q > \theta - (1 - 1/r) > 0$ and

$$W^{1,r}(I, W_D^{-1,p'}) \cap L^r(I, W_D^{1,p'}) \hookrightarrow L^q(I, (W_D^{-1,p'}, W_D^{1,p'})_{\theta,1}) \hookrightarrow L^q(I, H_D^{\zeta,p'}),$$

which means that B^* maps $W^{1,r}(I, W_D^{-1,p'}) \cap L^r(I, W_D^{1,p'})$ to $L^q(\Lambda)$. For the two embeddings we refer e.g. to Amann (2003, formula (1.2)). We will come back to this in Sect. 5.1: Given an estimate on the control in L^r, we have estimates for linearized state and adjoint state in $W^{1,r}(I, W_D^{-1,p}) \cap L^r(I, W_D^{1,p})$ and $W^{1,r}(I, W_D^{-1,p'}) \cap L^r(I, W_D^{1,p'})$ respectively. Application of B^* either yields an estimates for the control in L^q with some $q > r$ or in L^∞ if r already was large enough.

4.3 Some properties of $\mathscr{A}''$

Recall the definition of $\mathscr{A}''$ from the beginning of Sect. 3. For the proof of the L^2- and L^∞-stability results in Sect. 5.1 we need the following

Lemma 4.3 *It holds*

$$\|\mathscr{A}''(y)[v,\cdot]^*p\|_{L^r(I,W_D^{-1,p'})} \le C(\xi,\mu,y)\|p\|_{L^\infty(I,W_D^{1,p'})}\|y\|_{L^\infty(I,W_D^{1,p})}\|v\|_{L^r(I,W_D^{1,p})}.$$

The constant C can be chosen uniformly with respect to y for y's coming from a bounded subset of $W^{1,s}(I, W_D^{-1,p}) \cap L^s(I, W_D^{1,p})$.

Proof Estimate $\langle \mathscr{A}''(y)[v,\cdot]^* p, w\rangle = \langle \mathscr{A}''(y)[v,w], p\rangle$ for an arbitrary testfunction $w \in L^{r'}(I, W^{1,p}_D)$ utilizing Hölders inequality. □

In Lemma 4.3 we bounded the norm of $\mathscr{A}''(\bar{y})[v,\cdot]^*\bar{p}$ in the space $L^r(I, W^{-1,p'}_D)$ against the norm of v in the space $W^{1,r}(I, W^{1,p}_D) \cap L^r(I, W^{1,p}_D)$ for each $r \in [2, s]$ by estimating $\langle \mathscr{A}''(y)[v,w], p\rangle$ with arguments $v \in L^r(I, W^{1,p}_D)$ resp. $w \in L^{r'}(I, W^{1,p}_D)$. This generality will be necessary in the bootstrapping argument in the proof of the L^∞-stability, which was already mentioned in the previous Sect. 4.2. As explained in the remark after Lemma 4.4, this requires bounds for y in $L^\infty(I, W^{1,p}_D)$ and p in $L^\infty(I, W^{1,p'})$. However, in the next section we will require an estimate of $\langle \mathscr{A}''(y)[v,w], p\rangle$ directly (and not of $\mathscr{A}(y)''[v,\cdot]^* p$) which allows us to use the arguments v and w from the space $W^{1,2}(I, W^{-1,p}_D) \cap L^2(I, W^{1,p}_D)$ in Lemma 4.4. In that case we can exploit more regularity of v, w, which allows to relax the assumptions on y and p.

Lemma 4.4 *It holds*

$$|\langle \mathscr{A}''(y)[v,w], p\rangle| \le C(\xi,\mu,y)\|y\|_{L^s(I,W^{1,p}_D)}\|p\|_{L^s(I,W^{1,p'}_D)} \cdot\|v\|_{W^{1,2}(I,W^{1,p}_D)\cap L^2(I,W^{1,p}_D)}\|w\|_{W^{1,2}(I,W^{1,p}_D)\cap L^2(I,W^{1,p}_D)}.$$

The constant C can be chosen uniformly with respect to y for y's coming from a bounded subset of $W^{1,s}(I, W^{-1,p}_D) \cap L^s(I, W^{1,p}_D)$.

Proof The proof works similar as for Lemma 4.3, but now we try to exploit more regularity of v and w. Using embeddings due to Amann (2003, formula (1.2)) and Griepentrog et al. (2002, Theorem 3.5) we find

$$W^{1,2}(I, W^{-1,p}_D) \cap L^2(I, W^{1,p}_D) \hookrightarrow L^q(I, L^\infty),$$

with some $q \in (2, \infty)$ satisfying

$$\frac{2}{q} + \frac{2}{s} \le 1 \quad \text{and} \quad \frac{1}{q} + \frac{1}{2} + \frac{1}{s} \le 1. \tag{10}$$

Now, an application of Hölders inequality (the temporal integrability exponents match due to (10)) yields the desired result. The uniform choice of the constant with respect to y follows from the boundedness of ξ and its derivatives on bounded sets of $\mathbb{R}$ and the compactness of the embedding $W^{1,s}(I, W^{-1,p}_D) \cap L^s(I, W^{1,p}_D) \hookrightarrow \mathscr{C}^0(\overline{I \times \Omega})$. □

Remark 4.5 The difference in the regularities assumed for y and p in the two lemmas is essential: Lemma 4.3 will be applied in Sect. 5.1 only for $y = \bar{y}$ and $p = \bar{p}$, i.e. the required regularity is guaranteed by Lemma 4.1 for $\bar{p}$ and Theorem 7.1 (1), (2b) for $\bar{y}$, respectively. In Sect. 4.4 we will have to apply Lemma 4.4 for $y = y_k$, $p = p_k$ with y_k, p_k being iterates of the SQP method, i.e. y_k and p_k are solutions of the linearized state and adjoint equation. Hence, the regularity requirements for Lemma 4.4 are met, but not immediately those of Lemma 4.3.

Remark 4.6 *(Necessity of higher regularity for the adjoint state)* Note that Lemma 4.3 cannot be improved: The limiting factor is the summand

$$\int_{I\times\Omega} \xi'(y)w\nabla p\nabla v,$$

which has to be estimated for $v \in W^{1,r}(I, W_D^{-1,p}) \cap L^r(I, W_D^{1,p})$ and $w \in L^{r'}(I, W_D^{1,p})$, $r \in [2, s]$. The function w has temporal integrability r' and spatial integrability ∞, whereas ∇v has temporal integrability r and spatial integrability p, which is the best we can expect from the assumptions each. This implies that we require $p \in L^\infty(I, W_D^{1,p'})$ in order to be able to estimate the above integral.

4.4 Derivatives associated to (QP)

In this section we provide results on the first and second derivatives of the reduced objective functionals associated to the quadratic subproblems (**QP**). We will apply them in Sect. 6.3 briefly before obtaining our main result.

Recall the definition of the space X_s from Sect. 3.2 and denote by j_k: $L^2(\Lambda) \to \mathbb{R}$ the reduced functional associated with the linear quadratic optimal control problem (**QP**) at $(y_k, u_k, p_k) \in X_s$. In particular note that j_k'' is constant, because j_k is a quadratic functional, which makes us write j_k'' instead of $j_k''(v)$ for some v, because $v \mapsto j_k''(v)[\cdot,\cdot]$ is constant and hence independent of such v.

Proposition 4.7 *Let Assumptions 2.1–2.4 and 3.4 be satisfied. Then, it holds uniformly in $u \in L^2(\Lambda)$*

$$|(j_k'' - j''(\bar{u}))u^2| \lesssim \Big(\|y_k - \bar{y}\|_{W^{1,s}(I,W_D^{-1,p})\cap L^s(I,W_D^{1,p})} \\ + \|p_k - \bar{p}\|_{W^{1,s}(I,W_D^{-1,p'})\cap L^s(I,W_D^{1,p'})}\Big)\|u\|_{L^2}^2$$

as $y_k \to \bar{y}$, $p_k \to \bar{p}$ in the above norms.

Proof Recall by (7) that $j_k'' \cdot u^2 = L_{OCP}''(y_k, u_k, p_k)(y, u)^2$ with

$$e_y(y_k, u_k)y + e_u(y_k, u_k)u = 0, \tag{11}$$

holds. We expand this as

$$
\begin{aligned}
L''_{OCP}(y_k, u_k, p_k)(y, u)^2 = & \underbrace{L''_{OCP}(\bar{y}, \bar{u}, \bar{p})(\tilde{y}, u)^2}_{=:(I)} \\
& - \underbrace{\left(L''_{OCP}(\bar{y}, \bar{u}, \bar{p})(\tilde{y}, u)^2 - L''_{OCP}(\bar{y}, \bar{u}, \bar{p})(y, u)^2\right)}_{=:(II)} \\
& - \underbrace{\left(L''_{OCP}(\bar{y}, \bar{u}, \bar{p}) - L''_{OCP}(y_k, u_k, p_k)\right)(y, u)^2}_{=:(III)}
\end{aligned}
\tag{12}
$$

with $\tilde{y} \in W^{1,2}(I, W_D^{-1,p}) \cap L^2(I, W_D^{1,p})$ defined by

$$
e_y(\bar{y}, \bar{u})\tilde{y} + e_u(\bar{y}, \bar{u})u = 0. \tag{13}
$$

From the definition of the Lagrangian we know $(I) = j''(\bar{u})u^2$. Hence it remains to show that the contribution of (II) and (III) gets uniformly small as claimed above. By definition we have

$$
\begin{aligned}
(II) &= \underbrace{\|\tilde{y}\|^2 - \|y\|^2}_{=:(IIa)} - \underbrace{\langle \bar{p}, \mathscr{A}''(\bar{y})\tilde{y}^2 - \mathscr{A}''(\bar{y})y^2 \rangle}_{=:(IIb)}, \\
(III) &= \langle p_k, \mathscr{A}''(y_k)y^2 \rangle - \langle \bar{p}, \mathscr{A}''(\bar{y})y^2 \rangle \\
&= \underbrace{\langle p_k - \bar{p}, \mathscr{A}''(y_k)y^2 \rangle}_{=:(IIIa)} + \underbrace{\langle \bar{p}, (\mathscr{A}''(y_k) - \mathscr{A}''(\bar{y}))y^2 \rangle}_{=:(IIIb)},
\end{aligned}
$$

wherein the summands

$$
(IIa) = \langle \tilde{y} + y, \tilde{y} - y \rangle \qquad \text{and} \qquad (IIb) = \langle \bar{p}, \mathscr{A}''(\bar{y})[\tilde{y} + y, \tilde{y} - y] \rangle \tag{14}
$$

can be estimated using the boundedness of the solution operator of the linearized state equation (Bonifacius and Neitzel 2018, Proposition 4.4) and applying Lemma 4.4 and a similar argument as in the proof of Lemma 4.4. In particular recall Remark 4.5. In the same way one can treat (III) as well. □

For the gradient of j_k we find:

Proposition 4.8 *If* $(y_k, u_k, p_k) \to (\bar{y}, \bar{u}, \bar{p})$ *in* X_s, $v_k \to \bar{u}$ *in* L^s, *it holds*

$$
\nabla j_k(v_k) \to \nabla j(\bar{u}), \qquad \textit{strongly in } L^2(\Lambda).
$$

Proof We split

$$
\nabla j_k(v_k) - \nabla j(\bar{u}) = \underbrace{\nabla j_k(v_k) - \nabla j(v_k)}_{=:(A)} + \underbrace{\nabla j(v_k) - \nabla j(\bar{u})}_{=:(B)}
$$

and estimate both summands. For some $v \in U_{ad}$, e.g. $v = v_k$, introducing the following quantities will be helpful:

$$\begin{aligned} y(v) &\quad \text{state associated to } v \text{ w.r.t. (OCP)},\\ p(v) &\quad \text{adjoint state associated to } v \text{ w.r.t. (OCP)},\\ y_k(v) &\quad \text{state associated to } v \text{ w.r.t. (QP)}\\ p_k(v) &\quad \text{adjoint state associated to } v \text{ w.r.t. (QP)}. \end{aligned}$$

Regarding (B) we know from (Bonifacius and Neitzel 2018, Proposition 4.9) that

$$\|\nabla j(v_k) - \nabla j(\bar{u})\|_{L^2(\Lambda)} \le \gamma \|v_k - \bar{u}\|_{L^2} + \|B^*(p(v_k) \\ -p(\bar{u}))\|_{L^2} \to 0 \quad \text{as } v_k \to \bar{u} \text{ in } L^s,$$

holds, because the adjoint states $p(v_k)$ converge in $L^s(I, W_D^{1,p'})$ to $\bar{p}$. To estimate (A) first note that the states $y_k(v_k)$ of the quadratic problem converge to $\bar{y} = y(\bar{u})$ in $W^{1,2}(I, W_D^{-1,p}) \cap L^2(I, W_D^{1,p})$. This is shown using the convergence of the solution operators of the linearized state equation (Bonifacius and Neitzel 2018, Proposition 4.9). Utilizing similar techniques as before the desired result follows after some straight forward computations. We omit the details. □

5 Generalized Newton method on U_{ad}^{σ}

Following the standard arguments, see e.g. Tröltzsch (2000, 1999), Goldberg and Tröltzsch (1998), Alt et al. (2010), Griesse et al. (2010, 2008), Wachsmuth (2007) and Hintermüller and Hinze (2006), we show that the Newton–Josephy method applied to a modified version of the generalized equation (GE), see Sect. 3.2, converges. Our own contribution here is to verify that—under the correct choice of spaces and with help of suitable auxiliary results that have been achieved in the previous section—existing arguments apply to the quasilinear case as well. Proving convergence of this generalized Newton method is a central step towards showing convergence of the SQP method: The iterates of the generalized Newton method will be interpreted as iterates of the SQP method in Sect. 6.

From formula (9) in Sect. 3.3 recall the definition of the modified admissible set U_{ad}^{σ} for some $\sigma \in [0, \infty]$. We consider the generalized equation with this modified admissible set, i.e. we replace (GE) by

$$0 \in F(y, u, p) + N^{\sigma}(y, u, p), \tag{GE-σ}$$

where U_{ad} is replaced by U_{ad}^{σ} in the definition of the normal cone map N, i.e.

$$N^{\sigma}(y, u, p) := \left(\{0\}, \{0\}, \{0\}, \{0\}, N_{U_{ad}^{\sigma}}(u)\right)^T,$$

where $N_{U^\sigma_{ad}}(u)$ denotes the normal cone of U^σ_{ad} at u. The map $F\colon X_s \to Z_s$ as well as the spaces X_s, Z_s, see Sect. 3.2 for the definitions, do not change.

To prove convergence of the generalized Newton method strong regularity in the sense of Robinson has to be shown at an optimal point $(\bar y, \bar u, \bar p) \in X_s$, i.e. for every perturbation $d \in Z_s$ sufficiently close to 0 the generalized equation

$$d \in F(\bar y, \bar u, \bar p) + F'(\bar y, \bar u, \bar p)(y - \bar y, u - \bar u, p - \bar p) + N^\sigma(y, u, p) \qquad \text{(GE-}\sigma\text{-D)}$$

needs to have a unique solution that depends Lipschitz continuous on $d \in Z_s$. For the definition of strong regularity we refer e.g. to Robinson (1980), (Hinze et al. 2009, Definition 2.5).

Translating back this generalized equation for (y, u, p) into an optimal control problem yields

$$\begin{cases} \min\limits_{y,u} \dfrac{1}{2}\|y - y_d\|^2 + \dfrac{\gamma}{2}\|u\|^2 - \dfrac{1}{2}\langle \bar p, \mathscr{A}''(\bar y)[y - \bar y]^2\rangle \\ \quad + \langle d_T, \mathrm{tr}_T\, y\rangle - \langle d_u, u\rangle + \langle d_p, y\rangle \\ \text{subject to} \quad u \in U^\sigma_{ad} \\ \quad \text{and} \quad \begin{pmatrix} d_y \\ d_0 \end{pmatrix} = e_y(\bar y, \bar u)(y - \bar y) + e_u(\bar y, \bar u)(u - \bar u) \end{cases} \qquad \textbf{(QP-}\sigma\textbf{-D)}$$

for a given perturbation vector $d = (d_y, d_0, d_p, d_T, d_u) \in Z_s$ with components coming from the corresponding spaces. Note that (GE-σ-D) is indeed the first order necessary and (due to convexity) sufficient optimality condition for (**QP-σ-D**), because (**QP-σ-D**) is convex since only linear perturbation terms have been added to the convex objective function from (**QP-σ**). The perturbation in the corresponding affine linear state equation is only a constant and does not destroy convexity as well.

5.1 Stability of the quadratic problems (QP-σ)

We fix $d_0 = 0$ and $d_T = 0$, i.e. we assume that initial and final conditions are met exactly during the application of the SQP method, which is reasonable from the numerical point of view.

Proposition 5.1 *Let Assumptions 2.1–2.4 and 3.4 with some $\sigma \in [0, \infty]$ hold. Denote with $(y^i, u^i, p^i) \in X_s$, $i = 1, 2$, the solution of* (**QP-σ-D**) *for arbitrary perturbation vectors $d^i \in Z_s$. Then it holds*

$$\|u^2 - u^1\|^2_{L^2} \lesssim \|d^2_u - d^1_u\|^2_{L^2} + \|d^2_y - d^1_y\|^2_{L^2(I,W^{-1,p})} + \|d^2_p - d^1_p\|^2_{L^2(I,W^{-1,p'})}.$$

The hidden constant depends on the data of (**OCP**) *and $(\bar y, \bar u, \bar p)$, but not on d^i.*

To enhance clarity we state the KKT-system of the perturbed problems, that can easily be derived from (GE-σ-D) using (2) and (3)–(5), before starting the proof:

$$\begin{cases} \partial_t y^i + \mathscr{A}(\bar{y})y^i + \mathscr{A}'(\bar{y})y^i = Bu^i + \mathscr{A}'(\bar{y})\bar{y} + d_y^i \\ y^i(0) = y_0 \\ -\partial_t p^i + \mathscr{A}(\bar{y})^* p^i + \mathscr{A}'(\bar{y})^* p^i = y^i - y_d - \mathscr{A}''(\bar{y})[y^i - \bar{y}, \cdot]^* \bar{p} + d_p^i \\ p^i(T) = 0 \\ d_u^i \in \gamma u^i + B^* p^i + N_{U_{ad}^\sigma}(u^i). \end{cases} \tag{15}$$

In the following we use the short notation $\Delta_y := y^2 - y^1$, $\Delta_u := u^2 - u^1$, $\Delta_p := p^2 - p^1$ (and similarly for d_y, d_u, d_p). From (15) we derive:

$$\partial_t \Delta_y + \mathscr{A}(\bar{y})\Delta_y + \mathscr{A}'(\bar{y})\Delta_y = B\Delta_u + \Delta_{d_y}, \tag{16}$$

$$-\partial_t \Delta_p + \mathscr{A}(\bar{y})^* \Delta_p + \mathscr{A}'(\bar{y})^* \Delta_p = \Delta_y - \mathscr{A}''(\bar{y})[\Delta_y, \cdot]^* \bar{p} + \Delta_{d_p}, \tag{17}$$

with vanishing initial and final condition, respectively: $\Delta_y(0) = 0$ and $\Delta_p(T) = 0$.

Proof The proof relies on the linear quadratic structure of (**QP-σ-D**) and regularity results for the linearized state equation resp. the adjoint equation.

Hence it works completely analogous to Goldberg and Tröltzsch (1998) and we omit the details and only mention the required regularity results (Bonifacius and Neitzel 2018, Propositions 4.4 resp. 4.7) and that terms containing $\mathscr{A}''$ are estimated with help of Lemma 4.3. □

This shows L^2-stability of the quadratic problems (**QP-σ**) with respect to perturbations measured in corresponding norms. Utilizing a standard bootstrapping argument as e.g. in Tröltzsch (2000) we can show the corresponding L^s- resp. L^∞-stability result:

Theorem 5.2 *Let Assumptions 2.1–2.4 and 3.4 with some $\sigma \in [0, \infty]$ hold. Then, for the (y^i, u^i, p^i), $i = 1, 2$, from the previous proposition we have*

$$\|u^2 - u^1\|_{L^s} \lesssim \|d_u^2 - d_u^1\|_{L^s} + \|d_y^2 - d_y^1\|_{L^s(I, W^{-1,p})} + \|d_p^2 - d_p^1\|_{L^s(I, W^{-1,p'})},$$
$$\|u^2 - u^1\|_{L^\infty} \lesssim \|d_u^2 - d_u^1\|_{L^\infty} + \|d_y^2 - d_y^1\|_{L^s(I, W^{-1,p})} + \|d_p^2 - d_p^1\|_{L^s(I, W^{-1,p'})}$$

and

$$\|(y^1, u^1, p^1) - (y^2, u^2, p^2)\|_{X_s} \lesssim \|d^1 - d^2\|_{Z_s},$$
$$\|(y^1, u^1, p^1) - (y^2, u^2, p^2)\|_{X_\infty} \lesssim \|d^1 - d^2\|_{Z_\infty}.$$

In particular, the generalized equation (GE-σ) is strongly regular at its solution $(\bar{y}, \bar{u}, \bar{p})$ with respect to the spaces X_s, Z_s and X_∞, Z_∞.

Proof Again, the proof follows the techniques from Goldberg and Tröltzsch (1998); Tröltzsch (2000). From the projection formula $u^i = \mathrm{Proj}_{U_{ad}^{\sigma}}\Big(-\frac{1}{\gamma}(B^* p^i - d_u^i)\Big)$, $i = 1, 2$, we infer that

$$|\Delta_u| \le \frac{1}{\gamma}\left(|B^*\Delta_p| + |\Delta_{d_u}|\right)$$

holds pointwise on Λ. Thus, we can bound Δ_u in the $L^q(\Lambda)$-norm, if we can bound $B^*\Delta_p$ and Δ_{d_u} in the $L^q(\Lambda)$-norm. We apply a bootstrapping argument that relies on the property of B^* from Sect. 4.2: Assume that we already know

$$\|\Delta_u\|_{L^r} \lesssim \|\Delta_{d_u}\|_{L^r} + \|\Delta_{d_y}\|_{L^r(I,W^{-1,p})} + \|\Delta_{d_p}\|_{L^r(I,W^{-1,p'})}$$

for some $r \in [2, s)$. Using the regularity theory of the linearized state resp. adjoint equation for (16) resp. (17) we conclude

$$\|\Delta_p\|_{L^r(I,W_D^{-1,p'})} \lesssim \|\Delta_{d_u}\|_{L^r} + \|\Delta_{d_y}\|_{L^r(I,W^{-1,p})} + \|\Delta_{d_p}\|_{L^r(I,W^{-1,p'})}.$$

At this point we need the full strength of Lemma 4.3 to estimate the $\mathscr{A}''$-terms for different $r \in [2, s]$. Note that $\bar{p} \in L^\infty(I, W^{1,p'})$ holds due to Lemma 4.1. Our discussion of B^* from Sect. 4.2 shows that either

$$(\zeta + 1)/2 < 1 - 1/r, \text{ which implies } \|B^*\Delta_p\|_{L^\infty} \lesssim \|\Delta_p\|_{L^r(I,W^{-1,p'})}$$

or

$$(\zeta + 1)/2 > 1 - 1/r, \text{ which implies } \|B^*\Delta_p\|_{L^q} \lesssim \|\Delta_p\|_{L^r(I,W^{-1,p'})}$$

with some q fulfilling $1/q > 1/r + (\zeta - 1)/2$ holds. In the first case it follows

$$\|\Delta_u\|_{L^\infty} \lesssim \|\Delta_{d_u}\|_{L^\infty} + \|\Delta_{d_y}\|_{L^s(I,W_D^{-1,p})} + \|\Delta_{d_p}\|_{L^s(I,W_D^{-1,p'})}$$

and we are done. In the second case we have

$$\|\Delta_u\|_{L^q} \lesssim \|\Delta_{d_u}\|_{L^q} + \|\Delta_{d_y}\|_{L^q(I,W^{-1,p})} + \|\Delta_{d_p}\|_{L^q(I,W^{-1,p'})}$$

and we repeat the procedure with $r = q$ as long as the first holds, which is clearly the case due to Assumption 2.4 if $r = s$ is reached. Note that $(\zeta - 1)/2 < 0$ is fixed and that we can avoid q being equal to the exceptional cases of Lemma 4.2 due to the strict inequality that allows small perturbations. □

Remark 5.3 In addition to the case of purely time-dependent control Bonifacius and Neitzel (2018) discuss the case of distributed control, i.e. $U = L^s(I \times \Omega)$ in Assumption 2.4 and B is the embedding $L^s(I \times \Omega) \hookrightarrow L^s(I, H_D^{-\zeta,p})$.

The main difficulty when generalizing our results to the setting of distributed control lies in keeping the arguments for Proposition 5.1 and Theorem 5.2 working. In that case, B^* is the embedding $L^{s'}(I, W^{1,p'}) \hookrightarrow L^{s'}(I \times \Omega)$ and a similar discussion as in Sect. 4.2 has to be done. Sufficiently good estimates for Δ_p could be obtained using the regularity theorem from Sect. 7, whereas the corresponding estimates for Δ_y would require an analogous analysis of the linearized state equation on $H^{-\zeta,p}$-spaces, which is beyond scope and focus of this paper.

5.2 Convergence of the generalized Newton method

Invoking a general result on the convergence of generalized Newton methods, e.g. Hinze et al. (2009), Theorem 2.19, our previous results allow to derive the following

Theorem 5.4 *Let Assumptions 2.1–2.4 and 3.4 with some $\sigma \in [0, \infty]$ hold.*

1. *Then there is a radius $r_{Newton} > 0$ such that for any triple $(y_0, u_0, p_0) \in X_s$ fulfilling*

$$(y_0, u_0, p_0) \in \mathbb{B}^{X_s}_{r_{Newton}}((\bar{y}, \bar{u}, \bar{p}))$$

   the sequence of iterates generated by the Newton–Josephy method for equation (GE-σ) with (y_0, u_0, p_0) as start is well-defined, stays in the ball $\mathbb{B}^{X_s}_{r_{Newton}}((\bar{y}, \bar{u}, \bar{p}))$ and converges q-superlinearly to $(\bar{y}, \bar{u}, \bar{p})$ in X_s.
2. *The same result as in (1) holds with X_∞ instead of X_s.*

Proof The proof is standard, see e.g. Tröltzsch (1999), Goldberg and Tröltzsch (1998), Wachsmuth (2007), Hintermüller and Hinze (2006), Griesse et al. (2008, 2010). □

6 Convergence of the SQP method

The well-definedness of the iterates in Theorem 5.4 is so far only ensured by some generalized implicit function theorem and the strong regularity of (GE-σ) at $(\bar{y}, \bar{u}, \bar{p})$. Convexity of the quadratic subproblems (**QP**-σ) is so far only known in the case $(y_k, u_k, p_k) = (\bar{y}, \bar{u}, \bar{p})$, i.e. the relation of possible minimizers of (**QP**-σ) and solutions of (GE-σ) is unclear at the moment.

Therefore, this final section is devoted to an extended analysis of the generalized Newton method for (GE) and the interpretation of the Newton iterates as solutions of some linear quadratic optimal control problems. In order to make the flow of the argumentation more clear, we give a short summary of this section:

In a first step (Sect. 6.1) we consider the quadratic problems restricted to U^σ_{ad}, i.e. the set of those controls from U_{ad} that coincide with the optimal control $\bar{u}$ on the σ-active set of $\bar{u}$. The main argument here is that the quadratic problems sufficiently close to the true KKT-triple get strictly convex when restricted to U^σ_{ad}. Hence, their unique solution is characterized by the corresponding first order necessary optimality condition, which coincides with the generalized equation originating from the Newton method discussed in Sect. 5. The assumption to restrict to U^σ_{ad} can be slightly relaxed in case that (SSC-σ)

holds for a positive σ: The quadratic subproblems have to be restricted to $U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})}$ with some radius $\rho > 0$, as shown in Sect. 6.3, and the generalized Newton method for (GE) converges locally, even without further restrictions, see Sect. 6.2. That the restriction of the quadratic subproblems can be done in terms of L^2-balls around $\bar{u}$ (instead of L^∞-balls as in previous results) is—to our best knowledge—a new result that we obtain by careful application of the SSCs. The main steps of the argument are as follows: First, we establish convergence of the generalized Newton method for the corresponding set valued equation (GE) in Sect. 6.2, Theorem 6.10. Thereby the proof of strong regularity is the crucial part and essentially relies on the observation that L^2-local quadratic growth and L^2-local uniqueness of critical points implied by SSCs for certain quadratic problems also stays valid uniformly under perturbation (Proposition 6.7). This and the fact that the set of strongly active points behaves sufficiently well under perturbation (Lemma 6.6) allows to carry over results on U_{ad}^σ to $U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})}$ in Corollary 6.8. Finally, in Sect. 6.3 the iterates of the generalized Newton method are identified with the solutions of the quadratic subproblems, see Proposition 6.14. We start with the iterates of the SQP method with subproblems restricted to U_{ad}^σ from Sect. 6.1. Using perturbation arguments analogous to those from Sect. 6.2 it is shown that sufficiently close to the true KKT-triple these iterates can also be obtained as unique solution of the quadratic subproblems on $U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})}$ with appropriate $\rho > 0$, or as the unique local solution of the global quadratic subproblem that is contained in the aforementioned set, see Proposition 6.14.

Note that for theoretical reasons it is not possible to avoid technical restrictions as the above ones completely, even in finite dimensions, cf. the example given by Goldberg and Tröltzsch (1998, Section 6). In the infinite dimensional case an additional difficulty arises as pointed out by Tröltzsch (1999, final Remark): Unlike in finite dimensions we cannot assume that the possibly infinite set of active constraints is correctly identified after the first iteration, and therefore technical restrictions encoding some a-priori knowledge on the correct active set have to be imposed.

6.1 SQP method on U_{ad}^σ

In this section we relate the iterates of the Newton method from Sect. 5 to solutions of (**QP**-σ), see Sect. 3.3 for the definition of U_{ad}^σ and (**QP**-σ). To do so we will show that the formal optimality conditions for (**QP**-σ) encoded in the Newton equations for (GE-σ) are indeed sufficient optimality conditions for (**QP**-σ). Following again the work of Tröltzsch (1999), Goldberg and Tröltzsch (1998), and Wachsmuth (2007) this is done by showing strict convexity for (**QP**-σ) for (y_k, u_k, p_k) sufficiently close to $(\bar{y}, \bar{u}, \bar{p})$. We prove convergence of the SQP method under the technical restriction to replace U_{ad} by U_{ad}^σ. Assuming strongest possible SSCs, i.e. $U_{ad} = U_{ad}^\sigma$, this yields our first main result.

Recall the definition of the space X_s from Sect. 3.2. The following result corresponds to Lemma 6.2, Corollary 6.3 by Tröltzsch (1999).

Proposition 6.1 *Let Assumptions 2.1–2.4 and 3.4 with some $\sigma \in [0, \infty]$ be satisfied. Then, the linear quadratic SQP problem (**QP-**σ) is a strictly convex optimization problem as long as (y_k, u_k, p_k) is sufficiently close to $(\bar{y}, \bar{u}, \bar{p})$ in X_s.*

Proof The optimization problems (**QP-**σ) are of linear quadratic type. To show strict convexity it suffices to show coercivity, but the latter is an immediate consequence of the second order sufficient condition (SSC-σ) and the uniform estimate from Proposition 4.7. □

Now we are ready to show locally superlinear convergence of the SQP method with quadratic problems on U_{ad}^{σ}:

Theorem 6.2 *Let the assumptions of Theorem 5.4 be fulfilled.*

1. *There is a radius $r_{SQP-\sigma} > 0$ such that for any start triple $(y_0, u_0, p_0) \in X_s$ fulfilling*

$$(y_0, u_0, p_0) \in \mathbb{B}_{r_{SQP-\sigma}}^{X_s}((\bar{y}, \bar{u}, \bar{p}))$$

*the sequences of iterates generated by the generalized Newton method applied to (GE-σ) resp. generated by the SQP method with quadratic subproblems (**QP-**σ) are both well-defined, coincide, stay in the ball $\mathbb{B}_{r_{SQP\sigma}}^{X_s}((\bar{y}, \bar{u}, \bar{p}))$ and converge superlinearly to $(\bar{y}, \bar{u}, \bar{p})$ in X_s.*

2. *The statement analogous to (1) with X_s replaced by X_∞ is true, too.*
3. *There is a radius $\tilde{r}_{SQP-\sigma} > 0$ such that the SQP method with quadratic subproblems (**QP-**σ) and initial iterate (y_0, u_0, p_0) with*

$$\|y_0 - \bar{y}\|_{W^{1,s}(I, W_D^{-1,p}) \cap L^s(I, W_D^{1,p})} + \|p_0 - \bar{p}\|_{W^{1,s}(I, W_D^{-1,p'}) \cap L^s(I, W_D^{1,p'})} \leq \tilde{r}_{SQP-\sigma}$$

converges superlinearly in X_s and X_∞ to $(\bar{y}, \bar{u}, \bar{p})$. In particular we can choose

$$u_0 \in U_{ad}, \qquad \|u_0 - \bar{u}\|_{L^2(\Lambda)} \ \textit{sufficiently small},$$
$$y_0, p_0 \ \textit{state and adjoint state associated to}\ u_0.$$

Proof For (1) and (2) the proof works analogous to that of Theorem 6.4 in Tröltzsch (1999). For (3) note that (**QP-**σ) is actually independent of the current control iterate u_k, cf. also the remark after (5), which shows the first statement in (3). Since U_{ad} is bounded in L^∞ and $s > 2$ by Assumption 2.4 it holds

$$\|u_0 - \bar{u}\|_{L^s} \leq C \|u_0 - \bar{u}\|_{L^2}^{2/s} \qquad \forall u_0 \in U_{ad}$$

by the Riesz-Thorin interpolation theorem, cf. also the remark after the next theorem. Here, $C > 0$ is a constant depending only on the L^∞-bound of U_{ad}, i.e. on u_a and u_b. From this we conclude by continuity

$$\|(y_0, u_0, p_0) - (\bar{y}, \bar{u}, \bar{p})\|_{X_s} \lesssim \|u_0 - \bar{u}\|_{L^2}^{2/s},$$

which shows the second statement of (3). □

Assuming strongest possible second order sufficient conditions, i.e. coercivity of the second derivative of the Lagrangian on the whole space instead of only on a subspace, we are able to state our first main result. Note that it is possible to formulate all "closeness" required for convergence of the SQP method with respect to L^2-norms.

Theorem 6.3 *Let the Assumptions 2.1–2.4 be fulfilled and let the second order sufficient condition (SSC-σ) from Assumption 3.4 hold on the whole space $L^2(\Lambda)$ (i.e. $\sigma = \infty$). Then the SQP method for* (**OCP**) *started in $(y_0, u_0, p_0) \in X_s$,*

$$u_0 \in U_{ad}, \qquad \|u_0 - \bar{u}\|_{L^2(\Lambda)} \text{ sufficiently small,}$$
$$y_0, p_0 \text{ state and adjoint state associated to } u_0,$$

converges superlinearly in X_s and X_∞ to $(\bar{y}, \bar{u}, \bar{p})$.

Proof Use Theorem 6.2 (3) together with $U_{ad}^\sigma = U_{ad}$. □

Remark 6.4 That the topologies generated by the L^2- and the L^s-norm ($s > 2$), respectively, coincide on an L^∞-bounded set by the Riesz-Thorin interpolation theorem, is a well known fact. However, this observation is a key argument for many proofs concerning second order conditions without two norm gap, see e.g. (Casas and Tröltzsch 2012, Proposition 3.4) or (Bonifacius and Neitzel 2018, Theorem 4.14). Here, we made use of this technique in Theorem 6.2 (3) and 6.3 to tighten the unsatisfying gap between the quadratic growth condition for j implied by (SSC-σ)—this growth condition holds L^2-locally—and the L^s-local convergence of the SQP method.

For the rest of Sect. 6 we will be concerned with relaxing this rather abstract and technical condition towards a more natural restriction.

6.2 Generalized Newton method on U_{ad} resp. $U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})}$

Before showing convergence of the SQP method restricted to $U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})}$ we consider convergence of the Newton method for the associated generalized equation first. Our arguments follow in particular the presentation by Wachsmuth (2007), but similar results are also due to Goldberg and Tröltzsch (1998) and Tröltzsch (1999). To replace L^∞-locality by L^2-locality in the statements of Proposition 6.7 is—to our best knowledge—a new result. An analogous technique will be utilized afterwards in Sect. 6.3 to prove also convergence of the SQP method under certain localization conditions.

In the following we consider the perturbed generalized equation

$$d \in F(\bar{y}, \bar{u}, \bar{p}) + F'(\bar{y}, \bar{u}, \bar{p})(y - \bar{y}, u - \bar{u}, p - \bar{p}) + N(y, u, p). \tag{GE-D}$$

Note that we now use the normal cone map N associated with the true set of admissible controls U_{ad} instead of the normal cone map N^σ associated with the modified admissible set U_{ad}^σ that was used for the definition of (GE-σ-D) in the previous sections. Furthermore, note that (GE-σ-D) can be understood as generalized equation in the spaces X_s, Z_s resp. X_∞, Z_∞ both. For the definition of these spaces see Sect. 3.2. As before, the generalized equation (GE-σ-D) is the formal optimality system of the following perturbed optimal control problem:

$$\begin{cases} \min\limits_{y,u} \frac{1}{2}\|y-y_d\|^2 + \frac{\gamma}{2}\|u\|^2 - \frac{1}{2}\langle \bar{p}, \mathscr{A}''(\bar{y})[y-\bar{y}]^2\rangle - \langle d_u, u\rangle + \langle d_p, y\rangle \\ \text{subject to} \quad u \in U_{ad} \\ \quad \text{and} \quad \begin{pmatrix} d_y \\ 0 \end{pmatrix} = e_y(\bar{y},\bar{u})(y-\bar{y}) + e_u(\bar{y},\bar{u})(u-\bar{u}) \end{cases} \tag{QP-D}$$

The reduced objective function for (**QP-D**) will be denoted by j_d. Note that we did not discuss properties of this optimization problem so far. Further, we introduce the following notation for the strongly active sets:

$$\begin{aligned} A_d^\sigma(u) &:= \left\{x \in \Lambda \colon |\nabla j_d(u)|(x) = |B^* p + \gamma u - d_u|(x) > \sigma\right\}, \\ A^\sigma(u) &:= A_0^\sigma(u), \qquad \text{i.e. } d = 0 \text{ in the definition above.} \end{aligned}$$

Here, p denotes the adjoint state associated with u with respect to (**QP-D**) with perturbation vector d, see (15). Note that $A_0^\sigma(\bar{u})$ coincides with the strongly active set for $\bar{u}$ defined in Assumption 3.4.

In Sect. 5 we observed that under Assumptions 2.1–2.4 and 3.4 the restricted optimal control problem (**QP-σ-D**), i.e. problem (**QP-D**) restricted to U_{ad}^σ, is strictly convex and admits a unique solution $(\bar{y}_d, \bar{u}_d, \bar{p}_d)$. This holds true for arbitrarily large perturbation vectors d. In particular, the map $d = (d_y, d_p, d_u) \mapsto (\bar{y}_d, \bar{u}_d, \bar{p}_d)$ was shown to be Lipschitz from Z_∞ to X_∞ in Theorem 5.2, say with modulus $L' > 0$. It follows that the mapping

$$\begin{aligned} &Z_\infty \to L^\infty(\Lambda), \\ &\quad d \mapsto \gamma\bar{u}_d + B^*\bar{p}_d - d_u = \nabla j_d(\bar{u}_d) \end{aligned} \tag{18}$$

is Lipschitz as well, say with modulus $L > 0$.

Remark 6.5 Of course, even the map $Z_s \to X_s$, $d \mapsto (\bar{y}_d, \bar{u}_d, \bar{p}_d)$ is Lipschitz continuous as shown in Theorem 5.2, which implies that $d \mapsto \gamma\bar{u}_d + B^*\bar{p}_d - d_u$ is Lipschitz continuous from Z_s to $L^s(\Lambda)$. Unfortunately, we rely on L^∞-estimates in the following.

Assuming that (SSC-σ) holds for some $\sigma \in (0, \infty)$ we can draw some immediate conclusions from the Lipschitz continuity of (18) as done by Wachsmuth (2007, Corollaries 5.3 and 5.4):

Lemma 6.6 *Let Assumptions 2.1–2.4 and 3.4 with some $\sigma \in (0, \infty)$ hold and suppose that $\|d\|_{Z_\infty} < \frac{\sigma}{2L}$.*

1. *It holds $A^\sigma(\bar{u}) \subset A_d^{\sigma/2}(\bar{u}_d)$ and the signs of $\nabla j_d(\bar{u}_d)$ and $\nabla j_0(\bar{u})$ coincide on $A^\sigma(\bar{u})$.*

2. *The solution $(\bar{y}_d, \bar{u}_d, \bar{p}_d)$ of* (**QP-σ-D**) *is a solution of* (GE-D) *as well, i.e. it holds*

$$\langle \gamma \bar{u}_d + B^* \bar{p}_d - d_u, u - \bar{u}_d \rangle_{L^2(\Lambda)} \geq 0, \qquad \forall u \in U_{ad}.$$

Proof Completely analogous to Wachsmuth (2007). □

Lemma 6.6 shows that our solution of (**QP-σ-D**) that depends Z_∞-X_∞-Lipschitz on d is a solution of (GE-D) as well, if the perturbation d is small enough in Z_∞. To establish strong regularity of (GE) (with spaces X_∞, Z_∞) from this result we have to show that this solution is locally unique. This is done by proving that $(\bar{y}_d, \bar{u}_d, \bar{p}_d)$ is not only a global solution of (**QP-σ-D**) but even a local solution of (**QP-D**) fulfilling a quadratic growth condition on a ball around $(\bar{y}_d, \bar{u}_d, \bar{p}_d)$ with radius independent of d.

Proposition 6.7 *Let the assumptions of Lemma 6.6 be satisfied.*

1. *Then there exist $0 < \tilde{\epsilon} < \frac{\sigma}{2L}$ and $\tilde{\rho}, \eta > 0$, such that $(\bar{y}_d, \bar{u}_d, \bar{p}_d)$, i.e. the solution of* (**QP-σ-D**)*, is also a L^2-local solution of* (**QP-D**) *and satisfies the quadratic growth condition*

$$j_d(u) \geq j_d(\bar{u}_d) + \eta \|u - \bar{u}_d\|_{L^2}^2$$

for $\|u - \bar{u}_d\|_{L^2(\Lambda)} \leq \tilde{\rho}$, $u \in U_{ad}$, as long as $\|d\|_{Z_\infty} < \tilde{\epsilon}$.

2. *There are $0 < \hat{\epsilon} \leq \tilde{\epsilon}$, $0 < \hat{\rho} \leq \tilde{\rho}$ such that $(\bar{y}_d, \bar{u}_d, \bar{p}_d)$ is the only stationary*[1] *point for* (**QP-D**) *in $\overline{\mathbb{B}_{\hat{\rho}}^{L^2}}(\bar{u}_d)$.*

The first statement of this proposition corresponds to Theorem 5.5 (Wachsmuth 2007) with the L^∞-ball around $\bar{u}_d$ replaced by an L^2-ball. To establish quadratic growth L^∞-locally around $\bar{u}_d$, one could follow the direct proof of Theorem 5.17 (Tröltzsch 2010). Avoiding the two norm gap—which is our aim—can be done following ideas due to Casas and Tröltzsch (2012, Theorem 2.3), see also Tröltzsch and Wachsmuth (2006, Theorem 3.22), utilizing a proof by contradiction. We mention that similar arguments were also used by Casas and Tröltzsch Casas and Tröltzsch (2012) in the context of abstract finite element errors.

Note that for every single perturbation $d \in Z_\infty$, both properties in the proposition are directly implied by Theorem 2.3 resp. Corollary 2.6 from Casas and Tröltzsch (2012). The crucial point here is to guarantee that the radii of the respective balls can be chosen independently of the choice of d as long as $\|d\|_{Z_\infty}$ is small enough.

Proof For the proof of (1) we extended the technique presented by (Casas and Tröltzsch 2012, Theorem 2.3) to our needs. First, note that due to the quadratic structure of (**QP-D**) it holds $j_d''(\bar{u}_d)[v_1, v_2] = j''(\bar{u})[v_1, v_2]$. In particular, j_d'' is independent of d.

We are going to argue by contradiction and assume the contrary of our claim: There are sequences $(d_n)_n \subset Z_\infty$, $(h_n)_n \subset L^2(\Lambda)$ with

$$\|d_n\|_{Z_\infty} < \frac{1}{n}, \qquad \|h_n\|_{L^2} < \frac{1}{n} \text{ and } \bar{u}_{d_n} + h_n \in U_{ad}$$

[1] We call (y, u, p) stationary for (**QP-D**) if (y, u, p) fulfills the first order necessary conditions for (**QP-D**).

such that

$$j_{d_n}(\bar{u}_{d_n}+h_n)-j_{d_n}(\bar{u}_{d_n})<\frac{1}{n}\|h_n\|_{L^2}^2. \tag{19}$$

Define $v_n:=\frac{h_n}{\|h_n\|_{L^2}}$ and $\rho_n:=\|h_n\|_{L^2}$. It holds $d_n=(d_{y,n},d_{p,n},d_{u,n})\to 0$ strongly in Z_∞, which implies $\bar{u}_{d_n}\to\bar{u}$ and $\nabla j_{d_n}(\bar{u}_{d_n})\to\nabla j(\bar{u})$ strongly in $L^\infty(\Lambda)$. Due to $\|v_n\|_{L^2}=1$ for all $n\in\mathbb{N}$ we can w.l.o.g. assume that $v_n \rightharpoonup v_*$ weakly in $L^2(\Lambda)$ for some $v_*\in L^2(\Lambda)$.

Step 1: We prove $j'(\bar{u})v_*=0$. We have

$$\begin{aligned} j'(\bar{u})v_* &= \langle \text{strong-}\lim_{n\to\infty}\nabla j_{d_n}(u_{d_n}), \text{weak-}\lim_{n\to\infty} v_n\rangle_{L^2} \\ &= \lim_{n\to\infty}\langle\nabla j_{d_n}(u_{d_n}),v_n\rangle_{L^2}\geq 0, \end{aligned} \tag{20}$$

because $\langle\nabla j_{d_n}(u_{d_n}),v_n\rangle_{L^2}=\frac{1}{\rho_n}\langle\nabla j_{d_n}(u_{d_n}),h_n\rangle_{L^2}\geq 0$ holds for every n due to $\bar{u}_{d_n}+h_n\in U_{ad}$ and Lemma 6.6 (2), for which we can assume w.l.o.g. that $\|d_n\|_{Z_\infty}<\frac{\sigma}{2L}$. Further, using the mean value theorem there are $\theta_n\in(0,1)$ such that

$$\frac{j_{d_n}(u_{d_n}+\rho_n v_n)-j_{d_n}(\bar{u}_{d_n})}{\rho_n}=\langle\nabla j_{d_n}(\bar{u}_{d_n}+\theta_n\rho_n v_n),v_n\rangle_{L^2}.$$

Due to the structure of (**QP-D**)—see e.g. (16), (17) and use regularity results as in the proof of Theorem 5.2—we know that $\nabla j_{d_n}(\bar{u}_{d_n}+\theta_n\rho_n v_n)\to\nabla j(\bar{u})$ strongly in $L^2(\Lambda)$, which implies

$$\frac{j_{d_n}(u_{d_n}+\rho_n v_n)-j_{d_n}(\bar{u}_{d_n})}{\rho_n}\to j'(\bar{u})v_* \qquad \text{as } n\to\infty. \tag{21}$$

On the other hand it holds by assumption (19):

$$\frac{j_{d_n}(u_{d_n}+\rho_n v_n)-j_{d_n}(\bar{u}_{d_n})}{\rho_n}<\frac{1}{\rho_n}\cdot\frac{1}{n}\|h_n\|_{L^2}^2=\frac{\rho_n}{n}\to 0,$$

which together with (21) yields $j'(\bar{u})v_*\leq 0$ first and then together with (20):

$$j'(\bar{u})v_*=0. \tag{22}$$

Step 2: We want to show $v_*=0$ if $|\nabla j(\bar{u})|>0$. To do so we show $v_*\geq 0$ if $\nabla j(\bar{u})>0$ and $v_*\leq 0$ if $\nabla j(\bar{u})<0$, which implies together with Step 1 the desired property: For $\sigma'>0$ arbitrary define $A^{\sigma',a}(\bar{u}):=\{x\in\Lambda:\nabla j(\bar{u})>\sigma'\}$. As in the proof of Lemma 6.6 we conclude that $\nabla j_{d_n}(\bar{u}_{d_n})>0$ on $A^{\sigma',a}(\bar{u})$ for all sufficiently large n, which implies $h_n, v_n\geq 0$ on $A^{\sigma',a}(\bar{u})$ for all such n. Because weak convergence preserves signs we conclude $v_*\geq 0$ on $A^{\sigma',a}(\bar{u})$. Since $\sigma'>0$ was arbitrary it follows $v_*\geq 0$ whenever $\nabla j(\bar{u})>0$, as stated. The case $\nabla j(\bar{u})<0$ is shown in the same way.

Step 3: In Step 2 we have shown that $v_* \in C^0(\bar{u}) \subset C^\sigma(\bar{u})$ holds. For the definition of $C^0(\bar{u})$ and $C^\sigma(\bar{u})$ see Assumption 3.4. In the present step we will arrive at the final contradiction. First observe that by our assumption

$$\frac{\rho_n^2}{n} = \frac{1}{n}\|h_n\|_{L^2}^2 > j_{d_n}(\bar{u}_{d_n} + h_n) - j_{d_n}(\bar{u}_{d_n}) \overset{(\bigstar)}{=} j'_{d_n}(\bar{u}_{d_n})h_n + \frac{1}{2}j''(\bar{u})h_n$$
$$\overset{(\blacksquare)}{\geq} \frac{\rho_n^2}{2} j''(\bar{u})v_n^2,$$

where we used the linear quadratic structure of (**QP-D**) at ($\bigstar$) and the first order optimality condition at ($\blacksquare$). It follows

$$j''(\bar{u})v_*^2 \leq \liminf_{n\to\infty} j''(\bar{u})v_n^2 \leq \liminf_{n\to\infty} \frac{2}{n} = 0, \tag{23}$$

where the first inequality comes from the weak lower semicontinuity of $j''(\bar{u})$, see Proposition 4.10 (Bonifacius and Neitzel 2018). Since $v_* \in C^\sigma(\bar{u})$ we can apply (SSC-σ) and conclude from (23) that $v_* = 0$. Using property (4.11) by Bonifacius and Neitzel (2018) at ($\blacktriangle$) we obtain

$$\gamma = \gamma \liminf_{n\to} \|v_n\|_{L^2}^2 \overset{(\blacktriangle)}{\leq} \liminf_{n\to\infty} j''(\bar{u})v_n^2 \overset{(23)}{=} 0,$$

which is the desired contradiction. □

The second part of the proposition is shown similarly adapting the proof of Corollary 2.6 by Casas and Tröltzsch (2012). We leave the details to the reader. □

Given a radius $\rho > 0$ we introduce another modification of the perturbed linear quadratic problem (**QP-D**)

$$\begin{cases} \min\limits_{y,u} \dfrac{1}{2}\|y - y_d\|^2 + \dfrac{\gamma}{2}\|u\|^2 - \dfrac{1}{2}\langle \bar{p}, \mathscr{A}''(\bar{y})[y - \bar{y}]^2\rangle - \langle d_u, u\rangle + \langle d_p, y\rangle \\ \text{subject to} \quad u \in U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})} \\ \qquad \text{and} \quad \begin{pmatrix} d_y \\ 0 \end{pmatrix} = e_y(\bar{y}, \bar{u})(y - \bar{y}) + e_u(\bar{y}, \bar{u})(u - \bar{u}) \end{cases} \tag{QP-D-ρ}$$

for which the following result holds:

Corollary 6.8 *Let the assumptions of Lemma 6.6 be satisfied.*

1. *There are $\epsilon, \rho > 0$, such that the triple $(\bar{y}_d, \bar{u}_d, \bar{p}_d)$, i.e. the unique solution of* (**QP-σ-D**)*, is also the unique solution of* (**QP-D-ρ**) *if $\|d\|_{Z_\infty} < \epsilon$.*
2. *There are $\epsilon, \tau > 0$, such that for $\|d\|_{Z_\infty} < \epsilon$ the control $\bar{u}_d$ is the unique solution of* (GE-D) *that is contained in the set $U_{ad} \cap \overline{\mathbb{B}_\tau^{L^2}(\bar{u})}$.*

A result similar to (2)—but with L^∞- instead of L^2-balls—was proven by (Goldberg and Tröltzsch 1998, Theorem 5.4) using a different argument that relies on strongly active sets and continuity of (18).

Proof 1. Choose $\rho = \frac{2\tilde{\rho}}{3}$ and $\epsilon < \min\left(\tilde{\epsilon}, \frac{\tilde{\rho}}{3C}\right)$, where $C > 0$ is the Z_∞-L^2-Lipschitz constant for the map $d \mapsto \bar{u}_d$, cf. Theorem 5.2 for the Lipschitz continuity. Then, it holds in particular $\|d\|_{Z_\infty} < \tilde{\epsilon}$, i.e. the previous Proposition applies, and

$$\bar{u}_d \in U_{ad} \cap \overline{\mathbb{B}^{L^2}_\rho(\bar{u})} \subset U_{ad} \cap \overline{\mathbb{B}^{L^2}_{\tilde{\rho}}(\bar{u}_d)}$$

for all $\|d\|_{Z_\infty} < \epsilon$. Since $\bar{u}_d$ is the unique minimizer of (**QP-D**) restricted to $U_{ad} \cap \overline{\mathbb{B}^{L^2}_{\tilde{\rho}}(\bar{u}_d)}$ by quadratic growth (Proposition 6.7 (1)) and this minimizer is contained in the smaller set $U_{ad} \cap \overline{\mathbb{B}^{L^2}_\rho(\bar{u})}$, we finally proved that $\bar{u}_d$ is the unique minimizer of (**QP-D**) restricted to $U_{ad} \cap \overline{\mathbb{B}^{L^2}_\rho(\bar{u})}$, i.e. the unique minimizer of (**QP-D-ρ**).

2. Similarly as for (1). Now make use of Proposition 6.7 (2).

□

We introduce another variation of (GE):

$$0 \in F(y, u, p) + N^\rho(y, u, p), \tag{GE-ρ}$$

with the set valued map $N^\rho(y, u, p) := \{\{0\}, \{0\}, \{0\}, \{0\}, N_{U_{ad} \cap \overline{\mathbb{B}^{L^2}_\rho(\bar{u})}}(u)\}^T$, where $N_{U_{ad} \cap \overline{\mathbb{B}^{L^2}_\rho(\bar{u})}}(u)$ denotes the normal cone of the closed convex set $U_{ad} \cap \overline{\mathbb{B}^{L^2}_\rho(\bar{u})}$ at some point u. The first part of the following result is similar to Corollary 5.6 (Wachsmuth 2007), the second part to the observation on top of p. 240 by Goldberg and Tröltzsch (1998).

Theorem 6.9 *Let the assumptions of Lemma 6.6 be fulfilled. It holds:*

1. *The generalized equation* (GE) *in the spaces X_∞, Z_∞ is strongly regular at* $(\bar{y}, \bar{u}, \bar{p})$.
2. *There is a $\rho > 0$ such that the generalized equation* (GE-ρ) *in the spaces X_∞, Z_∞ is strongly regular at* $(\bar{y}, \bar{u}, \bar{p})$.

Proof Both statements are consequences of Corollary 6.8 resp. Theorem 5.2. The first part is proven in the same way as in Wachsmuth (2007). We have to use that the L^∞-norm is stronger than the L^2-norm. For the second part note that for all u in the L^2-interior of the ball $\overline{\mathbb{B}^{L^2}_\rho(\bar{u})}$, i.e. in particular for all u sufficiently close to $\bar{u}$ in the L^∞-norm, the equality $N_{U_{ad} \cap \overline{\mathbb{B}^{L^2}_\rho(\bar{u})}}(u) = N_{U_{ad}}(u)$ holds, as already mentioned by Goldberg and Tröltzsch (1998). □

The following result is an immediate consequence of an abstract result (Hinze et al. 2009, Theorem 2.19) and Theorem 6.9. The closed graph property for the normal cone map N^ρ is standard.

Theorem 6.10 *Let Assumptions 2.1–2.4 and 3.4 with some $\sigma \in (0, \infty)$ hold. For any (y_0, p_0) sufficiently close to $(\bar{y}, \bar{p})$ in the space*

$$W^{1,s}(I, W_D^{-1,p}) \cap L^s(I, W_D^{1,p}) \times W^{1,s}(I, W_D^{-1,p'}) \cap L^s(I, W_D^{1,p'})$$

it holds:

1. *The sequence of iterates generated by the Newton–Josephy method for (GE) with initial iterate (y_0, u_0, p_0) is well-defined and converges superlinearly in X_∞ to $(\bar{y}, \bar{u}, \bar{p})$.*
2. *The same holds true for the sequence of iterates generated by the Newton–Josephy method for (GE-ρ) with ρ from Theorem 6.9 (2).*

From Lemma 6.6 on we had to consider perturbations in Z_∞, i.e. we had to measure the control in $L^\infty(\Lambda)$. This is the reason why have to show strong regularity only in Z_∞, X_∞ and not in Z_s, X_s as well as we did before. That we impose no condition on u_0 is due to the fact that the Newton update equations for (GE) resp. (GE-ρ) are independent of the current u-iterate u_k, see the comment after equation (5).

6.3 SQP method on $U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})}$

Finally, we investigate how the iterates of the generalized Newton method from Theorem 6.10 can be computed by solving linear quadratic optimal control problems restricted on $U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})}$. For analogous results in the case of semilinear equations (but with L^∞- instead of L^2-balls) we refer to Tröltzsch (1999) and Goldberg and Tröltzsch (1998).

Lemma 6.11 *Let the assumptions of Theorem 6.10 hold. Let $(y_k, u_k, p_k) \in X_\infty$ be a given triple and consider the restricted quadratic subproblem (**QP-σ**) associated with this triple. There exists an X_∞-neighbourhood V_1 of $(\bar{y}, \bar{u}, \bar{p})$ such that the map*

$$(y_k, u_k, p_k) \mapsto (y_{k+1}^\sigma, u_{k+1}^\sigma, p_{k+1}^\sigma)$$

*is well-defined on V_1 and Lipschitz continuous, where $(y_{k+1}^\sigma, u_{k+1}^\sigma, p_{k+1}^\sigma)$ denotes the unique solution of (**QP-σ**).*

Proof Existence and uniqueness of a solution to (**QP-σ**) is established in Proposition 6.1 for (y_k, u_k, p_k) sufficiently close to $(\bar{y}, \bar{u}, \bar{p})$. Define $\tilde{V}$ to be such a neighbourhood of $(\bar{y}, \bar{u}, \bar{p})$. To see Lipschitz continuity, note that $(y_{k+1}^\sigma, u_{k+1}^\sigma, p_{k+1}^\sigma)$ is solution of the parametrized generalized equation

$$\begin{aligned} 0 &\in G((y_k, u_k, p_k), (y, u, p)) + N^\sigma(y, u, p) \\ &:= F(y_k, u_k, p_k) + F'(y_k, u_k, p_k)(y - y_k, u - u_k, p - p_k) + N^\sigma(y, u, p) \end{aligned}$$

–with (y_k, u_k, p_k) being the parameter—and that

$$\begin{aligned} 0 &\in G((\bar{y}, \bar{u}, \bar{p}), (y, u, p)) + N^\sigma(y, u, p) \\ &= F(\bar{y}, \bar{u}, \bar{p}) + F'(\bar{y}, \bar{u}, \bar{p})(y - \bar{y}, u - \bar{u}, p - \bar{p}) + N^\sigma(y, u, p) \end{aligned}$$

is strongly regular at its solution $(\bar{y}, \bar{u}, \bar{p})$ according to Theorem 5.2. Further, G and G', i.e. F and F', depend continuously on (y_k, u_k, p_k), because F: $X_\infty \to Z_\infty$ is continuously differentiable (Lemma 3.3). Hence, Theorem 2.18 (Hinze et al. 2009) implies the desired Lipschitz continuity of $(y_k, u_k, p_k) \mapsto (y^\sigma_{k+1}, u^\sigma_{k+1}, p^\sigma_{k+1})$ from X_∞ to X_∞ on a sufficiently small neighbourhood $\hat{V}$ of $(\bar{y}, \bar{u}, \bar{p})$. Now, $V_1 := \tilde{V} \cap \hat{V}$ yields the desired neighbourhood. □

With the previous lemma we have shown in particular that

$$\begin{aligned} X_\infty &\to L^\infty(\Lambda) \\ (y_k, u_k, p_k) &\mapsto \nabla j_k(u^\sigma_{k+1}) = \gamma u^\sigma_{k+1} + B^* p^\sigma_{k+1} \end{aligned} \tag{24}$$

is Lipschitz continuous on the X_∞-neighbourhood V_1 of $(\bar{y}, \bar{u}, \bar{p})$. With j_k we denoted the reduced functional of (**QP**-σ) and p^σ_{k+1} is the adjoint state (w.r.t. (**QP**-σ)) associated with the control u^σ_{k+1}, see Eqs. (3), (4). The same argument as for Lemma 6.6 now shows

Lemma 6.12 *Let the assumptions of Theorem 6.10 hold. There is an X_∞-neighbourhood V_2 of $(\bar{y}, \bar{u}, \bar{p})$ such that for all $(y_k, u_k, p_k) \in V_2$ the solution $(y^\sigma_{k+1}, u^\sigma_{k+1}, p^\sigma_{k+1})$ of (**QP**-σ) satisfies the first order necessary optimality conditions of (**QP**).*

Proof State- and adjoint equation of (**QP**) and (**QP**-σ) coincide. We only have to show that $(y^\sigma_{k+1}, u^\sigma_{k+1}, p^\sigma_{k+1})$ fulfills the variational inequality of (**QP**) as well and this works completely analogous to Lemma 6.6 replacing (18) with (24). □

Now, we can show the following result that is similar to Proposition 6.7:

Proposition 6.13 *Let the assumptions of Theorem 6.10 hold. There is an X_∞-neighbourhood V_3 of $(\bar{y}, \bar{u}, \bar{p})$ and there are $\rho, \eta > 0$ such that for all triples $(y_k, u_k, p_k) \in V_3$ the unique solution $(y_{k+1}, u_{k+1}, p_{k+1}) := (y^\sigma_{k+1}, u^\sigma_{k+1}, p^\sigma_{k+1})$ of (**QP**-σ)*

1. *is a L^2-local solution of (**QP**) satisfying the quadratic growth condition*

$$j_k(u) \geq j_k(u_{k+1}) + \eta \|u - u_{k+1}\|^2_{L^2}$$

   for $u \in U_{ad}$ such that $\|u - u_{k+1}\|_{L^2(\Lambda)} \leq \rho$.

2. *is the only stationary point for (**QP**) in $\overline{\mathbb{B}^{L^2}_\rho}(u_{k+1})$.*

Proof We proceed as in the proofs of Proposition 6.7 (1) and (2) and argue by contradiction. Instead of j_{d_n} and $\bar{u}_{d_n}$ we have to consider j_k and u_{k+1}. We only mention the essential ingredients that keep all the previous arguments working:

(i) For any sequence $(w_k) \subset U_{ad}$ such that $w_k \to \bar{u}$ in $L^2(\Lambda)$ it holds

$$\nabla j_k(w_k) \to \nabla j(\bar{u}), \qquad \text{strongly in } L^2(\Lambda).$$

This was shown in Proposition 4.8; use the Riesz-Thorin interpolation theorem (see Remark 6.4) to obtain the required L^s-convergence $w_k \to \bar{u}$ from the given L^2-convergence.

(ii) If $u_k \to \bar{u}$ strongly in L^2 and $v_k \rightharpoonup v_*$ weakly in L^2 we have:

$$j''(\bar{u})v_*^2 \leq \liminf_{k\to\infty} j_k''(u_k)v_k^2.$$

Using the boundedness of $(v_k)_k$, this is a consequence of Proposition 4.7 and the weak lower semicontinuity of j'', see Bonifacius and Neitzel (2018), (4.10):

$$\liminf_k j_k''(u_k)v_k^2 \geq \liminf_k \underbrace{\left(j_k''(u_k) - j''(u_k)\right) v_k^2}_{\to 0 \text{ uniformly in } v_k} + \liminf_k j''(u_k)v_k^2 \geq j''(\bar{u})v_*^2$$

(iii) If $v_* = 0$ in (2), then $\gamma \liminf_{k\to\infty} \|v_k\|_{L^2}^2 \leq \liminf_{k\to\infty} j_k''(u_k)v_k^2$: This is shown by the same argument as above.

□

Next, we obtain with the same argument as for Corollary 6.8:

Proposition 6.14 *Let the assumptions of Theorem 6.10 hold.*

1. *There is an X_∞-neighbourhood V_4 of $(\bar{y}, \bar{u}, \bar{p})$ and a radius $\rho > 0$ such that for all $(y_k, u_k, p_k) \in V_4$ the next SQP iterate $(y_{k+1}, u_{k+1}, p_{k+1})$ given by the unique solution of* (**QP-σ**) *is also the unique solution of* (**QP**) *with admissible set $U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})}$.*
2. *There is an X_∞-neighbourhood V_5 of $(\bar{y}, \bar{u}, \bar{p})$ and a radius $\rho > 0$, such that for all $(y_k, u_k, p_k) \in V_5$ the next SQP iterate $(y_{k+1}, u_{k+1}, p_{k+1})$ given by the unique solution of* (**QP-σ**) *is also the unique L^2-local solution of the global quadratic problem* (**QP**) *that is contained in $U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})}$.*

For convenience of the reader we write down the quadratic problem which we will refer to in our final theorem:

$$\begin{cases} \min\limits_{y,u} J_k(y,u) := \frac{1}{2}\|y - y_d\|^2 + \frac{\gamma}{2}\|u\|^2 - \frac{1}{2}\langle p_k, \mathscr{A}''(y_k)[y - y_k]^2\rangle \\ \text{subject to} \quad u \in U_{ad} \cap \overline{\mathbb{B}_\rho^{L^2}(\bar{u})}, \\ \qquad \text{and} \quad \begin{cases} \partial_t y + \mathscr{A}(y_k)y + \mathscr{A}'(y_k)y = Bu + \mathscr{A}'(y_k)y_k \\ y(0) = y_0 \end{cases} \end{cases} \tag{$\mathbf{QP}(\rho, y_k, p_k)$}$$

Theorem 6.15 *Let Assumptions* 2.1–2.4 *and* 3.4 *with some* $\sigma \in (0, \infty)$ *hold. Then there are radii* $\rho > 0$, $r_{SQP} > 0$ *such that for any initial guess*

$$(y_0, p_0) \in W^{1,s}(I, W_D^{-1,p}) \cap L^s(I, W_D^{1,p}) \times W^{1,s}(I, W_D^{-1,p'}) \cap L^s(I, W_D^{1,p'})$$

fulfilling

$$\|y_0 - \bar{y}\|_{W^{1,s}(I,W_D^{-1,p}) \cap L^s(I,W_D^{1,p})} + \|p_0 - \bar{p}\|_{W^{1,s}(I,W_D^{-1,p'}) \cap L^s(I,W_D^{1,p'})} \leq r_{SQP}$$

the sequence of iterates generated by the successive solution of the SQP subproblems (**QP**(ρ, y_k, p_k)) *converges superlinearly in* X_∞ *to* $(\bar{y}, \bar{u}, \bar{p})$.

A possible choice of y_0, p_0 *are state* y_0 *and adjoint state* p_0 *associated to some control* $u_0 \in U_{ad}$ *w.r.t.* (**OCP**) *if* $\|u_0 - \bar{u}\|_{L^2}$ *is chosen small enough.*

Proof Combine Proposition 6.14 with Theorem 5.4. □

This theorem is our main result. Note in particular that we tightened the gap between the L^2-local growth condition originating from the second order sufficient conditions and the "closeness"-conditions in the SQP method. The latter had been formulated with respect to L^∞ in the existing literature. Now, in Theorem 6.15 above all required "closeness" can be formulated with respect to the L^2-norm.

7 Regularity of the adjoint state

In this section we prove the regularity required for the adjoint state in our analysis. In (Bonifacius and Neitzel 2018, Proposition 4.7) it was shown that

$$\bar{p} \in W^{1,r}(I, W_D^{-1,p'}) \cap L^r(I, W_D^{1,p'}) \qquad \forall r \in [s', \infty),$$

whereas we need additional regularity $\bar{p} \in L^\infty(I, W^{1,p'})$ as explained in Remark 4.6. In fact, we will show even higher regularity for $\bar{p}$ in the theorem below than necessary.

To improve readability of our arguments, we start with a collection of results from Bonifacius and Neitzel (2018). As further reference for maximal parabolic regularity on $H^{-\zeta,p}$-spaces we mention the work of Haller-Dintelmann and Rehberg (2009). Some of the results cited below are originally due to them.

Theorem 7.1 1. *For every right hand side* $f \in L^s(I, H_D^{-\zeta,p})$ *there is a unique solution* $y \in W^{1,s}(I, H_D^{-\zeta,p}) \cap L^s(I, \mathscr{D})$ *to the nonlinear state equation*

$$\partial_t y + \mathscr{A}(y)y = f, \qquad y(0) = y_0. \tag{E}$$

2. *The following embeddings hold true:*

(a) $\mathscr{D} \hookrightarrow W_D^{1,p} \hookrightarrow_c L^p \hookrightarrow H_D^{-\zeta,p}$

(b) $W^{1,s}(I, H_D^{-\zeta,p}) \cap L^s(I, \mathscr{D}) \hookrightarrow_c \mathscr{C}^\alpha(I, W_D^{1,p})$ *for some* $\alpha > 0$.

3. *The linear map* $W_D^{1,p} \to \mathscr{L}(\mathscr{D}, H_D^{-\zeta,p})$, $\xi \mapsto -div(\xi\mu\nabla\cdot)$ *is continuous.*
4. *Let* y *be a solution of* (E). *Then it holds:*
   (a) $\mathscr{A}(y)$ *has maximal parabolic regularity on* $L^r(I, H_D^{-\zeta,p})$ *for* $r \in (1,\infty)$.
   (b) $\mathscr{A}(y) + \mathscr{A}'(y)$ *has maximal parabolic regularity on* $L^r(I, W_D^{-1,p})$ *for every* $r \in (1, s]$.
   (c) $\mathscr{A}(y)^\bullet + \mathscr{A}'(y)^\bullet$ *(where* $^\bullet$ *indicates taking adjoints and reversing time) has maximal parabolic regularity on* $L^{r'}(I, W_D^{-1,p'})$ *for every* $r' \in [s', \infty)$.
   (d) $\mathscr{A}(y)^*+\mathscr{A}'(y)^*$ *has maximal parabolic regularity on* $L^{r'}(I, W_D^{-1,p'})$ *for every* $r' \in [s', \infty)$.
5. *For* $\tau \in (\frac{1+\zeta}{2}, 1)$ *it holds* $(H_D^{-\zeta,p}, \mathscr{D})_{\tau,1} \hookrightarrow W_D^{1,p}$.

Proof 1. Bonifacius and Neitzel (2018, Theorem 3.20 for regularity, Proposition 3.5 for existence)
2. Bonifacius and Neitzel (2018, (a) below Proposition 3.6, (b) Corollary 3.7)
3. Bonifacius and Neitzel (2018) Proposition 3.6(ii).
4. See Bonifacius and Neitzel (2018, Theorem 3.20 for (a), Proposition 4.4 is comp (resp. text between formulas (4.4) and (4.5)) for (b), Proposition 4.7 for (c)). For (d): Bonifacius and Neitzel (2018, proof of Proposition 4.7) state that every autonomous operator $\mathscr{A}(y(t))^* + \mathscr{A}'(y(t))^*$ has maximal parabolic regularity on $W_D^{-1,p'}$. Since the map $t \mapsto \mathscr{A}(y(t))^* + \mathscr{A}'(y(t))^*$ is continuous from I to $\mathscr{L}(W_D^{1,p'}, W_D^{-1,p'})$ the nonautonomous operator inherits maximal parabolic regularity, see Amann (2004, Theorem 7.2)
5. Bonifacius and Neitzel (2018, Proposition 3.6(i)).

□

Now, we fix $y \in W^{1,s}(I, H_D^{-\zeta,p}) \cap L^s(I, \mathscr{D})$. In particular, y can be a solution of (E) for some right hand side $f \in L^s(I, H_D^{-\zeta,p})$. It was shown, see Theorem 7.1 (4c) and Amann (2004, Proposition 3.1) resp. Amann (2003, formula (6.2)), that

$$\begin{aligned}\left(-\partial_t + \mathscr{A}(y)^* + \mathscr{A}'(y)^*, \mathrm{tr}_T\right)\colon\, & W^{1,r}(I, W_D^{-1,p'}) \cap L^r(I, W_D^{1,p'}) \\ & \to L^r(I, W_D^{-1,p'}) \times (W_D^{-1,p'}, W_D^{1,p'})_{1-1/r,r} \end{aligned} \tag{25}$$

is a topological isomorphism for $r \in [s', \infty)$. In fact, this also holds for every $r \in (1,\infty)$ due to continuity of $t \mapsto \mathscr{A}(y)^* + \mathscr{A}'(y)^*$ as map $I \to \mathscr{L}(W^{1,p'}, W^{-1,p'})$ by Amann (2004, Proposition 7.1 and Theorem 7.1). The required continuity with respect to time is shown by Bonifacius and Neitzel (2018) in the proof of Proposition 4.7.

We want to obtain more regularity for the adjoint state and to do so we consider restrictions of the above isomorphism onto smaller spaces of more regular functions. First, note that a short computation shows $\mathscr{A}(y)^*|_{L^r(I,W_D^{1,p})} = \mathscr{A}(y)|_{L^r(I,W_D^{1,p})}$ and similarly we can express $\mathscr{A}'(y)^*$ restricted to $L^r(I, W_D^{1,p})$ as first order differential operator $\mathscr{A}'(y)^*\varphi = \xi'(y)\mu\nabla y\nabla\varphi$. Standard Sobolev embeddings imply that under

Assumption 2.4 it holds

$$L^{p/2} \hookrightarrow H_D^{-\zeta,p}. \tag{26}$$

We already know by Theorem 7.1 (4a) that $-\mathrm{div}(\xi(y)\mu\nabla\cdot)$ has maximal parabolic regularity on $L^r(I, H_D^{-\zeta,p})$ and that $t \mapsto -\mathrm{div}(\xi(y)\mu\nabla\cdot)$ is continuous as map $I \to \mathscr{L}(\mathscr{D}, H_D^{-\zeta,p})$, which follows from Theorem 7.1 (2a) and (3). As above, we infer from Amann (2004) that

$$\begin{aligned}
&(-\partial_t - \mathrm{div}(\xi(y)\mu\nabla\cdot), \mathrm{tr}_T)\colon W^{1,r}(I, H_D^{-\zeta,p}) \cap L^r(I, \mathscr{D})\\
&\quad\to L^r(I, H_D^{-\zeta,p}) \times (H_D^{-\zeta,p}, \mathscr{D})_{1-1/r,r}
\end{aligned}$$

is a topological isomorphism. Now, choose $\frac{1+\zeta}{2} < \theta < 1$ such that $\frac{1}{r} > 1 - \theta$. It follows by (Amann 2003, formula (1.2)) and Theorem 7.1 (2a), (5) that

$$W^{1,r}(I, H_D^{-\zeta,p}) \cap L^r(I, \mathscr{D}) \hookrightarrow_c L^r(I, (H_D^{-\zeta,p}, \mathscr{D})_{\theta,1}) \hookrightarrow L^r(I, W_D^{1,p})$$

holds. Hence, the operator

$$\begin{aligned}
&\mathscr{A}'(y)^*\colon W^{1,r}(I, H_D^{-\zeta,p}) \cap L^r(I, \mathscr{D}) \hookrightarrow_c L^r(I, W_D^{1,p})\\
&\quad\to L^r(I, L^{p/2}) \hookrightarrow L^r(I, H_D^{-\zeta,p}),\\
&z \mapsto \xi'(y)\mu\nabla y\nabla z
\end{aligned}$$

is compact as it can be expressed as composition of linear operators of which one is a compact embedding. We conclude that the sum

$$\begin{aligned}
&(-\partial_t - \mathrm{div}(\xi(y)\mu\nabla\cdot) + \xi'(y)\mu\nabla y\nabla\cdot, \mathrm{tr}_T)\colon W^{1,r}(I, H_D^{-\zeta,p})\\
&\cap L^r(I, \mathscr{D})\\
&\to L^r(I, H_D^{-\zeta,p}) \times (H_D^{-\zeta,p}, \mathscr{D})_{1-1/r,r}
\end{aligned}$$

is a Fredholm-operator of index 0 for every $r \in (1, \infty)$. Since it is the restriction of the isomorphism (25) above, its kernel is trivial and therefore we actually have an isomorphism. To sum this up we have shown the following regularity result:

Theorem 7.2 *Given $y \in W^{1,s}(I, H_D^{-\zeta,p}) \cap L^s(I, \mathscr{D})$ the map*

$$\begin{aligned}
&(-\partial_t - div(\xi(y)\mu\nabla\cdot) + \xi'(y)\mu\nabla y\nabla\cdot, tr_T)\colon W^{1,r}(I, H_D^{-\zeta,p}) \cap L^r(I, \mathscr{D})\\
&\to L^r(I, H_D^{-\zeta,p}) \times (H_D^{-\zeta,p}, \mathscr{D})_{1-1/r,r}
\end{aligned}$$

is a topological isomorphism for every $r \in (1, \infty)$, i.e. the adjoint equation

$$\begin{aligned}
-\partial_t z - div(\xi(y)\mu\nabla z) + \xi'(y)\mu\nabla y\nabla z &= w,\\
z(T) &= w_T
\end{aligned}$$

admits a unique solution $z \in W^{1,r}(I, H_D^{-\zeta,p}) \cap L^r(I, \mathscr{D})$ *provided that* $w \in L^r(I, H_D^{-\zeta,p})$ *and* $w_T \in (H_D^{-\zeta,p}, \mathscr{D})_{1-1/r,r}$.

Remark 7.3 Note that we did not need more assumptions than Bonifacius and Neitzel (2018) except for the slightly higher integrability of y_d. In the framework of maximal parabolic regularity on $W_D^{-1,p}$-spaces they discuss first order necessary and second order sufficient optimality conditions, but in order to deal with the adjoint equation in the maximal parabolic regularity context (Bonifacius and Neitzel 2018, Lemma 4.6, Proposition 4.7) they required states in $\mathscr{C}^\alpha(I, W_D^{1,p})$ which was achieved by consideration of the state equation on $H_D^{-\zeta,p}$ spaces. Since we aim at SQP methods having an adjoint equation with corresponding regularity theory is necessary anyway and therefore restriction to the $H_D^{-\zeta,p}$-setting is not superfluous.

Remark 7.4 Since μ was assumed to be symmetric we could identify $\mathscr{A}(y)^*$ with $\mathscr{A}(y)$ etc. directly. In fact, all arguments go through if we postulate the same assumptions for μ^T as already done for μ.

8 Numerical Examples

In this final section we present numerical examples in order to illustrate our theoretical results. To do so we have constructed so-called manufactured solution examples, i.e. an optimal control problem with analytically known solution, see (Tröltzsch 2010, Section 2.9) for the construction of such examples. Further, we test with an example based on real-world parameters, cf. Sect. 8.2.

We implemented the SQP algorithm in `python` using an optimize-then-discretize approach and `FEniCS` (Alnæs et al. 2015; Logg et al. 2012) for the finite element discretization of the problem. Following the approach of Hintermüller and Hinze (2006) the algorithm implemented consists of three nested loops: The outermost iteration is given by the SQP method. The quadratic subproblem of each SQP iteration is solved iteratively by application of the semismooth Newton method (SSN), see e.g. Ulbrich (2011). Finally, the innermost loop consists of the iterative solution of the Newton-update equation by the CG method in every semismooth Newton iteration.

In order to solve the quadratic subproblems accurately enough we choose the relative tolerance for SSN to be 10^{-5}, i.e. the solver of the quadratic subproblems either terminates when the L^2-norm of projection residual (of the subproblem) is reduced by at least 10^{-5} or the maximum of 20 SSN iterations is reached. To avoid problems in case of already very small initial residual norms, the SSN iteration also ends when the residual norm gets smaller than 10^{-12} (absolute tolerance). Similarly, the CG method terminates if the intial CG-residual is decreased by factor at least 10^{-2}. To enhance stability, SSN is combined with Armijo linesearch with the squared L^2-norm of projection residual (of the subproblem) as merit function.

As observed in the existing literature the restriction of the quadratic subproblems to L^∞- or—in our case—L^2-balls is only required to prove convergence of the algorithm in function space. Fortunately, we can omit this additional constraint in practice and solve the quadratic subproblems on U_{ad} without loosing convergence, i.e. the subproblems in our implementation are given by (**QP**), cf. the end of Sect. 3.2.

Initial guess for the SQP method is in all three examples $(y_0, u_0, p_0) := (0, 0, 0)$. To measure optimality of some iterate u_k we compute the L^2-norm of the residual of the projection formula

$$\mathrm{res}_k := \left\| u_k - \mathrm{Proj}_{U_{ad}} \left(-\gamma^{-1} B^* p(u_k) \right) \right\|_{L^2},$$

where the adjoint state $p(u_k)$ associated to u_k is computed using the implicit Euler scheme. The nonlinear equations appearing at each timestep during the solution of the state equation are solved by the built-in nonlinear solver of `FEniCS`. Convergence of the SQP-Algorithm is measured by the increments

$$\begin{aligned} \mathrm{incr}_k^\infty &:= \|y_{k+1} - y_k\|_{L^\infty} + \|u_{k+1} - u_k\|_{L^\infty} + \|p_{k+1} - p_k\|_{L^\infty}, \\ \mathrm{incr}_k^2 &:= \|y_{k+1} - y_k\|_{L^2(I, H_D^1)} + \|y_{k+1} - y_k\|_{W^{1,2}(I, H_D^{-1})} + \|u_{k+1} - u_k\|_{L^\infty} \\ &\quad + \|p_{k+1} - p_k\|_{L^2(I, H_D^1)} + \|p_{k+1} - p_k\|_{W^{1,2}(I, H_D^{-1})}. \end{aligned}$$

Note that we do not compute the norm of the increments with respect to the norms appearing in Theorem 6.15 because we do not have the abstract exponents p, s at hand in a practical context. To illustrate our theoretical results, we show for all examples both increments and residuals for different discretizations. Convergence behaviour of the SQP method uniform with respect to sufficiently fine discretization strongly indicates convergence in function space.

8.1 Manufactured solution examples

8.1.1 Example 1

For $I = [0, 1]$ and $\Omega = [0, 1]$ we consider the problem

$$\begin{cases} \min\limits_{y,u} J(y, u) := \frac{1}{2} \|y - y_d\|_{L^2(I \times \Omega)}^2 + 10^{-3} \cdot \|u\|_{L^2([0,1])}^2 \\ \text{subject to} \quad u \in \left\{ v \in L^2([0, 1]) : -\frac{9}{10} \leq v(x) \leq \frac{\sqrt{2}}{2} \quad \text{a.e.} \right\}, \\ \quad \text{and} \quad \begin{cases} \partial_t y - \mathrm{div}(\xi(y)\nabla y) = b \cdot u + f \quad \text{on } I \times \Omega, \\ y = 0 \quad \text{on } I \times \partial\Omega, \\ y(0) = \sin(\pi x_1), \end{cases} \end{cases} \tag{27}$$

and choose

$$\begin{aligned}\bar{y}(t,x) &= \cos(2\pi t)\sin(\pi x),\\ \bar{p}(t,x) &= \frac{1}{100}\sin(2\pi t)\sin(\pi x),\\ b(x) &= \mathbf{1}_{[1/3,2/3]}(x),\\ \xi(z) &= \frac{1}{2} + \frac{1}{1+\exp(-5z)}.\end{aligned}$$

With help of `Wolfram Mathematica` we compute the remaining quantities y_d, f, $\bar{u}$ such that the optimality system for (27) is fulfilled. In particular it holds

$$\bar{u}(t) = \min\left(\frac{\sqrt{2}}{2}, \max\left(-\frac{9}{10}, -\frac{10}{\pi}\sin(2\pi t)\right)\right).$$

Note that all our theoretical results remain true for a problem of type (27) since addition of the term f to the model problem (**OCP**) does not change its structural properties.

Discretization of spatial functions is done with piecewise linear finite elements on a equidistant partition of $\Omega = [0,1]$ into N_h subintervals. For time discretization we apply an implicit Euler discretization with $N_t = N_h^2$ timesteps, whereby the number of timesteps is chosen in order to roughly balance spatial and temporal discretization errors, cf. Casas and Chrysafinos (2019). The behaviour of the increments incr_k^∞ and incr_k^2 during the SQP iteration is shown in Table 1, whereas L^2-residuals and errors of the SQP-iterates with respect to the interpolated true KKT-triple are shown in Table 2. Note that increments (Table 1a, b) and their decrease factors (Table 1c, d) indicate superlinear convergence and behave uniform with respect to the different discretization levels, which illustrates superlinear convergence in function space. Also, the residuals (Table 2a) and errors (Table 2b–f) seem to behave uniformly, at least until their convergence stagnates due to the limited accuracy given by discretization.

8.1.2 Example 2

For $I = [0,1]$ and $\Omega = [0,1]^2$ we consider a problem of the same structure as the 1D manufactured solution example (27), but now with

$$\begin{aligned}y_0(x) &= \sin(\pi x_1)\sin(\pi x_2),\\ \bar{y}(t,x) &= \cos(2\pi t)\sin(\pi x_1)\sin(\pi x_2),\\ \bar{p}(t,x) &= \frac{1}{100}\sin(2\pi t)\sin(\pi x_1)\sin(\pi x_2),\\ b(x) &= \pi^2 \cdot \mathbf{1}_{[1/3,2/3]^2}(x)\end{aligned}$$

and the regularization parameter $\gamma = 2\cdot 10^{-3}$ in (27) replaced by $\gamma = 10^{-2}$. As before, the remaining quantities are computed utilizing `Wolfram Mathematica` and the optimal control is given by

Table 1 Increments during the SQP method applied to Example 1 (Manufactured Solution in 1D, Sect. 8.1.1)

k	$N_h = 32$	$N_h = 64$	$N_h = 128$	$N_h = 32$	$N_h = 64$	$N_h = 128$
	(a) Increments incr_k^∞			*(b) Increments* incr_k^2		
0	2.15e+00	2.15e+00	2.15e+00	3.54e+00	3.54e+00	3.54e+00
1	1.89e+00	1.89e+00	1.89e+00	2.22e+00	2.22e+00	2.22e+00
2	1.46e−01	1.46e−01	1.46e−01	1.54e−01	1.54e−01	1.54e−01
3	9.00e−05	9.29e−05	9.16e−05	9.81e−05	1.01e−04	1.00e−04
4	4.96e−10	4.98e−10	5.15e−10	5.13e−10	5.17e−10	5.33e−10
5	1.52e−15	2.73e−15	4.75e−15	9.27e−15	1.82e−14	3.64e−14
	(c) Decrease of increments $\frac{\mathrm{incr}_{k+1}^\infty}{\mathrm{incr}_k^\infty}$			*(d) Decrease of increments* $\frac{\mathrm{incr}_{k+1}^2}{\mathrm{incr}_k^2}$		
0	8.79e−01	8.79e−01	8.79e−01	6.27e−01	6.28e−01	6.27e−01
1	7.74e−02	7.71e−02	7.73e−02	6.96e−02	6.93e−02	6.94e−02
2	6.15e−04	6.37e−04	6.27e−04	6.36e−04	6.58e−04	6.48e−04
3	5.51e−06	5.36e−06	5.62e−06	5.23e−06	5.10e−06	5.33e−06
4	3.06e−06	5.47e−06	9.22e−06	1.81e−05	3.52e−05	6.82e−05

$$\bar{u}(t) = \min\left(\frac{\sqrt{2}}{2}, \max\left(-\frac{9}{10}, -\sin(2\pi t)\right)\right).$$

Discretization of spatial functions is now done with piecewise linear finite elements on a triangular mesh generated by `mshr`, the mesh-generation tool of `FEniCS`, with maximum element diameter $h_{\max}$. For time discretization we apply an implicit Euler discretization with N_t timesteps, whereby the size of timesteps $\tau = N_t^{-1} \approx h_{\max}^2$ is chosen in order to roughly balance spatial and temporal discretization errors, cf. Casas and Chrysafinos (2019). Maximum element diameter and number of timesteps of the four different discretization levels used in our numerical experiments can be found in Table 3. In Table 4 we display increments and their decrease rates during the SQP iteration. Similarly to the 1D manufactured solution example these quantities behave uniform with respect to different discretization levels, which illustrates convergence in function space. Moreover, residuals (Table 5a) and errors of the iterates with respect to the interpolated true KKT-triple (Table 5b–f) show uniform behaviour until stagnation due to the respective discretization occurs.

8.2 Example 3

This final example is chosen to demonstrate that our assumptions also cover an example with real-world parameters. We consider the following problem related to heat conduction in a block of silicon modelled according to Selberherr (1984):

Table 2 Residuals and errors of the iterates during the SQP method applied to Example 1 (Manufactured Solution in 1D, Sect. 8.1.1). We use the abbreviation $\|\cdot\|_W := \|\cdot\|_{L^2(I,H_D^1)} + \|\cdot\|_{W^{1,2}(I,H_D^{-1})}$

k	$N_h = 32$	$N_h = 64$	$N_h = 128$	$N_h = 32$	$N_h = 64$	$N_h = 128$
	(a) Residuals res_k			*(b) Error in the control* $\|u_k - \bar{u}\|_{L^\infty}$		
0	7.60e−01	7.56e−01	7.58e−01	9.00e−01	9.00e−01	9.00e−01
1	9.24e−01	9.23e−01	9.26e−01	1.61e+00	1.61e+00	1.61e+00
2	1.13e−01	1.13e−01	1.13e−01	1.70e−01	1.48e−01	1.43e−01
3	3.29e−05	3.85e−05	3.78e−05	3.48e−02	1.59e−02	6.09e−03
4	3.60e−06	3.29e−07	2.51e−07	3.48e−02	1.60e−02	6.04e−03
5	3.60e−06	3.29e−07	2.51e−07	3.48e−02	1.60e−02	6.04e−03
6	3.60e−06	3.29e−07	2.51e−07	3.48e−02	1.60e−02	6.04e−03
	(c) Error in the state $\|y_k - \bar{y}\|_{L^\infty}$			*(d) Error in the state* $\|y_k - \bar{y}\|_W$		
0	1.00e+00	1.00e+00	1.00e+00	2.61e+00	2.61e+00	2.61e+00
1	2.56e−01	2.57e−01	2.56e−01	5.63e−01	5.69e−01	5.68e−01
2	6.11e−03	6.68e−03	6.65e−03	1.50e−02	1.40e−02	1.40e−02
3	1.85e−03	6.31e−04	2.06e−04	4.73e−03	1.23e−03	4.29e−04
4	1.85e−03	6.32e−04	2.05e−04	4.73e−03	1.23e−03	4.28e−04
5	1.85e−03	6.32e−04	2.05e−04	4.73e−03	1.23e−03	4.28e−04
6	1.85e−03	6.32e−04	2.05e−04	4.73e−03	1.23e−03	4.28e−04
	(e) Error in the adjoint state $\|p_k - \bar{p}\|_{L^\infty}$			*(f) Error in the adjoint state* $\|p_k - \bar{p}\|_W$		
0	1.00e−02	1.00e−02	1.00e−02	2.61e−02	2.61e−02	2.61e−02
1	2.15e−02	2.16e−02	2.15e−02	4.42e−02	4.47e−02	4.46e−02
2	1.10e−03	1.09e−03	1.07e−03	2.28e−03	2.22e−03	2.15e−03
3	1.32e−04	4.29e−05	1.12e−05	3.36e−04	9.45e−05	2.50e−05
4	1.31e−04	4.27e−05	1.15e−05	3.36e−04	9.41e−05	2.48e−05
5	1.31e−04	4.27e−05	1.15e−05	3.36e−04	9.41e−05	2.48e−05
6	1.31e−04	4.27e−05	1.15e−05	3.36e−04	9.41e−05	2.48e−05

Table 3 Discretization levels for Example 2 (Manufactured Solution in 2D, Sect. 8.1.2)

	Level 1	Level 2	Level 3
$h_{\max}$	$7.95 \cdot 10^{-2}$	$3.98 \cdot 10^{-2}$	$1.99 \cdot 10^{-2}$
N_t	158	632	2529

Table 4 Increments during the SQP method applied to Example 2 (Manufactured Solution in 2D, Sect. 8.1.2)

k	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
	(a) Increments incr_k^∞			*(b) Increments* incr_k^2		
0	2.15e+00	2.16e+00	2.16e+00	2.75e+00	2.67e+00	2.61e+00
1	1.04e+00	1.05e+00	1.06e+00	1.10e+00	1.11e+00	1.11e+00
2	2.57e−02	2.38e−02	2.22e−02	2.70e−02	2.52e−02	2.35e−02
3	6.32e−06	7.45e−06	7.99e−06	6.09e−06	7.21e−06	7.86e−06
4	1.84e−11	2.03e−12	1.93e−12	1.90e−11	1.10e−12	1.04e−12
	(*c*) *Decrease of increments* $\frac{\mathrm{incr}_{k+1}^\infty}{\mathrm{incr}_k^\infty}$			*(d) Decrease of increments* $\frac{\mathrm{incr}_{k+1}^2}{\mathrm{incr}_k^2}$		
0	4.82e−01	4.88e−01	4.89e−011	4.00e−01	4.17e−01	4.26e−01
1	2.49e−02	2.26e−02	2.10e−02	2.46e−02	2.26e−02	2.11e−02
2	2.45e−04	3.13e−04	3.60e−04	2.25e−04	2.87e−04	3.34e−04
3	2.91e−06	2.72e−07	2.42e−07	3.13e−06	1.53e−07	1.32e−07

$$\left\{\begin{array}{l} \min\limits_{y,u} J(y,u) := \dfrac{1}{2}\|y-y_d\|^2_{L^2(I\times\Omega)} + 10^{-2}\cdot\|u\|^2_{L^2(I)} \\ \text{subject to} \quad u \in \Big\{v \in L^2(I) : 2.9 \le v(t) \le 10\Big\} \\ \qquad\quad \text{and} \quad \left\{\begin{array}{c} \partial_t y - \operatorname{div}(\xi(y)\nabla y) = 0 \quad \text{on } I\times\Omega, \\ \xi(y)\partial_{n_\Omega} y + \alpha y = \alpha u \quad \text{on } I\times\partial\Omega, \\ y(0) = 10. \end{array}\right. \end{array}\right. \tag{28}$$

The spatial domain is

$$\Omega = [-2,2]\times[-0.5,0.5]\times[-1,0] \cup [-0.5,0.5]\times[-2,2]\times[0,1] \subset \mathbb{R}^3,$$

consisting of two crossed beams, the time interval $I = [0,T] = [0,40]$, the desired state

$$y_d(t,x) = 10 - \frac{71}{400}t,$$

the nonlinearity given by

$$\xi(y) := \frac{1}{a+by+cy^2}, \qquad a = 0.0818292, \quad b = 0.4255118, \quad c = 0.0450061,$$

Table 5 Residuals and errors of the iterates during the SQP method applied to Example 2 (Manufactured Solution in 2D, Sect. 8.1.2). We use the abbreviation $\|\cdot\|_W := \|\cdot\|_{L^2(I,H^1_D)} + \|\cdot\|_{W^{1,2}(I,H^{-1}_D)}$

k	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
	(a) Residuals res_k			*(b) Error in the control* $\|u_k - \bar{u}\|_{L^\infty}$		
0	7.60e−01	7.50e−01	7.43e−01	9.00e−01	9.00e−01	9.00e−01
1	8.05e−01	7.99e−01	7.94e−01	8.23e−01	8.13e−01	8.12e−01
2	2.21e−02	2.36e−02	2.45e−02	1.33e−01	9.08e−02	5.54e−02
3	5.72e−06	1.39e−06	1.13e−06	1.37e−01	8.09e−02	4.48e−02
4	6.88e−06	3.80e−07	1.05e−06	1.37e−01	8.09e−02	4.48e−02
5	6.88e−06	3.80e−07	1.05e−06	1.37e−01	8.09e−02	4.48e−02
	(c) Error in the state $\|y_k - \bar{y}\|_{L^\infty}$			*(d) Error in the state* $\|y_k - \bar{y}\|_W$		
0	1.00e+00	1.00e+00	1.00e+00	1.81e+00	1.72e+00	1.67e+00
1	2.33e−01	2.40e−01	2.45e−01	2.82e−01	2.93e−01	2.93e−01
2	3.02e−02	1.62e−02	1.02e−02	4.15e−02	2.06e−02	1.13e−02
3	2.93e−02	1.49e−02	7.48e−03	4.06e−02	1.93e−02	9.27e−03
4	2.93e−02	1.49e−02	7.47e−03	4.06e−02	1.93e−02	9.27e−03
5	2.93e−02	1.49e−02	7.47e−03	4.06e−02	1.93e−02	9.27e−03
	(e) Error in the adjoint state $\|p_k - \bar{p}\|_{L^\infty}$			*(f) Error in the adjoint state* $\|p_k - \bar{p}\|_W$		
0	1.00e−02	1.00e−02	1.00e−02	1.81e−02	1.72e−02	1.67e−02
1	7.66e−03	8.72e−03	9.12e−03	1.17e−02	1.22e−02	1.22e−02
2	1.08e−03	5.52e−04	3.10e−04	1.39e−03	6.26e−04	3.15e−04
3	1.06e−03	6.07e−04	3.20e−04	1.46e−03	6.71e−04	3.18e−04
4	1.06e−03	6.07e−04	3.20e−04	1.46e−03	6.71e−04	3.18e−04
5	1.06e−03	6.07e−04	3.20e−04	1.46e−03	6.71e−04	3.18e−04

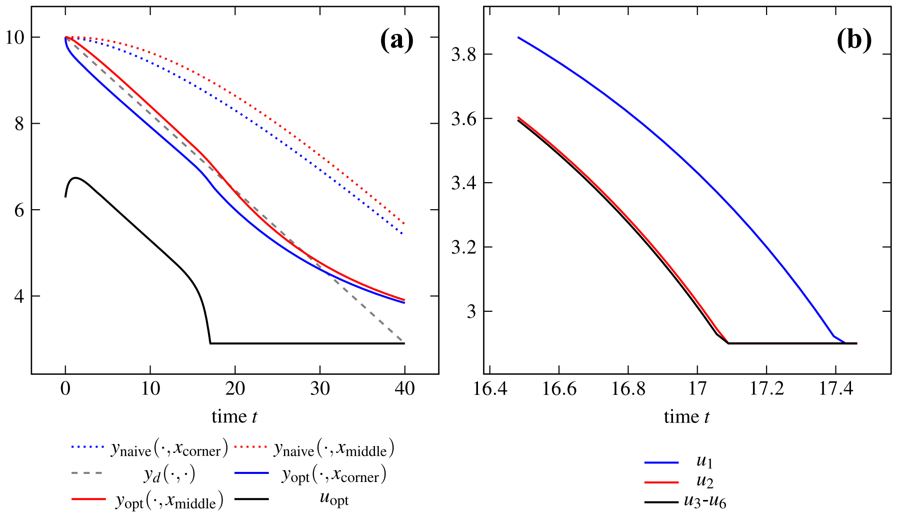

Fig. 1 Example 3 (Sect. 8.2) on the finest discretization level: **a** Optimal control and optimal state y_{opt} evaluated at the points $x_{\text{corner}} = (-0, 5, 2, 0)$ and $x_{\text{middle}} = (0, 0, 0)$ (left hand side plot). As comparison we also display the state y_{naive} associated with the "naive" first guess $u_{\text{naive}}(t) := 10 - \frac{71}{400}t$ at the same points and the desired trajectory y_d. **b** Control iterates during the SQP method on a certain subinterval of I

Table 6 Discretization levels for Example 3 (Sect. 8.2)

	Level 1	Level 2	Level 3
$h_{\max}$	$5.90 \cdot 10^{-1}$	$2.84 \cdot 10^{-1}$	$1.84 \cdot 10^{-1}$
N_t	115	495	1182

and $\alpha = 0.0146647$. In order to make ξ formally fulfill Assumption 2.2 we can choose a C^2-continuous uniformly bounded from below and above continuation of the above ξ outside the relevant values of y.

Measuring temperature in units of 100K, length in 0.1m and time in 60s, the state equation of (28) describes the evolution of the temperature y of a block Ω of silicon with initial temperature 1000K when the temperature of the surrounding air is given by the control variable u. Hence, the optimal control problems aims at finding the optimal temperature trajectory for the ambient air in order to cool down the block Ω following the desired temperature trajectory y_d from 1000K to room temperature 290K. Density, specific heat, and temperature-dependent thermal conductivity are taken from Selberherr (1984, Chapters 2.5 and 4.3) and rescaled according to the abovementioned units. For the heat transfer coefficient between silicon and air (forced convection) we guess the value 40W$\text{m}^{-2}\text{K}^{-1}$ which results in the value given for α.

As pointed out after Assumption 2.1 the domain under consideration fulfills our assumptions although not being a domain with Lipschitz boundary. The Robin boundary condition in (28) is not covered by our assumptions, but since it differs from Neumann boundary conditions only by a linear term, this can be tackled by straightforward modifications of our arguments, cf. also Meinlschmidt et al. (2017a, b).

Table 7 Increments during the SQP method applied to Example 3 (Sect. 8.2)

k	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
	(a) Increments incr_k^∞			*(b) Increments* incr_k^2		
0	1.71e+01	1.72e+01	1.72e+01	1.91e+02	1.85e+02	1.75e+02
1	1.16e+00	1.15e+00	1.17e+00	1.08e+01	1.09e+01	1.07e+01
2	2.24e−02	2.52e−02	2.58e−02	1.42e−01	1.50e−01	1.49e−01
3	5.23e−06	1.15e−05	6.77e−06	3.78e−05	4.51e−05	4.01e−05
4	2.78e−11	2.40e−10	3.37e−10	4.05e−11	2.69e−10	3.74e−10
5	–	1.03e−13	2.80e−13	-	1.29e−12	2.12e−12
	(c) Decrease of increments $\frac{\mathrm{incr}_{k+1}^\infty}{\mathrm{incr}_k^\infty}$			*(d) Decrease of increments* $\frac{\mathrm{incr}_{k+1}^2}{\mathrm{incr}_k^2}$		
0	6.79e−02	6.70e−02	6.80e−02	5.68e−02	5.88e−02	6.10e−02
1	1.93e−02	2.19e−02	2.21e−02	1.31e−02	1.38e−02	1.40e−02
2	2.33e−04	4.57e−04	2.62e−04	2.65e−04	3.01e−04	2.68e−04
3	5.31e−06	2.09e−05	4.98e−05	1.07e−06	5.96e−06	9.33e−06
4	–	4.30e−04	8.31e−04	–	4.81e−03	5.67e−03

Table 8 Residuals during the SQP-method applied to Example 3 (Sect. 8.2)

k	Residuals res_k		
	Level 1	Level 2	Level 3
0	2.58e+01	2.59e+01	2.59e+01
1	1.15e+01	1.16e+01	1.16e+01
2	1.69e+00	1.72e+00	1.72e+00
3	4.23e−04	4.28e−04	4.28e−04
4	2.09e−08	1.10e−09	1.44e−08
5	2.09e−08	1.06e−09	1.45e−08
6	–	1.06e−09	1.45e−08

All computations were performed on tetrahedral meshes generated by `mshr` with maximal cell diameter $h_{\max}$ and N_t implicit Euler timesteps, see Table 6 for the different discretization levels. The numerically determined optimal control and associated optimal state are shown in Figure 1 a). Due to the three-dimensionality of the problem we were not able to choose discretization as fine as in the previous examples and therefore the behaviour the increments (Table 7) and residuals (Table 8) is not as illustrative as in 1D or 2D.

Figure 1 b) shows an enlarged section of the control iterates near the change from inactive to active set at $t \approx 17.1$: It can be seen that once the correct active set is identified after the third iteration, convergence is so fast that there is no visible difference between the further iterates. This might be seen as an illustration of the importance of detection of the correct active sets in infinite dimensions that has been discussed at the beginning of Sect. 6. The small kinks in the plots at the border between active and inactive set are due to the fact that time discretization (size of timesteps $\tau \approx 3.38 \cdot 10^{-2}$) does not resolve the active/inactive sets exactly.

Acknowledgements Open Access funding provided by Projekt DEAL. Open Access funding provided by Projekt DEAL. The authors are grateful to Hannes Meinlschmidt (RICAM, Linz) for suggesting a more elegant and shorter proof for Theorem 7.2 than our original argument. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 211504053 - SFB 1060.

References

Alnæs MS, Blechta J, Hake J, Johansson A, Kehlet B, Logg A, Richardson C, Ring J, Rognes ME, Wells GN (2015) The fenics project version 1.5. Arch Numer Softw 3(100):9–23. https://doi.org/10.11588/ans.2015.100.20553

Alt W (1990) The Lagrange–Newton method for infinite-dimensional optimization problems. Numer Funct Anal Optim 11(3–4):201–224. https://doi.org/10.1080/01630569008816371

Alt W, Griesse R, Metla N, Rösch A (2010) Lipschitz stability for elliptic optimal control problems with mixed control-state constraints. Optimization 59(5–6):833–849. https://doi.org/10.1080/02331930902863749

Amann H (2003) Nonautonomous parabolic equations involving measures. Zap Nauchn Sem S-Peterburg Otdel Mat Inst Steklov (POMI) 306(Kraev. Zadachi Mat. Fiz. i Smezh. Vopr. Teor. Funktsii. 34):16–52, 229. https://doi.org/10.1007/s10958-005-0376-8

Amann H (2004) Maximal regularity for nonautonomous evolution equations. Adv Nonlinear Stud 4(4):417–430. https://doi.org/10.1515/ans-2004-0404

Amann H (2005) Quasilinear parabolic problems via maximal regularity. Adv Differ Equ 10(10):1081–1110

Bonifacius L, Neitzel I (2018) Second order optimality conditions for optimal control of quasilinear parabolic equations. Math Control Relat Fields 8(1):1–34. https://doi.org/10.3934/mcrf.2018001

Bonnans JF (1998) Second-order analysis for control constrained optimal control problems of semilinear elliptic systems. Appl Math Optim 38(3):303–325. https://doi.org/10.1007/s002459900093

Casas E, Chrysafinos K (2018) Analysis and optimal control of some quasilinear parabolic equations. Math Control Relat Fields 8(3–4):607–623

Casas E, Chrysafinos K (2019) Numerical analysis of quasilinear parabolic equations under low regularity assumptions. Numer Math. https://doi.org/10.1007/s00211-019-01071-5

Casas E, Dhamo V (2011) Optimality conditions for a class of optimal boundary control problems with quasilinear elliptic equations. Control Cybern 40(2):457–490

Casas E, Tröltzsch F (2009) First- and second-order optimality conditions for a class of optimal control problems with quasilinear elliptic equations. SIAM J Control Optim 48(2):688–718. https://doi.org/10.1137/080720048

Casas E, Tröltzsch F (2011) Numerical analysis of some optimal control problems governed by a class of quasilinear elliptic equations. ESAIM Control Optim Calc Var 17(3):771–800. https://doi.org/10.1051/cocv/2010025

Casas E, Tröltzsch F (2012) A general theorem on error estimates with application to a quasilinear elliptic optimal control problem. Comput Optim Appl 53(1):173–206. https://doi.org/10.1007/s10589-011-9453-8

Casas E, Tröltzsch F (2012) Second order analysis for optimal control problems: improving results expected from abstract theory. SIAM J Optim 22(1):261–279. https://doi.org/10.1137/110840406

Casas E, Tröltzsch F (2015) Second order optimality conditions and their role in PDE control. Jahresber Dtsch Math-Ver 117(1):3–44. https://doi.org/10.1365/s13291-014-0109-3

de Los Reyes JC, Dhamo V (2016) Error estimates for optimal control problems of a class of quasilinear equations arising in variable viscosity fluid flow. Numer Math 132(4):691–720. https://doi.org/10.1007/s00211-015-0737-2

de Los Reyes JC, Merino P, Rehberg J, Tröltzsch F (2008) Optimality conditions for state-constrained PDE control problems with time-dependent controls. Control Cybern 37(1):5–38

Dontchev AL (1996) Local analysis of a Newton-type method based on partial linearization. In: The mathematics of numerical analysis (Park City, UT, 1995), Lectures in Applied Mathematics, vol 32. American Mathematical Society, Providence, RI, pp 295–306

Feldhordt H (2017) Boundary control of a chemotaxis system. Dissertation, Univeristät Duisburg-Essen. https://nbn-resolving.org/urn:nbn:de:hbz:464-20170809-110745-3

Goldberg H, Tröltzsch F (1989) Second order optimality conditions for a class of control problems governed by nonlinear integral equations with application to parabolic boundary control. Optimization 20(5):687–698. https://doi.org/10.1080/02331938908843489

Goldberg H, Tröltzsch F (1998) On a Lagrange-Newton method for a nonlinear parabolic boundary control problem. Optim Methods Softw 8(3–4):225–247. https://doi.org/10.1080/10556789808805678

Griepentrog JA, Gröger K, Kaiser HC, Rehberg J (2002) Interpolation for function spaces related to mixed boundary value problems. Math Nachr 241:110–120. https://doi.org/10.1002/1522-2616(200210)244:1<110::AID-MANA110>3.0.CO;2-S

Griesse R, Metla N, Rösch A (2008) Convergence analysis of the SQP method for nonlinear mixed-constrained elliptic optimal control problems. ZAMM Z Angew Math Mech 88(10):776–792. https://doi.org/10.1002/zamm.200800036

Griesse R, Metla N, Rösch A (2010) Local quadratic convergence of SQP for elliptic optimal control problems with mixed control-state constraints. Control Cybern 39(3):717–738

Gröger K (1989) A $W^{1,p}$-estimate for solutions to mixed boundary value problems for second order elliptic differential equations. Math Ann 283(4):679–687. https://doi.org/10.1007/BF01442860

Haller-Dintelmann R, Rehberg J (2009) Maximal parabolic regularity for divergence operators including mixed boundary conditions. J Differ Equ 247(5):1354–1396. https://doi.org/10.1016/j.jde.2009.06.001

Haller-Dintelmann R, Meyer C, Rehberg J, Schiela A (2009) Hölder continuity and optimal control for nonsmooth elliptic problems. Appl Math Optim 60(3):397–428. https://doi.org/10.1007/s00245-009-9077-x

Heinkenschloss M, Tröltzsch F (1999) Analysis of the Lagrange–SQP–Newton method for the control of a phase field equation. Control Cybern 28(2):177–211

Hintermüller M, Hinze M (2006) A SQP-semismooth Newton-type algorithm applied to control of the instationary Navier–Stokes system subject to control constraints. SIAM J Optim 16(4):1177–1200. https://doi.org/10.1137/030601259

Hintermüller M, Volkwein S, Diwoky F (2007) Fast solution techniques in constrained optimal boundary control of the semilinear heat equation. In: Kunisch K, Sprekels J, Leugering G, Tröltzsch F (eds) Control of coupled partial differential equations, vol 155. International series of numerical mathematics. Birkhäuser, Basel, pp 119–147

Hinze M, Kunisch K (2001) Second order methods for optimal control of time-dependent fluid flow. SIAM J Control Optim 40(3):925–946. https://doi.org/10.1137/S0363012999361810

Hinze M, Pinnau R, Ulbrich M, Ulbrich S (2009) Optimization with PDE constraints, vol 23. Mathematical modelling: theory and applications. Springer, New York

Ioffe AD (1979) Necessary and sufficient conditions for a local minimum. III. Second order conditions and augmented duality. SIAM J Control Optim 17(2):266–288. https://doi.org/10.1137/0317021

Ito K, Kunisch K (2004) The primal-dual active set method for nonlinear optimal control problems with bilateral constraints. SIAM J Control Optim 43(1):357–376. https://doi.org/10.1137/S0363012902411015

Josephy N (1979) Newton's method for generalized equations. Technical report

Lions JL (1971) Optimal control of systems governed by partial differential equations. Translated from the French by S. K. Mitter. Die Grundlehren der mathematischen Wissenschaften, Band 170, Springer-Verlag, New York-Berlin

Logg A, Mardal KA, Wells GN et al (2012) Automated solution of differential equations by the finite element method. Springer, Berlin. https://doi.org/10.1007/978-3-642-23099-8

Meinlschmidt H, Rehberg J (2016) Hölder-estimates for non-autonomous parabolic problems with rough data. Evol Equ Control Theory 5(1):147–184. https://doi.org/10.3934/eect.2016.5.147

Meinlschmidt H, Meyer C, Rehberg J (2017a) Optimal control of the thermistor problem in three spatial dimensions, part 1: existence of optimal solutions. SIAM J Control Optim 55(5):2876–2904. https://doi.org/10.1137/16M1072644

Meinlschmidt H, Meyer C, Rehberg J (2017b) Optimal control of the thermistor problem in three spatial dimensions, part 2: optimality conditions. SIAM J Control Optim 55(4):2368–2392. https://doi.org/10.1137/16M1072656

Nicaise S, Tröltzsch F (2017) Optimal control of some quasilinear Maxwell equations of parabolic type. Discrete Contin Dyn Syst Ser S 10(6):1375–1391. https://doi.org/10.3934/dcdss.2017073

Robinson SM (1980) Strongly regular generalized equations. Math Oper Res 5(1):43–62. https://doi.org/10.1287/moor.5.1.43

Selberherr S (1984) Analysis and simulation of semiconductor devices. Springer, Berlin

Tröltzsch F (1999) On the Lagrange-Newton-SQP method for the optimal control of semilinear parabolic equations. SIAM J Control Optim 38(1):294–312. https://doi.org/10.1137/S0363012998341423

Tröltzsch F (2000) Lipschitz stability of solutions of linear-quadratic parabolic control problems with respect to perturbations. Dyn Contin Discrete Impuls Syst 7(2):289–306

Tröltzsch F (2010) Optimal control of partial differential equations: theory. methods and applications. Graduate studies in mathematics, vol 112. American Mathematical Society, Providence, RI. https://doi.org/10.1090/gsm/112. Translated from the 2005 German original by Jürgen Sprekels

Tröltzsch F, Wachsmuth D (2006) Second-order sufficient optimality conditions for the optimal control of Navier–Stokes equations. ESAIM Control Optim Calc Var 12(1):93–119. https://doi.org/10.1051/cocv:2005029

Ulbrich M (2011) Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces, vol 11. MOS-SIAM series on optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia. https://doi.org/10.1137/1.9781611970692

Ulbrich S, Ziems JC (2017) Adaptive multilevel trust-region methods for time-dependent PDE-constrained optimization. Port Math 74(1):37–67. https://doi.org/10.4171/PM/1992

Wachsmuth D (2007) Analysis of the SQP-method for optimal control problems governed by the nonstationary Navier–Stokes equations based on L^p-theory. SIAM J Control Optim 46(3):1133–1153. https://doi.org/10.1137/S0363012904443506

Yousept I (2013) Optimal control of quasilinear h(curl)-elliptic partial differential equations in magnetostatic field problems. SIAM J Control Optim 51(5):3624–3651. https://doi.org/10.1137/120904299

Ziems JC (2013) Adaptive multilevel inexact SQP-methods for PDE-constrained optimization with control constraints. SIAM J Optim 23(2):1257–1283. https://doi.org/10.1137/110848645

Ziems JC, Ulbrich S (2011) Adaptive multilevel inexact SQP methods for PDE-constrained optimization. SIAM J Optim 21(1):1–40. https://doi.org/10.1137/080743160