ON THE LAGRANGE-NEWTON-SQP METHOD FOR THE OPTIMAL CONTROL OF SEMILINEAR PARABOLIC EQUATIONS*

FREDI TRÖLTZSCH †

Abstract. A class of Lagrange-Newton-SQP methods is investigated for optimal control problems governed by semilinear parabolic initial- boundary value problems. Distributed and boundary controls are given, restricted by pointwise upper and lower bounds. The convergence of the method is discussed in appropriate Banach spaces. Based on a weak second order sufficient optimality condition for the reference solution, local quadratic convergence is proved. The proof is based on the theory of Newton methods for generalized equations in Banach spaces.

Key words. optimal control, parabolic equation, semilinear equation, sequential quadratic programming, Lagrange-Newton method, convergence analysis

AMS subject classifications. 49J20,49M15,65K10,49K20

1. Introduction. This paper is concerned with the numerical analysis of a Sequential Quadratic Programming Method for optimal control problems governed by semilinear parabolic equations. We extend convergence results obtained in the author's papers [31] and [32] for simplified cases. Here, we allow for distributed and boundary control. Moreover, terminal, distributed, and boundary observation are included in the objective functional. In contrast to the former papers, where a semigroup approach was chosen to deal with the parabolic equations, the theory is now presented in the framework of weak solutions relying on papers by Casas [7], Raymond and Zidani [28], and Schmidt [30]. We refer also to Heinkenschloss and Tröltzsch [15], where the convergence of an SQP method is proved for the optimal control of a phase field model. Including first order sufficient optimality conditions in the considerations, we are able to essentially weaken the second order sufficient optimality conditions needed to prove the convergence of the method. These sufficient conditions tighten the gap to the associated necessary ones. However, the approach requires a quite extensive analysis.

SQP methods for the optimal control of ODEs have already been the subject of many papers. We refer, for instance, to the discussion of quadratic convergence and the associated numerical examples by Alt [1], [2], Alt and Malanowski [5], [6], to the mesh independence principle in Alt [3], and to the numerical application by Machielsen [27]. Moreover, we refer to the more extensive references therein. For a paper standing in some sense between the control of ODEs and PDEs we refer to Alt, Sontag and Tröltzsch [4], who investigated the control of weakly singular Hammerstein integral equations.

Following recent developments for ordinary differential equations, we adopt here the relation between the SQP method and a generalized Newton method. This approach makes the whole theory more transparent. We are able to apply known results on the convergence of generalized Newton methods in Banach spaces assuming the so called strong regularity at the optimal reference point. In this way, the convergence analysis is shorter, and we are able to concentrate on specific questions arising from the presence of partial differential equations.

* This research was partially supported by Deutsche Forschungsgemeinschaft, under Project number Tr 302/1-2.

† Fakultät für Mathematik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany

Once the convergence of the Newton method is shown, we still need an extensive analysis to make the theory complete. We have to ensure the strong regularity by sufficient conditions and to show that the Newton steps can be performed by solving linear-quadratic control problems (SQP-method). This interplay between the Newton method and the SQP method is a specific feature, which cannot be derived from general results in Banach spaces, since we have to discuss pointwise relations.
We should underline that this paper does not aim to discuss the numerical application of the method. Any computation has to be connected with a discretization of the problem. This gives rise to consider approximation errors, stability estimates, the interplay between mesh adaption and precision (particularly delicate for PDEs) and the numerical implementation. Besides the fact that some of these questions are still unsolved, the presentation of the associated theory would go far beyond the scope of one paper. We understand the analysis of our paper as a general line applicable to any proof of convergence for these numerical methods. Some test examples close to this paper were presented by Goldberg and Tröltzsch [11], [12]. The fast convergence of the SQP method is demonstrated there by examples in spatial domains of dimension one and two relying on a fine discretization of the problems. Lagrange-Newton type methods were also discussed for partial differential equations by Heinkenschloss and Sachs [14], Ito and Kunisch [16], [17], Kelley and Sachs [19], [20], [21], Kupfer and Sachs [23], Heinkenschloss [13], and Kunisch and Volkwein [22] who report in much more detail on the numerical details needed for an effective implementation.
The paper is organized as follows. Section 2 is concerned with existence and uniqueness of weak solutions for the equation of state. After stating the problem and associated necessary and sufficient optimality conditions in section 3, the generalized Newton method is established in section 4. The strong stability of the generalized equation is discussed in Section 5, while section 6 is concerned with performing the Newton steps by SQP steps.

2. The equation of state. The dynamics of our control system is described by the semilinear parabolic initial-boundary value problem

$$\begin{aligned} y_t(x,t) + \operatorname{div}\,(\mathcal{A}(x)\,\mathrm{grad}_x y(x,t)) + d(x,t,y(x,t),v(x,t)) &= 0 \quad \text{in } Q \\ \partial_\nu y(x,t) + b(x,t,y(x,t),u(x,t)) &= 0 \quad \text{on } \Sigma \\ y(x,0) - y_0(x) &= 0 \quad \text{on } \Omega\,. \end{aligned} \tag{2.1}$$

This system is considered in $Q = \Omega \times (0,T)$, where $\Omega \subset \mathbb{R}^N (N \geq 2)$ is a bounded domain and $T > 0$ a fixed time. By ∂_ν the co-normal derivative $\partial y/\partial \nu_A = -\nu^\top \mathcal{A} \nabla y$ is denoted, where ν is the outward normal on Γ. The functions u, v denote *boundary* and *distributed control*, $\Sigma = \Gamma \times (0,T)$, $\Gamma = \partial\Omega$, and y_0 is a fixed initial state function. Following [7] and [28] we impose the following assumptions on the data:

(A1) Γ is of class $C^{2,\alpha}$ for some $\alpha \in (0,1]$. The coefficients a_{ij} of the matrix $\mathcal{A} = (a_{ij})_{i,j=1,\ldots,N}$ belong to $C^{1,\alpha}(\overline{\Omega})$, and there is $m_0 > 0$ such that

$$-\xi^\top \mathcal{A}(x)\,\xi \geq m_0\,|\xi|^2 \quad \forall \xi \in \mathbb{R}^N, \forall x \in \overline{\Omega}. \tag{2.2}$$

$\mathcal{A}(x)$ is (w.l.o.g.) symmetric .

(A2) The "distributed" nonlinearity $d = d(x,t,y,v)$ is defined on $\overline{Q} \times \mathbb{R}^2$ and satisfies the following Carathéodory type condition:

- **(i)** For all $(y,v) \in \mathbb{R}^2$, $d(\cdot,\cdot,y,v)$ is Lebesgue measurable on Q.
- **(ii)** For almost all $(x,t) \in Q$, $d(x,t,\cdot,\cdot)$ is of class $C^{2,1}(\mathbb{R}^2)$.

The "boundary" nonlinearity $b = b(x, t, y, u)$ is defined on $\Sigma \times \mathbb{R}^2$ and is supposed to fulfill (i), (ii) with Σ substituted for Q.

In our setting, the controls u, v will be uniformly bounded by a certain constant K.

(A3) The functions d, b fulfill the *assumptions of boundedness*

(i)

$$|d(x,t,0,v)| \le d_K(x,t) \quad \forall (x,t) \in Q, |v| \le K, \tag{2.3}$$

where $d_K \in L^q(Q)$ and $q > \frac{N}{2} + 1$. There is a number $c_0 \in \mathbb{R}$, and a non-decreasing function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$c_0 \le d_y(x,t,y,v) \le \eta(|y|) \tag{2.4}$$

for a.e. $(x,t) \in Q$, all $y \in \mathbb{R}$, all $|v| \le K$.

(ii)

$$|b(x,t,0,u)| \le b_K(x,t) \quad \forall (x,t) \in \Sigma, |u| \le K \tag{2.5}$$

and

$$c_0 \le b_y(x,t,y,u) \le \eta(|y|) \tag{2.6}$$

for a.e. $(x,t) \in \Sigma$, all $y \in \mathbb{R}$, all $|u| \le K$, where $b_K \in L^r(\Sigma)$, $r > N+1$.

The assumptions imply those supposed in [7], [28], since our controls are uniformly bounded. The $C^{2,1}$-assumption on d, b is not necessary for the discussion of the equation of state. We shall need it for the Lagrange-Newton method. Although the discussion of existence and uniqueness for the nonlinear system (2.1) is not necessary for our analysis we quote the following result from [7], [28]:

THEOREM 2.1. *Suppose that (A1)-(A3) are satisfied, $y_0 \in C(\overline{\Omega})$, $v \in L^\infty(Q)$, $u \in L^\infty(\Sigma)$. Then the system (2.1) admits a unique weak solution $y \in L^2(0,T;H^1(\Omega)) \cap C(\overline{\Omega})$.*

A *weak solution* of (2.1) is a function y of $L^2(0,T;H^1(\Omega)) \cap C(\overline{Q})$ such that

$$\begin{aligned} &-\int_Q (y \cdot p_t + (\nabla_x y)^\top \mathcal{A}(x) \nabla_x p)\, dxdt + \int_Q d(x,t,y,v)\; p\; dxdt + \\ &\quad + \int_\Sigma b(x,t,y,u)\; p\; dSdt - \int_\Omega y_0(x) p(x,0) dx = 0 \end{aligned} \tag{2.7}$$

holds for all $p \in W_2^{1,1}(Q)$ satisfying $p(x,T) = 0$. In (2.7 we have assumed that $y \in C(\overline{Q})$) to make the nonlinearities d, b well defined. Theorem 2.1 was shown by a detailed discussion of regularity for an associated linear equation. This linear version of Theorem 2.1 is more important for our analysis. In what follows, we shall use the symbol $A = \operatorname{div} \mathcal{A} \operatorname{grad} y$. Moreover, we need the space $W(0,T) = \{y \in L^2(0,T;H^1(\Omega)) | y_t \in L^2(0,T;H^1(\Omega)')\}$. Regard the linear initial-boundary value problem

$$\begin{aligned} y_t + A\,y + a\,y &= v \quad &&\text{on } Q \\ \partial_\nu y + b\,y &= u \quad &&\text{on } \Sigma \\ y(0) &= y_0 \quad &&\text{on } \Omega. \end{aligned} \tag{2.8}$$

THEOREM 2.2. *Suppose that $a \in L^\infty(Q)$, $b \in L^\infty(\Sigma)$, $q > N/2 + 1$, $r > N + 1$, $a(x,t) \ge c_0$, $b(x,t) \ge c_0$ a.e. on Q and Σ, respectively, and $y_0 \in C(\overline{\Omega})$. Then there is a constant $c_l = c(c_0, q, r, m_0, \Omega, T)$ not depending on a, b, v, u, y_0 such that*

$$||y||_{L^2(0,T;H^1(\Omega))} + ||y||_{C(\overline{Q})} \le c_l\,(||v||_{L^q(Q)} + ||u||_{L^r(\Sigma)} + ||y_0||_{C(\overline{\Omega})}) \tag{2.9}$$

holds for the weak solution of the linear system (2.8).

For the proof we refer to [7] or [28]. (2.9) yields a similar estimate for $b \cdot y$. Regarding the linear system (2.8) with right hand sides $v - ay$, $u - by$, y_0, respectively, the L^2-theory of linear parabolic equations applies to derive

$$\|y\|_{W(0,T)} \le c'(\|v\|_{L^q(Q)} + \|u\|_{L^r(\Sigma)} + \|y_0\|_{C(\overline{\Omega})}), \tag{2.10}$$

where c_l' depends also on $\|a\|_{L^\infty(Q)}, \|b\|_{L^\infty(\Sigma)}$. We shall work in the state space $Y = \{y \in W(0,T) \,|\, y_t + Ay \in L^q(Q),\ \partial_\nu y \in L^p(\Sigma),\ y(0) \in C(\overline{\Omega})\}$ endowed with the norm $\|y\|_Y := \|y\|_{W(0,T)} + \|y_t + Ay\|_{L^q(Q)} + \|\partial_\nu y\|_{L^p(\Sigma)} + \|y(0)\|_{C(\overline{\Omega})}$. Y is known to be continuously embedded into $C(\overline{Q})$. From (2.9), (2.10) we get

$$\|y\|_Y \le \tilde{c}_l\,(\|v\|_{L^q(Q)} + \|u\|_{L^r(\Sigma)} + \|y_0\|_{C(\overline{\Omega})}) \tag{2.11}$$

where $\tilde{c}_l$ depends on $c_0, q, r, m_0, \Omega, T, \|a\|_{L^\infty(Q)}, \|b\|_{L^\infty(\Sigma)}$. We shall furtheron need the Hilbert space $H = W(0,T) \times L^2(\Omega) \times L^2(\Sigma)$ equipped with the norm $\|(y,v,u)\|_H := (\|v\|^2_{W(0,T)} + \|v\|^2_{L^q(Q)} + \|u\|^2_{L^r(\Sigma)})^{1/2}$.

3. Optimal control problem and SQP method. Let $\varphi : \Omega \times \mathbb{R} \to R, f : Q \times \mathbb{R}^2 \to \mathbb{R}$, and $g : \Sigma \times \mathbb{R}^2 \to \mathbb{R}$ be given functions specified below. Consider the problem **(P)** to minimize

$$J(y,v,u) = \int_\Omega \varphi(x, y(x,T))dx + \int_Q f(x,t,y,v)dxdt + \int_\Sigma g(x,t,y,u)dSdt \tag{3.1}$$

subject to the state-equation (2.1) and to the pointwise constraints on the control

$$v_a \le v(x,t) \le v_b \quad \text{a.e. on } Q \tag{3.2}$$

$$u_a \le u(x,t) \le u_b \quad \text{a.e. on } \Sigma, \tag{3.3}$$

where v_a, v_b, u_a, u_b are given functions of $L^\infty(Q)$ and $L^\infty(\Sigma)$, respectively, such that $v_a \le v_b$, a.e. on Q and $u_a \le u_b$ a.e. on Σ. The *controls* v and u belong to the *sets of admissible controls*

$$V_{ad} = \{v \in L^\infty(Q) \,|\, v \text{ satisfies (3.2)}\}, \quad U_{ad} = \{u \in L^\infty(\Sigma) \,|\, u \text{ satisfies (3.3)}\}.$$

(P) is a non-convex programming problem, hence different local minima will possibly occur. Numerical methods will deliver a local minimum close to their starting point. Therefore, we do not restrict our investigations to global solutions of (P). We will assume later that a fixed *reference solution* is given satisfying certain first and second order optimality conditions (ensuring local optimality of the solution). For the same reason, we shall not discuss the problem of existence of global (optimal) solutions for (P).

In the next assumptions, D^2 will denote Hessian matrices of functions. The functions $\varphi, f,$ and g are assumed to satisfy the following assumptions on smoothness and growth:

(A4) For all $x \in \Omega$, $\varphi(x,\cdot)$ belongs to $C^{2,1}(\mathbb{R})$ with respect to $y \in \mathbb{R}$, while $\varphi(\cdot,y)$, $\varphi_y(\cdot,y)$, $\varphi_{yy}(\cdot,y)$ are bounded and measurable on Ω. There is a constant $c_K > 0$ such that

$$|\varphi_{yy}(x,y_1) - \varphi_{yy}(x,y_2)| \le c_K\,|y_1 - y_2| \tag{3.4}$$

holds for all $y_i \in \mathbb{R}$ such that $|y_j| \le K$, $i = 1,2$.
For all $(x,t) \in Q$, $f(x,t,\cdot,\cdot)$ is of class $C^{2,1}(\mathbb{R}^2)$ with respect to $(y,v) \in \mathbb{R}^2$, while f, f_y, f_v, f_{yy}, f_{yv}, and f_{vv}, all depending on $(\cdot,\cdot,y,v)$ are bounded and measurable w.r. to $(x,t) \in Q$. There is a constant $f_K > 0$ such that

$$(3.5)\quad \|D^2 f(x,t,y_1,v_1) - D^2 f(x,t,y_2,v_2)\| \le f_K\,(|y_1-y_2| + |v_1-v_2|)$$

holds for all y_i, v_i satisfying $|y_i| \le K$, $|v_i| \le K$, $i = 1,2$ and almost all $(x,t) \in Q$. Here, $\|\cdot\|$ denotes any useful norm for 2×2-matrices.
The function g satisfies analogous assumptions on $\Sigma \times \mathbb{R}^2$. In particular,

$$(3.6)\quad \|D^2 g(x,t,y_1,u_1) - D^2 g(x,t,y_2,u_2)\| \le g_K\,(|y_1-y_2| + |u_1-u_2|)$$

holds for all y_i, u_i satisfying $|y_i| \le K$, $|u_i| \le K$, $i = 1,2$ and almost all $(x,t) \in \Sigma$.

Let us recall the known standard *first order necessary optimality system* for a local minimizer (y,v,u) of (P). The triplet (y,v,u) has to satisfy together with an *adjoint state* $p \in W(0,T)$ the state system (2.1), the constraints $v \in V_{ad}$, $u \in U_{ad}$, the *adjoint equation*

$$(3.7)\qquad \begin{aligned} -p_t + A\,p + d_y(x,t,y,v)\,p &= f_y(x,t,y,v) && \text{in } Q\\ \partial_\nu p + b_y(x,t,y,u)\;p &= g_y(x,t,y,v) && \text{on } \Sigma\\ p\,(x,T) &= \varphi_y(x,y(x,T)) && \text{in } \Omega, \end{aligned}$$

and the *variational inequalities*

$$(3.8)\qquad \int_Q (f_v(x,t,y,v) - d_v(x,t,y,v)\cdot p)(z-v)\,dxdt \ge 0 \quad \forall z \in V_{ad}$$

$$(3.9)\qquad \int_\Sigma (g_u(x,t,y,u) - b_u(x,t,y,u)\cdot p)(z-u)\,dSdt \ge 0 \quad \forall z \in U_{ad}.$$

We introduce for convenience the *Lagrange function* $\mathcal{L}$,

$$(3.10)\qquad \begin{aligned} \mathcal{L}(y,v,u;p) = J(y,v,u) &- \int_Q \{(y_t + Ay + d(x,t,y,v)\}\,p\,dxdt\\ &- \int_\Sigma \{\partial_\nu y + b(x,t,y,v)\}\,p\,dSdt \end{aligned}$$

defined on $Y \times L^\infty(Q) \times L^\infty(\Sigma) \times W(0,T)$. $\mathcal{L}$ is of class $C^{2,1}$ w.r. to (y,v,u) in $Y \times L^\infty(Q) \times L^\infty(\Sigma)$. Moreover, we define the *Hamilton functions*

$$(3.11)\qquad H^Q = H^Q(x,t,y,p,v) = f(x,t,y,v) - p\,d(x,t,y,v)$$

$$(3.12)\qquad H^\Sigma = H^\Sigma(x,t,y,p,u) = g(x,t,y,u) - p\,b(x,t,y,u),$$

containing the "nondifferential" parts of $\mathcal{L}$. Then the relations (3.7) - (3.9) imply

$$(3.13)\qquad \mathcal{L}_y(y,v,u;p)h = 0 \quad \forall h \in W(0,T) \text{ satisfying } h(0) = 0,$$

$$(3.14)\quad \mathcal{L}_v(y,v,u;p)(z-v) = \int_Q H_v^Q(x,t,y,p,v)(z-v)\,dxdt \ge 0 \quad \forall z \in V_{ad},$$

$$(3.15)\quad \mathcal{L}_u(y,v,u;p)(z-u) = \int_\Sigma H_u^\Sigma(x,t,y,p,u)(z-u)dSdt \ge 0 \quad \forall z \in U_{ad}.$$

Let us suppose once and for all that a fixed *reference triplet* $(\overline{y}, \overline{v}, \overline{u}) \in Y \times L^\infty(Q) \times L^\infty(\Sigma)$ is given satisfying together with $\overline{p} \in W(0,T)$ the optimality system. This system is not sufficient for local optimality. Therefore, we shall assume some kind of *second order sufficient conditions.* We have to consider them along with a *first order sufficient condition.* Following Dontchev, Hager, Poore and Yang [10], the sets

$$Q(\sigma) = \{(x,t) \in Q \mid | H_v^Q(x,t,\overline{y}(x,t),\overline{v}(x,t),\overline{p}(x,t)) | \geq \sigma\} \tag{3.16}$$

$$\Sigma(\sigma) = \{(x,t) \in \Sigma \mid | H_u^\Sigma(x,t,\overline{y}(x,t),\overline{u}(x,t),\overline{p}(x,t)) | \geq \sigma\} \tag{3.17}$$

are defined for arbitrarily small but fixed $\sigma > 0$. $Q(\sigma)$ and $\Sigma(\sigma)$ contain the points, where the control constraints are strongly active enough. Here we are able to avoid second order sufficient conditions, since first order sufficiency applies. D^2H^Q and D^2H^Σ denote the Hessian matrices of H^Q, H^Σ w.r. to (y,v) and (y,u) respectively, taken at the reference point. For instance,

$$D^2H^Q(x,t) = \begin{pmatrix} H_{yy}^Q(x,t,\overline{y}(x,t),\overline{v}(x,t),\overline{p}(x,t)) & H_{yv}^Q(x,t,\overline{y}(x,t),\overline{v}(x,t),\overline{p}(x,t)) \\ H_{vy}^Q(x,t,\overline{y}(x,t),\overline{v}(x,t),\overline{p}(x,t)) & H_{vv}^Q(x,t,\overline{y}(x,t),\overline{v}(x,t),\overline{p}(x,t)) \end{pmatrix}.$$

D^2H^Σ is defined analogously. Moreover, we introduce a quadratic form B depending on $h_i = (y_i, v_i, u_i) \in Y \times L^\infty(Q) \times L^\infty(\Sigma)$, $i = 1,2$, by

$$\begin{aligned} B[h_1,h_2] = &\int_\Omega \varphi_{yy}(x,\overline{y}(x,T)) y_1(x,T) y_2(x,T)\, dx + \int_Q (y_1,v_1) D^2H^Q (y_2,v_2)^\top \, dxdt \\ &+ \int_\Sigma (y_1,u_1) D^2H^\Sigma (y_2,u_2)^\top \, dSdt. \end{aligned} \tag{3.18}$$

The *second order sufficient optimality condition is defined as follows:*

(SSC) There are $\delta > 0, \sigma > 0$ such that

$$B[h,h] \geq \delta \cdot \|h\|_H^2 \tag{3.19}$$

holds for all $h = (y,v,u) \in W(0,T) \times L^2(Q) \times L^2(\Sigma)$, where $v \in V_{ad}$, $v(x,t) = 0$ on $Q(\sigma)$, $u \in U_{ad}$, $u = 0$ on $\Sigma(\sigma)$, and y is the associated weak solution of the linearized equation

$$\begin{aligned} y_t + A\,y + d_y(\overline{y},\overline{v})\,y + d_v(\overline{y},\overline{v})\,v &= 0 \\ \partial_\nu y + b_y(\overline{y},\overline{u})\,y + b_u(\overline{y},\overline{u})\,u &= 0 \\ y(0) &= 0. \end{aligned} \tag{3.20}$$

Next we introduce the SQP method to solve the problem (P) iteratively. Let us first assume that the controls are unrestricted, that is $V_{ad} = L^\infty(Q)$, $U_{ad} = L^\infty(\Sigma)$. Then the optimality system (2.1), (3.7), (3.8), (3.9) is a nonlinear system of equations for the unknown functions v, p, y, u, which can be treated by the Newton method. In each step of the method, a linear system of equations is to be solved. This linear system is the optimality system of a linear-quadratic optimal control problem without constraints on the controls, which can be solved instead of the linear system of equations.

In the case of constraints on the controls, the optimality system is no longer a system of equations. However, there is no difficulty to generalize the linear-quadratic control problems by adding the control-constraints. This idea leads to the following iterative method: Suppose that (y_i, p_i, v_i, u_i), $i = 1,..,n$, have already been determined. Then $(y_{n+1}, v_{n+1}, u_{n+1})$ is computed by solving the following linear-quadratic optimal control problem (QP_n):

(QP_n) *Minimize*

$$J_n(y,v,u) = \int\limits_\Omega \varphi_y^n \cdot y(T)dx + \int\limits_Q (f_y^n \cdot y + f_v^n \cdot v)\, dxdt + \int\limits_\Sigma (g_y^n\ y + g_u^n\ u)dSdt$$
$$+\frac{1}{2}\int\limits_\Omega \varphi_{yy}^n\ (y(T)-y_n(T))^2 dx + \frac{1}{2}\int\limits_Q (y-y_n, v-v_n) D^2 H^{Q,n} \begin{pmatrix} y-y_n \\ v-v_n \end{pmatrix} dxdt$$
$$+\frac{1}{2}\int\limits_\Sigma (y-y_n, u-u_n) D^2 H^{\Sigma,n} \begin{pmatrix} y-y_n \\ u-u_n \end{pmatrix} dSdt \tag{3.21}$$

subject to

$$\begin{aligned} y_t + A\,y + d^n + d_y^n\,(y-y_n) + d_v^n\,(v-v_n) &= 0 \\ \partial_\nu y + b^n + b_y^n\,(y-y_n) + b_u^n\,(u-u_n) &= 0 \\ y(0) &= y_0 \end{aligned} \tag{3.22}$$

and to

$$v \in V_{ad},\, u \in U_{ad}. \tag{3.23}$$

In this setting, the notation $\varphi_y^n = \varphi_y(x, y_n(x,T))$, $\varphi_{yy}^n = \varphi_{yy}^n(x, y_n(x,T))$, $f_y^n = f_y^n(x,t,y_n(x,t),v_n(x,t))$, $D^2H^{Q,n} = D^2H_{(y,v,u)}(x,t,y_n(x,t),v_n(x,t),p_n(x,t))$ etc., was used. The associated adjoint state p_{n+1} is determined from

$$\begin{aligned} -p_t + A\,p + d_y^n\,(p-p_n) &= H_y^{Q,n} + H_{yy}^{Q,n}(y_{n+1}-y_n) + H_{yv}^{Q,n}(v_{n+1}-v_n) \\ p(T) &= \varphi_y^n + \varphi_{yy}^n\ (y_{n+1}-y_n)(T) \\ \partial_\nu p + b_y^n\,(p-p_n) &= H_y^{\Sigma,n} + H_{yy}^{\Sigma,n}(y_{n+1}-y_n) + H_{yu}^{\Sigma,n}(u_{n+1}-u_n). \end{aligned} \tag{3.24}$$

In this way, a sequence of quadratic optimization problems is to be solved, giving the method the name Sequential Quadratic Programming (SQP-) method. The main aim of this paper is to show that this process exhibits a local quadratic convergence. We shall transform the optimality system into a *generalized equation*. Then we are able to interprete the SQP method as a Newton method for a generalized equation. This approach gives direct access to known results on the convergence of Newton methods. In the analysis, a specific difficulty arises from the fact that (QP_n) might be non-convex. It therefore may have multiple local minima. We shall have to restrict the control set to a sufficiently small neighbourhood around the reference solution.

4. Generalized equation and Newton method. To transform the optimality system into a generalized equation, we re-formulate the variational inequalities (3.8)-(3.9) as generalized equations, too. Therefore, we define the *normal cones*

$$N^Q(v) = \begin{cases} \{z \in L^\infty(Q) \mid \int\limits_Q z(\tilde v - v)dxdt \le 0 \quad \forall \tilde v \in V_{ad}\}, \text{ if } v \in V_{ad} \\ \emptyset, \text{ if } v \ni V_{ad} \end{cases} \tag{4.1}$$

$$N^\Sigma(u) = \begin{cases} \{z \in L^\infty(\Sigma) \mid \int\limits_\Sigma z(\tilde u - u)dSdt \le 0 \quad \forall \tilde u \in U_{ad}\}, \text{ if } u \in U_{ad} \\ \emptyset, \text{ if } u \ni U_{ad}. \end{cases} \tag{4.2}$$

Then (3.8), (3.9) read $-H_v^Q(y,p,v) \in N^Q(v)$, $-H_u^\Sigma(y,p,u) \in N^\Sigma(u)$, or

$$0 \in H_v^Q(y,p,v) + N^Q(v) \tag{4.3}$$
$$0 \in H_u^\Sigma(y,p,u) + N^\Sigma(u) \tag{4.4}$$

(H_v^Q and H_u^Σ are Nemytskii operators defined analogously to H_y^Q, H_y^Σ). The set-valued mappings $T_1 : v \mapsto N^Q(v)$ from $L^\infty(Q)$ to $2^{L^\infty(Q)}$ and $T_2 : u \mapsto N^\Sigma(u)$ from $L^\infty(\Sigma)$ to $2^{L^\infty(\Sigma)}$ have closed graph.
We introduce now the space $E = (L^\infty(Q) \times L^\infty(\Sigma) \times C(\overline{\Omega}))^2 \times L^\infty(Q) \times L^\infty(\Sigma)$ with elements $\eta = (e_Q, e_\Sigma, 0, \gamma_Q, \gamma_\Sigma, \gamma_\Omega, \gamma_v, \gamma_u)$, endowed with the norm $\|\eta\|_E = \|e_Q\|_{L^\infty(Q)} + \|e_\Sigma\|_{L^\infty(\Sigma)} + \|\gamma_Q\|_{L^\infty(Q)} + \|\gamma_\Sigma\|_{L^\infty(\Sigma)} + \|\gamma_\Omega\|_{C(\overline{\Omega})} + \|\gamma_v\|_{L^\infty(Q)} + \|\gamma_u\|_{L^\infty(\Sigma)}$, and the space $W = Y \times Y \times L^\infty(Q) \times L^\infty(\Sigma)$ equipped with the norm $\|(y,p,v,u)\|_W = \|y\|_Y + \|p\|_Y + \|v\|_{L^\infty(\Omega)} + \|u\|_{L^\infty(\Sigma)}$. Moreover, define the set-valued mapping $T : W \to 2^E$ by

$$T(w) = (\{0\}, \{0\}, \{0\}, \{0\}, \{0\}, \{0\}, N^Q(v), N^\Sigma(u)),$$

and $F : W \to E$ by $F(w) = (F_1(w), ..., F_8(w))$, where

$$\begin{aligned}
F_1(w) &= y_t + A\,y + d(y,v)\\
F_2(w) &= \partial_\nu y + b(y,u)\\
F_3(w) &= y(0) - y_0\\
F_4(w) &= -p_t + A\,p - H_y^Q(y,p,v)\\
F_5(w) &= \partial_\nu p - H_y^\Sigma(y,p,u)\\
F_6(w) &= p(T) - \varphi_y(y(T))\\
F_7(w) &= H_v^Q(y,p,v)\\
F_8(w) &= H_u^\Sigma(y,p,u).
\end{aligned}$$

In the definition of E, the third component is vanishing, since it will correspond to the initial condition $y(0) - y_0 = 0$, which is kept fixed in the generalized Newton method. The optimality system is easily seen to be equivalent to the generalized equation

$$0 \in F(w) + T(w), \tag{4.5}$$

where F is of class $C^{1,1}$, and the set-valued mapping T has closed graph. Obviously, the reference solution $\overline{w} = (\overline{y}, \overline{p}, \overline{v}, \overline{u})$ satisfies (4.5). The *generalized Newton method* for solving (4.5) is similar to the Newton method for equations in Banach spaces. Suppose that we have already computed $w_1, ..., w_n$. Then w_{n+1} is to be determined by the generalized equation

$$0 \in F(w_n) + F'(w_n)(w - w_n) + T(w). \tag{4.6}$$

The convergence analysis of this method is closely related to the notion of *strong regularity* of (4.5) going back to Robinson [29]. The generalized equation (4.5) is said to be *strongly regular at* $\overline{w}$, if there are constants $r_1 > 0$, $r_2 > 0$, and $c_L > 0$ such that for all perturbations $e \in \mathcal{B}_{r_1}(0_E)$ the *linearized equation*

$$e \in F(\overline{w}) + F'(\overline{w})(w - \overline{w}) + T(w) \tag{4.7}$$

has in $\mathcal{B}_{r_2}(\overline{w})$ a unique solution $w = w(e)$, and the Lipschitz property

$$\|w(e_1) - w(e_2)\|_W \le c_L \|e_1 - e_2\|_E \tag{4.8}$$

holds for all $e_1, e_2 \in \mathcal{B}_{r_1}(0_E)$. In the case of an equation $F(w) = 0$, we have $F(\overline{w}) = 0, T(w) = \{0\}$, and strong regularity means the existence and boundedness of $(F'(\overline{w}))^{-1}$. The following result gives a first answer to the convergence analysis of the generalized Newton method.

THEOREM 4.1. *Suppose that (4.5) is strongly regular at $\overline{w}$. Then there are $r_{\mathcal{N}} > 0$ and $c_{\mathcal{N}} > 0$ such that for each starting element $w_1 \in \mathcal{B}_{r_{\mathcal{N}}}(\overline{w})$ the generalized Newton method generates a unique sequence $\{w_n\}_{n=1}^{\infty}$. This sequence remains in $\mathcal{B}_{\|w_1-\overline{w}\|_W}(\overline{w})$, and it holds*

$$\|w_{n+1} - \overline{w}\|_W \le c_{\mathcal{N}} \|w_n - \overline{w}\|_W^2 \qquad \forall n \in \mathbb{N}. \tag{4.9}$$

This result was apparently shown first by Josephy [18]. Generalizations can be found in Dontchev [8] and Alt [1], [2]. We refer in particular to the recent publication by Alt [3], where a mesh-independence principle was shown for numerical approximation of (4.5). We shall verify that the second order condition (SSC) implies strong regularity of the generalized equation at $\overline{w} = (\overline{y}, \overline{p}, \overline{v}, \overline{u})$ in certain subsets $\widehat{V}_{ad} \subset V_{ad}$, $\widehat{U}_{ad} \subset U_{ad}$. Then Theorem 4.1 yields the quadratic convergence of the generalized Newton method in these subsets.

5. Strong regularity. To investigate the strong regularity of the generalized equation (4.5) at $\bar{w}$ we have to consider the perturbed generalized equation (4.7). Once again, we are able to interprete this equation as the optimality system of a linear-quadratic control problem. This problem is not necessarily convex, therefore we study the behaviour of the following auxiliary linear-quadratic problem associated with the perturbation e:

($\widehat{\mathbf{QP}}_e$) *Minimize*

$$\begin{aligned} J_e(y,v,u) = &\int_\Omega (\bar{\varphi}_y + \gamma_\Omega)\, y(T)\, dx + \int_Q (\bar{f}_y + \gamma_Q)\, y\, dxdt + \int_Q (\bar{f}_v + \gamma_v)\, v\, dxdt \\ &+ \int_\Sigma (\bar{g}_y + \gamma_\Sigma)\, v\, dSdt + \int_\Sigma (\bar{g}_u + \gamma_u)\, u\, dSdt + \tfrac{1}{2}\int_\Omega \bar{\varphi}_{yy} (y(T) - \bar{y}(T))^2 dx \\ &+ \tfrac{1}{2}\int_Q \begin{pmatrix} y-\bar{y} \\ v-\bar{v}\end{pmatrix}^{\top} D^2 \bar{H}^Q \begin{pmatrix} y-\bar{y} \\ v-\bar{v}\end{pmatrix} dxdt + \tfrac{1}{2}\int_\Sigma \begin{pmatrix} y-\bar{y} \\ u-\bar{u}\end{pmatrix}^{\top} D^2 \bar{H}^\Sigma \begin{pmatrix} y-\bar{y} \\ u-\bar{u}\end{pmatrix} dSdt \end{aligned} \tag{5.1}$$

subject to

$$\begin{aligned} y_t + A\,y + d(\bar{y},\bar{v}) + \bar{d}_y\,(y-\bar{y}) + \bar{d}_v\,(v-\bar{v}) &= e_Q &&\text{in } Q \\ \partial_\nu y + b(\bar{y},\bar{u}) + \bar{b}_y\,(y-\bar{y}) + \bar{b}_u\,(u-\bar{u}) &= e_\Sigma &&\text{on } \Sigma \\ y(0) &= y_0 &&\text{in } \Omega, \end{aligned} \tag{5.2}$$

and to the constraints on the control

$$\begin{aligned} v \in \widehat{V}_{ad} &= \{v \in V_{ad} \mid v(x,t) = \bar{v}(x,t) \text{ on } Q(\sigma)\} \\ u \in \widehat{U}_{ad} &= \{u \in U_{ad} \mid u(x,t) = \bar{u}(x,t) \text{ on } \Sigma(\sigma)\}. \end{aligned} \tag{5.3}$$

In this setting, the perturbation vector $e = (e_Q, e_\Sigma, 0, \gamma_Q, \gamma_\Sigma, \gamma_\Omega, \gamma_v, \gamma_u)$ belongs to E. The hat in ($\widehat{\mathrm{QP}}_e$) indicates that v and u are taken equal to $\bar{v}$ and $\bar{u}$ on the strongly active sets $Q(\sigma)$, $\Sigma(\sigma)$, respectively.

Remark: The generalized equation (4.7) is equivalent to the optimality system of the problem (QP_e) obtained from ($\widehat{\mathrm{QP}}_e$) on substituting V_{ad} for $\widehat{V}_{ad}$ and U_{ad} for $\widehat{U}_{ad}$, respectively.

In the space of perturbations E we need another norm

$$\begin{aligned} \|e\|_2 = &\|e_Q\|_{L^2(Q)} + \|e_\Sigma\|_{L^2(\Sigma)} + \|\gamma_Q\|_{L^2(Q)} + \|\gamma_\Sigma\|_{L^2(\Sigma)} + \\ &+ \|\gamma_\Omega\|_{L^2(\Omega)} + \|\gamma_v\|_{L^2(Q)} + \|\gamma_u\|_{L^2(\Sigma)}. \end{aligned}$$

Moreover, in W we shall also use the norm

$$\|(y,p,v,u)\|_2 = \|y\|_{W(0,T)} + \|p\|_{W(0,T)} + \|v\|_{L^2(Q)} + \|u\|_{L^2(\Sigma)}.$$

The following results are known from the author's paper [33]:
LEMMA 5.1. *Suppose that the second order sufficient optimality condition (SSC) is satisfied at* $(\bar{y},\bar{v},\bar{u})$ *with associated adjoint state* $\bar{p}$. *Then for each* $e \in E$, *the problem* $(\widehat{QP}_e)$ *has a unique solution* (y_e, v_e, u_e) *with associated adjoint state* p_e. *Let* (y_i, v_i, u_i) *and* $p_i, i = 1,2$, *be the solutions to* $e_i \in E, i = 1,2$. *There is a constant* $l_2 > 0$, *not depending on* e_i, *such that*

$$\|(y_1,p_1,v_1,u_1) - (y_2,p_2,v_2,u_2)\|_2 \le l_2 \|e_1 - e_2\|_2 \tag{5.4}$$

holds for all $e_i \in E, i = 1,2$.
By continuity, (5.4) extends to perturbations e_i of L^2. It was shown in [33] that the second order condition (SSC) implies the following strong *Legendre - Clebsch* condition:

$$\textbf{(LC)} \qquad \begin{array}{ll} H^Q_{vv}(x,t,\bar{y}(x,t),\bar{v}(x,t),\bar{p}(x,t)) \ge \delta & a.e. \text{ on } Q \\ H^\Sigma_{uu}(x,t,\bar{y}(x,t),\bar{u}(x,t)\bar{p}(x,t)) \ge \delta & a.e. \text{ on } \Sigma. \end{array}$$

THEOREM 5.2. *Let the assumptions of Lemma 5.1 be satisfied. Then there is a constant* $l_\infty > 0$, *not depending on* e_i, *such that*

$$\|(y_1,p_1,v_1,u_1) - (y_2,p_2,v_2,u_2)\|_W \le l_\infty \|e_1 - e_2\|_E \tag{5.5}$$

holds for (y_i, v_i, u_i, p_i) *and* $e_i, i = 1,2$, *introduced in Lemma 6.1.*
This Theorem follows from [33], Thm. 5.2 (notice that $v_i = \bar{v}$ and $u_i = \bar{u}$ on $Q(\sigma)$ and $\Sigma(\sigma)$, respectively. This can be expressed by taking $u_a := u_b := \bar{u}$ and $v_a := v_b := \bar{v}$ on these sets. Then [33], Thm 5.2 is easy to apply).
Unfortunately, (5.5) holds only for $\widehat{V}_{ad}$ and $\widehat{U}_{ad}$. We are not able to prove (5.5) in V_{ad}, U_{ad}. In this case, J_e might be nonconvex and (QP_e) may have multiple solutions, if solvable at all. However, formulating Theorem 5.2 in the context of our generalized equation, we already have obtained the following result on strong regularity:
THEOREM 5.3. *Suppose that* $\bar{w} = (\bar{y},\bar{p},\bar{v},\bar{u})$ *satisfies the first order optimality system (2.1), (3.2) - (3.3), (3.7) - (3.9) together with the second order sufficient condition (SSC). Then the generalized equation (4.5) is strongly regular at* $\bar{w}$, *provided that the control sets* $\widehat{V}_{ad}, \widehat{U}_{ad}$ *are substituted for* V_{ad}, U_{ad} *in the definition of* $T(w)$.
Remark: The last assumption means that the normal cones $N^Q(v), N^\Sigma(u)$ are defined on using $\widehat{V}_{ad}$ and $\widehat{U}_{ad}$, respectively.
To complete the discussion of the Newton method, the following questions have to be answered yet: How we can solve the generalized equation (4.6) in $\widehat{V}_{ad}, \widehat{U}_{ad}$, and how we get rid of the artificial restriction $v = \bar{v}$ on $Q(\sigma)$, $u = \bar{u}$ on $\Sigma(\sigma)\Gamma$
We shall show that the SQP method, restricted to a sufficiently small neighbourhood around $\bar{v}$ and $\bar{u}$, will solve both the problems: If the region is small enough, then the SQP method delivers a unique solution $w_n = (y_n, p_n, v_n, u_n)$, where $v_n = \bar{v}, u_n = \bar{u}$ is automatically satisfied on $Q(\sigma), \Sigma(\sigma)$. Moreover, this w_n is a solution of the generalized equation (4.5), that is, a solution of the optimality system for (P).

6. The linear-quadratic subproblems (QP$_n$). The presentation of the SQP method is still quite formal. We do not know whether the quadratic subproblem (QP_n) defined by (3.21) - (3.23) is solvable at all. Moreover, if solutions exist, we are not able to show their uniqueness. There might exist multiple stationary solutions, i.e. solutions satisfying the optimality system for (QP_n). Notice that the objective J_n of (QP_n) is only convex on a subspace. Owing to this, we have to restrict (QP_n) to a sufficiently small neighbourhood around the reference solution $(\bar v, \bar u)$. This region is defined by

$$\begin{aligned} V_{ad}^{\varrho} &= \{v \in V_{ad} \,|\, \|v - \bar v\|_{L^\infty(Q)} \le \varrho\} \\ U_{ad}^{\varrho} &= \{u \in U_{ad} \,|\, \|u - \bar u\|_{L^\infty(\Sigma)} \le \varrho\}, \end{aligned}$$

where $\varrho > 0$ is a sufficiently small radius. To avoid the unknown reference solution $(\bar v, \bar u)$ in the definition of the neighbourhood, we shall later replace this neighborhood by a ball around the initial iterate (v_1, u_1) .
Let us denote by $(\mathrm{QP}_n^{\varrho})$ the problem (QP_n) restricted to V_{ad}^{ϱ}, U_{ad}^{ϱ} and by $(\widehat{\mathrm{QP}}_n)$ the same problem restricted to $\widehat V_{ad}$, $\widehat U_{ad}$, respectively. To analyze $(\widehat{\mathrm{QP}}_n)$ in a first step, we need some auxiliary results.
LEMMA 6.1. *For all $K > 0$ there is a constant $c_L = c_L(K)$ such that*

$$\mathcal{E} \le c_L(K) \|w_n - \bar w\|_W \tag{6.1}$$

holds for all $w_n \in W$ with $\|w_n - \bar w\|_W \le K$, where the expression $\mathcal{E}$ is defined by

$$\begin{aligned} \mathcal{E} = max\,\{ & \|f_v^n - \bar f_v\|_{L^\infty(Q)}, \|f_y^n - \bar f_y\|_{L^\infty(Q)}, \|g_v^u - \bar g_u\|_{L^\infty(\Sigma)}, \|g_y^n - \bar g_y\|_{L^\infty(\Sigma)}, \\ & \|d_y^n - \bar d_y\|_{L^\infty(Q)}, \|d_v^n - \bar d_v\|_{L^\infty(Q)}, \|b_y^n - \bar b_y\|_{L^\infty(\Sigma)}, \|b_u^n - \bar b_u\|_{L^\infty(\Sigma)}, \|\varphi_y^n - \bar\varphi_y\|_{C(\bar\Omega)}, \\ & \|\varphi_{yy}^n - \bar\varphi_{yy}\|_{C(\bar\Omega)}, \|D^2H^{Q,n} - D^2\bar H^Q\|_{L^\infty(Q)}, \|D^2H^{\Sigma,n} - D^2\bar H^\Sigma\|_{L^\infty(Q)}\}. \end{aligned}$$

Proof. The estimate follows from the assumptions (A2)–(A4) imposed on the functions f, g, φ, b, d in section 2 and 3. For instance, the mean value theorem yields

$$\begin{aligned} \|f_v^n - \bar f_v\|_{L^\infty(Q)} &= \sup_{(x,t)\in Q} \mathrm{ess} |f_{vy}(y^\vartheta, v^\vartheta)(y_n - \bar y) + f_{vv}(v^\vartheta, v^\vartheta)(v_n - \bar v)| \\ &\le c(K) \sup_{(x,t)\in Q} \mathrm{ess}(|y_n - \bar y| + |v_n - \bar v|) \end{aligned}$$

by (3.5), where $y^\vartheta = \bar y + \vartheta(y_n - \bar y)$, $v^\vartheta = \bar v + \vartheta(v_n - \bar v)$ and $\vartheta = \vartheta(x,t)$ belongs to $(0,1)$. (Consider for example the estimation

$$\begin{aligned} |f_{vy}(y^\vartheta, v^\vartheta)| &\le |f_{yv}(0,0)| + |f_{vy}(y^\vartheta, v^\vartheta) - f_{vy}(0,0)| \le c_1 + c(K)\,(|y^\vartheta| + |v^\vartheta|) \\ &\le c_1 + c(K) \cdot K, \end{aligned}$$

which follows from (3.5)). The other terms in $\mathcal{E}$ are handled analogously. ☐
We shall denote the quadratic part of the functional J_n by

$$\begin{aligned} B_n[(y_1, v_1, u_1), (y_2, v_2, u_2)] = & \int_\Omega \varphi_{yy}^n \; y_1(T) y_2(T)\, dx + \int_Q (y_1, v_1) D^2H^{Q,n} (y_2, v_2)^\top \, dxdt \\ & + \int_\Sigma \; (y_1, u_1) D^2 H^{\Sigma,n} (y_2, u_2)^\top \, dSdt \end{aligned} \tag{6.2}$$

and write for short $B_n[(y,v,u),(y,v,u)] = B_n[y,v,u]^2$.
LEMMA 6.2. *Suppose that the second order sufficient optimality conditon (SSC) is satisfied. Then there is $\varrho_1 > 0$ with the following property: If $\|w_n - \bar w\|_W \le \varrho_1$, then*

$$B_n[y,v,u]^2 \ge \frac{\delta}{2} \|(y,v,u)\|_H^2 \tag{6.3}$$

holds for all $(y,v,u)\in H$ *satisfying* $v=0$ *on* $Q(\sigma)$, $u=0$ *on* $I^u(\sigma)$ *together with*

$$\text{(6.4)}\qquad \begin{aligned} y_t + A\,y + d_y^n\; y + d_v^n\; v &= 0\\ \partial_\nu y + b_y^n\; y + b_u^n\; u &= 0\\ y(0) &= 0. \end{aligned}$$

Proof. Let z denote the weak solution of the parabolic equation obtained from (6.4) on substituting $\bar d_y, \bar d_v, \bar b_y, \bar b_u$ for d_y^n, d_v^n, b_y^n, b_u^n, respectively. Then

$$\begin{aligned} (y-z)_t + A\,(y-z)) + \bar d_y\,(y-z) &= (\bar d_y - d_y^n)\,y + (\bar d_v - d_v^n)\,v\\ \partial_\nu(y-z) + \bar b_y\,(y-z) &= (\bar b_y - b_y^n)\,y + (\bar b_u - b_u^n)\,u\\ (y-z)(0) &= 0. \end{aligned}$$

We have $\bar d_y \ge c_0, \bar b_y \ge c_0$. The differences on the right hand sides can be estimated by Lemma 6.1, where $K = \|\bar w\|_W + \varrho_1$, hence parabolic L^2 - regularity yields

$$\text{(6.5)}\qquad \begin{aligned} \|y-z\|_{W(0,T)} &\le c\,(\|\bar d_y - d_y^n\|_{L^\infty(Q)}\|y\|_{L^2(Q)} + \|\bar d_y - d_v^n\|_{L^\infty(Q)}\|v\|_{L^2(Q)}\\ &\quad + \|\bar b_y - b_y^n\|_{L^\infty(\Sigma)}\|y\|_{L^2(\Sigma)} + \|\bar b_u - b_u^n\|\|u\|_{L^2(\Sigma)})\\ &\le c\,\varrho_1(\|y\|_{W(0,T)} + \|v\|_{L^2(Q)} + \|u\|_{L^2(\Sigma)}) \le c\,\varrho_1\|(y,v,u)\|_H. \end{aligned}$$

Substituting $y = z + (y-z)$ in B_n,

$$\begin{aligned} B_n[y,v,u]^2 &= B_n[z+(y-z),v,u]^2\\ &= B[z,v,u]^2 + (B_n - B)[z,v,u]^2 + 2B_n[(z,v,u),(y-z,0,0)]\\ &\quad + B_n[y-z,0,0]^2 \end{aligned}$$

is obtained. (SSC) applies to the first expression B, while the second is estimated by Lemma 6.1. In the remaining two parts, we use the uniform boundedness of all coefficients. Therefore, by (6.5)

$$\begin{aligned} B_n[y,v,u]^2 &\ge \delta\,\|(z,v,u)\|_H^2 - c\,\varrho_1\|(z,v,u)\|_H^2 - c\,\|(z,v,u)\|_H\|y-z\|_{W(0,T)}\\ &\quad - c\,\|y-z\|_{W(0,T)}^2\\ &\ge \frac{3}{4}\delta\,\|(z,v,u)\|_H^2 - c\varrho_1\|(z,v,u)\|_H\|(y,v,u)\|_H - c\varrho_1^2\|(y,v,u)\|_H^2, \end{aligned}$$

if ϱ_1 is sufficiently small. Next we re-substitute $z = y + (z-y)$ and apply (6.5) again. In this way, the desired estimate (6.3) is easily verified for sufficiently small $\varrho_1 > 0$. □

COROLLARY 6.3. *If* $\|w_n - \bar w\|_W \le \varrho_1$ *and (SSC) is satisfied at* $\bar w$, *then* $(\widehat{QP}_n)$ *has a unique optimal pair of controls* $(\hat v, \hat u)$ *with associated state* $\hat y$.

Proof. The functional J_n to be minimized in $(\widehat{QP}_n)$ has the form (see (3.21))

$$J_n(y,v,u) = a_n(y,v,u) + \frac{1}{2}B_n[y-y_n, v-v_n, u-u_n]^2,$$

where a_n is a linear integral functional. J_n is uniformly convex on the feasible region of $(\widehat{QP}_n)$. By Lemma 6.2, the sets $\widehat V_{ad}, \widehat U_{ad}$ are weakly compact in $L^2(Q)$ and $L^2(\Sigma)$, respectively. Therefore, the Corollary follows from standard arguments. □

Let us return to the discussion of the relation between Newton method and SQP method. In what follows, we shall denote by $\hat w_n = (\hat y_n, \hat p_n, \hat v_n, \hat u_n)$ the sequence of iterates generated by the SQP method performed in $\widehat V_{ad}$, $\widehat U_{ad}$ (provided that this sequence is well defined). The iterates of the generalized Newton method are denoted by w_n. Consider now both methods initiating from the same element $w_n = \hat{w}_n$.

If $\|w_n - \bar{w}\|_W \leq \varrho_1$, then Corollary 6.3 shows the existence of a unique solution $(\hat{y}_{n+1}, \hat{v}_{n+1}, \hat{u}_{n+1})$ of $(\widehat{QP}_n)$ having the associated adjoint state $\hat{p}_{n+1}$. The element $\hat{w}_{n+1}$ solves the optimality system corresponding to $(\widehat{QP}_n)$. By convexity (Lemma 6.2), any other solution of this system solves $(\widehat{QP}_n)$, hence it is equal to is $\hat{w}_{n+1}$. On the other hand, the optimality system is equivalent to the generalized equation (4.6) at w_n (based on the sets $\widehat{V}_{ad}, \widehat{U}_{ad}$). For $\|w_n - \bar{w}\|_W \leq r_{\mathcal{N}}$, one step of the generalized Newton method delivers the unique solution w_{n+1} of (4.6). As w_{n+1} solves the optimality system for $(\widehat{QP}_n)$, it has to coincide with $\hat{w}_{n+1}$. Suppose further that $\|w_n - \bar{w}\|_W \leq min\,\{r_{\mathcal{N}}, \varrho_1\}$. Then Theorem 4.1 implies that $w_{n+1} = \hat{w}_{n+1}$ remains in $\mathcal{B}_{min\,\{r_{\mathcal{N}}, \varrho_1\}}(\overline{w})$, so that $\|\hat{w}_{n+1} - \bar{w}\|_W \leq min\,\{r_{\mathcal{N}}, \varrho_1\}$. Consequently, we are able to perform the next step in both the methods. Moreover, in $\widehat{V}_{ad}, \widehat{U}_{ad}$ each step of the Newton method is equivalent to solving $(\widehat{QP}_n)$, which always has a unique solution. In other words, Newton method and SQP method are identical in $\widehat{V}_{ad}, \widehat{U}_{ad}$:

THEOREM 6.4. *Let $\bar{w} = (\bar{y}, \bar{p}, \bar{v}, \bar{u})$ satisfy the first order optimality system (2.1), (3.2) - (3.3), (3.7) - (3.9) together with the second order sufficient optimality conditions (SSC). Suppose that $w_1 = (y_1, p_1, v_1, u_1) \in W$ is given such that $\|w_1 - \bar{w}\|_W \leq min\,\{\varrho_1, r_{\mathcal{N}}\}$, $v_1 \in \widehat{V}_{ad}$, and $u_1 \in \widehat{U}_{ad}$. Then in $\widehat{V}_{ad}, \widehat{U}_{ad}$ the generalized Newton method is equivalent to the SQP method: The solution of the generalized equation (4.6) is given by the unique solution of $(\widehat{QP}_n)$ along with the associated adjoint state.*

The result follows from Theorem 5.3 (strong regularity) and the considerations above.

Remark: It is easy to verify that $\hat{w}_n$, the solution of $(\widehat{QP}_n)$, obeys the optimality system for (P) in the original sets V_{ad}, U_{ad} (cf. also Corollary 6.9).

Next, we discuss the optimality system for $(\widehat{QP}_n)$ and (QP_n^{ϱ}). Let us denote the associated Hamilton functions by $\tilde{H}$ to distinguish them from H, which belongs to (P):

$$\begin{aligned}
\tilde{H}^Q(x,t,y,p,v) &= f_y^n(y-y_n) + f_v^n(v-v_n) - p\,\left(d^n + d_y^n(y-y_n) + d_v^n(v-v_n)\right) \\
&\quad + \tfrac{1}{2}(y-y_n, v-v_n)D^2H^{Q,n}(y-y_n, v-v_n)^\top \\
\tilde{H}^\Sigma(x,t,y,p,u) &= g_y^n(y-y_n) + g_u^n(u-u_n) - p\,\left(b^n + b_y^n(y-y_n) + b_u^n(u-u_n)\right) \\
&\quad + \tfrac{1}{2}(y-y_n, u-u_n)D^2H^{\Sigma,n}(y-y_n, u-u_n)^\top,
\end{aligned}$$

where y, v, p, u are real numbers and (x,t) appears in the quantities depending on n. Notice that these Hamiltonians coincide for $(\widehat{QP}_n), (QP_n^{\varrho})$ and (QP_n), since these problems differ only in the underlying sets of admissible controls. We consider the problems defined at $w_n = (y_n, p_n, v_n, u_n)$. In what follows, we denote solutions of the optimality system corresponding to (QP_n^{ϱ}) by (y^+, v^+, u^+). The optimality system for (QP_n^{ϱ}) consists of

$$\int_Q \tilde{H}_v^Q(y^+, p^+, v^+)(v - v^+)\,dxdt \geq 0 \quad \forall v \in V_{ad}^{\varrho} \tag{6.6}$$

$$\int_\Sigma \tilde{H}_u^\Sigma(y^+, p^+, u^+)(u - u^+)\,dSdt \geq 0 \quad \forall u \in U_{ad}^{\varrho}, \tag{6.7}$$

where the associated adjoint state p^+ is defined by

$$(6.8)\qquad \begin{aligned} -p_t^+ + A\,p^+ &= \tilde H_y^Q = f_y^n + H_{yy}^{Q,n}(y^+ - y_n) + H_{yv}^{Q,n}(v^+ - v_n) - d_y^n p^+ \\ p(T) &= \varphi_y^n + \varphi_{yy}^n (y^+(T) - y(T)) \\ \partial_\nu p &= \tilde H_y^\Sigma = g_y^n + H_{yy}^{\Sigma,n}(y^+ - y_n) + H_{yv}^{\Sigma,n}(u^+ - u_n) - b_y^n p^+. \end{aligned}$$

The state-equation (3.22) for y^+ and the constraints $v^+ \in V_{ad}^\varrho, u^+ \in U_{ad}^\varrho$ are included in the optimality system, too. The optimality system of $(\widehat{QP}_n)$ has the same principal form as (6.6) - (6.8) and is obtained on replacing (y^+, p^+, v^+, u^+) by $(\hat y_{n+1}, \hat p_{n+1}, \hat v_{n+1}, \hat u_{n+1})$. Moreover, $\widehat V_{ad}$, $\widehat U_{ad}$ is to be substituted for V_{ad}^ϱ, U_{ad}^ϱ there. In the further analysis, we shall perform the following steps: First we prove by a sequence of results that the solution $(\hat v_n, \hat u_n)$ of $(\widehat{QP}_n)$ satisfies the optimality system of (QP_n^ϱ) for sufficiently small ϱ. Moreover, we prove that (QP_n^ϱ) has at least one optimal pair, if w_n is sufficiently close to $\bar w$. Finally, relying on (SSC), we verify uniqueness for the optimality system of (QP_n^ϱ). Therefore, $(\hat v_n, \hat u_n)$ can be obtained as the unique global solution of (QP_n^ϱ). Notice that (QP_n^ϱ) might be non-convex, hence the optimality of $(\hat v_n, \hat u_n)$ does not follow directly from fulfilling the optimality system.

LEMMA 6.5. *There is $\varrho_2 < 0$ with the following property: If $\varrho \le \varrho_2$, $w_n \in W$ fullfils $\|w_n - \bar w\|_W \le \varrho_2$, and (y^+, v^+, u^+) satisfies the constraints of (QP_n^ϱ) with associated adjoint state p^+, then*

$$(6.9)\qquad sign\ \tilde H_v^Q(y^+, p^+, v^+)(x,t) = sign\ H_v^Q(\bar y, \bar p, \bar v)(x,t) \quad a.e.\ on\ \ Q(\sigma)$$

$$(6.10)\qquad sign\ \tilde H_u^\Sigma(y^+, p^+, u^+)(x,t) = sign\ H_u^\Sigma(\bar y, \bar p, \bar u)(x,t) \quad a.e.\ on\ \ \Sigma(\sigma)$$

$$(6.11)\qquad |\tilde H_v^Q(y^+, p^+, v^+)(x,t)| \ge \frac{\sigma}{2} \quad a.e.\ on\ \ Q(\sigma)$$

$$(6.12)\qquad |\tilde H_u^\Sigma(y^+, p^+, u^+)(x,t)| \ge \frac{\sigma}{2} \quad a.e.\ on\ \ \Sigma(\sigma).$$

Proof. Let us discuss $\tilde H_v^Q$, the proof is analogous for $\tilde H_u^\Sigma$. We have

$$\begin{aligned} \tilde H_v^Q &= f_v^n + H_{yv}^{Q,n}(y^+ - y_n) + H_{vv}^{Q,n}(v^+ - v_n) - p^+ d_v^n \\ &= \bar f_v - \bar p \bar d_v + \{f_v^n - \bar f_v + (f_{yv}^n - p_n d_{yv}^n)(y^+ - y_n) \\ &\quad + (f_{vv}^n - p_n d_{vv}^n)(v^+ - v_n) + (\bar p \bar d_v - p^+ d_v^n)\} = \bar H_v^Q + \{...\} \ge \sigma - |\{...\}| \end{aligned}$$

a.e. on $Q(\sigma)$. Lemma 6.1 applies to estimate $|\{...\}| \le c \cdot \varrho_2$, where c does not depend on w_n, y^+, p^+, u^+, v^+, provided that we are able to prove that $\|p^+ - \bar p\|_{C(\bar Q)} \le c\ \varrho_2$ and $\|y^+ - \bar y\|_{C(\bar Q)} \le c\ \varrho_2$ holds with an associated constant c. Let us sketch the estimation of $y^+ - \bar y =: y$. This function satisfies

$$\begin{aligned} y_t + A\,y + d_y^n y &= -d_v^n(v^+ - \bar v) + (d_y^n - d_y^\vartheta)(y_n - \bar y) + (d_v^n - d_v^\vartheta)(v_n - \bar v) \\ \partial_\nu y + b_y^n y &= -d_u^n(u^+ - \bar u) + (b_y^n - b_y^\vartheta)(y_n - \bar y) + (b_u^n - b_u^\vartheta)(u_n - \bar u) \\ y(0) &= 0, \end{aligned}$$

where $d_y^\vartheta = d_y(\bar y + \vartheta(y_n - \bar y), \bar v + \vartheta(v_n - \bar v))$, $\vartheta = \vartheta(x,t) \in (0,1)$, and the other quantities are defined accordingly. We have $max\,\{\|v^+ - \bar v\|_{L^\infty(Q)}, \|u^+ - \bar u\|_{L^\infty(\Sigma)}\} \le \varrho$, $max\,\{\|y_n - \bar y\|_{C(\bar Q)}, \|u_n - \bar u\|_{L^\infty(\Sigma)}, \|v_n - \bar v\|_{L^\infty(Q)}\} \le \varrho_2$. Thus the right hand sides of the PDE and its boundary condition are estimated by $c \cdot \varrho_2$. The estimate for

$\|y^+ - \bar{y}\|$ follows from Theorem 2.2. The difference $p^+ - \bar{p}$ is handled in the same way. □

COROLLARY 6.6. *If max* $\{\|w_n - \bar{w}\|_W, \varrho\} \le \varrho_2$*, then the relations*

$$\begin{array}{ll} v^+(x,t) = \bar{v}(x,t) & a.e. \text{ on } Q(\sigma) \\ u^+(x,t) = \bar{u}(x,t) & a.e. \text{ on } \Sigma(\sigma) \end{array}$$

hold for all controls (v^+, u^+) *of* (QP_n^ϱ) *satisfying together with the associated state* y^+ *and the adjoint state* p^+ *the optimality system (6.6) - (6.8), (3.22).*

Proof. On $Q(\sigma)$ we have $\bar{v}(x,t) = v_b$, where $\bar{H}_v^Q(x,t) \le -\sigma$, and $\bar{v}(x,t) = v_a$, where $\bar{H}_v^Q(x,t) \ge \sigma$. Therefore, $v^+ \in V_{ad}^\varrho$ means $v(x,t) \in [v_b - \varrho, v_b]$ or $v(x,t) \in [v_a, v_a + \varrho]$, respectively. Lemma 6.5 yields $\tilde{H}_v^Q \le -\sigma/2$ or $\tilde{H}_v^Q \ge \sigma/2$ on $Q(\sigma)$, hence the variational inequality (6.6) gives $v^+ = v_b$ or $v^+ = v_a$, respectively. In this way, we have shown $v^+ = \bar{v}$ on $Q(\sigma)$; u^+ is handled analogously. □

COROLLARY 6.7. *Let the assumptions of Theorem 6.4 be satisfied and suppose that* $\|w_1 - \bar{w}\|_W \le \varrho := min\{r_{\mathcal{N}}, \varrho_1, \varrho_2\}$*. Then* $\|\hat{w}_n - \bar{w}\|_W \le \varrho$ *holds for all* $n \in N$*. In particular,* $\hat{v}_n \in V_{ad}^\varrho$, $\hat{u}_n \in U_{ad}^\varrho$.

This is obtained by Theorem 4.1 and the convergence estimate (4.9).

COROLLARY 6.8. *Under the assumptions of Corollary 6.7, the sign-conditions (6.9) - (6.12) hold true for* $(y^+, p^+, v^+, u^+) := (\hat{y}_n, \hat{p}_n, \hat{v}_n, \hat{u}_n)$.

(Corollary 6.7 yields $\hat{v}_n \in V_{ad}^{\varrho_2}$, $\hat{u}_n \in U_{ad}^{\varrho_2}$, hence the result follows from Lemma 6.5.)

COROLLARY 6.9. *Under the assumptions of Corollary 6.7, the solution* $(\hat{v}_n, \hat{u}_n)$ *of* $(\widehat{QP}_n)$ *satisfies the optimality system of* (QP_n)*, too.*

Proof. The optimality systems for $(\widehat{QP}_n)$ and (QP_n) differ only in the variational inequalities. From the optimality system of $(\widehat{QP}_n)$ we know that

$$\int_Q \tilde{H}_v^Q(\hat{y}_n, \hat{p}_n, \hat{v}_n)(v - \hat{v}_n)\, dxdt \ge 0 \quad \forall v \in \widehat{V}_{ad}. \tag{6.13}$$

On $Q(\sigma), \hat{v}_n = \bar{v} = v_a$, if $\bar{H}_v^Q \ge \sigma$ and $\hat{v}_n = \bar{v} = v_b$, if $\bar{H}_v^Q \le -\sigma$. Lemma 6.5 and Corollary 6.8 yield $\tilde{H}_v^Q(\hat{y}_n, \hat{p}_n, \hat{v}_n) \ge \sigma/2$ or $\tilde{H}_v^Q(\hat{y}_n, \hat{p}_n, \hat{v}_n) \le -\sigma/2$, respectively. Therefore, $\tilde{H}_v^Q(\hat{y}_n, \hat{v}_n, \hat{p}_n)(v - \hat{v}_n) \ge 0$ holds on $Q(\sigma)$ for all real numbers $v \in [v_a, v_b]$. On the complement $Q \setminus Q(\sigma)$, the controls of $\widehat{V}_{ad}$ are not restricted to be equal to $\bar{v}$, hence in (6.13) v was arbitrary in $[u_a, u_b]$. This yields

$$\int_Q \tilde{H}_v^Q (v - \hat{v}_n)\, dxdt = \int_{Q\setminus Q(\sigma)} \tilde{H}_v^Q (v - \hat{v}_n)\, dxdt + \int_{Q(\sigma)} \tilde{H}_v^Q (v - \hat{v}_n)\, dxdt \qquad \forall v \in V_{ad},$$

where the nonnegativity of the first term follows from (6.13). The variational inequality for $\hat{u}_n$ is discussed in the same way. □

COROLLARY 6.10. *Let the assumptions of Corollary 6.7 be fulfilled. Then* $(\hat{v}_n, \hat{u}_n)$*, the solution of* $(\widehat{QP}_n)$*, satisfies the optimality system for* (QP_n^ϱ).

Proof. By Corollary 6.9, $(\hat{v}_n, \hat{u}_n)$ satisfies the variational inequality (6.13) for all $v \in V_{ad}, u \in U_{ad}$, in particular for all $v \in V_{ad}^\varrho, u \in U_{ad}^\varrho$. Moreover, $\hat{v}_n \in V_{ad}^\varrho$, $\hat{u}_n \in U_{ad}^\varrho$ is granted by Corollary 6.9. □

LEMMA 6.11. *Assume that* $\bar{w} = (\bar{y}, \bar{p}, \bar{v}, \bar{u})$ *satisfies the second order condition (SSC). If* $\varrho_3 > 0$ *is taken sufficiently small, and* $\|w_n - \bar{w}\|_W \le \varrho_3$*, then for all* $\varrho > 0$ *the problem* (QP_n^ϱ) *has at least one pair of (globally) optimal controls* (v, u).

Proof. If $\|w_n - \bar{w}\|_W \le \varrho_3$ and $\varrho_3 > 0$ is sufficiently small, then

$$H^Q_{vv}(x,t,y_n(x,t),p_n(x,t),v_n(x,t)) \ge \frac{\delta}{2} \quad \text{a.e. on } Q \tag{6.14}$$

$$H^\Sigma_{uu}(x,t,y_n(x,t),p_n(x,t),u_n(x,t)) \ge \frac{\delta}{2} \quad \text{a.e. on } \Sigma, \tag{6.15}$$

follows from (LC), $\|y_n - \bar{y}\|_{C(\bar{Q})} + \|p_n - \bar{p}\|_{C(\bar{Q})} + \|v_n - \bar{v}\|_{L^\infty(Q)} + \|u_n - \bar{u}\|_{L^\infty(\Sigma)} \le \varrho_3$ and the Lipschitz properties of H^Q_{vv}, H^Σ_{vv}. Notice that w_n belongs to a set of diameter $K := \|\bar{w}\|_W + \varrho_3$, hence the Lipschitz estimates (3.5) and (3.6) apply. Therefore, (QP^ϱ_n) has the following properties: It is a linear-quadratic problem with linear equation of state. In the objective, the controls appear linearly and convex-quadratically (with convexity following from (6.14) - (6.15)). The control-state mapping $(v,u) \mapsto y$ is compact from $L^2(Q) \times L^2(\Sigma)$ to Y. Moreover, $V^\varrho_{ad}, U^\varrho_{ad}$ are non-empty weakly compact sets of L^2. Now the existence of at least one optimal pair of controls follows by standard arguments. Here, it is essential that the quadratic control-part of J_n is weakly l.s.c. with respect to the controls and that products of the type $y \cdot v$ or $y \cdot u$ lead to sequences of the type "strongly convergent times weakly convergent sequence", so that $y_n \to y$ and $v_n \rightharpoonup v$ implies $y_n v_n \rightharpoonup yv$. □

Remark: Alternatively, this result can be deduced also from the fact that $(\hat{y}_n, \hat{v}_n, \hat{u}_n)$ satisfies together with $\hat{p}_n$ the first and second order necessary conditions for (QP^ϱ_n) and that the optimality system of (QP^ϱ_n) is uniquely solvable (cf. Thm. 6.12).

THEOREM 6.12. *Let $\bar{w} = (\bar{y}, \bar{p}, \bar{v}, \bar{u})$ fulfil the first order necessary conditions (2.1), (3.2) - (3.3), (3.7) - (3.9) together with the second order sufficient optimality condition (SSC). If $w_n = (y_n, p_n, v_n, u_n) \in W$ is given such that max$\{\|w_n - \bar{w}\|_W, \varrho\} \le$ min$\{r_\mathcal{N}, \varrho_1, \varrho_2, \varrho_3\}$, then the solution $(\hat{v}_n, \hat{u}_n)$ of $(\widehat{QP}_n)$ is (globally) optimal for (QP^ϱ_n). Together with $\hat{y}_n, \hat{p}_n$ it delivers the unique solution of the optimality system of (QP^ϱ_n).*

Proof. Denote by (v^+, u^+) the solution of (QP^ϱ_n), which exists according to Lemma 6.11. Therefore, $(y^+, p^+, v^+, u^+) = w^+$ has to satisfy the associated optimality system. On the other hand, also $\hat{w}_n = (\hat{y}_n, \hat{p}_n, \hat{v}_n, \hat{u}_n)$ fulfils this optimality system by Corollary 6.10. We show that the solution of the optimality system is unique, then the Theorem is proven.

Let us assume that another $\hat{w} = (\hat{y}, \hat{p}, \hat{v}, \hat{u})$ obeys the optimality system, too. Inserting $(\hat{v}, \hat{u})$ in the variational inequalities for (v^+, u^+), while (v^+, u^+) is inserted in the corresponding ones for $(\hat{v}, \hat{u})$, we arrive at

$$\begin{aligned} &\int_Q \{\tilde{H}^Q_v(y^+, p^+, v^+)(\hat{v} - v^+) + \tilde{H}^Q_v(\hat{y}, \hat{p}, \hat{v})(v^+ - \hat{v})\}\, dxdt + \\ &+ \int_\Sigma \{\tilde{H}^\Sigma_u(y^+, p^+, u^+)(\hat{u} - u^+) + \tilde{H}^\Sigma_u(\hat{y}, \hat{p}, \hat{u})(u^+ - \hat{u})\}\, dSdt \ge 0. \end{aligned} \tag{6.16}$$

The expressions under the integral over Q in (6.16) have the form

$$\begin{aligned} &f^n_v(\hat{v} - v^+) + H^{Q,n}_{yv}(y^+ - y_n)(\hat{v} - v^+) + H^{Q,n}_{vv}(v^+ - v_n)(\hat{v} - v^+) - p^+ d^n_v(\hat{v} - v^+) \\ &+ f^n_v(\hat{v} - v^+) + H^{Q,n}_{yv}(y^- y_n)(\hat{v} - v^+) + H^{Q,n}_{vv}(v^+ - v_n)(\hat{v} - v^+) - p^+ d^n_v(\hat{v} - v^+), \end{aligned}$$

the other terms look similarly. Simplifying (6.16) we get after setting $y = \hat{y} - y^+$, $v = \hat{v} - v^+$, $u = \hat{u} - u^+$, $p = \hat{p} - p^+$

$$\begin{aligned} 0 \le &- \int_Q \{H^{Q,n}_{yv} yv + H^{Q,n}_{vv} v^2 + p\ d^n_v\ v\}\, dxdt \\ &- \int_\Sigma \{H^{\Sigma,n}_{yu} yu + H^{\Sigma,n}_{uu} u^2 + p\ b^n_u\ u\}\, dSdt. \end{aligned} \tag{6.17}$$

The difference $p = \hat{p} - p^+$ obeys

$$
(6.18) \qquad \begin{aligned} -p_t + A\,p &= H_{yy}^{Q,n} y + H_{yv}^{Q,n} v - d_y^n p \\ \partial_\nu p &= H_{yy}^{\Sigma,n} y + H_{yu}^{\Sigma,n} u - b_y^n p \\ p(T) &= \varphi_{yy}^n y(T). \end{aligned}
$$

Multiplying the PDE in (6.18) by y and integrating over Q we find after an integration by parts

$$
(6.19) \qquad \begin{aligned} &-\int_\Omega p(T)y(T)dx + \int_0^T (y_t, p)_{H^1(\Omega)', H^1(\Omega)} dt + \int_Q < \mathcal{A}\nabla p, \nabla y > dxdt \\ &= \int_Q (H_{yy}^{Q,n} y^2 + H_{yv}^{Q,n} yv - d_y^n\ p\, y)\, dxdt + \int_\Sigma (H_{yy}^{\Sigma,n} y^2 + H_{yu}^{\Sigma,n} yu - b_y^n\ p\ y)\, dSdt. \end{aligned}
$$

This description of the procedure was formal, as the definition of the weak solution of (6.18) requires the test function y to be zero at $t = T$. To make (6.19) precise we have to use the information that $p \in W(0,T), y \in W(0,T)$ along with the integration by parts formula

$$
\int_0^T (p_t, y)_{H^1(\Omega)', H^1(\Omega)} dt = \int_\Omega (p(T)y(T) - p(0)y(0))dt - \int_0^T (y_t, p)_{H^1(\Omega)', H^1(\Omega)} dt.
$$

Next, we invoke the state equation for $y = \hat{y} - y^+$ and the condition for $p(T)$ to obtain from (6.19)

$$
(6.20) \qquad \begin{aligned} &-\int_\Omega \varphi_{yy}^n y(T)^2 dx - \int_Q (H_{yy}^{Q,n} y^2 + H_{yv}^{Q,n} yv)\, dxdt \\ &\quad - \int_\Sigma (H_{yy}^{\Sigma,n} y^2 + H_{yu}^{\Sigma,n} yu)\, dSdt = \int_Q d_v^n\ v\ p\ dxdt + \int_\Sigma d_u^n\ u\ p\ dSdt. \end{aligned}
$$

Adding (6.20) to (6.17) yields

$$
0 \le -\int_\Omega \varphi_{yy}^n y(T)^2 dx - \int_Q (y,v) D^2 H^{Q,n} (y,v)^\top dxdt - \int_\Sigma (y,u) D^2 H^{\Sigma,n}(y,u)^\top dSdt,
$$

that is $0 \le -Q^n[y,v,u]^2$. As $max\{\, \|w_n - \bar{w}\|_W, \varrho\} \le \varrho_2$, Corollary 6.6 yields $v = 0$ on $Q(\sigma)$ and $u = 0$ on $\Sigma(\sigma)$. Therefore, Lemma 6.2 applies to conclude $\delta/2\, \|(y,v,u)\|_H^2 \le 0$, i.e. $v = 0, u = 0$. In other words, $\hat{v} = v^+, \hat{u} = u^+$, completing the proof. □

Now we are able to formulate the main result of this paper:

THEOREM 6.13. *Let $\bar{w} = (\bar{y}, \bar{p}, \bar{v}, \bar{u})$ satisfy the assumptions of Theorem 6.12 and define $\varrho_\mathcal{N} = min\{r_\mathcal{N}, \varrho_1, \varrho_2, \varrho_3\}$. If $max\,\{\varrho, \|w_1 - \bar{w}\|\} \le \varrho_\mathcal{N}$ then the sequence $\{w_n\} = \{(y_n, p_n, v_n, u_n)\}$ generated by the SQP method by solving (QP_n^ϱ) coincides with the sequence $\hat{w}_n$ obtained by solving $(\widehat{QP}_n)$. Therefore, w_n converges q-quadratically to $\bar{w}$ according to the convergence estimate (4.9).*

Thanks to this Theorem, we are justified to solve (QP_n^ϱ) instead of $(\widehat{QP}_n)$ to obtain the same (unique) solution. This result is still not completely satisfactory, as the unknown element $\bar{w}$ was used to define (QP_n^ϱ).

However, an analysis of this section reveals that any convex, closed set $\tilde{V}_{ad}, \tilde{U}_{ad}$ can be taken instead of $V_{ad}^\varrho, U_{ad}^\varrho$, if the following properties are satisfied:
$\tilde{V}_{ad} \subset V_{ad}^{\varrho_\mathcal{N}}, \tilde{U}_{ad} \subset U_{ad}^{\varrho_\mathcal{N}}$, and $\tilde{V}_{ad} \supset V_{ad}^{\varrho_0}, \tilde{U}_{ad} \supset U_{ad}^{\varrho_0}$ for some $\varrho_0 > 0$ (the last condition is needed to guarantee $\hat{v}_n = \bar{v}$ on $Q(\sigma), \hat{u}_n = \bar{u}$ on $\Sigma(\sigma)$ and, last but not least, to make the convergence $\hat{v}_n \to \bar{v}, \hat{u}_n \to \bar{u}$ possible).

Define, for instance, $\varrho_0 = \|\bar{w} - w_1\|_W$, where $\varrho_0 \le \frac{1}{3}\varrho_{\mathcal{N}}$,

$$\tilde{V}_{ad} = \{v \in V_{ad} |\ \|v - v_1\|_{L^\infty(Q)} \le 2\varrho_0\}$$

$$\tilde{U}_{ad} = \{u \in U_{ad} |\ \|u - u_1\|_{L^\infty(\Sigma)} \le 2\varrho_0\},$$

where $\varrho_0 = \|\bar{w} - w_1\|_W$ is the distance of the starting element of the SQP method to $\bar{w}$. Then $V_{ad}^{\varrho_0} \subset \tilde{V}_{ad} \subset V_{ad}^{\varrho_{\mathcal{N}}}$. The same property holds for $\tilde{U}_{ad}$. Then case the SQP method will deliver the same solution in $\tilde{V}_{ad}, \tilde{Q}_{ad}$ as in $V_{ad}^{\varrho_{\mathcal{N}}}, U_{ad}^{\varrho_{\mathcal{N}}}$. This however, is the solution in $\widehat{V}_{ad}, \widehat{U}_{ad}$.
Remark: The restriction of the admissible sets to V_{ad}^{ϱ}, U_{ad}^{ϱ} might appear artificial, since restrictions of this type are not known from the theory of SQP methods in spaces of finite dimension. However, it is indispensible. In finite dimensions, the set of active constraints is detected after one step, provided that the starting value was chosen sufficiently close to the reference solution. The further analysis can rely on this. Here, we cannot determine the active set in finitely many steps unless we assume this a-priorily as in the definition of $\widehat{QP}_n$.

REFERENCES

[1] W. Alt. The Lagrange-Newton method for infinite-dimensional optimization problems. *Numer. Funct. Anal. and Optim.* 11:201–224, 1990. Sequential quadratic programming in Banach spaces.

[2] W. Alt. The Lagrange Newton method for infinite-dimensional optimization problems. *Control and Cybernetics*, 23:87–106, 1994.

[3] W. Alt. Discretization and mesh independence of Newton's method for generalized equations. In A.K. Fiacco, Ed. *Mathematical Programming with Data Perturbation.* Lecture Notes in Pure and Appl. Math., Vol. 195, Marcel Dekker, New York, 1–30.

[4] W. Alt, R. Sontag, and F. Tröltzsch. An SQP Method for Optimal Control of a Weakly Singular Hammerstein Integral Equation. *Appl. Math. Opt.*, 33:227–252, 1996.

[5] W. Alt and K. Malanowski. The Lagrange–Newton method for nonlinear optimal control problems. *Computational Optimiz. and Appl.*, 2:77–100, 1993.

[6] W. Alt and K. Malanowski. The Lagrange–Newton method for state-constrained optimal control problems. *Computational Optimiz. and Appl.*, 4:217–239, 1995.

[7] E. Casas. Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations. *SIAM J. Control Optim.*, 35:1297–1327, 1997.

[8] A.L. Dontchev. Local analysis of a Newton–type method based on partial linearization. In J. Renegar, M. Shub, and S. Smale, Eds. *Proc. of the Summer Seminar "Mathematics of numerical analysis: Real Number algorithms"*, Park City, UT July 17– August 11, 1995, to appear.

[9] A.L. Dontchev and W.W. Hager. Lipschitzian stability in nonlinear control and optimization. *SIAM J. Contr. Optim.*, 31:569–603, 1993.

[10] A.L. Dontchev, W.W. Hager, A.B. Poore, and B. Yang. Optimality, stability, and convergence in optimal control. *Appl. Math. Optim.*, 31:297–326, 1995.

[11] H. Goldberg and F. Tröltzsch. On a Lagrange–Newton method for a nonlinear parabolic boundary control problem. *Optimization Methods and Software*, 8:225-247, 1998.

[12] H. Goldberg and F. Tröltzsch. On a SQP-Multigrid technique for nonlinear parabolic boundary control problems. In W.W Hager and P.M. Pardalos, Eds., *Optimal Control: Theory, Algorithms, and Applications*, pp. 154–177. Kluwer Academic Publishers B.V. 1998.

[13] M. Heinkenschloss. The numerical solution of a control problem governed by a phase field model. *Optimization Methods and Software*, 7:211–263, 1997.

[14] M. Heinkenschloss and E. W. Sachs. Numerical solution of a constrained control problem for a phase field model. *Control and Estimation of Distributed Parameter Systems, Int. Ser. Num. Math.*, 118:171–188, 1994.

[15] M. Heinkenschloss and F. Tröltzsch. Analysis of the Lagrange-SQP-Newton method for the control of a Phase field equation *Virginia Polytechnic Institute and State Unversity*, ICAM Report 95-03-01, submitted.

[16] K. Ito and K. Kunisch. Augmented Lagrangian-SQP methods for nonlinear optimal control problems of tracking type. *SIAM J. Control Optim.*, 34:874–891, 1996.

[17] K. Ito and K. Kunisch. Augmented Lagrangian-SQP methods in Hilbert spaces and application to control in the coefficients problems. *SIAM J. Optim.*, 6:96–125, 1996.

[18] N.H. Josephy. Newton's method for generalized equations. *Techn. Summary Report No. 1965, Mathematics Research Center, University of Wisconsin-Madison*, 1979.

[19] C.T. Kelley and E.W. Sachs. Fast Algorithms for Compact Fixed Point Problems with Inexact Function Evaluations. *SIAM J. Scientific and Stat. Computing* 12:725–742, 1991.

[20] C.T. Kelley and E.W. Sachs. Multilevel algorithms for constrained compact fixed point problems. *SIAM J. Scientific and Stat. Computing* 15:645–667, 1994.

[21] C.T. Kelley and E.W. Sachs. Solution of optimal control problems by a pointwise projected Newton method. *SIAM J. Contr. Optimization*, 33:1731–1757, 1995.

[22] K. Kunisch, S. Volkwein. Augmented Lagrangian-SQP techniques and their approximations. *Contemporary Mathematics*, 209:147–159, 1997.

[23] S.F. Kupfer and E.W. Sachs. Numerical solution of a nonlinear parabolic control problem by a reduced SQP method. *Computational Optimization and Applications*, 1:113–135, 1992.

[24] O.A. Ladyženskaya, V.A. Solonnikov, and N.N. Ural'ceva. Linear and quasilinear equations of parabolic type. Transl. of Math. Monographs, Vol. 23, Amer. Math. Soc., Providence, R.I. 1968.

[25] J.L. Lions. *Contrôle optimal des systèmes gouvernés par des equations aux derivées partielles.* Dunod, Gauthier-Villars Paris, 1968.

[26] J.L. Lions and E. Magenes. *Problèmes aux limites non homogènes et applications*, volume 1–3. Dunod, Paris, 1968.

[27] K. Machielsen. Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space. CWI Tract, 53, Amsterdam, 1987.

[28] J.P. Raymond and H. Zidani. Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations, to appear.

[29] S.M. Robinson. Strongly regular generalized equations. *Math. of Operations Research*, 5:43–62, 1980.

[30] E.J.P.G. Schmidt. Boundary control for the heat equation with nonlinear boundary condition. *J. Diff. Equat.*, 78:89–121, 1989.

[31] F. Tröltzsch. An SQP method for the optimal control of a nonlinear heat equation. *Control and Cybernetics*, 23(1/2):267–288, 1994.

[32] F. Tröltzsch. Convergence of an SQP–Method for a class of nonlinear parabolic boundary control problems. In W. Desch, F. Kappel, K. Kunisch, eds., *Control and Estimation of Distributed Parameter Systems. Nonlinear Phenomena.* Int. Series of Num. Mathematics, Vol. 118, Birkhäuser Verlag, Basel 1994, pp. 343–358.

[33] F. Tröltzsch. Lipschitz stability of solutions to linear-quadratic parabolic control problems with respect to perturbations. Preprint-Series Fac. of Math., TU Chemnitz-Zwickau, Report 97–12, *Dynamics of Continuous, Discrete and Impulsive Systems*, accepted.