# Convergence of an SQP–method for a class of nonlinear parabolic boundary control problems

Fredi Tröltzsch

Technical University of Chemnitz-Zwickau,
Department of Mathematics
09009 Chemnitz, PSF 964, Germany

**Abstract.** We investigate local convergence of an SQP method for an optimal control problem governed by a parabolic equation with nonlinear boundary condition. Sufficient conditions for local quadratic convergence of the method are discussed.

**AMS subject classification.** 49M05, 49M40, 49K24.

**Keywords.** Sequential quadratic programming, optimal control, parabolic equation, nonlinear boundary condition, control constraints.

## 1 Introduction

In this paper, we investigate the behaviour of a Sequential Quadratic Programming (SQP)–method applied to the following very simplified model problem (P):

(P) Minimize

$$\frac{1}{2}\int_0^T \int_\Gamma \left\{(w(t,\xi)-q(t,\xi))^2+\lambda u(t,\xi)^2\right\} dS_\xi dt \tag{1.1}$$

subject to

$$\begin{array}{rcll} w_t(t,\xi) & = & (\Delta_\xi w - w)(t,\xi) & in \quad \Omega \\ w(0,\xi) & = & 0 & in \quad \Omega \\ \dfrac{\partial w}{\partial n}(t,\xi) & = & b(w(t,\xi))+u(t,\xi) & on \quad \Gamma \end{array} \tag{1.2}$$

and to the constraints on the control

$$|u(t,\xi)| \leq 1, \tag{1.3}$$

$t \in [0,T]$. The control $u$ is looked upon in $L_\infty((0,T)\times\Gamma)$, while the state $w$ is defined as mild solution of (1.2) (cf. section 2).

In this setting, a bounded domain $\Omega \subset R^n (n \geq 2)$ with $C^\infty$–boundary $\Gamma$, positive constants $\lambda, T$, and functions $b \in C^2(I\!R), q \in L_\infty((0,T)\times\Gamma)$ are given. By $n$ and $dS$ the outward normal vector and the surface measure on $\Gamma$, respectively, are denoted.

Pointwise constraints of the type (1.3) are often imposed for a correct modelling of the underlying process. They reflect technical limitations to the possible choice of the control and cannot be realized by a smooth penalization. The term $\lambda\|u\|^2$ in (1.1) is to enhance continuity of the optimal control (it may express the cost for the control, too).

It is known since several years that the SQP algorithm, applied to mathematical programming problems in finite-dimensional spaces, exhibits local quadratic convergence. The method can be easily extended to infinite-dimensional optimization problems such as optimal control problems. We refer, for instance, to the works by Alt [1], [2], Alt and Malanowski [3], Kelley and Wright [7], or Levitin and Polyak [10]. In the context of nonlinear parabolic control problems without control constraint we mention the numerical work by Kupfer and Sachs [8].

Recently, Alt, Sontag and Tröltzsch [4] proved the local quadratic convergence of the SQP method for the optimal control of a weakly singular Hammerstein integral equation with pointwise constraints on the control. In the author's paper [12], the proof of convergence was transferred to the one-dimensional heat equation with nonlinear boundary condition. The aim of this note is to extend the convergence result to a parabolic equation in a domain of dimension $n$.

We assume that $b$ and its derivatives up to the order 2 are uniformly bounded and Lipschitz: There are constants $c_B, c_l$:

$$|b^{(i)}(w)| \leq c_B \qquad |b^{(i)}(w_1) - b^{(i)}(w_2)| \leq c_l, \tag{1.4}$$

for all $w, w_1, w_2 \in I\!R, i = 0,1,2$. We may weaken the conditions (1.4) to local ones. However, this would lead to difficulties, as the solution $w$ of (1.2) could blow up in finite time. For convenience we consider only the very special type of boundary condition in (1.2). This enables us to work out the principal behaviour of the SQP–method and to avoid tedious technical estimates. The case of more general boundary conditions having the form $\partial w/\partial n = b_1(w) + b_2(w)\,u$ is discussed in [4].

## 2 Integral equation method

Let us define $A : L_2(\Omega) \supset D(A) \to L_2(\Omega)$ by $D(A) = \{w \in W_2^2(\Omega) : \partial w/\partial n = 0 \ on \ \Gamma\}$, $Aw = -\Delta w + w$ for $w \in D(A)$. $-A$ is known to generate an analytic semigroup $\{S(t)\}, t \geq 0$, of continuous linear operators in $L_2(\Omega)$. Moreover, we introduce the Neumann operator $N : L_2(\Gamma) \to W_2^s(\Omega)(s < 1 + 1/2)$ by $N : g \mapsto w, \Delta w - w = 0, \partial w/\partial n = g$. Next, we fix $\sigma, p \in I\!R$ by $p > n + 1$ and $n/p < \sigma < 1 + 1/p$. The last two inequalities have a non–void intersection for $p > n - 1$, while $p > n + 1$ is needed to work with states being continuous w.r. to t. In the same way we may introduce operators $A_r, S_r(t), N_r$ just by substituting above the order of integrability r for 2, where $1 < r < \infty$. Restricting $AS(t)N$ to $L_r(\Gamma)(r \geq 2)$, we obtain $A_r S_r(t) N_r$. Therefore, we shall use in the paper the same symbol $AS(t)N$ regarded in different spaces $L_r$. To continue our preparations we define a Nemytskiĭ operator $\mathcal{B} : C(\Gamma) \to C(\Gamma)$ by $(\mathcal{B}x)(\xi) = b(x(\xi))$.

A function $w \in C([0,T], W_p^\sigma(\Omega))$ is said to be a *mild solution* of (1.2), if the Bochner integral equation

$$w(t) = \int_0^t AS(t-s)N(\mathcal{B}(\tau w(s)) + u(s))ds \tag{2.1}$$

holds on $[0,T]$ ($\tau$: trace operator). The expression on the right hand side makes sense, as $u \in L_\infty((0,T)\times\Gamma) \subset L_p((0,T)\times\Gamma) = L_p(0,T;L_p(\Gamma))$ and $W_p^\sigma(\Omega) \subset C(\overline{\Omega})$ by $n/p < \sigma$ (here we regard $AS(t)N$ as operator from $L_p(\Gamma)$ to $W_p^\sigma(\Omega), t > 0$). Owing to the strong assumption (1.4), to each $u \in L_\infty((0,T)\times\Gamma)$ a unique global solution $w$ of (2.1) exists (cf. Tröltzsch [11]). Turning over to the trace $x(t) = \tau w(t)$ in (2.1) we arrive at the integral equation

$$x(t) = \int_0^t \tau AS(t-s)N(\mathcal{B}(x(s)) + u(s))ds \tag{2.2}$$

for $x \in C([0,T], C(\Gamma))$.

The estimate (see Amann [4])

$$\|AS(t)N\|_{L_r(\Gamma)\to W_r^\sigma(\Omega)} \leq c\, t^{-\alpha}, \tag{2.3}$$

where $\alpha = 1 - (\sigma' - \sigma)/2$ and $0 < \sigma < \sigma' < 1 + 1/r$, turns out to be essential for investigating properties of the integral operator $K$,

$$(Kz)(t) = \int_0^t \tau AS(t-s)Nz(s)ds. \tag{2.4}$$

Let us briefly discuss (2.3) for $r := p$: Taking $\sigma = n/p+\varepsilon, \sigma' = 1+1/p-\varepsilon(\varepsilon > 0)$ we find that (2.3) holds for all $\alpha > 0.5+(n-1)/2p$. $K$ maps continuously $L_p(0,T;L_p(\Gamma))$ into $C([0,T],C(\Gamma))$ provided that $p > 1/(1-\alpha)$. This holds together with the last inequality for $\alpha$, if $p > n+1$. For $p \downarrow n+1$ we may take $\alpha \downarrow n/(n+1)$.

For convenience we regard $K$ between different spaces: $K$ may be viewed as operator in $L_r(0,T;L_r(\Gamma))$ for all $1 < r < \infty$. Let its adjoint $K^*$ be defined for $r = 2$. It is known that

$$(K^*z)(t) = \int_t^T \tau A S(s-t) N z(s) ds,$$

hence $K^*$ has the same transformation properties as $K$. By means of these prerequisites we are able to write (P) as

$$\text{(P)} \qquad f(x,u) = \frac{1}{2}\int_0^T \{\|x(t)-q(t)\|^2_{L_2(\Gamma)} + \lambda\|u(t)\|^2_{L_2(\Gamma)}\}dt = min!$$

subject to

$$x = K(B(x)+u), \quad u \in U^{ad}. \tag{2.5}$$

Here we have introduced $B$ in $C([0,T],C(\Gamma))$ by $(Bx)(t,\cdot) = \mathcal{B}(x(t,\cdot))$ and $U^{ad} := \{u \in L_\infty((0,T)\times\Gamma) : |u(t,\xi)| \le 1\}$. The equation in (2.5) is well defined, as $K$ maps $L_p(0,T;L_p(\Gamma))$ into $C([0,T],C(\Gamma))$.
In the paper, the following notation is used: We write $L_r = L_r(0,T;L_r(\Gamma)), 1 \le r < \infty, L_\infty = L_\infty((0,T)\times\Gamma), C = C([0,T],C(\Gamma)) = C([0,T]\times\Gamma)$ and endow the spaces with their natural norms $\|\cdot\|_r$ and $\|\cdot\|_\infty$, respectively. The natural norm of $L_\alpha(0,T;L_\beta(\Gamma)), 1 < \alpha,\beta < \infty$, is denoted by $\|\cdot\|_{L_{\alpha,\beta}}$ . For $\alpha = \beta = \infty$ we set $\|\cdot\|_{L_{\infty,\infty}} := \|\cdot\|_\infty$. In product spaces of this type, the norm is defined as the sum of the corresponding norms. In $C \times L_r, \|(x,u)\|_{\infty,r} = \|x\|_\infty + \|u\|_r, 1 \le r \le \infty$, and $\|(x,u)\|_r := \|(x,u)\|_{r,r}$, in $C \times L_r \times L_r : \|(x,u,y)\|_{\infty,r} = \|x\|_\infty + \|(u,y)\|_r, \|(x,u,y)\|_\infty := \|(x,u,y)\|_{\infty,\infty}$. An "inner product" is defined formally by

$$< x,y > = \int_0^T \int_\Gamma x(t,\xi)y(t,\xi)\, dS_\xi dt$$

just denoting integration of $x\,y$ over $[0,T]\times\Gamma$.

# 3 Known optimality conditions

The functional $f : C \times L_p \to I\!R$ and the mapping $(x,u) \mapsto B(x)+u$ from $C \times L_p$ to $L_p$ are twice continuously Fréchet differentiable. This enables us to apply lateron second order methods to $(P)$. Owing to the convexity of $f$ and the linear appearance of $u$ in (1.2), standard methods show the existence of at least one optimal control $u_0$ for $(P)$. Let $x_0$ be the corresponding state. For $y \in L_\infty$ the *Lagrange function*

$$\mathcal{L}(x,u,y) = f(x,u) - < y, x - K(B(x)+u) >$$

is defined. From $\mathcal{L}_x = 0$ we obtain formally the equation $y_0 = f_x + B'(x_0)^* K^* y_0$ for the Lagrange multiplier $y_0$. A careful discussion (taking derivatives in $C \times L_\infty$ and regarding the derivatives as linear operators in $L_p$) justifies this (cf. Tröltzsch [11]):

$$y_0(t) = x_0(t) - q(t) + B'(x_0(t)) \int_t^T \tau A S(s-t) N \, y_0(s) ds. \tag{3.1}$$

Here, we took advantage of $(\mathcal{B}'(x_0(t,\cdot))h(t,\cdot))(\xi) = b'(x_0(t,\xi)) \cdot h(t,\xi)$, hence $B'(x_0)$ is formally self–adjoint. The variational inequality $< \mathcal{L}_u, u - u_0 > \geq 0 \ \forall u \in U^{ad}$ yields $< \lambda u_0 + K^* y_0, u - u_0 > \geq 0$. After a standard pointwise discussion we arrive at

$$u_0(t,\xi) = P_{[-1,1]}\{-\lambda^{-1}(K^* y_0)(t,\xi)\}, \tag{3.2}$$

where $P_{[-1,1]} : I\!R \to [-1,1]$ denotes projection onto $[-1,1]$. We assume that in addition to the first order necessary conditions (3.1), (3.2) the following *second order sufficient optimality condition* is satisfied: There is a $\delta > 0$ such that

(SSC) $$\mathcal{L}_{vv}(x_0,u_0,y_0)[v-v_0, v-v_0] \geq \delta \|v - v_0\|_2^2 \tag{3.3}$$

for all $v = (x,u)$ satisfying the linearized equation

$$x - x_0 = K(B'(x_0)(x-x_0) + u - u_0). \tag{3.4}$$

In (3.3), $\mathcal{L}_{vv}$ denotes the second order $F$-derivative of $\mathcal{L}$ w.r. to $v = (x,u)$ in $C \times L_p$ at $v_0 := (x_0,u_0)$. The sufficiency of (SSC) was discussed in [6]. We finish this section by the following very useful Lemma:

**Lemma 3.1** *Let $1 \leq r, \rho \leq \infty, D$ be the linear continuous operator in $L_{r,\rho}$ defined by $(Dx)(t,\xi) = \beta(t,\xi)x(t,\xi)$, where $\beta \in L_\infty$. Then $(I - DK)^{-1}$ and $(I-KD)^{-1}$ exist as linear continuous operators in $L_{r,\rho}$ and their norm is bounded by a constant $c_{r\rho}$ which depends only $\|\beta\|_\infty, r,$ and $\rho$.*

**Proof.** The invertibility follows from a standard application of the Banach fixed point theorem. Let $z \in L_{r,\rho}$ be given. We estimate the solution of the equation $x = BKx + z$. Then

$$\begin{aligned}\|x(t)\|_{L_\rho(\Gamma)} &\leq \int_0^t \|\beta\|_\infty \|\tau A S(t-s) N\|_{L_\rho(\Gamma)\to L_\rho(\Gamma)} \|x(s)\|_{L_\rho(\Gamma)} ds + \|z(t)\|_{L_\rho(\Gamma)} \\ &\leq c \int_0^t (t-s)^{-\alpha} \|x(s)\|_{L_\rho(\Gamma)} ds + \|z(t)\|_{L_\rho(\Gamma)}.\end{aligned}$$

Here, $c > 0$ depends only on $\rho$, $\|\beta\|_\infty$, and $\alpha \in (0,1)$. $\|x(t)\|$ is majorized by the real function $\varphi(t)$ solving the corresponding weakly singular integral equation. Therefore

$$\|x\|_{L_{r,\rho}} = (\int_0^T \|x(t)\|^r_{L_\rho(\Gamma)} dt)^{1/r} \leq \|\varphi\|_{L_r(0,T)} \leq c(\int_0^T \|z(t)\|^r_{L_\rho(\Gamma)} dt)^{1/r} = c\|z\|_{L_{r,\rho}}.$$

In this way the Lemma is shown for $I - DK$. The arguing for $I - KD$ is identical. □

# 4 The SQP method, Hölder estimates

Initiating from a starting point $(x_1, u_1, y_1)$ in $C \times L_\infty \times L_\infty$ the (full) SQP method generates sequences $\{x_n\}, \{u_n\}, \{y_n\}$ by solving certain quadratic programs. Adopting the notation by Alt [1], one step of the method can be described as follows: Let $w = (x_w, u_w, y_w)$ be the result of the last iteration. As before, we write $v_w = (x_w, u_w), v = (x, u)$. The next iterate $\overline{v}_w = (\overline{x}_w, \overline{u}_w)$ is obtained as the solution of

$$(QP)_w \qquad F(v; w) = f'(v_w)(v - v_w) + \frac{1}{2}\mathcal{L}_{vv}(v_w,\ y_w)[v - v_w, v - v_w] = min! \tag{4.1}$$

$$x = K(B'(x_w)(x - x_w) + B(x_w) + u), \quad u \in U^{ad}. \tag{4.2}$$

$\overline{y}_w$ is the corresponding Lagrange multiplier.

*Remark.* Define $g(v) = g(x, u) = x - K(B(x) + u)$. Then the state equation (2.5) reads $g(v) = 0$. (4.2) is its linearization $g(v_w) + g'(v_w)(v - v_w) = 0$, which simplifies by linearity w.r. to $u$.

The Lagrange function $\tilde{\mathcal{L}}$ to $(QP)_w$ is

$$\tilde{\mathcal{L}}(v, y) = F(v; w) - \ < y, x - K(B'(x_w)(x - x_w) + B(x_w) + u) > .$$

$\tilde{\mathcal{L}}_x = 0$ leads to the (adjoint) equation

$$\begin{array}{rcl} y & = & B'(x_w)K^*y + f_x(v_w) + f_{xx}(v_w)[\overline{x}_w - x_w, \cdot] \\ & & +B''(x_w)[\overline{x}_w - x_w, \cdot]K^*y_w \end{array} \tag{4.3}$$

for $y = \overline{y}_w$. After a simple calculation we find

$$\begin{array}{rcl} \overline{y}_w(t) & = & \mathcal{B}'(x_w(t)) \int_t^T \tau AS(s-t)N\overline{y}_w(s)ds + \overline{x}_w(t) - q(t) \\ & & +\mathcal{B}''(x_w(t))[\overline{x}_w(t) - x_w(t), \cdot] \int_t^T \tau AS(s-t)Ny_w(s)ds. \end{array} \tag{4.4}$$

The variational inequality determining $\overline{u}_w, < \tilde{\mathcal{L}}_u, u - \overline{u}_w > \geq 0 \quad \forall u \in U^{ad}$, gives

$$\overline{u}_w(t,\xi) = P_{[-1,1]}\{-\lambda^{-1}(K^*\overline{y}_w)(t,\xi)\}. \tag{4.5}$$

A straightforward calculation by means of (SSC) yields

$$F(v;w_0) \geq \delta\|v - v_0\|_2^2 = F(v_0;w_0) + \delta\|v - v_0\|_2^2 \tag{4.6}$$

for all $v$ being admissible for $(QP)_{w_0}$ (where $w_0 = (v_0, y_0) = (x_0, u_0, y_0)$).

**Lemma 4.1** *For all $w$ in a sufficiently small $C \times L_p \times L_p$–neighbourhood $N_1(w_0)$, $(QP)_w$ admits a unique solution $(\overline{x}_w, \overline{u}_w) = \overline{v}_w$ with associated Lagrange multiplier $\overline{y}_w$. There is a constant $c_H$ not depending on $w$, such that*

$$\|\overline{v}_w - v_0\|_2 \leq c_H\|w - w_0\|_2^{1/2} \tag{4.7}$$

$$\|\overline{y}_w\|_\infty \leq c_H \tag{4.8}$$

*for all $w \in N_1(w_0)$.*

**Proof**. We shall only briefly sketch the proof, which is along the lines of Alt [1] or Alt, Sontag and Tröltzsch [3]. A first step initiates from the simple observation $F(\overline{v}_w;w) \leq F(\tilde{v}_w;w)$, where $\tilde{v}_w = (\tilde{x}_w, u_w)$ and $\tilde{x}_w$ is the state obtained from (4.2) for $u = u_w$. By means of Lemma 3.1 the *upper estimate*

$$F(\overline{v}_w;w) \leq c\|w - w_0\|_2 \tag{4.9}$$

can be derived. A *lower estimate*

$$F(\overline{v}_w;w) \geq \frac{\delta}{2}\|\overline{v}_w - v_0\|_2^2 - c\|w - w_0\|_2 \tag{4.10}$$

follows from re-writing terms like $f'(v_w)$ or $\mathcal{L}_{vv}(v_w, y_w)$ in terms of $f'(v_0), \mathcal{L}_{vv}(v_0, y_0)$ etc., estimating the correction parts and exploiting (SSC) and Lemma 3.1. We omit the details of the lengthy computations. (4.9) and (4.10) yield (4.7). As regards (4.8), we first observe that $\overline{u}_w$ is uniformly bounded, hence $\|\overline{x}\|_\infty$ is bounded, too (this is a consequence of (4.2) for $u = \overline{u}_w : \|x_w\|_\infty$ is bounded, as $w \in N_1(w_0)$; apply Lemma 3.1 to $x = \overline{x}_w$ in (4.2)). Now the uniform boundedness of $\overline{y}_w$ is an immediate conclusion of (4.4) and Lemma 3.1. □

**Corollary 4.2** *The estimate (4.7) holds true in the form*

$$\|\overline{v}_w - v_0\|_{\infty,p} \leq c'_H \|w - w_0\|_{\infty,p}^{1/p} \tag{4.11}$$

*for all $w \in N'_1(w_0) \subset N_1(w_0)$.*

**Proof.** (4.7) means in particular $\|\overline{u}_w - u_0\|_2 \leq c_H \|w - w_0\|_2^{1/2}$. Exploiting $|\overline{u}_w - u_0| \leq 2$ it is easy to show that

$$\|\overline{u}_w - u_0\|_p \leq c\|w - w_0\|_2^{1/p} \leq c'\|w - w_0\|_{\infty,p}^{1/p}. \tag{4.12}$$

Subtracting the equations for $\overline{x}_2$ and $x_0$ we get

$$\begin{aligned}&(\overline{x}_w - x_0) - KB'(x_w)(\overline{x}_w - x_0) = \\ &K(B(x_w) - B(x_0) + B'(x_w)(x_0 - x_w) + \overline{u}_w - u_0).\end{aligned}$$

Therefore, the $L_\infty$–norm of the left–hand side is less or equal $\|K\|_{L_p \to C}(c_1\|x_w - x_0\|_\infty + c_2\|\overline{u}_w - u_0\|_p) \leq c\|w - w_0\|_{\infty,p}^{1/p}$ by (4.12) (provided that $\|w - w_0\|_{\infty,p} \leq 1$). Now Lemma 3.1 applies to the left–hand side,

$$\|\overline{x}_w - x_0\|_\infty \leq c''\|w - w_0\|_{\infty,p}^{1/p},$$

implying (4.11). □

In the same way, subtraction of the equations for $y_0, \overline{y}_w$ yields

**Corollary 4.3** *There is a constant $c''_H > 0$ such that*

$$\|\overline{y}_w - y_0\|_\infty \leq c''_H \|w - w_0\|_{\infty,p}^{1/p} \tag{4.13}$$

*for all $w \in N''_1(w_0) \subset N_1(w_0)$.*

We omit the proof. In what follows let $N_1(w_0)$ denote the intersection $N_1(w_0) \cap N'_1(w_0) \cap N''_1(w_0)$.

# 5 Right hand side perturbations, Lipschitz estimate

Following Alt [1], [2], we consider now the close relationship between the stability of $(QP)_w$ and certain perturbations of $(QP)_{w_0}$. We discuss the perturbed problem

$$(QS)_\pi \qquad f'(v_0)(v-v_0)+\tfrac{1}{2}\mathcal{L}_{vv}(v_0,y_0)[v-v_0,v-v_0]-<d,v-v_0> = min!$$

$$x = e + K(B'(x_0)(x-x_0)+B(x_0)+u),\ u\in U^{ad} \tag{5.1}$$

belonging to the perturbation $\pi=(d,e)=(d_x,d_u,e)\in L_\infty\times L_\infty\times C$. For $\pi=0$ this problem has the unique solution $v_0=(x_0,u_0)$. In $(QS)_\pi$ we regard $x,u$ in $L_2$, although the constraint $u\in U^{ad}$ automatically generates only $L_\infty-$solutions.

$(QS)_\pi$ is a linear–quadratic parabolic control problem, where the theory is already widely investigated. Owing to $(SSC)$, the following result is therefore standard: There is a neighbourhood $N_2(0)$, and a positive constant $c_h>0$ such that for all $e\in N_2(0)$ and all $d\in L_\infty\times L_\infty$ problem $(QS)_\pi$ admits a unique solution $v_\pi=(x_\pi,u_\pi)$ and

$$\|v_\pi - v_0\|_2 \le c_h\|\pi\|_2^{1/2} \tag{5.2}$$

for all $\pi=(d,e)$ such that $e\in N_2(0)$. We are able to improve this estimate in Theorem 5.3.

**Lemma 5.1** *Let $y_\pi$ be the Lagrange multiplier belonging to $v_\pi$ and $2\le\alpha,\beta\le\infty$. Then there is a constant $c_{\alpha\beta}>0$ such that*

$$\|y_\pi - y_0\|_{L_{\alpha,\beta}} \le c_{\alpha\beta}\left(\|x_\pi - x_0\|_{L_{\alpha,\beta}}+\|\pi\|_{L_{\alpha,\beta}}\right) \tag{5.3}$$

*for all $\pi\in L_\infty\times L_\infty\times C$.*

**Proof.** The adjoint equations defining $y_0, y_\pi$ are (3.1) and

$$y_\pi = (x_\pi - q) - d_x + B'(x_0)K^*y_\pi + B''(x_0)[x_\pi - x_0,\cdot]K^*y_0. \tag{5.4}$$

Subtraction of these equations yields after some estimations

$$\begin{aligned}&\|(y_0-y_\pi)-B'(x_0)K^*(y_0-y_\pi)\|_{L_{\alpha,\beta}}\\ &=\|d_x+x_\pi-x_0+B''(x_0)[x_\pi-x_0,\cdot]K^*y_0\|_{L_{\alpha,\beta}}.\end{aligned}$$

Applying Lemma 3.1

$$\|y_0-y_\pi\|_{L_{\alpha,\beta}} \le c_1\|d_x\|_{L_{\alpha,\beta}} + c_2\|x_\pi - x_0\|_{L_{\alpha,\beta}}$$

is obtained. This implies (5.3). □

One of the decisive steps for showing quadratic convergence is the following Lipschitz estimate improving (5.2):

**Theorem 5.2** *There is a constant $c_L > 0$ such that*

$$\|v_\pi - v_0\|_2 \le c_L \|\pi\|_2 \tag{5.5}$$

*for all $\pi \in L_\infty \times L_\infty \times C$.*

**Proof.** We outline the main steps of the proof. The first order condition for $v_\pi$ as a solution of $(QS)_\pi$ is

$$0 \le < \tilde{\mathcal{L}}_v(v_\pi, y_\pi),\ v - v_\pi > \quad \forall v \in L_2 \times U^{ad},$$

i.e.

$$\begin{aligned} 0 \quad \le \quad & f'(v_0)(v - v_\pi) + \mathcal{L}_{vv}(v_0, y_0)[v_\pi - v_0, v - v_\pi] \\ & - < d, v - v_\pi > - < y_\pi, (x - x_\pi) - K(B'(x_0)(x - x_\pi) + u - u_\pi) > \end{aligned}$$

for all $x \in L_2, u \in U^{ad}$. Now we insert $x = x_0, v = (x_0, u_0), u = u_0$ and find after exploiting the first order necessary optimality conditions for $v_0$ as a solution for $(P)$

$$\begin{aligned} \mathcal{L}_{vv}(v_0, y_0)[v_\pi - v_0, v_\pi - v_0] \quad & \le \quad - < e, y_0 - y_\pi > - < d, v_0 - v_\pi > \\ & \le \quad \|y_0 - y_\pi\|_2 \, \|e\|_2 + \ \|d\|_2 \, \|v_0 - v_\pi\|_2 \\ & \le \quad c\, \|\pi\|_2^2 \ + \ c\, \|\pi\|_2 \, \|v_0 - v_\pi\|_2 \end{aligned} \tag{5.6}$$

by Lemma 5.1. The difference $\xi = v_\pi - v_0 = (x_\pi - x_0, u_\pi - u_0) = (\xi_x, \xi_u)$ solves $\xi_x - K(B'(x_0)\xi_x + \xi_u) = e$, hence $\xi$ does not satisfy the linearized equation (3.4), where (SSC) applies. Define $\hat{\xi} = (\hat{\xi}_x, \xi_u)$, where $\hat{\xi}_x$ is the solution of $\hat{\xi}_x - K(B'(x_0)\hat{\xi}_x + \xi_u) = 0$. Then $\xi_x = \hat{\xi}_x + \Delta$, and $\|\Delta\|_2 \le c\, \|\pi\|_2$ (apply Lemma 3.1). $(SSC)$ is valid for $\hat{\xi}_x$, hence simple estimations yield

$$\begin{aligned} \mathcal{L}_{vv}(v_0, y_0)[\xi, \xi] \quad & \ge \quad \delta \|\hat{\xi}\|_2^2 \ - 2c\, \|\hat{\xi}\|_2 \, \|\Delta\|_2 \ - c\, \|\Delta\|_2^2 \\ & \ge \quad \delta\, \|\xi\|_2^2 \ - c(\|\xi\|_2 \, \|\Delta\|_2 \ + \|\Delta\|_2^2). \end{aligned} \tag{5.7}$$

Taking into account (5.6) and $\|\Delta\|_2 \le c\, \|\pi\|_2$ we easily find

$$\|\xi\|_2^2 \ \le c(\|\xi\|_2 \, \|\pi\|_2 \ + \|\pi\|_2^2) \le c\, \|\xi\|_2 \, \|\pi\|_2,$$

if $\|\pi\| \le \|\xi\|$. This implies (5.5), if $\|\xi\| \ge \|\pi\|$. Thus (5.5) holds for $c_L := max(1, c)$. □

The estimate (5.5) in the $L_2-$norm is not sufficient for our purposes. However, we are able to show

**Theorem 5.3** *There is a constant $c'_L > 0$ such that*

$$\|v_\pi - v_0\|_\infty \;\le\; c'_L \|\pi\|_{\infty,p} \tag{5.8}$$

*for all $\pi \in L_\infty \times L_\infty \times C$.*

**Proof.** We start with the equation for $x_\pi - x_0$,

$$x_\pi - x_0 - KB'(x_0)(x_\pi - x_0) = e + K(u_\pi - u_0). \tag{5.9}$$

We have $K(u_\pi - u_0) = \tau w$, where $w$ solves the $PDE$ (1.2) with boundary condition $\partial w / \partial n = u_\pi - u_0$. By $L_2-$regularity,

$$\|w\|_{L_2(0,T;H^{3/2-\varepsilon}(\Omega))} \;\le c\, \|u_\pi - u_0\|_2,$$

($\varepsilon > 0$ fixed sufficiently small), hence

$$\|K(u_\pi - u_0)\|_{L_2(0,T;H^{1-\varepsilon}(\Gamma))} \le c\|u_\pi - u_0\|_2. \tag{5.10}$$

Sobolev embedding theorems yield $H^{1-\varepsilon}(\Gamma) \subset L_{\frac{2(n-1)}{n-1-2(1-\varepsilon)}}(\Gamma) = L_{\frac{n-1}{(n-1)/2-(1-\varepsilon)}}(\Gamma) =:$ $L_{p_1}(\Gamma)$. Denote the left hand side of (5.9) by $E$. Thus $\|E\|_{L_{2,p_1}} \le \; \|e\|_{L_{2,p_1}}$ $+ c\|u_\pi - u_0\|_2 \le \; \|e\|_{L_{2,p_1}} + c\|\pi\|_2$ (Theorem 5.2) $\le c\|\pi\|_{L_{2,p_1}}$.

By Lemma 3.1,

$$\|x_\pi - x_0\|_{L_{2,p_1}} \;\le c\, \|E\|_{L_{2,p_1}} \;\le c\, \|\pi\|_{L_{2,p_1}}. \tag{5.11}$$

Invoking the first order necessary conditions for $u_\pi$, $u_\pi = P_{[-1,1]}\{-\lambda^{-1}(K^* y_\pi - d_u)\}$, a simple estimation yields

$$\begin{aligned} \|u_\pi - u_0\|_{L_{2,p_1}} &\le \lambda^{-1}\|K^*\|\, \|y_\pi - y_0\|_{L_{2,p_1}} + \lambda^{-1}\|d_u\|_{L_{2,p_1}} \\ &\le c\, \|x_\pi - x_0\|_{L_{2,p_1}} + c\, \|\pi\|_{L_{2,p_1}} \le c\, \|\pi\|_{L_{2,p_1}} \end{aligned} \tag{5.12}$$

by Lemma 5.1 and (5.11). In this way, we have already extended (5.5) to the $L_2(0,T;L_{p_1}(\Gamma))-$norm performing one step of a bootstrapping argument. Now we continue estimating (5.9) by means of the $L_{p_1}-$regularity of parabolic equations. The solution $w$ can be estimated in the $L_2(0,T;W_{p_1}^{1+1/p_1-\varepsilon}(\Omega))-$norm, hence we have for its trace $\tau w = K(u_\pi - u_0)$

$$\|K(u_\pi - u_0)\|_{L_2(0,T;W_{p_1}^{1-\varepsilon}(\Gamma))} \le c\, \|u_\pi - u_0\, \|_{L_{2,p_1}}.$$

Embedding $W^{1-\varepsilon}_{p_1}(\Gamma) \subset L_{\frac{n-1}{(n-1)/p_1-(1-\varepsilon)}}(\Gamma) = L_{\frac{n-1}{(n-1)/2-2(1-\varepsilon)}}(\Gamma) = L_{p_2}(\Gamma)$

$$\|x_\pi - x_0\|_{L_{2,p_2}} \le c\,\|\pi\|_{L_{2,p_2}}$$

is obtained as above. Proceeding in the same way we arrive after at most $[(n-1)/2]+1$ steps at the case $(n-1)/2 - k(1-\varepsilon) < 0$, while $(n-1)/2 - (k-1)(1-\varepsilon) > 0$ (provided $\varepsilon > 0$ is sufficiently small). Here we end up with the possibility of an estimate in the norm of $L_2(0,T;C(\Gamma))$. However, we use only

$$\begin{aligned}\|K(u_\pi - u_0)\|_{L_{2,p}} &\le c\|K(u_\pi - u_0)\|_{L_2(0,T;W^{1-\varepsilon}_{p_{k-1}}(\Gamma))}\\ &\le c\|\pi\|_{L_{2,p_{k-1}}} \le c\|\pi\|_{L_{2,p}}.\end{aligned}$$

provided that $p_{k-1} \le p$. If $p_{k-1} > p$, then we use the argument

$$\|K(u_\pi - u_0)\|_{L_{2,p}} \le c\|K(u_\pi - u_0)\|_{L_{2,p_{k-1}}} \le c\|\pi\|_{L_{2,p_{k-2}}} \le c\|\pi\|_{L_{2,p}}$$

(note that $p_{k-2} < p$ must hold, as $L_{2,p_{k-2}}$ is still not transformed into $C$).

Thus finally (invoking (5.9) and the optimality conditions for $u_\pi, u_0$)

$$\|v_\pi - v_0\|_{L_{2,p}} \le c\|\pi\|_{L_{2,p}} \tag{5.13}$$

can be derived.

It remains to lift the regularity with respect to the time $t$. From Krasnosel'skiĭ a.o. [9] it is known that a weakly singular integral operator with weak singularity $\alpha \in (0,1)$ maps continuously $L_p(0,T)$ into $L_{p'}(0,T)$, if $1/p' > 1/p + \alpha - 1$. Put $\delta = 1 - \alpha > 0$ and take $\lambda \in (0,1)$. Then $K$ transforms $L_2(0,T;L_p(\Gamma))$ into $L_{\beta_1}(0,T;L_p(\Gamma))$, where $1/\beta_1 = 1/2 - \lambda\delta$. Arguing as in the first part of the proof,

$$\|u_0 - u_\pi\|_{L_{\beta_1,p}} \le c\|\pi\|_{L_{\beta_1,p}} \tag{5.14}$$

is obtained. $K$ transforms $L_{\beta_1}(0,T;L_p(\Gamma))$ into $L_{\beta_2}(0,T;L_p(\Gamma))$, provided that $1/\beta_2 = 1/\beta_1 - \lambda\delta = 1/2 - 2\lambda\delta$. Therefore, the estimate (5.14) can be derived in the norm of $L_{\beta_2}(0,T;L_p(\Gamma))$. Proceeding in this way, (5.14) is seen to hold in the $\|\cdot\|_p$−norm after finitely many steps. (5.8) follows easily, as $K$ transforms $L_p$ into $C$. □

The next two results are standard. We refer to the proofs given by Alt in [1], which simplify considerably for our model problem (P).

**Lemma 5.4** *There is a $C \times L_p \times L_p-$neighbourhood $N_3(w_0)$, such that for all $w \in N_3(w_0)$ the following equivalence holds true: If $w \in N_3(w_0)$, then the solution $\overline{v}_w = (\overline{x}_w, \overline{u}_w)$ is also the unique solution of $(QS)_\pi$ for the following choice of $\pi = (d, e) = (d_x, d_u, e) : d_u = 0$,*

$$\begin{aligned} d_x &= B''(x_0)[\overline{x}_w - x_0, \cdot]K^* y_0 - B''(x_w)[\overline{x}_w - x_w, \cdot]K^* y_w \\ &\quad -(B'(x_w) - B'(x_0))K^* \overline{y}_w \end{aligned} \tag{5.15}$$

$$e = K(B'(x_w)(\overline{x}_w - x_w) - B'(x_0)(\overline{x}_w - x_0) + B(x_w) - B(x_0)). \tag{5.16}$$

**Proof.** We know that $(QS)_\pi$ is a convex problem with a solution determined uniquely by the conditions (5.1), (5.4), and

$$u_\pi = P_{[-1,1]}\{-\lambda^{-1}(K^* y_\pi - d_u)\}. \tag{5.17}$$

Thus it suffices to show that the triplet $(\overline{x}_w, \overline{u}_w, \overline{y}_w)$ fulfils these relations for an appropriate $\pi$ and $y_\pi := \overline{y}_w$. As regards $\overline{x}_w$, it is a solution of

$$\overline{x}_w = K(B'(x_w)(\overline{x}_w - x_w) + B(x_w) + \overline{u}_w).$$

In order to comply with (5.1), it must hold

$$\overline{x}_w = K(B'(x_0)(\overline{x}_w - x_0) + B(x_0) + \overline{u}_w) + e.$$

Subtracting the last equations we end up with (5.16). The adjoint state $\overline{y}_w$ is defined by (4.4). Comparing this with (5.4),

$$\overline{y}_w = \overline{x}_w - q - d_x + B'(x_0)K^* \overline{y}_w + B''(x_0)[\overline{x}_w - x_0, \cdot]K^* y_0,$$

we easily arrive at formula (5.15). Obviously, $\overline{u}_w$ satisfies (4.5) together with (5.17) iff $d_u = 0$. $\square$

**Lemma 5.5** *Define $d$ and $e$ according to (5.15), (5.16). Then for all $w \in N_4(w_0)$*

$$\|e\|_\infty \leq c_T(\|x_0 - x_w\|_\infty^2 + \|x_w - x_0\|_\infty \|\overline{x}_w - x_0\|_\infty) \tag{5.18}$$

$$\begin{aligned} \|d\|_\infty &\leq c_T(\|\overline{y}_w\|_p \|x_0 - x_w\|_\infty^2 + \|\overline{x}_w - x_0\|_\infty(\|x_w - x_0\|_\infty \\ &+ \|y_w - y_0\|_p) + \|x_w - x_0\|_\infty(\|y_w - y_0\|_p + \|\overline{y}_w - y_0\|_p)) \end{aligned} \tag{5.19}$$

*with a certain constant $c_T$ not depending on $w$.*

The proof follows completely analogous to [4] from re–arranging and estimating (5.15) – (5.16).

## 6 Quadratic convergence of the SQP–method

**Theorem 6.1** *There is a $C \times L_p \times L_p$–neighbourhood $N_5(w_0)$, and a positive constant $\nu$ such that for all $w \in N_5(w_0)$ the solution $\overline{v}_w$ of $(QP)_w$ and the corresponding Lagrange multiplier $\overline{y}_w$ satisfy*

$$\|(\overline{v}_w, \overline{y}_w) - (v_0, y_0)\|_{\infty,p} \le \nu \|w - w_0\|^2_{\infty,p}. \tag{6.1}$$

**Proof.** We take at first $N(w_0) \subset N_1(w_0) \cap N_4(w_0)$ such that the radius of $N(w_0)$ is less than 1. According to Corollary 4.2., $\|\overline{x}_w - x_0\|_\infty$ and $\|\overline{y}_w\|_p$ remain bounded by a constant $c > 0$ for all $w \in N(w_0)$. From (5.15) - (5.16)

$$\begin{aligned} max(\|e\|_\infty,\ \|d\|_\infty) \ &\le\ c(\|v_0 - v_w\|^2_\infty\ +\ \|w_0 - w_w\|_{\infty,p}) \\ &\le\ c\|w - w_0\|_{\infty,p} \end{aligned} \tag{6.2}$$

as the diameter of $N(w_0)$ is less than 1.

Thus on $N(w_0)$,

$$\|\pi\|_\infty \ \le\ c\|w - w_0\|_{\infty,p}. \tag{6.3}$$

On the other hand, Lemma 5.4. and Theorem 5.3. yield now $y_\pi = y_{\overline{w}}$ and

$$\|\overline{v}_w - v_0\|_{\infty,p} \ \le\ c\|\pi\|_{\infty,p} \ \le\ c\|w - w_0\|_{\infty,p}. \tag{6.4}$$

Analogously we find

$$\|\overline{y}_w - y_0\|_p \ \le\ c\|w - w_0\|_{\infty,p} \tag{6.5}$$

by Lemma 5.1. and (6.3), (6.4). Inserting (6.4) - (6.5) in (5.18) - (5.19) we obtain

$$\|\pi\|_{\infty,p} \ \le\ c\|\pi\|_\infty \ =\ \|e\|_\infty \ +\ \|d\|_\infty \ \le\ c\|w - w_0\|^2_{\infty,p}, \tag{6.6}$$

implying together with (6.4), (5.3) the relation (6.1) □

Now we reformulate the $SQP$–method and state the result on its local convergence. The $SQP-$method runs as follows.

**(SQP)**: Choose a starting point $w_1 = (v_1, y_1)$. Having $w_k = (v_k, y_k)$ compute $w_{k+1} = (v_{k+1}, y_{k+1})$ to be the solution and the associated Lagrange multiplier of the quadratic optimization problem $(QP)_{w_k}$.

Using Theorem 6.1. it follows now by standard proof techniques that this method converges quadratically to $w_0 = (x_0, u_0, y_0)$, if the starting point $w_1$ is chosen sufficiently close to $w_0$ (see [2], Theorem 5.1.). Let $\nu$ be defined by Theorem 6.1. Denote by $B_\delta(w_0)$ the open ball around $w_0$ with radius $r$ in the sense of $C \times L_p \times L_p$.

**Theorem 6.2** *Suppose that the assumptions (1.4) and $(SSC)$ are satisfied. Choose $p > n+1$ and let $\gamma > 0$ be such that $\delta := \nu\gamma < 1$, and $B_{\gamma\delta}(w_0) \subset N_5(w_0)$. Then for any starting point $w_1 \in B_{\gamma\delta}(w_0)$ the SQP method computes a unique sequence $w_k$ with*

$$\|w_k - w_0\|_{\infty,p} \leq \gamma\, \delta^{2^k - 1},$$

*and $w_k \in B_{\gamma\delta}(w_0)$ for $k \geq 2$.*

The proof is identical to that given in [2].

Thus we have local quadratic convergence of the $SQP$ method in $(x, u, y)$. More precisely, Theorem 6.2 expresses r-quadratic convergence, while Theorem 6.1 shows q-quadratic convergence of the method.

# References


[1] Alt, W.: The Lagrange–Newton method for infinite–dimensional optimization problems. Numer. Funct. Anal. and Optimization 11 (1990), 201–224.

[2] Alt, W.: Sequential quadratic programming in Banach spaces. In W. Oettli and D. Pallaschke, editors, Advances in Optimization, number 382 in Lecture Notes in Economics and Mathematical Systems, 281–301. Springer Verlag, 1992.

[3] Alt, W. and K. Malanowski: The Lagrange–Newton method for nonlinear optimal control problems. Computational Optimization and Applications, 2 (1993), 77–100.

[4] Alt,W., Sontag, R., and F. Tröltzsch: An SQP method for optimal control of weakly singular Hammerstein integral equations. DFG–Schwerpunktprogramm ,,Anwendungsbezogene Optimierung und Steuerung ", Report No. 423 (1993), to appear in Appl. Math. Optimization.

[5] Amann, H.: Parabolic evolution equations with nonlinear boundary conditions. In F.E. Browder, editor, Nonlinear Functional Analysis Proc. Sympos. Pure Math., Vol. 45, Part I, 17–27, 1986.

[6] Goldberg, H. and F. Tröltzsch: Second order optimality conditions for nonlinear parabolic boundary control problems. SIAM J. Contr. Opt. 31 (1993), 1007–1025.

[7] Kelley, C.T. and S.J. Wright: Sequential quadratic programming for certain parameter identification problems. Math. Programming 51 (1991) , 281–305.

[8] Kupfer, S.-F. and Sachs, E.W.: Numerical solution of a nonlinear parabolic control problem by a reduced sqp method. Computational Optimization and Applications 1 (1992), 113–135.

[9] Krasnoselskiĭ, M.A. et al.: Linear operators in spaces of summable functions (in Russian). Nauka, Moscow, 1966.

[10] Levitin, E.S. and B.T. Polyak: Constrained minimization methods. USSR J. Comput. Math. and Math. Phys. 6 (1966), 1–50.

[11] Tröltzsch, F.: On the semigroup approach for the optimal control of semilinear parabolic equations including distributed an boundary control. Z. für Analysis und Anwendungen 8 (1989), 431–443.

[12] Tröltzsch, F.: An SQP Method for Optimal Control of a Nonlinear Heat Equation (1993), to appear in Control Cybernetics.