An SQP Method for the Optimal Control of a Nonlinear Heat Equation

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 the one-dimensional heat equation with nonlinear boundary condition. Sufficient conditions for local quadratic convergence of the method based are discussed.

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

Keywords. Lagrange-Newton method, sequential quadratic programming, optimal control, heat equation, nonlinear boundary condition, control constraints.

1 Introduction

The Lagrange–Newton method is obtained by applying Newton's method or a generalized version of it to find a stationary point of the Lagrangian function associated to a nonlinear optimization problem. If a constraint qualification and a strong second order sufficiency condition are satisfied, the Lagrange-Newton method defines a sequential quadratic programming (SQP) algorithm. It is known since several years that the SQP algorithm exhibits local quadratic convergence in finite-dimensional spaces. 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]. Their results were formulated for Banach or Hilbert spaces and focused mainly on the application to optimal control problems governed by nonlinear ODE's, while Kupfer and Sachs [9] consider the numerical application to parabolic control problems with nonlinear equality constraints. In this context, we also mention Heinkenschloß [6], who applies Newton type methods to nonlinear parabolic control problems without constraints.

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. The aim of this paper is to transfer their convergence result to nonlinear parabolic boundary control problems. It is quite obvious that this is possible by means of the integral equation method. However, we shall develop the theory directly in the context of weak solutions of the parabolic system rather than by reducing the problem to one for a Hammerstein integral equation. In this way, we shed more light on the specific aspects connected with PDE's. Moreover, our presentation is self-contained and may serve as a guide to handle parabolic PDE in domains of higher dimensions.

We shall consider the following optimal control problem.

(P) Minimize

$$f(\theta, u) = \frac{1}{2}\int_0^T (\theta(t,1) - q(t))^2 dt + \frac{\lambda}{2}\int_0^T u(t)^2 dt \tag{1.1}$$

subject to the initial-boundary value problem

$$\begin{aligned} \theta_t(t,x) &= \theta_{xx}(t,x) \\ \theta(0,x) &= 0 \\ \theta_x(t,0) &= 0 \\ \theta_x(t,1) &= b(\theta(t,1)) + u(t) \end{aligned} \tag{1.2}$$

for $x \in (0,1)$, $t \in (0,T]$, and subject to the constraint on the control

$$|u(t)| \le 1 \quad \text{on} \quad [0,T]. \tag{1.3}$$

In (P), the *control* u is looked upon in $L_\infty(0,T)$, while the *state* θ is defined as weak solution of (1.2) (cf. section 2). Moreover, constants $\lambda > 0$, $T > 0$, and functions $b \in C^2(I\!R)$, and $q \in L_\infty(0,T)$ are given.

We assume that b and its derivatives up to the order 2 are uniformly bounded and Lipschitz on $I\!R$: There are constants c_b, c_l such that

$$|b^{(i)}(\theta)| \le c_b, \quad |b^{(i)}(\theta_1) - b^{(i)}(\theta_2)| \le c_l \tag{1.4}$$

for all θ, θ_1, $\theta_2 \in I\!R$, $i = 0,1,2$. These very strong assumptions may be sligthly weakened to local estimates. Moreover, we are able to discuss more general nonlinear functionals than (1.1) and more general nonlinearities in the boundary condition of the heat equation (1.2). However, we confine ourselves to the simplest case containing the typical difficulties for proving convergence of the SQP method. In this way we avoid many technicalities as well as notational complexity. In this paper, we shall use the following

Notations: $L_r = L_r(0,T)$, $1 \le r \le \infty$, $C = C[0,T]$, endowed with natural norms $\|\cdot\|_r$, and $\|\cdot\|_\infty$, respectively. $\|\cdot\|_\infty$ will be used also for the norm in any space of continuous functions. Other norms are denoted by appropriate subscripts. In product spaces $X \times Y$ we introduce the norm by $\|\cdot\|_{X\times Y} = \|\cdot\|_X + \|\cdot\|_Y$. By $(\cdot\,;\,\cdot)$ the inner product of $H = L_2(0,1)$ is denoted, while $(\cdot,\cdot)$ is used for ordered pairs of elements. Moreover, $U^{ad} = \{u \in L_\infty : \|u\|_\infty \leq 1\}$, $Q = (0,T) \times (0,1)$. Traces of functions in $H^1(0,1)$ at $x = 1$ will be indicated by τ, for instance, $\tau\theta = \theta(\cdot,1)$. Within proofs, c denotes a generic constant.

2 Weak solutions and integral equation method

Let us regard at first the linear counterpart of (1.2),

$$\begin{aligned} \theta_t(t,x) &= \theta_{xx}(t,x) \\ \theta(0,x) &= 0 \\ \theta_x(t,0) &= 0 \\ \theta_x(t,1) &= g(t) \end{aligned} \tag{2.1}$$

for $x \in (0,1)$, $t \in (0,T]$, where $g \in L_2(0,T)$. We introduce $V = H^1(0,1)$ and

$$W(0,T) = \{\theta \in L_2(0,T;V) : \theta_t \in \mathrm{L}_2(0,T;V^*)\},$$

where V^* is the dual space to V, and θ_t is the derivative of θ in the sense of vector-valued distributions. A function $\theta \in W(0,T)$ is said to be a *weak solution* of (2.1), if

$$\begin{aligned} (\theta_t(t);v) + (\nabla\theta(t);\nabla v) &= g(t)v(1) \\ \theta(0) &= 0 \end{aligned} \tag{2.2}$$

for almost all $t \in [0,T]$ and all $v \in V$. In (2.2), $(\cdot;\cdot)$ denotes the pairing between V^* and V as well as the natural inner product of $L_2(0,1)$. Note that $\theta_t(t)$ is a function of $L_2(0,T;V^*)$. It is known (cf. [12]) that for each $g \in L_2$ a unique weak solution $\theta \in W(0,T)$ of (2.2) exists. The mapping $g \mapsto \theta$ from L_2 to $W(0,T)$ is continuous. Moreover, we can assume $\theta \in C([0,T],H)$.

In order to gain L_p–estimates we derive a representation of weak solutions by an equivalent integral equation. Regard the Sturm-Liouville eigenvalue problem

$$\begin{aligned} -v''(x) &= cv(x) \\ v'(0) &= v'(1) = 0. \end{aligned} \tag{2.3}$$

The non-negative eigenvalues are $c_o = 0$, $c_n = n^2\pi^2$, $n = 1,2,\ldots$, with normalized eigenfunctions $v_o(x) = 1$, $v_n(x) = \sqrt{2}\cos n\pi x$. The system $\{v_n\}_{n=0}^\infty$ forms an orthonormal basis of H. Expanding the weak solution $\theta(t)$, for each fixed t, into a Fourier series $\theta(t) = \sum_{n=0}^\infty \phi_n(t)v_n$ we end up with the integral representation

$$\theta(t,x) = \sum_{n=0}^\infty v_n(x)v_n(1)\int_0^t e^{-n^2\pi^2(t-s)}g(s)\,ds. \tag{2.4}$$

Lemma 2.1 *θ is a weak solution of (2.1) with boundary data $g \in L_2$, if and only if θ satisfies (2.4). If $g \in L_p$, $p > 2$, then θ is continuous on $\bar{Q}$. There is a constant $c = c(p)$ not depending on g such that for all $g \in L_p$*

$$\|\theta\|_\infty \le c\,\|g\|_p. \tag{2.5}$$

Proof: As (2.4) is standard, we show only (2.5). It holds

$$\begin{array}{rcl} |v_n(x)v_n(1)\int_0^t e^{-n^2\pi^2(t-s)}g(s)\,ds| & \le & 2(\int_0^t e^{-qn^2\pi^2(t-s)}ds)^{1/q}\|g\|_p \\ & \le & c\,n^{-2/q}, \end{array} \tag{2.6}$$

where $1/q + 1/p = 1$. From $p > 2$ we obtain $q < 2$, hence $\sum_{n=1}^\infty n^{-2/q}$ is a convergent majorant for (2.4). The statement follows from the Weierstraß theorem. □

Note that $g(t) := b(\theta(t,1)) + u(t)$ belongs to L_∞ by (1.4). Invoking (2.6) and the Lebesgue dominated convergence theorem it is easy to verify that (2.4) is equivalent to

$$\theta(t,x) = \int_0^t \sum_{n=0}^\infty v_n(x)v_n(1)e^{-n^2\pi^2(t-s)}g(s)\,ds \tag{2.7}$$

for $g \in L_p$, $p > 2$. At $x = 1$,

$$\theta(t,1) = \int_0^t k(t-s)g(s)\,ds, \tag{2.8}$$

where

$$k(t) = \sum_{n=0}^\infty v_n(1)^2 e^{-n^2\pi^2 t} = 1 + 2\sum_{n=1}^\infty e^{-n^2\pi^2 t}. \tag{2.9}$$

The kernel $k(t)$ is weakly singular at $t = 0$, as

$$\sum_{n=0}^\infty e^{-n^2\pi^2 t} \le \int_0^\infty e^{-\pi^2 t x^2}dx = \frac{1}{2\sqrt{\pi t}} \quad (t > 0),$$

hence

$$k(t) \le c\,t^{-1/2}, \tag{2.10}$$

$t \in (0,T]$. For convenience we introduce also the *Green's function*

$$G(x,\xi,t) = \sum_{n=0}^\infty v_n(x)v_n(\xi)e^{-n^2\pi^2 t}.$$

Now we return to the nonlinear equation (1.2). A function $\theta \in W(0,T)$ is said to be a *weak solution* of (1.2), if

$$\begin{array}{rcl} (\theta_t(t); v) + (\nabla\theta(t); \nabla v) & = & (b(\theta(t,1)) + u(t))v(1) \\ \theta(0) & = & 0 \end{array} \tag{2.11}$$

for almost all $t \in [0,T]$ and all $v \in V$.

Lemma 2.2 *If $\theta \in W(0,T)$ is a weak solution of (1.2), then $z(t) := \theta(t,1)$ is a continuous solution of the integral equation*

$$z(t) = \int_0^t k(t-s)(b(z(s)) + u(s))\, ds. \tag{2.12}$$

Conversely,

$$\theta(t,x) = \int_0^t G(x,1,t-s)(b(z(s)) + u(s))\, ds \tag{2.13}$$

is a weak solution of (1.2), if $z \in C[0,T]$ satisfies (2.12).

Proof: Let θ be a weak solution of (1.2) and put $g(t) = b(\theta(t,1)) + u(t)$. Then θ is also a weak solution of (2.1) for this $g(t)$. Moreover, g is bounded and measurable. By Lemma 2.1 and (2.8) we see that $z(t)$ satisfies the integral equation (2.12). The weakly singular integral operator in (2.12) transforms bounded and measurable functions into continuous functions. As $g \in L_\infty$, we have $z \in C[0,T]$.

Conversely, let z solve (2.12) and define θ by (2.13). Setting $x = 1$ in this equation shows $\theta(t,1) = z(t)$, thus

$$\theta(t,x) = \int_0^t G(x,1,t-s)(b(\theta(s,1)) + u(s))\, ds.$$

The last statement is an immediate consequence of Lemma 2.1. □

Lemma 2.3 *For each $u \in L_\infty$ the parabolic initial-boundary value problem (1.2) admits a unique weak solution $\theta \in W(0,T)$.*

Proof: Owing to the strong assumptions (1.4) on b, the integral equation (2.12) has for all $u \in L_\infty$ a unique solution $z \in C[0,T]$. The existence of θ is a conclusion of Lemma 2.2. By $\theta(t,1) = z(t)$ and Lemma 2.1 applied to $g(t) = b(\theta(t,1)) + u(t)$ we obtain immediately the uniqueness of θ. □

Corollary 2.4 *The weak solution θ of (1.2) is continuous on $\bar{Q}$.*

(Apply Lemma 2.1 to $g(t) = b(\theta(t,1)) + u(t)$.)

3 Optimality conditions

It can be shown by standard methods that (P) possesses at least one optimal control u_o. We now fix one optimal control u_o as reference control for all what follows. Let θ_o denote the corresponding state, obtained as solution of (1.2). The following result is well known: Define the *adjoint state* $y_o \in W(0,T)$ to be the weak solution of the *adjoint system*

$$\begin{aligned} -y_t(t,x) &= y_{xx}(t,x) \\ y(T,x) &= 0 \\ y_x(t,0) &= 0 \\ y_x(t,1) &= b'(\theta_o(t,1))y(t,1) + \theta_o(t,1) - q(t) \end{aligned} \tag{3.1}$$

for $x \in (0,1)$, $t \in [0,T)$. The definition of weak solutions of this system as well as investigations concerning existence and uniqueness can be transferred to (1.2) performing the transformation $t' = T - t$. By means of the Green's function we find

$$y(t,x) = \int_t^T G(x,1,s-t)[b'(\theta_o(s,1))y(s,1) + \theta_o(s,1) - q(s)]ds. \qquad (3.2)$$

It is easy to show that this equation has a unique solution at x=1. On account of this, the existence of a unique weak solution to (3.1) is an immediate consequence. The following result is known.

Theorem 3.1 *Let u_o be optimal for (P) with associated state θ_o and adjoint state y_o. Then*

$$\int_0^T (\lambda u_o(t) + y_o(t,1))(u(t) - u_o(t))\, dt \geq 0 \qquad (3.3)$$

for all $u \in U^{ad}$.

For the proof, which can be carried out by the integral equation method, we refer to [5], [13]. Formally, we are able to derive this result by means of the *Lagrange function*

$$\begin{aligned} L(\theta, u, y) \;&=\; f(\theta, u) - \textstyle\int_0^T (\theta_t(t) - \theta_{xx}(t); y(t))\, dt \\ &\quad + \textstyle\int_0^T \{b(\theta(t,1)) + u(t) - \theta_x(t,1)\} y(t,1)\, dt. \end{aligned} \qquad (3.4)$$

The adjoint system (3.1) follows from $L_\theta(\theta_o, u_o, y) = 0 \quad \forall \theta \in W(0,T)$ after an integration by parts. $L_u(\theta_o, u_o, y)(u - u_o) \geq 0 \quad \forall u \in U^{ad}$ gives the variational inequality (3.3). This is very formal, since the differentiability of the mapping $\theta(\cdot,1) \mapsto b(\theta(\cdot,1))$ is quite delicate. Moreover, θ_{xx} would need a further explanation. However, this formal use of L is a reliable guide to establish our SQP method. Performing an integration by parts we shall later make this well defined.

In addition to the first order necessary optimality conditions (3.1), (3.3) we suppose the following *second order sufficient optimality condition.*

(SSC) There is a $\delta > 0$ such that

$$\int_0^T \{1 + y_o(t,1)\, b''(\theta_o(t,1))\} \theta(t,1)^2 dt + \lambda \|u\|_2^2 \geq \delta \|u\|_2^2$$

for all $u \in L_2$ and $\theta \in W(0,T)$ satisfying the *linearized equation*

$$\begin{aligned} \theta_t(t,x) &= \theta_{xx}(t,x) \\ \theta(0,x) &= 0 \\ \theta_x(t,0) &= 0 \\ \theta_x(t,1) &= b'(\theta_o(t,1))\theta(t,1) + u(t). \end{aligned} \qquad (3.5)$$

Formally, the left hand side of (SSC) is the second derivative of the Lagrange function with respect to $v = (\theta, u)$. This derivative can be given a precise meaning as follows. Integrating by parts in (3.4) we obtain $L = \mathcal{L}$, where

$$\begin{aligned}\mathcal{L} &= f(\theta, u) - \textstyle\int_0^T [(\theta_t(t); y(t)) + (\nabla\theta(t); \nabla y(t))]\, dt \\ &\quad + \textstyle\int_0^T (b(\theta(t,1)) + u(t)) y(t,1)\, dt \\ &= f - \mathcal{L}_1 + \mathcal{L}_2.\end{aligned}$$

In what follows, we shall define the Lagrange function in this way. The quadratic functional f is twice continuously Fréchet differentiable on $W(0,T) \times L_2$, and

$$f''(\theta_o, u_o)[v_1, v_2] = \int_0^T \{\theta_1(t,1)\theta_2(t,1) + \lambda u_1(t) u_2(t)\}\, dt,$$

$v_1 = (\theta_1, u_1)$, $v_2 = (\theta_2, u_2)$. $\mathcal{L}_1$ is linear and continuous with respect to θ, hence twice continuously differentiable on $W(0,T) \times L_2$, too (with vanishing second order derivative). In $\mathcal{L}_2$, we shall regard $\theta(t,1)$ as function of $C[0,T]$. In this sense, $\mathcal{L}_2$ is twice continuously differentiable on $C \times L_2$, where

$$\mathcal{L}_{2,vv}(\theta, u, y)[v_1, v_2] = \int_0^T b''(\theta(t,1))\theta_1(t,1)\theta_2(t,1)\, dt.$$

Therefore, we define

$$\mathcal{L}_{vv}(\theta, u, y)[v_1, v_2] = f''(\theta, u)[v_1, v_2] + \mathcal{L}_{2,vv}(\theta, u, y)[v_1, v_2]. \tag{3.6}$$

In this way we are able to expand $\mathcal{L}$ into a Taylor series up to the order 2 with increments belonging to $C[0,T]$ with respect to $\theta(t,1)$.

Corollary 3.2 *It holds*

$$u_o(t) = P_{[-1,1]}\{-\lambda^{-1} y_o(t,1)\}, \tag{3.7}$$

where $P_{[-1,1]}$ *denotes the projection operator from* $\mathbb{R}$ *onto* $[-1,1]$.

(This result is obtained after a standard discussion of (3.3).)

To finish this section we establish an estimate for solutions of the system

$$\begin{aligned}\theta_t(t,x) &= \theta_{xx}(t,x) \\ \theta(0,x) &= 0 \\ \theta_x(t,0) &= 0 \\ \theta_x(t,1) &= \beta(t)\theta(t,1) + u(t)\end{aligned} \tag{3.8}$$

on $[0,T] \times [0,1]$.

Lemma 3.3 *Let* $\beta \in L_\infty(0,T)$ *be given fixed and* θ *be the weak solution of (3.8) associated to* $u \in L_\infty(0,T)$. *There exist constants* $c_{r,2}$ *and* $c_{\infty,r}$ *depending only on* r *and* $\|\beta\|_\infty$ *but not on* $u \in L_\infty$ *such that*

$$\begin{aligned}\|\tau\theta\|_r &\le c_{r,2}\, \|u\|_2 \qquad \forall r \in [2,\infty) \\ \|\tau\theta\|_\infty &\le c_{\infty,r}\, \|u\|_r \qquad \forall r \in (2,\infty].\end{aligned}$$

Proof: According to Lemma 2.2, $z(t) = \theta(t,1)$ is the solution of

$$z(t) = \int_0^t k(t-s)\beta(s)z(s)\,ds + \int_0^t k(t-s)u(s)\,ds, \tag{3.9}$$

hence

$$|z(t)| \le \int_0^t k(t-s)\|\beta\|_\infty |z(s)|ds + \int_0^t k(t-s)|u(s)|ds. \tag{3.10}$$

This is a weakly singular integral inequality for $|z(t)|$ with positive kernel k. Therefore, it holds $|z(t)| \le \phi(t)$, where ϕ is the unique solution of the associated integral equation. Let K denote the weakly singular integral operator generated by $k(t-s)$. K (having the order of singularity 1/2) is known to transform continuously L_r into $L_{r'}$ provided that $1/r' > 1/r - 1/2$. Thus $K : L_2 \to L_r, r < \infty$, and $L_r \to L_\infty$, $r > 2$ (we refer to Krasnoselskiĭ and others [8]). ϕ satisfies

$$\phi = \|\beta\|_\infty K\phi + K\,|u|.$$

The assertion of the Lemma follows now easily. For instance,

$$\begin{aligned} \|z\|_r \le \|\phi\|_r &\le \|(I - \|\beta\|_\infty K)^{-1}\|_{L_r \to L_r} \|K\|_{L_2 \to L_r} \|u\|_2 \\ &\le c\,\|u\|_2. \end{aligned} \tag{3.11}$$

□

4 SQP method and Hölder estimate

Initiating from a starting point (θ_1, u_1, y_1) in $W(0,T) \times \mathrm{L}_\infty \times W(0,T)$ the (full) SQP method generates sequences $\{\theta_n\}$, $\{u_n\}$, $\{y_n\}$ by solving certain quadratic programs. Adopting the notation used by Alt [1], one step of the method can be described as follows:

Let $w := (\theta, u, y)$ be the result of the last iteration, serving as a starting point. To indicate this, we write $w = (\theta_w, u_w, y_w)$. As before, we put $v = (\theta, u), v_w = (\theta_w, u_w)$. The next iterate $\bar{v}_w = (\bar{\theta}_w, \bar{u}_w)$ is obtained as the solution of the problem

$(QP)_w$ Minimize

$$F(v,w) = f'(v_w)(v - v_w) + \frac{1}{2}\mathcal{L}_{vv}(\theta_w, u_w, y_w)[v - v_w, v - v_w] \tag{4.1}$$

subject to the linearized equation

$$\begin{aligned} \theta_t(t,x) &= \theta_{xx}(t,x) \\ \theta(0,x) &= 0 \\ \theta_x(t,0) &= 0 \\ \theta_x(t,1) &= b(\theta_w(t,1)) + b'(\theta_w(t,1))(\theta - \theta_w)(t,1) + u(t) \end{aligned} \tag{4.2}$$

for $x \in (0,1)$, $t \in (0,T]$, and subject to the constraint $u \in U^{ad}$.

In what follows, the bar indicates solutions of $(QP)_w$. In detail, $F(v,w)$ is

$$\begin{aligned} F(v,w) &= \textstyle\int_0^T \{(\theta_w - q)(\theta - \theta_w) + \lambda u_w(u-u_w) \\ &\quad +1/2 b''(\theta_w) y_w (\theta-\theta_w)^2 \\ &\quad +1/2((\theta-\theta_w)^2 + \lambda(u-u_w)^2)\}\, dt, \end{aligned} \tag{4.3}$$

where y_w, θ, and θ_w are to be taken at $x=1$. $(QP)_w$ is a linear-quadratic parabolic boundary control problem. The corresponding theory of optimality conditions is standard. We refer to Lions [11] . The Lagrange function $\tilde{\mathcal{L}}$ of $(QP)_w$ is

$$\begin{aligned} \tilde{\mathcal{L}}(v,y) = \tilde{\mathcal{L}}(\theta,u,y) &= F(v,w) - \textstyle\int_0^T \{(\theta_t(t); y(t)) + (\nabla\theta(t); \nabla y(t))\}\, dt \\ &\quad + \textstyle\int_0^T y(t,1)\{b(\theta_w) + b'(\theta_w)(\theta-\theta_w))(t,1) + u(t)\}\, dt. \end{aligned}$$

From $\tilde{\mathcal{L}}_\theta = 0$ we get the *adjoint system*

$$\begin{aligned} -y_t(t,x) &= y_{xx}(t,x) \\ y(T,x) &= 0 \\ y_x(t,0) &= 0 \\ y_x(t,1) &= [b'(\theta_w)\, y + b''(\theta_w)\, y_w\, (\bar{\theta}_w - \theta_w)](t,1) + \bar{\theta}_w(t,1) - q(t) \end{aligned} \tag{4.4}$$

for the adjoint state $y = \bar{y}_w$ of $(QP)_w$. Completely analogous to (3.7) the relation

$$\bar{u}_w(t) = P_{[-1,1]}\{-\lambda^{-1}\bar{y}_w(t,1)\} \tag{4.5}$$

is derived from $\tilde{\mathcal{L}}_u(u-\bar{u}) \geq 0$. Moreover, the following relation for F is useful:

$$F(v,w_o) \geq \delta\|u-u_o\|_2^2 = F(v_o,w_o) + \delta\|u-u_o\|_2^2. \tag{4.6}$$

($F(v_o,w_o) = 0$ is trivial. The inequality follows from the first order condition and (SSC).)

It is important to note that (SSC) is not stable with respect to L_2-perturbations of the optimal triplet (θ_o, u_o, y_o). However, it remains stable under L_∞-perturbations.

For the following statements it is convenient to introduce the trace of w by $\tau w(t) := (\theta(t,1), u(t), y(t,1))$, thus $\tau w = (\tau\theta, u, \tau y)$.

Lemma 4.1 *There is a constant $c_\rho > 0$ such that*

$$\|\theta\|_\infty \leq c_\rho \tag{4.7}$$

for all $\theta \in W(0,T)$ satisfying the state-equation (4.2) of $(QP)_w$, independently of how θ_w with $\|\theta_w\|_\infty \leq \rho$ and $u \in U^{ad}$ are chosen.

Proof: Let θ satisfy (4.2). Then

$$\theta_x(t,1) - \beta(t)\theta(t,1) = b(\theta_w(t,1)) - \beta(t)\theta_w(t,1) + u(t),$$

where $\beta(t) = b'(\theta_w(t,1))$. We have $\|\beta\|_\infty \le c_b$. Moreover, the right hand side is bounded. From Lemma 3.3 we obtain $\|\tau\theta\|_\infty \le c$. (4.7) follows from Lemma 2.1 with $g := \beta(\theta - \theta_w) + b(\theta_w) + u$. $\square$

In the sequel, the following auxiliary system of equations is frequently referred to.

$$\theta_t(t,x) = \theta_{xx}(t,x), \quad \theta(0,x) = 0, \quad \theta_x(t,0) = 0. \tag{4.8}$$

Lemma 4.2 *There is a $C \times L_\infty \times C$-neighbourhood $N_1(\tau w_o)$ such that for all $w = (\theta_w, u_w, y_w) \in W(0,T) \times L_\infty \times W(0,T)$ with $\tau w \in N_1(\tau w_o)$*

$$\mathcal{L}_{vv}(v_w, y_w)[v,v] \ge \delta/2\|u\|_2^2, \tag{4.9}$$

provided that $v = (\theta, u)$ satisfies (4.8) together with the boundary condition

$$\theta_x(t,1) - b'(\theta_w(t,1))\theta(t,1) = u(t). \tag{4.10}$$

Proof: From Lemma 3.3, the assumption (1.4) on boundedness, and (4.10) we infer

$$\|\tau\theta\|_2 \le c\,\|u\|_2. \tag{4.11}$$

(4.10) is equivalent with,

$$\theta_x - b'(\theta_o)\theta = (b'(\theta_w) - b'(\theta_o))\theta + u.$$

Let ϕ be the solution of (4.8) together with the boundary condition

$$\phi_x - b'(\theta_o)\phi = u.$$

Then at $x = 1$

$$(\theta - \phi)_x - b'(\theta_o)(\theta - \phi) = (b'(\theta_w) - b'(\theta_o))\theta.$$

Applying (1.4) and Lemma 3.3 again,

$$\|\tau(\theta - \phi)\|_2 \le \|\tau(\theta_w - \theta)\|_\infty \|\tau\theta\|_2. \tag{4.12}$$

Re-writing L,

$$\begin{array}{rcl} \mathcal{L}_{vv}(v_w, y_w)[v,v] & = & \int_0^T (b''(\theta_o)y_o\theta^2 + \theta^2)dt + \lambda\|u\|_2^2 \\ & & + \int_0^T (b''(\theta_w)y_w - b''(\theta_o)y_o)\theta^2 dt \\ & = & \mathcal{L}_{vv}(v_o, y_o)[v,v] + R. \end{array} \tag{4.13}$$

In contrary to θ, ϕ satisfies the linearized equation (3.5), where (SSC) holds. Inserting $\theta = \phi + (\theta - \phi)$ in (4.13), (4.9) is easy to show by means of (4.11) and (4.12) for sufficiently small $\|\tau(\theta - \theta_o)\|_\infty \le \|\tau(w - w_o)\|_\infty$. $\square$

Corollary 4.3 *$F(v,w)$ is strictly convex with respect to v on the feasible set of $(QP)_w$, if w satisfies the assumptions of Lemma 4.2.*

Proof: The feasible pairs (θ, u) for $(QP)_w$ satisfy $u \in U^{ad}$, (4.8), and

$$\theta_x(t,1) - b'(\theta_w(t,1))\theta(t,1) - u(t) = b(\theta_w(t,1)) - b'(\theta_w(t,1))\theta_w(t,1).$$

Let $\tilde{\theta}$ be the solution of (4.8) subject to

$$\tilde{\theta}_x(t,1) - b'(\theta_w(t,1))\tilde{\theta}(t,1) = b(\theta_w(t,1)) - b'(\theta_w(t,1))\theta_w(t,1).$$

All feasible states θ can be represented in the form $\theta = \theta_1 + \tilde{\theta}$, where θ_1 fulfils the boundary condition (4.10). Owing to the last result, F is strictly convex with respect to (θ_1, u). This strict convexity is preserved under the shift $\tilde{\theta}$. □

Lemma 4.4 *(Hölder estimate) There is a $C \times L_\infty \times C$-neighbourhood $N_2(\tau w_o)$ with the following properties: $(QP)_w$ admits a unique solution $\bar{v}_w = (\bar{\theta}_w, \bar{u}_w)$ for all $w \in W(0,T) \times L_\infty \times W(0,T)$ having boundary data τw in $N_2(\tau w_o)$. Moreover, there is a constant $c_H > 0$ not depending on w such that*

$$\|\bar{v}_w - v_o\|_{W(0,T)\times L_2} \leq c_H \|\tau(w - w_o)\|_2^{1/2} \tag{4.14}$$

$$\|\tau \bar{y}_w\|_\infty \leq c_H \tag{4.15}$$

for all w mentioned above.

Proof: a) Existence and uniqueness of $\bar{v}_w$ follow by standard methods from Corollary 4.3. Therefore we confine ourselves to showing (4.14), (4.15).

b) *Upper estimate*: We write for short $F(v,w) =: F(v)$. Obviously, $F(\bar{v}_w) \leq F(v)$, where v is taken as $v = (\theta, u_w)$, and θ is the state associated to $u = u_w$. θ is defined by (4.2) for $u = u_w$, hence θ solves (4.8) subject to the boundary condition

$$\theta_x = b(\theta_w) + b'(\theta_w)(\theta - \theta_w) + u_w,$$

while θ_o satisfies the same system with

$$(\theta_o)_x = b(\theta_o) + u_o$$

at $x = 1$. Subtraction yields

$$(\theta - \theta_o)_x - b'(\theta_w)(\theta - \theta_o) = b(\theta_w) - b(\theta_o) + b'(\theta_w)(\theta_o - \theta_w) + u_w - u_o.$$

The L_2-norm of the right hand side of this equation is less or equal than $c\,\|\tau(w - w_o)\|_2$. Lemma 3.3 applies again together with (4.1) to show

$$\|\tau(\theta - \theta_o)\|_2 \leq c\,\|\tau(w - w_o)\|_2,$$

hence

$$\|\tau(\theta - \theta_w)\|_2 \leq \|\tau(\theta - \theta_o)\|_2 + \|\tau(\theta_o - \theta_w)\|_2 \leq c\,\|\tau(w - w_o)\|_2,$$

too. From (4.3) with $u = u_w$ it is easy to see that

$$\begin{array}{rcl} F(\bar{v}_w) \leq F(v) & \leq & c_1\|\tau(\theta - \theta_w)\|_2 + c_2\|\tau(\theta - \theta_w)\|_2^2 \\ & \leq & c_1\|\tau(w - w_o)\|_2 + c_2\|\tau(w - w_o)\|_2^2 \\ & \leq & c\,\|\tau(w - w_o)\|_2 \end{array} \qquad (4.16)$$

provided that $\|\tau(w - w_o)\|_2 \leq 1$.

c) *Lower estimate*: Initiating from (4.1) we write

$$F(\bar{v}_w) = f'(v_w)(\bar{v}_w - v_o) + f'(v_w)(v_o - v_w) + 1/2\mathcal{L}_{vv}(v_w, y_w)[\bar{v}_w \mp v_o - v_w, \bar{v}_w \mp v_o - v_w].$$

Simple calculations yield

$$F(\bar{v}_w) \geq f'(v_w)(\bar{v}_w - v_o) + 1/2\mathcal{L}_{vv}(v_w, y_w)[\bar{v}_w - v_o, \bar{v}_w - v_o] - c\,\|\tau(w - w_o)\|_2 \quad (4.17)$$

for sufficiently small $\|\tau(w - w_o)\|_2 \leq 1$.

First, we show

$$\mathcal{L}_{vv}(v_w, y_w)[\bar{v}_w - v_o, \bar{v}_w - v_o] \geq \delta/2\|\bar{u}_w - u_o\|_2^2 - c\,\|\tau(\theta_w - \theta_o)\|_2. \qquad (4.18)$$

To see this, we subtract the equations defining $\bar{\theta}_w$ and θ_o. $\bar{\theta}_w - \theta_o$ satisfies (4.8) together with

$$(\bar{\theta}_w - \theta_o)_x - b'(\theta_w)(\bar{\theta}_w - \theta_o) - (\bar{u}_w - u_o) = b'(\theta_w)(\theta_o - \theta_w) + b(\theta_w) - b(\theta_o) = \Delta. \quad (4.19)$$

We have $\|\Delta\|_2 \leq c\,\|\tau(\theta_w - \theta_o)\|_2$. Now let ϕ denote the solution of (4.8) subject to

$$\phi_x - b'(\theta_w)\phi - (\bar{u}_w - u_o) = 0,$$

hence ϕ solves (4.19) for $\Delta = 0$. By Lemma 3.3,

$$\|\tau(\bar{\theta}_w - \theta_o - \phi)\|_2 \leq c\,\|\Delta\|_2 \leq c\,\|\tau(\theta_w - \theta_o)\|_2.$$

$\mathcal{L}_{vv}(v_w, y_w)$ is coercive with respect to $(\phi, \bar{u}_w - u_o)$ (Lemma 4.2). Therefore, (4.18) is an easy conclusion (note that $\bar{v}_w - v_o$ is bounded by Lemma 4.1).

Second, we have

$$f'(v_w)(\bar{v}_w - v_o) \geq -c\,\|\tau(\theta_w - \theta_o)\|_2. \qquad (4.20)$$

In fact,

$$\begin{array}{rcl} f'(v_w)(\bar{v}_w - v_o) & = & f'(v_o)(\bar{v}_w - v_o) + (f'(v_w) - f'(v_o))(\bar{v}_w - v_o) \\ & & \geq f'(v_o)(\tilde{v}_w - v_o + (\bar{v}_w - \tilde{v}_w)) - c\,\|\tau(\theta_w - \theta_o)\|_2, \end{array} \qquad (4.21)$$

where $\tilde{v}_w = (\tilde{\theta}_w, \bar{u}_w)$, and $\tilde{\theta}_w - \theta_o$ is the solution of (4.8) subject to

$$(\tilde{\theta}_w - \theta_o)_x - b'(\theta_o)(\tilde{\theta}_w - \theta_o) - (\bar{u}_w - u_o) = 0. \qquad (4.22)$$

By means of the same techniques as before, the L_2-norm of $\bar{\theta}_w - \tilde{\theta}_w$ can be estimated by $c\,\|\tau(\theta_w - \theta_o)\|_2$ (compare (4.19) with (4.22)). Moreover, $f'(v_o)(\tilde{v}_w - v_o) \geq 0$ is known from the theory of first order necessary conditions. Thus, (4.20) is a simple conclusion of (4.21). (4.16), (4.17), and (4.20) imply

$$\|\bar{u}_w - u_o\|_2 \leq c\,\|\tau(w - w_o)\|_2^{1/2}.$$

(4.14) is an immediate consequence. (4.15) follows from (4.4), Lemma 3.3, and Lemma 4.1 formulated for equations backward in time. □

Corollary 4.5 *For $p > 2$ it holds*

$$\|\bar{v}_w - v_o\|_\infty \leq c_H'\,\|\tau(w - w_o)\|_\infty^{1/p} \tag{4.23}$$

for all w having traces τw in $N_2'(\tau w_o) \subset N_2(\tau w_o)$.

Proof: (4.14) implies in particular $\|\bar{u}_w - u_o\|_2 \leq c_H\,\|\tau(w - w_o)\|_2^{1/2}$. From $|\bar{u}_w - u_o| \leq 2$ it is easy to conclude

$$\|\bar{u}_w - u_o\|_p \leq c\,\|\bar{u}_w - u_o\|_2^{2/p} \leq c\,\|\tau(w - w_o)\|_2^{1/p} \leq c\,\|\tau(w - w_o)\|_\infty^{1/p}. \tag{4.24}$$

$\bar{\theta}_w - \theta_o$ solves (4.8) together with

$$(\bar{\theta}_w - \theta_o)_x - b'(\theta_w)(\bar{\theta}_w - \theta_o) = (\bar{u}_w - u_o) + b'(\theta_w)(\theta_o - \theta_w) + b(\theta_w) - b(\theta_o).$$

at $x = 1$. The L_p-norm of the right hand side is less or equal than $\|\bar{u}_w - u_o\|_p + c\,\|\tau(\theta - \theta_o)\|_\infty$. Invoking Lemma 3.3 and (4.24),

$$\begin{aligned} \|\tau(\bar{\theta}_w - \theta_o)\|_\infty &\leq c\,\|\bar{u}_w - u_o\|_p + c\,\|\tau(\theta - \theta_o)\|_\infty \\ &\leq c\,\|\tau(w - w_o)\|_\infty^{1/p}, \end{aligned} \tag{4.25}$$

if additionally $\|\tau(w - w_o)\|_\infty \leq 1$. In the same way we arrive from the adjoint equation (4.4) at

$$\|\tau(\bar{y}_w - y_o)\|_\infty \leq c\,\|\tau(w - w_o)\|_\infty^{1/p}, \tag{4.26}$$

if $\|\tau(w - w_o)\|_\infty \leq 1$. Now we apply the optimality conditions (4.5), (3.7). Thus

$$\begin{aligned} |\bar{u}_w(t) - u_o(t)| &\leq |P_{[-1,1]}\{-\lambda^{-1}\bar{y}_w(t,1)\} - P_{[-1,1]}\{-\lambda^{-1}y_o(t,1)\}| \\ &\leq \lambda^{-1}|\bar{y}_w(t,1) - y_o(t,1)|. \end{aligned} \tag{4.27}$$

Thus we infer from (4.26) that $\|\bar{u}_w - u_o\|_\infty$ satisfies the same estimate as $\|\tau(\bar{\theta}_w - \theta_o)\|_\infty$ in (4.25). (4.23) is now obvious (apply Lemma 2.1 to extend the estimate for $\theta_w - \theta_o$ from $x = 1$ to [0,1]). □

The previous investigations show that u_o, θ_o, y_o, $\bar{u}_w$, $\bar{\theta}_w$ and $\bar{y}_w$ are continuous on their domains. On account of this, the strong distinction between θ, y, and their traces $\tau\theta$, τy is no longer necessary.

In what follows let $p > 2$ be a fixed real number. By means of Corollary 4.5 we finally arrive at

Theorem 4.6 *There is a sufficiently small $C(\bar{Q}) \times C[0,T] \times C(\bar{Q})$-neighbourhood $N_3(w_0)$ and a constant $c''_H > 0$ such that*

$$\|\bar{v}_w - v_o\|_\infty \le c''_H \, \|w - w_o\|_\infty^{1/p} \qquad \forall w \in N_3(w_o). \tag{4.28}$$

We recall that $\|\bar{v}_w - v_o\|_\infty = \|\bar{\theta}_w - \theta_o\|_\infty + \|\bar{u}_w - u_o\|_\infty$, $\|w - w_o\|_\infty = \|\theta_w - \theta_o\|_\infty + \|u_w - u_o\|_\infty + \|y_w - y_o\|_\infty$.

An analogous result is true for the adjoint state:

Corollary 4.7 *It holds*

$$\|\bar{y}_w - y_o\|_\infty \le c_y \, \|w - w_o\|_\infty^{1/p} \qquad \forall w \in N_3(w_o) \tag{4.29}$$

with a certain constant c_y not depending on w.

Proof: From the adjoint equations (3.1) and (4.4) defining y_o and $\bar{y}_w$ we get

$$\begin{aligned} (\bar{y}_w - y_o)_x - b'(\theta_w)(\bar{y}_w - y_o) &= (\bar{u}_w - u_o) + (b'(\theta_w) - b'(\theta_o))y_o + (\bar{\theta}_w - \theta_o) \\ &\quad + b''(\theta_w)y_w((\bar{\theta}_w - \theta_o) + (\theta_0 - \theta_w)) = R \end{aligned}$$

The L_∞-norm of R is less or equal than $c_1\|v_w - v_o\|_\infty + c_2\|\bar{v}_w - v_o\|_\infty$, hence Theorem 4.6 yields

$$\|R\|_\infty \le c\,\|w - w_o\|_\infty^{1/p}.$$

As before, (4.29) is obtained now by the backward variant of Lemma 3.3. □

5 Right hand side perturbations, Lipschitz estimate

Following Alt [1] we consider in this section the close relationship between the stability of $(QP)_w$ and certain perturbations of $(QP)_{w_o}$. We discuss the perturbed problem

$(QS)_\pi$ Minimize the functional

$$F(v, w_o) - d(v - v_o) \tag{5.1}$$

subject to the perturbed initial-boundary value problem

$$\begin{aligned} \theta_t(t,x) &= \theta_{xx}(t,x) \\ \theta(0,x) &= 0 \\ \theta_x(t,0) &= 0 \\ \theta_x(t,1) &= (b(\theta_o) + b'(\theta_o)(\theta - \theta_o))(t,1) + u(t) + e(t), \end{aligned} \tag{5.2}$$

for $x \in (0,1)$, $t \in (0,T]$, and subject to the constraint $u \in U^{ad}$, where $F(v, w_o)$ is defined by (4.1) taken at (v_o, y_o) substituted for (v_w, y_w). Moreover,

$$d(v) := \int_0^T (d_\theta(t)\theta(t,1) + d_u(t)u(t))\, dt, \tag{5.3}$$

is a linear and continuous functional on $W(0,T)\times L_2$. We regard $\theta \in W(0,T), u \in L_2$ (although feasible u are automatically bounded and measurable). The perturbation is the vector-function $\pi = (d_\theta, d_u, e) \in L_\infty(0,T)^3$. Later we shall recognize that d_u can be taken as zero. In the case $\pi = 0$ we verify by means of (4.6) that $v_o = (\theta_o, u_o)$ is the unique solution of $(QS)_0$. It should be underlined that neither d nor e destroy the strict convexity of the functional (5.1) on the feasible domain of $(QS)_\pi$. Owing to this, for each $\pi \in L_\infty^3$ the linear-quadratic boundary control problem $(QS)_\pi$ has a unique solution. We shall denote it by $v_\pi = (\theta_\pi, u_\pi)$ with associated adjoint state y_π.

The *adjoint equation* defining $y = y_\pi$ is

$$\begin{aligned} -y_t(t,x) &= y_{xx}(t,x) \\ y(T,x) &= 0 \\ y_x(t,0) &= 0 \\ y_x(t,1) &= [b'(\theta_o)y + b''(\theta_o)y_o(\theta_\pi - \theta_o) + \theta_\pi](t,1) - q(t) - d_\theta(t), \end{aligned} \tag{5.4}$$

while the necessary conditions for u_π admit the form

$$u_\pi(t) = P_{[-1,1]}\{-\lambda^{-1}(y_\pi(t,1) - d_u(t))\}. \tag{5.5}$$

The system $u \in U^{ad}$, (5.2), (5.4), (5.5) forms the set of necessary and sufficient optimality conditions for $(\theta_\pi, u_\pi, y_\pi)$.

It can be shown along the lines of the preceding section that there is a $L_\infty{}^3$-neighbourhood $N_4(0)$ and a constant $c_h > 0$ such that problem $(QS)_\pi$ admits for all $\pi \in N_4(0)$ a unique solution $v_\pi = (\theta_\pi, u_\pi)$, and

$$\|v_\pi - v_o\|_{W(0,T)\times L_2} \le c_h \|\pi\|_2^{1/2}. \tag{5.6}$$

However, we shall considerably improve this estimate in Theorem 5.2. At first we state the following counterpart of Corollary 4.7.

Lemma 5.1 *There is a constant $c_r = c(r) > 0$, such that for all $2 \le r \le \infty$*

$$\|\tau(y_\pi - y_o)\|_r \le c_r\,(\|\tau(\theta_\pi - \theta_o)\|_r + \|\pi\|_r). \tag{5.7}$$

Proof: Subtraction of the adjoint equations (5.4) and (3.1) shows that $y_\pi - y_o$ satisfies the first three equations of (5.4) together with

$$(y_\pi - y_o)_x - b'(\theta_o)(y_\pi - y_o) = \theta_\pi - \theta_o - d_\theta + b''(\theta_o)y_o(\theta_\pi - \theta_o)$$

at $x = 1$. Applying Lemma 3.3 in its backward version,

$$\|\tau(y_\pi - y_o)\|_r \le c\,\|\tau(\theta_\pi - \theta_o) - d_\theta\|_r \le c\,(\|\tau(\theta_\pi - \theta_o)\|_r + \|\pi\|_r). \qquad \square$$

One of the decisive steps for showing local quadratic convergence of the SQP method is to establish the following Lipschitz estimate.

Theorem 5.2 *(Lipschitz estimate) There is constant $c_L > 0$ such that*

$$\|v_\pi - v_o\|_{W(0,T)\times L_2} \le c_L \|\pi\|_2 \tag{5.8}$$

for all $\pi \in L^3_\infty$.

Proof: The Lagrange function for $(QS)_\pi$ is

$$\begin{aligned}
\tilde{\mathcal{L}}(v, y_\pi) &= F(v, w_o) - d(v - v_o) - \textstyle\int_0^T \{(\theta_t; y_\pi) + (\nabla\theta; \nabla y_\pi)\}\, dt \\
&\quad + \textstyle\int_0^T \{b(\theta_o) + b'(\theta_o)(\theta - \theta_o) + u + e\} y_\pi(t,1)\, dt \\
&= f'(v_o)(v - v_o) + 1/2 \mathcal{L}_{vv}(v_o, y_o)[v - v_o, v - v_o] - d(v - v_o) \\
&\quad - \textstyle\int_0^T \{(\theta_t; y_\pi) + (\nabla\theta; \nabla y_\pi)\}\, dt \\
&\quad + \textstyle\int_0^T \{b(\theta_o) + b'(\theta_o)(\theta - \theta_o) + u + e\} y_\pi(t,1)\, dt.
\end{aligned}$$

The first order necessary conditions for $v_\pi = (\theta_\pi, u_\pi)$ are $\tilde{\mathcal{L}}_v(v_\pi, y_\pi)(v - v_\pi) \ge 0$, hence

$$\begin{aligned}
0 \le\ & f'(v_o)(v_o - v_\pi) - d(v_o - v_\pi) + \mathcal{L}_{vv}(v_o, y_o)[v_\pi - v_o, v_o - v_\pi] \\
& + \textstyle\int_0^T \{-((\theta_o - \theta_\pi)_t; y_\pi) - (\nabla(\theta_o - \theta_\pi); \nabla y_\pi) \\
& + (b'(\theta_o)(\theta_o - \theta_\pi) + u_o - u_\pi) y_\pi(t,1)\}\, dt.
\end{aligned} \tag{5.9}$$

$\theta_\pi - \theta_o$ satisfies

$$((\theta_\pi - \theta_o)_t; y) + (\nabla(\theta_\pi - \theta_o); \nabla y) = (b'(\theta_o)(\theta_\pi - \theta_o) + u_\pi - u_o + e) y. \tag{5.10}$$

Thus the part in the curled brackets equals $e\, y_\pi$ and

$$\mathcal{L}_{vv}(v_o, y_o)[v_\pi - v_o, v_\pi - v_o] \le f'(v_o)(v_o - v_\pi) - d(v_o - v_\pi) + \int_0^T e\, y_\pi\, dt. \tag{5.11}$$

On the other hand, $v_o = (\theta_o, u_o)$ fulfils the first order necessary conditions for (P), in particular $\mathcal{L}_v(v_o, y_o)(v_\pi - v_o) \ge 0$:

$$\begin{aligned}
f'(v_o)(v_\pi - v_o) \quad + \quad & \textstyle\int_0^T \{-((\theta_\pi - \theta_o)_t; y_o) - (\nabla(\theta_\pi - \theta_o); \nabla y_o) \\
& + (b'(\theta_o)(\theta_\pi - \theta_o) + u_\pi - u_o) y_o(t,1)\}\, dt \ge 0.
\end{aligned}$$

Owing to (5.10), the integral part is $-\int_0^T e\, y_o\, dt$. Thus $f'(v_o)(v_o - v_\pi) \le -\int_0^T e\, y_o\, dt$. Inserting this in (5.11),

$$\begin{aligned}
\mathcal{L}_{vv}(v_o, y_o)[v_\pi - v_o, v_\pi - v_o] &\le -d(v_o - v_\pi) - \textstyle\int_0^T e\,(y_o - y_\pi)\, dt \\
&\le \|d\|_2 \|\tau(v_o - v_\pi)\|_2 + \|e\|_2 \|\tau(y_o - y_\pi)\|_2 \\
&\le c\, \|\pi\|_2 \|\tau(v_o - v_\pi)\|_2 + c\, \|\pi\|_2^2
\end{aligned} \tag{5.12}$$

by Lemma 5.1. According to the definition,

$$(\theta_\pi - \theta_o)_x - b'(\theta_o)(\theta_\pi - \theta_o) - (u_\pi - u_o) = e \tag{5.13}$$

at x = 1, hence $v = (\theta_\pi - \theta_o, u_\pi - u_o)$ does not satisfy the linearized system (3.5), where (SSC) applies. Define ϕ as solution of (4.8) subject to

$$\phi_x - b'(\theta_o)\phi - (u_\pi - u_o) = 0. \tag{5.14}$$

Repeating our standard estimation technique we obtain by means of Lemma 3.3 that

$$\|\Delta\|_2 = \|\tau(\theta_\pi - \theta_o) - \tau\phi\|_2 \le c\,\|e\|_2 \le c\,\|\pi\|_2.$$

As (SSC) holds true for $\hat{v} = (\phi, u_\pi - u_o)$, we have $\mathcal{L}_{vv}(v_o, y_o)[\hat{v}, \hat{v}] \ge \delta\|u_\pi - u_o\|_2^2$. We write $\tau\phi = \tau(\theta_\pi - \theta_o + (\phi - (\theta_\pi - \theta_o))) = \tau(\theta_\pi - \theta_o) - \Delta$. After a simple computation,

$$\mathcal{L}_{vv}(v_o, y_o)[v_\pi - v_o, v_\pi - v_o] \ge \delta\|u_\pi - u_o\|_2^2 - c_1\|\tau(\theta_\pi - \theta_o)\|_2\|\Delta\|_2 - c_2\|\Delta\|_2^2.$$

Owing to $\|\Delta\|_2 \le c\,\|\pi\|_2$ and (5.12),

$$\|u_\pi - u_o\|_2^2 \le c(\|\pi\|_2^2 + \|\pi\|_2\|\tau(v_\pi - v_o)\|_2) \tag{5.15}$$

is obtained. Moreover, $\|\tau(\theta_\pi - \theta_o)\|_2 \le c(\|u_\pi - u_o\|_2 + \|e\|_2)$ follows from (5.13) by Lemma 3.3. Inserting this into (5.15) we arrive at

$$\|u_\pi - u_o\|_2^2 \le c(\|\pi\|_2^2 + \|\pi\|_2\|u_\pi - u_o\|_2)$$

(note that $\|e\|_2 \le \|\pi\|_2$). Therefore, $\|u_\pi - u_o\|_2 \le c\,\|\pi\|_2$, if $\|\pi\|_2 \le \|u_\pi - u_o\|_2$. In the opposite case, $\|u_\pi - u_o\|_2 < 1 \cdot \|\pi\|_2$. Thus

$$\|u_\pi - u_o\|_2 \le max(c, 1)\|\pi\|_2. \tag{5.16}$$

(5.8) follows from $\|\theta_\pi - \theta_o\|_{W(0,T)} \le c\,\|u_\pi - u_o\|_2$. □

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

$$\|v_\pi - v_o\|_\infty \le c_L'\|\pi\|_\infty \tag{5.17}$$

for all $\pi \in L_\infty^3$.

Proof: $\theta_\pi - \theta_o$ satisfies (4.8) together with the boundary condition (5.13). Making use of Lemma 3.3,

$$\begin{aligned} \|\tau(\theta_\pi - \theta_o)\|_p \le c_{p,2}(\|u_\pi - u_o\|_2 + \|e\|_2) &\le c_{p,2}\|u_\pi - u_o\|_2 + c\,\|\pi\|_p \\ &\le c\,\|\pi\|_p \end{aligned} \tag{5.18}$$

by (5.8), where $p > 2$ is taken fixed. From Lemma 5.1

$$\|\tau(y_\pi - y_o)\|_p \le c\,\|\pi\|_p$$

is found. Proceeding as in the proof of Corollary 4.5,

$$|u_\pi(t) - u_o(t)| \le \lambda^{-1}|y_\pi(t, 1) - y_o(t, 1)| + \lambda^{-1}|d_u(t)|$$

follows from (3.7), (5.5). Thus

$$\|u_\pi - u_o\|_p \le c(\|\tau(y_\pi - y_o)\|_p + \|\pi\|_p) \le c\,\|\pi\|_p. \tag{5.19}$$

Now we repeat this procedure, beginning with the estimate

$$\begin{aligned}\|\tau(\theta_\pi - \theta_o)\|_\infty &\le c_{\infty,p}(\|u_\pi - u_o\|_p + \|e\|_p) \\ &\le c\,\|\pi\|_p \le c\,\|\pi\|_\infty,\end{aligned}$$

instead of (5.18). We end up with (5.19) for $p = \infty$. The final outcome is

$$\|u_\pi - u_o\|_\infty \le c\,\|\pi\|_\infty,$$

implying $\|\theta_\pi - \theta_o\|_\infty \le c\,\|\pi\|_\infty$ by Lemma 2.1. Therefore, (5.17) holds true. □

The next statement links the problem $(QS)_\pi$ with $(QP)_w$. It turns out that $\bar v_w = (\bar\theta_w, \bar u_w)$ is the solution of $(QS)_\pi$ for an appropriate choice of π:

Lemma 5.4 *For all $w \in W(0,T) \times L_\infty \times W(0,T)$ with $\tau w \in N_2(\tau w_o)$ the following equivalence holds true: The solution $\bar v_w$ of $(QP)_w$ is also the unique solution of $(QS)_\pi$ for the following choice of $\pi = (d_u, d_\theta, e)$: $d_u = 0$,*

$$\begin{aligned} d_\theta &= b''(\theta_o)y_o(\bar\theta_w - \theta_o) - b''(\theta_w)y_w(\bar\theta_w - \theta_w) - \bar y_w(b'(\theta_w) - b'(\theta_o)) && (5.20)\\ e &= b(\theta_w) - b(\theta_o) + b'(\theta_w)(\bar\theta_w - \theta_w) - b'(\theta_o)(\bar\theta_w - \theta_o). && (5.21)\end{aligned}$$

Proof: We know that $(QS)_\pi$ has a unique solution satisfying the necessary and sufficient conditions (5.2), (5.4), and (5.5). Thus it suffices to show that $\bar v_w$ fulfils these relations for a suitable π and adjoint state $y_\pi := \bar y_w$. As regards $\bar\theta_w$, it is a solution of (4.8) subject to the boundary conditions

$$(\bar\theta_w)_x - b'(\theta_w)\bar\theta_w = b(\theta_w) - b'(\theta_w)\theta_w + \bar u_w.$$

In order to comply with the constraints of $(QS)_\pi$ with control $\bar u_w$, $\bar\theta_w$ must satisfy

$$(\bar\theta_w)_x = b(\theta_o) + b'(\theta_o)(\bar\theta_w - \theta_o) + \bar u_w + e.$$

Comparing $(\bar\theta_w)_x$ in the last two equations,

$$b(\theta_o) + b'(\theta_o)(\bar\theta_w - \theta_o) + e = b'(\theta_w)\bar\theta_w + b(\theta_w) - b'(\theta_w)\theta_w$$

is obtained, being equivalent with (5.21).

$\bar y_w$ satisfies the boundary condition

$$(\bar y_w)_x = b'(\theta_w)\bar y_w + \bar\theta_w - q + b''(\theta_w)y_w(\bar\theta_w - \theta_w).$$

To be the adjoint state for $(QS)_\pi$ with optimal state $\theta_\pi = \bar\theta_w$, $\bar y_w$ must solve

$$(\bar y_w)_x = b'(\theta_o)\bar y_w - d_\theta + \bar\theta_w - q + b''(\theta_o)y_o(\bar\theta_w - \theta_o).$$

Subtracting these equations, we obtain the form (5.20) for d_θ. Clearly, $\bar u_w$ satisfies (5.5) together with (4.5) for $y_\pi = \bar y_w$ iff $d_u = 0$. □

The next result shows that e and d vanish of the order two.

Lemma 5.5 *Define d and e according to (5.20), (5.21). Then there exists a constant $c_T > 0$ such that*

$$\begin{aligned} \|e\|_\infty &\le c_T(\|\theta_w - \theta_o\|_\infty^2 + \|\theta_w - \theta_o\|_\infty \|\bar\theta_w - \theta_o\|_\infty) && (5.22) \\ \|d\|_\infty &\le c_T(\|\bar y_w\|_\infty \|\theta_w - \theta_o\|_\infty^2 + \|\bar\theta_w - \theta_o\|_\infty(\|\theta_w - \theta_o\|_\infty + \|y_w - y_o\|_\infty) \\ &\quad + \|\theta_w - \theta_o\|_\infty(\|y_w - y_o\|_\infty + \|\bar y_w - y_o\|_\infty)) && (5.23) \end{aligned}$$

for all w satisfying the assumptions of Lemma 5.4 .

Proof: (5.22) and (5.23) can be derived by simple estimates from (5.20), (5.21). We show (5.23). Re-writing the expression for d_θ,

$$\begin{aligned} d_\theta &= \bar y_w(b'(\theta_o) - b'(\theta_w) + b''(\theta_w)(\theta_w - \theta_o)) \\ &\quad + (b''(\theta_w)y_w - b''(\theta_o)y_o)(\theta_o - \bar\theta_w) + b''(\theta_w)(y_w - \bar y_w)(\theta_w - \theta_o). \end{aligned}$$

Therefore,

$$\begin{aligned} \|d_\theta\|_\infty &\le c_1\|\bar y_w\|_\infty \|\theta_w - \theta_o\|_\infty^2 + c_2(\|\theta_w - \theta_o\|_\infty + \|y_w - y_o\|_\infty)\|\bar\theta_w - \theta_o\|_\infty \\ &\quad + c_3\|\theta_w - \theta_o\|_\infty(\|y_w - y_o\|_\infty + \|\bar y_w - y_o\|_\infty). \end{aligned}$$

implying (5.23). (5.22) can be derived analogously. □

6 Quadratic convergence of the SQP–method

Theorem 6.1 *There is a $C(\bar Q) \times L_\infty \times C(\bar Q)$–neighbourhood $N_5(w_o)$ and a constant $\nu > 0$ such that for all $w \in W(0,T) \times L_\infty \times W(0,T)$ belonging to $N_5(w_o)$ the solution $\bar v_w$ of $(QP)_w$ satisfies together with the associated Lagrange multiplier $\bar y_w$*

$$\|(\bar v_w, \bar y_w) - (v_o, y_o)\|_\infty \le \nu \|w - w_o\|_\infty^2. \tag{6.1}$$

Proof: We take $N_5(w_o) \subset N_3(w_o)$ such that $w \in N_5(w_o)$ implies also $\tau w \in N_2(\tau w_o)$. Moreover we assume the diameter of $N_5(w_o)$ to be less than one. Let $w \in N_5(w_o)$ be given. From Lemma 4.4 we obtain the existence of a unique solution $\bar v_w$ of $(QP)_w$ with associated adjoint state $\bar y_w$. Define $\pi = (d, e)$ according to (5.20), (5.21). Owing to Theorem 4.6 and Corollary 4.7, $\|\bar v_w - v_o\|_\infty$ and $\|\bar y_w\|_\infty$ remain uniformly bounded for all $w \in N_5(w_o)$. From (5.22), (5.23)

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

as the diameter of $N_5(w_o)$ is less than one. Thus on $N_5(w_o)$

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

On the other hand, Lemma 5.4 ensures that $\bar{v}_w$ is a solution of $(QS)_\pi$ with Lagrange multiplier $y_\pi = \bar{y}_w$. Therefore, Theorem 5.3 and (6.3) yield

$$\|\bar{v}_w - v_o\|_\infty \le c\,\|\pi\|_\infty \le c\,\|w - w_o\|_\infty . \tag{6.4}$$

Analogously we find

$$\|\bar{y}_w - y_o\|_\infty \le c\,\|w - w_o\|_\infty \tag{6.5}$$

by Lemma 5.1 and (6.3), (6.4). Note that an L_∞-estimate for $\tau(\bar{y}_w - y_o)$ implies an estimate of $\bar{y}_w - y_o$ in the norm of $C(\bar{Q})$ (cf. Lemma 2.1). Inserting (6.4), (6.5) in (5.22), (5.23) we end up with

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

implying together with (6.4), (6.5) the estimate (6.1). □

Now we reformulate the SQP method and state a 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 techniques that the SQP method converges quadratically to $w_o = (\theta_o, u_o, y_o)$, if the starting point $w_1 = (\theta_1, u_1, y_1)$ is choosen sufficiently close to w_o:

Let ν be defined by Theorem 6.1. Let $B_{\gamma\delta}(w_o)$ denote the ball of $C(\bar{Q}) \times L_\infty \times C(\bar{Q})$ with radius $\gamma\delta$ around w_o.

Theorem 6.2 *Suppose that Assumptions (1.4) and (SSC) are satisfied. Let $\gamma > o$ be such that $\delta := \nu\gamma < 1$, and $B_{\gamma\delta}(w_o) \subset N_5(w_o)$. Then for any starting point $w_1 \in W(0,T) \times L_\infty \times W(0,T)$ belonging to $B_{\gamma\delta}(w_o)$ the SQP method computes a unique sequence w_k with*

$$\|w_k - w_o\|_\infty \le \gamma\,\delta^{2^k - 1},$$

$w_k \in W(0,T) \times L_\infty \times W(0,T)$, and $w_k \in B_{\gamma\delta}(w_o)$ for $k \ge 2$.

Proof: We follow the proof by Alt [2]. Theorem 6.1 implies

$$\|w_2 - w_o\|_\infty \le \nu \|w_1 - w_o\|_\infty^2 \le \nu\gamma^2\delta^2 = \gamma\delta^3 = \gamma\delta^{2^k - 1}.$$

Since $\delta < 1$, we have $w_2 \in B_{\gamma\delta}(w_o)$. Thus the theorem holds true for $k = 2$. By induction,

$$\|w_{k+1} - w_o\|_\infty \le \nu \|w_k - w_o\|_\infty^2 \le \nu\gamma^2\delta^{2^{k+1} - 2} = \gamma\delta^{2^{k+1} - 1}.$$

Since $\delta^{2^{k+1} - 1} < \delta$, w_{k+1} belongs to $B_{\gamma\delta}(w_o)$, too. This completes the proof. □

In this way, we have shown local quadratic convergence of the method.

References

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

[2] Alt, W.: Sequential Quadratic Programming in Banach Spaces. In: Oettli, W. and Pallaschke, D. (Eds.), Advances in Optimization. Lecture Notes in Economics and Mathematical Systems Vol. 382, Springer Verlag 1992, 281–301.

[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] Goldberg, H. and F. Tröltzsch: Second order optimality conditions for a class of control problems governed by non–linear integral equations with application to parabolic boundary control. Optimization 20 (1989), 687-698

[6] Heinkenschloß, M.: Numerical Solution of a Semilinear Parabolic Control Problem. Universität Trier, Report No. 92–24, 1992.

[7] Kelley, C.T. and S.J. Wright: Sequential quadratic programming for certain parameter identification problems. Preprint, 1990.

[8] Krasnoselskiĭ, M. A. and others: Linear operators in spaces of summable functions (in Russian). Nauka, Moscow 1966.

[9] 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.

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

[11] Lions, J. L.: Contrôle optimal de systèmes gouvernès par des équations aux dérivées partielles. Dunod, Paris 1968.

[12] Lions, J.L. and E. Magenes: Problèmes aux limites non homogènes et applications, vol. 1–3, Dunod, Paris 1968.

[13] Tröltzsch, F.: Optimality conditions for parabolic control problems and applications. Teubner–Texte zur Mathematik, vol. 62, Teubner–Verlag, Leipzig 1984.