Control and Cybernetics

vol. **23** (1994) No. 1/2

Local convergence of the Lagrange–Newton method with applications to optimal control

by

Walter Alt

Mathematisches Institut,
Universität Bayreuth,
D-95440 Bayreuth.

This paper investigates local convergence properties of the Lagrange Newton method for optimization problems in Banach spaces. Explicit estimates for the radius of convergence are derived. The results are applied to optimal control problems.
Keywords: Lagrange–Newton method, sequential quadratic programming, infinite–dimensional optimization.

1. Introduction

Sequential quadratic programming (SQP) methods are very efficient algorithms to solve nonlinear programming problems. These methods can be obtained by applying Newton's method to find a stationary point of the Lagrangian function, and are therefore also referred to as the *Lagrange–Newton* method. The theory of local convergence of SQP methods is fairly well developed for finite dimensional optimization problems (compare Fletcher 1987, Stoer 1985).

The SQP method resulting from the Lagrange–Newton method can be easily extended to infinite–dimensional optimization problems. For state constrained optimal control problems this method has been successfully implemented by Machielsen 1987. However, he does not give any results on convergence. For optimization problems in Hilbert spaces with convex constraints an extension of Newton's method can be found in Levitin, Polyak 1966. Kelley, Wright 1990, analyzed an SQP method for equality constrained optimization problems in Hilbert spaces. In this special case a stationary point of the Lagrange function is defined by a system of equations only, and therefore the Lagrange–Newton method is equivalent to the ordinary Newton method to solve this system.

For finite dimensional optimization problems the usual assumptions to show local quadratic convergence of the Lagrange–Newton method are linear independence of the gradients of binding constraints, a second–order sufficient optimality condition, and strict complementary slackness (compare Stoer 1985), where the last two assumptions can be replaced by a strong second-order optimality condition, Fletcher 1987. It is well-known (Stoer 1985) that these are the standard assumptions for stability and sensitivity analysis of nonlinear programming problems (compare Fiacco 1983). This indicates that there is close relationship between stability of the problem to be solved and convergence of the Lagrange–Newton method.

Instead of the original problem the Lagrange–Newton method solves a special quadratic optimization problem, which is a second-order approximation of the original problem and has the same solution as the original problem, if the standard assumptions mentioned above are satisfied. If we call this problem the reference problem, then the relationship to stability of optimization problems under perturbations arises from the fact that the Lagrange–Newton iteration defines a sequence of quadratic optimization problems, which can be interpreted as perturbations of the reference problem. Therefore, suitable stability results for perturbed optimization problems can be used to show local convergence of the Lagrange–Newton method. It seems that the first result of this type has been obtained in Robinson 1974, where a stability theorem for stationary points is used in the finite dimensional case to obtain rates of convergence for SQP methods. In Robinson 1980, Robinson proved a more general implicit-function theorem for strongly regular generalized equations. Applications to a Newton method for generalized equations can be found in Robinson 1983.

Following the approach of Robinson 1974, and using the implicit-function theorem of Robinson 1980, local convergence of the Lagrange–Newton method for infinite-dimensional optimization problems with nonlinear equality and inequality constraints has been invesitgated in Alt 1990. In Alt 1991, Alt 1992, the convergence theory has been extended to more general classes of optimization problems. This theory can be applied to various optimization problems. For finite dimensional problems one obtains the well-known convergence results (see Section 3 of Alt 1990 and Example 6.1 of Alt 1992). For parameter estimation problems the assumptions used in Ito, Kunisch 1992 to investigate stability of solutions to such problems are sufficient to apply the convergence theory (see Example 6.4 in Alt 1992).

In optimal control the convergence theory can only be applied to a restricted class of problems, where the cost functional is quadratic and the constraints are linear with respect to the control variable (see Section 5 of Alt 1990 and Examples 6.5 and 6.6 of Alt 1992). The reason is that, typically, optimal control problems nonlinear in the control variable are differentiable in the L^∞-norm, while the required second-order optimality condition holds only in the weaker L^2-norm. This fact is known as *two-norm discrepancy* (see e.g. Maurer 1981). Two-norm discrepancy creates serious difficulties in stability analysis of infinite dimensional optimization problems. We refer to Malanowski 1993B, Malanowski 1994, Malanowski 1993A, where stability results for such problems can be found. In the particular situation of optimal control problems with no state constraints two-norm discrepancy can be compensated for by regularity results associated with the necessary conditions and the solution for the optimal control problem. The first results in this area were obtained by Hager 1990, who analysed convergence for multiplier methods. Local convergence results for the Lagrange–Newton method applied to nonlinear optimal control problems subject to control constraints are derived in Alt, Malanowski 1993.

In Alt, Sontag, Tröltzsch 1993 an SQP method for optimal control problems governed by a weakly singular Hammerstein integral equation is considered. The convergence theory for these problems requires even a four–norm technique.

In the present paper we investigate the same class of optimization problems as in Alt, Malanowski 1993. In Section 2 we prove local convergence of the Lagrange–Newton method adapting the implicit–function theorem of Robinson 1980 to the special situation considered here. In this way we can derive an explicit bound for the radius of convergence of the method. In Section 3 this result is applied to a class of control problems with nonlinear cost, linear state equation and convex control constraints. Possible applications to perturbed and discretized control problems are indicated.

Notation: The Fréchet derivative of a map f is denoted by f', the partial Fréchet derivative with respect to the variable x is denoted by a subscript x, e.g. f_x. By $B_X(x,r)$ we denote the closed ball with radius r around x in the space X. B_X denotes the closed unit ball in the normed space X, and 0_X denotes the zero element of X. For two Banach spaces Z and Y, $L(Z,Y)$ will denote the space of linear continuous mappings from Z into Y supplied with the usual norm denoted by $\|\cdot\|_{Z\to Y}$.

2. The Lagrange–Newton method

We consider optimization problems in Banach spaces which can formally be described as follows. Let Z, Y be Banach spaces, Ω an open subset of Z, $C \subset Z$, $K \subset Y$ closed convex sets. Further let $f\colon \Omega \to \mathbb{R}$, $g\colon \Omega \to Y$. Then we consider the optimization problem

$$\text{(P)} \qquad \text{Minimize } f(z) \qquad \text{subject to} \quad z \in C\,,\ g(z) \in K\,.$$

Let $V \subset Z^*$ and $\Lambda \subset Y^*$ be subspaces with dense and continuous embeddings, and denote

$$\begin{aligned} X &:= V \times Y\,, & \|(v,y)\|_X &:= \max\{\,\|v\|_V, \|y\|_Y\}\,,\\ W &:= Z \times \Lambda\,, & \|(z,\lambda)\|_W &:= \max\{\,\|z\|_Z, \|\lambda\|_\Lambda\}\,. \end{aligned}$$

We assume that

(A1) There exists a (local) solution $\tilde{z}$ of (P).

(A2) The mappings f, g are two times Fréchet differentiable on Ω and

$$\begin{aligned} &f'(z) \in V\,, && \lambda\, g'(z) \in V\,,\\ &f''(z) \in L(Z,V)\,, && \lambda\, g''(z) \in L(Z,V) \end{aligned}$$

for all $z \in \Omega$ and all $\lambda \in \Lambda$. Moreover, there exist constants $r_1 > 0$ and $c_L > 0$ such that

$$\begin{aligned}
&\|f'(z_1) - f'(z_2)\|_V \le c_L \|z_1 - z_2\|_Z \,, \\
&\|(f''(z_1) - f''(z_2))(z)\|_V \le c_L \|z_1 - z_2\|_Z \|z\|_Z \,, \\
&\|g'(z_1) - g'(z_2)\|_{Z\to Y} \le c_L \|z_1 - z_2\|_Z \,, \\
&\|\lambda(g'(z_1) - g'(z_2))\|_V \le c_L \|\lambda\|_\Lambda \|z_1 - z_2\|_Z \,, \\
&\|\lambda(g''(z_1) - g''(z_2))\|_{Z\to V} \le c_L \|\lambda\|_\Lambda \|z_1 - z_2\|_Z \,, \\
&\|(\lambda_1 - \lambda_2)(g''(z))]\|_{Z\to V} \le c_L \|\lambda_1 - \lambda_2\|_\Lambda
\end{aligned}$$

for all $z, z_1, z_2 \in B_Z(\tilde{z}, r_1)$ and all $\lambda, \lambda_1, \lambda_2 \in \Lambda$.

REMARK. If Z and Y are Hilbert spaces, $V = Z^*$, and $\Lambda = Y^*$, then Assumption (A2) is satisfied, if f'' and g'' are Lipschitz on $B_Z(\tilde{z}, r_1)$.

A continuous linear functional $\lambda \in Y^*$ is called a *Lagrange multiplier* for $z \in Z$, if

$$\begin{aligned}
(f'(z) - \lambda g'(z))(c - z) &\ge 0 \quad \text{for all} \quad c \in C, \\
\lambda(y - g(z)) &\ge 0 \quad \text{for all} \quad y \in K.
\end{aligned}$$

For $z \in Z$, $\lambda \in Y^*$ the *Lagrangian function* is defined by

$$\mathcal{L}(z, \lambda) = f(z) - \lambda(g(z)).$$

The following sequential quadratic programming method is a straightforward extension of Wilson's method (see Robinson 1974, Wilson 1963) to the infinite–dimensional Problem (P).

(LNM): *Having* (z_k, λ_k), *compute* z_{k+1} *to be the solution the quadratic optimization problem*

$$\begin{aligned}
(\mathrm{QP})_k \qquad &\mathrm{Min}_{z\in Z} \quad f'(z_k)(z - z_k) + \tfrac{1}{2}\mathcal{L}_{zz}(z_k, \lambda_k)(z - z_k, z - z_k) \\
&\text{subject to} \quad z \in C\,,\ g(z_k) + g'(z_k)(z - z_k) \in K\,,
\end{aligned}$$

and let λ_{k+1} be an associated Lagrange multiplier.

This sequential quadratic programming method can be obtained by applying Newton's method to find a stationary point of the Lagrangian function, and is therefore referred to as the *Lagrange–Newton* method (compare Fletcher 1987, Alt 1990).

In the following we want to give sufficient conditions such that the Lagrange–Newton method is locally well–defined and that the sequence of iterates (z_k, λ_k), $k = 1, 2, \ldots$, converges locally to $(\tilde{z}, \tilde{\lambda})$, where $\tilde{\lambda}$ is a Lagrange multiplier associated with $\tilde{z}$. As in Robinson 1974 and Alt 1990 we consider the Problems $(\mathrm{QP})_k$ as perturbed optimization problems depending on the parameter (z_k, λ_k), and we use stability results for such problems in the convergence analysis.

Let $w = (z_w, \lambda_w) \in W$. Then we get a family of perturbed quadratic optimization problems

$$(\mathrm{QP})_w \qquad \mathrm{Min}_{z \in Z} \quad F(z,w) = f'(z_w)(z - z_w) + \tfrac{1}{2}\mathcal{L}_{zz}(w)(z - z_w, z - z_w)$$
$$\text{subject to} \quad z \in C\,,\ G(z,w) = g(z_w) + g'(z_w)(z - z_w) \in K\,,$$

with $\tilde{w} = (\tilde{z}, \tilde{\lambda})$ defining the *reference problem*. Now define $w_k = (z_k, \lambda_k)$. Then the Problems $(\mathrm{QP})_k$ and $(\mathrm{QP})_{w_k}$ are identical. Therefore, the Problems $(\mathrm{QP})_k$ can be interpreted as special perturbations of the Problem $(\mathrm{QP})_{\tilde{w}}$.

REMARK. Let $w = (z_w, \lambda_w) \in W$. Suppose $\bar{z}_w$ is a solution to $(\mathrm{QP})_w$. Then a continuous linear functional $\lambda \in Y^*$ is a Lagrange multiplier for $\bar{z}_w$, if

$$(f'(z_w) + \mathcal{L}_{zz}(z_w, \lambda_w)(\bar{z}_w - z_w) - \lambda\, g'(z_w))(c - \bar{z}_w) \geq 0 \tag{2.1}$$

for all $c \in C$, and

$$\lambda(y - g(z_w) - g'(z_w)(\bar{z}_w - z_w)) \geq 0 \tag{2.2}$$

for all $y \in K$. The pair $(\bar{z}_w, \lambda)$ is called a Kuhn–Tucker point for $(\mathrm{QP})_w$.

It is known from results of Robinson 1980, Robinson 1987 (compare also Alt 1990, Malanowski 1992) that the stability of solutions and Lagrange multipliers of the Problems $(\mathrm{QP})_w$ with respect to the parameter w is closely related to stability with respect to the parameter $p = (a, b) \in X$ of the linear–quadratic optimization problem

$$(\mathrm{QS})_p \qquad \mathrm{Min}_{z \in Z} \quad F(z, \tilde{w}) - a(z - \tilde{z})$$
$$\text{subject to} \quad z \in C\,,\ G(z, \tilde{z}) - b \in K\,.$$

We now introduce the assumptions required for the convergence analysis. The first assumption concerns constraint regularity (see Robinson 1976).

(A3) The local solution $\tilde{z}$ is a regular point of the system

$$z \in C\,, \quad g(z) \in K\,,$$

i.e., $0 \in \operatorname{int}\{\, g(\tilde{z}) + g'(\tilde{z})(z - \tilde{z}) - y \mid z \in C,\ y \in K\,\}$.

Assumption (A3) implies the existence of a Lagrange multiplier $\tilde{\lambda} \in Y^*$ associated with $\tilde{z}$. We assume higher regularity of $\tilde{\lambda}$, namely

(A4) $\tilde{\lambda} \in \Lambda$.

As in Alt 1990, Alt 1991 and Malanowski 1992, Malanowski 1993B, our results are based on an implicit–function theorem for generalized equations (Theorem 2.1 of Robinson 1980). This theorem requires Lipschitz continuity with respect to the parameter p of the solutions z_p and the Lagrange multipliers λ_p of the Problems $(\mathrm{QS})_p$. Therefore, we assume

(A5) There exist a ball $B_X(0, \sigma)$ of radius $\sigma > 0$ around 0 and a constant $c_Q > 0$ such that for any $p \in B_X(0, \sigma)$ Problem $(\mathrm{QS})_p$ has a unique solution z_p and a unique associated Lagrange multiplier λ_p with $(z_p, \lambda_p) \in W$ and

$$\|z_p - z_q\|_Z\,,\ \|\lambda_p - \lambda_q\|_\Lambda \leq c_Q\, \|p - q\|_X$$

for all $p, q \in B_X(0, \sigma)$.

REMARK. Assumption (A5) especially requires, that $z_0 = \tilde{z}$ is the unique solution of $(\mathrm{QS})_0$ and $\lambda_0 = \tilde{\lambda}$ is the unique associated Lagrange multiplier. Moreover, Assumption (A5) requires higher regularity of the Lagrange multipliers, since it assumes that the multipliers belong the space Λ.

By Assumption (A5) we can define a Lipschitz continuous function of p

$$\mathcal{S}_Q: B_X(0,\sigma) \to W\,, \quad p \mapsto (z_p, \lambda_p)\,,$$

where z_p is the unique solution of $(\mathrm{QS})_p$ and λ_p is the unique associated Lagrange multiplier.

Define $W_0 = \Omega \times \Lambda$. In the same way as in the proof of Theorem 2.1 in Robinson 1980, for $w = (z_w, \lambda_w)$, $\xi = (z, \lambda) \in X$ we define the function

$$\ell: W_0 \times W_0 \to X\,, \quad \ell(\xi, w) = (\,\ell_1(\xi, w), \ell_2(\xi, w)\,)\,,$$

where

$$\begin{aligned}\ell_1(\xi, w) &= f'(\tilde{z}) - \lambda\, g'(\tilde{z}) + \mathcal{L}_{zz}(\tilde{w})(z - \tilde{z}) \\ &\quad - f'(z_w) + \lambda\, g'(z_w) - \mathcal{L}_{zz}(w)(z - z_w)\,, \\ \ell_2(\xi, w) &= g(z_w) + g'(z_w)(z - z_w) - g(\tilde{z}) - g'(\tilde{z})(z - \tilde{z})\,.\end{aligned}$$

Analogous to Lemma 3.2 of Alt 1990 we show

LEMMA 2.1 *Suppose Assumption* (A2) *holds, and define*

$$c_\ell = \frac{c_L}{2}(3 + \|\tilde{\lambda}\|_\Lambda)\,. \tag{2.3}$$

Then

$$\|\ell(\tilde{w}, w)\|_X \le c_\ell\, \|w - \tilde{w}\|_W^2 \tag{2.4}$$

for all $w = (z_w, \lambda_w)$ *with* $z_w \in B_Z(\tilde{z}, r_1)$ *and* $\lambda_w \in \Lambda$.

PROOF. Let $w = (z_w, \lambda_w) \in W$ with $z_w \in B_Z(\tilde{z}, r_1)$ and $\lambda_w \in \Lambda$. By the definition of ℓ we have

$$\begin{aligned}\ell_1(\tilde{w}, w) \;=\; & f'(\tilde{z}) - f'(z_w) - f''(z_w)(\tilde{z} - z_w) + (\lambda_w - \tilde{\lambda})g''(z_w)(\tilde{z} - z_w) \\ & - \tilde{\lambda}\,(g'(\tilde{z}) - g'(z_w) - g''(z_w)(\tilde{z} - z_w))\,.\end{aligned}$$

By Assumption (A2) we therefore obtain

$$\begin{aligned}\|\ell_1(\tilde{w}, w)\|_V \;&\le\; \frac{c_L}{2}(1 + \|\tilde{\lambda}\|_\Lambda)\|\tilde{z} - z_w\|_Z^2 + c_L\|\tilde{\lambda} - \lambda_w\|_\Lambda\, \|\tilde{z} - z_w\|_Z \\ &\le\; c_\ell\|\tilde{w} - w\|_W^2\,.\end{aligned}$$

Again by the definition of ℓ we have

$$\ell_2(\tilde{w}, w) = -g(\tilde{z}) + g(z_w) + g'(z_w)(\tilde{z} - z_w)\,,$$

By Assumption (A2) we therefore obtain

$$\|\ell_2(\tilde{w}, w)\|_Y \leq \frac{c_L}{2}\|\tilde{z} - z_w\|_Z^2 \leq c_\ell \|\tilde{w} - w\|_W^2 .$$

This proves the assertion of the lemma. ■

Based on Assumption (A2) we further show analogous to Lemma 4.4 of Alt 1992

LEMMA 2.2 *Suppose Assumption* (A2) *holds, and define*

$$r_2 = \min\left\{1, r_1, \left(\frac{2\sigma}{(10 + 7\|\tilde{\lambda}\|_\Lambda)c_L}\right)^{\frac{1}{2}}\right\} . \tag{2.5}$$

Then

$$\|\ell(\xi, w)\|_X \leq \sigma \tag{2.6}$$

for all $w = (z_w, \lambda_w), \xi = (z, \lambda) \in B_W(\tilde{w}, r_2)$.

PROOF. Let $w = (z_w, \lambda_w), \xi = (z, \lambda) \in B_W(\tilde{w}, r_2)$ be given. By the definition of ℓ we have

$$\begin{aligned}\ell_1(\xi, w) = {} & f'(\tilde{z}) - f'(z_w) - f''(z_w)(\tilde{z} - z_w) + (f''(\tilde{z}) - f''(z_w))(z - \tilde{z}) \\ & - \lambda (g'(\tilde{z}) - g'(z_w) - g''(z_w)(\tilde{z} - z_w)) \\ & + \lambda (g''(z_w) - g''(\tilde{z}))(z_w - \tilde{z}) + \left(\lambda - \tilde{\lambda}\right) g''(\tilde{z})(z_w - \tilde{z}) \\ & + \tilde{\lambda} (g''(z_w) - g''(\tilde{z}))(z - z_w) + \left(\lambda_w - \tilde{\lambda}\right) g''(z_w)(z - z_w) .\end{aligned}$$

By Assumption (A2) we therefore obtain

$$\begin{aligned}\|\ell_1(\xi, w)\|_V \leq {} & \frac{c_L}{2}\|\tilde{z} - z_w\|_Z^2 + c_L\|\tilde{z} - z_w\|_Z \|z - \tilde{z}\|_Z \\ & + \frac{3c_L}{2}\|\lambda\|_\Lambda \|\tilde{z} - z_w\|_Z^2 + c_L\|\lambda - \tilde{\lambda}\|_\Lambda \|\tilde{z} - z_w\|_Z \\ & + c_L\|\tilde{\lambda}\|_\Lambda \|\tilde{z} - z_w\|_Z \|z - z_w\|_Z + c_L\|\lambda_w - \tilde{\lambda}\|_\Lambda \|\tilde{z} - z_w\|_Z .\end{aligned}$$

Since $\|z - z_w\| \leq 2r_2$ and $\|\lambda\|_\Lambda \leq 1 + \|\tilde{\lambda}\|_\Lambda$, we further obtain

$$\|\ell_1(\xi, w)\|_V \leq \frac{c_L}{2}(10 + 7\|\tilde{\lambda}\|_\Lambda)r_2^2 \leq \sigma .$$

Again by the definition of ℓ we have

$$\ell_2(\xi, w) = g(z_w) - g(\tilde{z}) - g'(\tilde{z})(z_w - \tilde{z}) + (g'(z_w) - g'(\tilde{z})(z - z_w) .$$

By Assumption (A2) we therefore obtain

$$\|\ell_2(\xi, w)\|_Y \leq \frac{c_L}{2}\|z_w - \tilde{z}\|_Z^2 + c_L\|z_w - \tilde{z}\|_Z \|z - z_w\|_Z .$$

Since $\|z - z_w\|_Z \leq 2r_2$ we further obtain

$$\|\ell_2(\xi, w)\|_Y \leq \frac{5c_L}{2} r_2^2 \leq \sigma .$$

This proves the assertion of the lemma. ∎

Finally, again based on Assumption (A2) we obtain

LEMMA 2.3 *Suppose Assumption* (A2) *holds, and define*

$$r_3 = \min\left\{1, r_1, \frac{2}{3(4 + \|\tilde{\lambda}\|_\Lambda)c_L c_Q}\right\} . \tag{2.7}$$

Then

$$\|\ell(\xi_1, w) - \ell(\xi_2, w)\|_X \leq \frac{2}{3c_Q}\|\xi_1 - \xi_2\|_W \tag{2.8}$$

for all $\xi_1 = (z_1, \lambda_1), \xi_2 = (z_2, \lambda_2) \in W$ *with* $z_1, z_2 \in B_Z(\tilde{z}, r_1)$ *and for all* $w = (z_w, \lambda_w) \in B_W(\tilde{w}, r_3)$.

PROOF. Let $\xi_1 = (z_1, \lambda_1), \xi_2 = (z_2, \lambda_2) \in W$ with $z_1, z_2 \in B_Z(\tilde{z}, r_1)$ and $w = (z_w, \lambda_w) \in B_W(\tilde{w}, r_3)$ be given. By the definition of ℓ we have

$$\begin{aligned}\ell_1(\xi_1, w) - \ell_1(\xi_2, w) &= (f''(\tilde{z}) - f''(z_w))(z_1 - z_2) + (\lambda_1 - \lambda_2)(g'(z_w) - g'(\tilde{z})) \\ &\quad + \Big((\lambda_w - \tilde{\lambda})g''(\tilde{z}) + \lambda_w(g''(z_w) - g''(\tilde{z}))\Big)(z_1 - z_2) .\end{aligned}$$

By Assumption (A2) we therefore obtain

$$\begin{aligned}\|\ell_1(\xi_1, w) - \ell_1(\xi_2, w)\|_V &\leq c_L\|\tilde{z} - z_w\|_Z\,\|z_1 - z_2\|_Z + c_L\|\tilde{z} - z_w\|_Z\,\|\lambda_1 - \lambda_2\|_\Lambda \\ &\quad + c_L\|\lambda_w - \tilde{\lambda}\|_\Lambda\,\|z_1 - z_2\|_Z \\ &\quad + c_L\left(\|\tilde{\lambda}\|_\Lambda + 1\right)\|\tilde{z} - z_w\|_Z\;\|z_1 - z_2\|_Z \\ &\leq c_L\left(4 + \|\tilde{\lambda}\|_\Lambda\right) r_3\,\|\xi_1 - \xi_2\|_W \\ &\leq \frac{2}{3c_Q}\|\xi_1 - \xi_2\|_W .\end{aligned}$$

Again by the definition of ℓ we have

$$\ell_2(\xi_1, w) - \ell_2(\xi_2, w) = (g'(z_w) - g'(\tilde{z})(z_1 - z_2) .$$

By Assumption (A2) we therefore obtain

$$\begin{aligned}\|\ell_2(\xi_1, w) - \ell_2(\xi_2, w)\|_Y &\leq c_L\|z_w - \tilde{z}\|_Z\,\|z_1 - z_2\|_Z \\ &\leq c_L r_3\,\|z_1 - z_2\|_Z \\ &\leq \frac{2}{3c_Q}\|\xi_1 - \xi_2\|_W .\end{aligned}$$

This proves the assertion of the lemma. ∎

By Lemma 2.2, $\mathcal{S}_Q(\ell(\xi, w))$ is well-defined for all $w = (z_w, \lambda_w), \xi = (z, \lambda) \in B_W(\tilde{w}, r_2)$. Hence, for $w \in B_W(\tilde{w}, r_2)$ we can define a mapping

$$\mathcal{S}_w : B_W(\tilde{w}, r_2) \to W\,, \quad \xi \mapsto \mathcal{S}_Q(\ell(\xi, w))\,.$$

REMARK. Since $\ell(\tilde{w}, \tilde{w}) = 0_X$ we have $\mathcal{S}_{\tilde{w}}(\tilde{w}) = \mathcal{S}_Q(0_X) = \tilde{w}$, i.e., $\tilde{w}$ is a fixed point of $\mathcal{S}_{\tilde{w}}$.

Following the argument in the proof of Theorem 2.1 in Robinson 1980 we show

THEOREM 2.4 *Suppose Assumptions* (A1)–(A5) *are satisfied. Choose*

$$r_4 = \min\left\{ r_2, r_3, \frac{1}{3c_\ell c_Q} \right\}\,. \tag{2.9}$$

Then there exists a single-valued function $\mathcal{S}: B_W(\tilde{w}, r_4) \to B_W(\tilde{w}, r_4)$ *such that for each* $w \in B_W(\tilde{w}, r_4)$, $\mathcal{S}(w)$ *is the unique fixed point in* $B_W(\tilde{w}, r_4)$ *of* $\mathcal{S}_w$, *and*

$$\|\mathcal{S}(w) - \tilde{w}\|_W \le 3c_Q \,\|\ell(\tilde{w}, w)\|_X \le 3c_Q c_\ell \,\|w - \tilde{w}\|_W^2\,. \tag{2.10}$$

PROOF. Define $\delta = \frac{2}{3}c_Q$, and choose any $w \in B_W(\tilde{w}, r_4)$. By Lemma 2.1 and the definition of r_4 we have

$$c_Q \,\|\ell(\tilde{w}, w)\|_X \le c_Q c_\ell r_4^2 \le \frac{1}{3} r_4 = (1 - c_Q \delta) r_4\,. \tag{2.11}$$

Let $\xi_1 = (z_1, \lambda_1), \xi_2 = (z_2, \lambda_2) \in B_W(\tilde{w}, r_4)$ be given. By Lemma 2.2, Assumption (A5) and Lemma 2.3 it follows that

$$\begin{aligned} \|\mathcal{S}_w(\xi_1) - \mathcal{S}_w(\xi_2)\|_W &= \|\mathcal{S}_Q(\ell(\xi_1, w)) - \mathcal{S}_Q(\ell(\xi_2, w))\|_W \\ &\le c_Q \,\|\ell(\xi_1, w) - \ell(\xi_2, w)\|_X \\ &\le \frac{2}{3} \,\|\xi_1 - \xi_2\|_W\,. \end{aligned} \tag{2.12}$$

Hence, $\mathcal{S}_w$ is strongly contractive on $B_W(\tilde{w}, r_4)$. Since $\tilde{w} = \mathcal{S}_Q(\ell(\tilde{w}, \tilde{w})) = \mathcal{S}_Q(0_X)$ and $\mathcal{S}_w(\tilde{w}) = \mathcal{S}_Q(\ell(\tilde{w}, w))$ we have by (2.11)

$$\|\mathcal{S}_w(\tilde{w}) - \tilde{w}\|_W \le c_Q \,\|\ell(\tilde{w}, w)\|_X \le \frac{1}{3} r_4\,,$$

and therefore for any $\xi \in B_W(\tilde{w}, r_4)$ by (2.12)

$$\begin{aligned} \|\mathcal{S}_w(\xi) - \tilde{w}\|_W &\le \|\mathcal{S}_w(\xi) - \mathcal{S}_w(\tilde{w})\|_W + \|\mathcal{S}_w(\tilde{w}) - \tilde{w}\|_W \\ &\le \frac{2}{3} \,\|\xi - \tilde{w}\|_W + \frac{1}{3} r_4 \le r_4\,, \end{aligned}$$

so that $\mathcal{S}_w$ is a self-map on $B_W(\tilde{w}, r_4)$. By the contraction principle, $\mathcal{S}_w$ has a unique fixed point $\mathcal{S}(w) \in B_W(\tilde{w}, r_4)$. Thus, we have established the existence of the function $\mathcal{S}: B_W(\tilde{w}, r_4) \to B_W(\tilde{w}, r_4)$. In addition, by the contraction principle for each $\xi \in B_W(\tilde{w}, r_4)$ one has the bound

$$\|\mathcal{S}(w) - \xi\|_W \leq (1 - c_Q\delta)^{-1} \|\mathcal{S}_w(\xi) - \xi\|_W = 3\,\|\mathcal{S}_w(\xi) - \xi\|_W \,. \tag{2.13}$$

To obtain the bound (2.10), we take any $w \in B_W(\tilde{w}, r_4)$, and apply (2.13) with $\xi = \tilde{w} = \mathcal{S}(\tilde{w})$ to get

$$\|\mathcal{S}(w) - \mathcal{S}(\tilde{w})\|_W \leq 3\,\|\mathcal{S}_w(\tilde{w}) - \mathcal{S}(\tilde{w})\|_W \,.$$

If we now recall that $\ell(\tilde{w}, \tilde{w}) = 0_X$, $\mathcal{S}(\tilde{w}) = \mathcal{S}_{\tilde{w}}(\mathcal{S}(\tilde{w}))$ and employ the bound

$$\begin{aligned} \|\mathcal{S}_w(\tilde{w}) - \mathcal{S}_{\tilde{w}}(\tilde{w})\|_W &= \|\mathcal{S}_Q(\ell(\tilde{w}, w)) - \mathcal{S}_Q(\ell(\tilde{w}, \tilde{w}))\|_W \\ &\leq c_Q\,\|\ell(\tilde{w}, w) - \ell(\tilde{w}, \tilde{w})\|_W \,, \end{aligned}$$

we immediately obtain the first inequality of (2.10). The second inequality follows from (2.4). ■

A simple computation shows that $\xi = (z, \lambda)$ is a fixed point of $\mathcal{S}_w$, i.e., $\xi = \mathcal{S}_Q(\ell(\xi, w))$, iff (2.1) and (2.2) are satisfied, i.e., iff the pair (z, λ) is a Kuhn–Tucker point for $(\mathrm{QP})_w$. Theorem 2.4 therefore shows that for $w \in B_W(\tilde{w}, r_4)$ there exists a unique Kuhn–Tucker point $\mathcal{S}(w) = (\bar{z}_w, \bar{\lambda}_w) \in B_W(\tilde{w}, r_4)$ of $(\mathrm{QP})_w$. Since we are interested in solutions of $(\mathrm{QP})_w$ we additionally assume

(A6) A sufficient optimality condition is satisfied such that if $w \in B_W(\tilde{w}, r_5)$ then $\mathcal{S}(w) = (\bar{z}_w, \bar{\lambda}_w)$ defines a global solution $\bar{z}_w$ of $(\mathrm{QP})_w$.

By Theorem 2.4 we get

THEOREM 2.5 *Suppose Assumptions* (A1)–(A6) *are satisfied. Choose*

$$\rho = \min\{r_4, r_5\} \,. \tag{2.14}$$

Then for each $w \in B_W(\tilde{w}, \rho)$ there exist a unique solution $\bar{z}_w$ of $(\mathrm{QP})_w$ and a unique associated Lagrange multiplier $\bar{\lambda}_w$, and

$$\|(\bar{z}_w, \bar{\lambda}_w) - (\tilde{z}, \tilde{\lambda})\|_W \leq 3c_Q\,\|\ell(\tilde{w}, w)\|_X \leq 3c_Q c_\ell\,\|w - \tilde{w}\|_W^2 \,. \tag{2.15}$$

In the same way as Theorem 3.3 in Alt 1990 we can now prove the following result on local quadratic convergence of the Lagrange–Newton method (LNM).

THEOREM 2.6 *Suppose that Assumptions* (A1)–(A6) *are satisfied. Choose*

$$\gamma < \min\{\rho, \frac{1}{3c_Q c_\ell}\} \,.$$

Then $\delta := 3c_Q c_\ell \gamma < 1$, and for any starting point $w_0 \in B_W(\tilde{w}, \gamma\delta)$ the Lagrange-Newton method generates a unique sequence $\{w_k\}$, $w_k = (z_k, \lambda_k)$ convergent to $\tilde{w}$. Moreover, z_k is a global solution of Problem $(\mathrm{QP})_k$, λ_k is the unique associated Lagrange multiplier and

$$\|w_k - \tilde{w}\|_W \leq \gamma\,\delta^{2^k - 1} \,,$$

for $k \geq 2$.

PROOF. For $k = 1$ we get from Theorem 2.5

$$\|w_1 - \tilde{w}\|_W = \|\mathcal{S}(w_0) - \mathcal{S}(\tilde{w})\|_W$$
$$\leq 3c_Q c_\ell \|w_0 - \tilde{w}\|_W^2 \leq 3c_Q c_\ell (\gamma\delta)^2 = \gamma\,\delta^3 = \gamma\,\delta^{2^k - 1}.$$

Since $\delta^3 < \delta$ we get $w_1 \in B_W(\tilde{w}, \gamma\delta)$, and since $w_1 = S(w_0)$, w_1 is uniquely determined. This proves the assertion for $k = 1$. Now, it follows by induction

$$\|w_{k+1} - w_0\|_W = \|S(w_k) - S(\tilde{w})\|_W \leq 3c_Q c_\ell \|w_k - \tilde{w}\|_W^2$$
$$\leq 3c_Q c_\ell \gamma^2 \delta^{2^{k+1}-2} = \gamma\,\delta^{2^{k+1}-1}.$$

Since $\delta^{2^{k+1}-1} < \delta$ we get $w_{k+1} \in B_W(\tilde{w}, \gamma\delta)$, and since $w_{k+1} = S(w_k)$, w_{k+1} is uniquely determined. This completes the proof. ■

REMARK. Assumptions (A5) and (A6) are formulated here in a very general form. In each application they have to be verified, which is the most difficult step in the application of Theorem 2.6. Conditions under which (A5) and (A6) are satisfied involve stronger constraint qualifications and strengthened second-order sufficient optimality conditions. They are discussed e.g., in Alt 1990, Alt 1991, Alt 1992, Hager 1990, Malanowski 1992, Malanowski 1993B, Malanowski 1994, Malanowski 1993A.

3. Optimal control problems with convex control constraints

In this section we apply the convergence theory of Section 2 to a class of non-linear optimal control problems. For sake of simplicity, we only consider control problems with linear state equation and convex control constraints, but with nonlinear cost. A slightly more general class of control problems has been investigated in Alt, Malanowski 1993.

All needed stability results for the quadratic control problems defined by the Lagrange–Newton method are taken from Hager's paper, Hager 1990, where they were used for the convergence analysis of multiplier methods.

The following notation will be used:

$\mathbb{R}^n$ is the n-dimensional Euclidean space with the inner product denoted by $\langle x, y\rangle$ and the norm $|x| = \langle x, x\rangle^{\frac{1}{2}}$.

$L^2(0,T;\mathbb{R}^n)$ is the Hilbert space of square integrable vector functions, with the inner product

$$(x, y) = \int_0^T \langle x(t), y(t)\rangle\, dt\,,$$

and the norm

$$\|x\|_2 = (x, x)^{\frac{1}{2}}.$$

$L^\infty(0,T;\mathbb{R}^n)$ is the Banach space of essentially bounded vector functions with the norm

$$\|x\|_\infty = \max_i \operatorname*{ess\,sup}_{t\in[0,T]} |x^i(t)|.$$

$W^{1,p}(0,T;\mathbb{R}^n)$ are the Sobolev spaces of absolutely continuous functions with derivatives in $L^p(0,T;\mathbb{R}^n)$, $p = 2, \infty$.

In $W^{1,2}(0,T;\mathbb{R}^n)$ there is defined the inner product

$$(x,y)_{1,2} = \langle x(0), y(0)\rangle + (\dot{x}, \dot{y}),$$

which induces the norm

$$\|x\|_{1,2} = (x,x)_{1,2}^{\frac{1}{2}},$$

whereas the norm in $W^{1,\infty}(0,T;\mathbb{R}^n)$ is given by

$$\|x\|_{1,\infty} = \max\{|x(0)|, \|\dot{x}\|_\infty\}.$$

The inner product $(\cdot,\cdot)$ will be extended by continuity to $L^\infty(0,T;\mathbb{R}^n) \times (L^\infty(0,T;\mathbb{R}^n))^*$, or to $W^{1,\infty}(0,T;\mathbb{R}^n) \times \left(W^{1,\infty}(0,T;\mathbb{R}^n)\right)^*$.

Let the space $Z = Z_1 \times Z_2$ be defined by

$$Z_1 = L^\infty(0,T;\mathbb{R}^m), \quad Z_2 = W^{1,\infty}(0,T;\mathbb{R}^n). \tag{3.16}$$

We consider the following optimal control problem:

$$\text{(O)} \qquad \mathrm{Min}_{(u,x)\in Z_1\times Z_2} \quad f(u,x) = \int_0^T f^0(u(t),x(t))\,dt$$

$$\begin{aligned} &\text{subject to} \\ &\dot{x}(t) = A(t)x(t) + B(t)u(t) \quad \forall' t \in [0,T], \\ &x(0) = x_0, \\ &u(t) \in U \quad \forall' t \in [0,T], \end{aligned}$$

where $f^0(\cdot,\cdot) : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, $A \in L^\infty(0,T;\mathbb{R}^{n\times n})$, $B \in L^\infty(0,T;\mathbb{R}^{m\times n})$, $U \subset \mathbb{R}^m$ is nonempty, closed and convex, and x_0 is the given initial state.

In order to put Problem (O) in the general framework of Section 2, we further define

$$\begin{aligned} &Y = L^\infty(0,T;\mathbb{R}^n), \\ &C = \{(u,x) \in Z \mid x(0) = x_0,\ u(t) \in U \text{ for a.a. } t \in [0,T]\}, \end{aligned}$$

and $g : Z \to Y$ by

$$g(u,x) = \dot{x} - Ax - Bu.$$

Along with the spaces Z and Y we define $\hat{Z} = \hat{Z}_1 \times \hat{Z}_2$ and $\hat{Y}$ by

$$\hat{Z}_1 = L^2(0,T;\mathbb{R}^m), \quad \hat{Z}_2 = W^{1,2}(0,T;\mathbb{R}^n), \quad \hat{Y} = L^2(0,T;\mathbb{R}^n). \tag{3.17}$$

The spaces V and Λ introduced in Section 2 are defined as follows:

$$V = L^\infty(0,T;\mathbb{R}^m) \times L^\infty(0,T;\mathbb{R}^n), \quad \Lambda = W^{1,\infty}(0,T;\mathbb{R}^n).$$

It is assumed that:

(B1) There exists a (local) solution $(\tilde{u}, \tilde{x})$ of (O).
(B2) $f^0(\cdot,\cdot)$ is two times Fréchet differentiable in all arguments, and the respective derivatives are locally Lipschitz continuous.

Certainly, by (B1) and the linearity of the state equation, Assumption (A1) of Section 2 is satisfied. It is easy to see that, by the definition of the spaces V and Λ and by (B2), Assumption (A2) holds for arbitrary $r_1 > 0$, where c_L of course depends on the choice of r_1. In view of the definition of r_2 and r_3 we choose $r_1 = 1$.

We are going to introduce some regularity conditions. Let us start with the constraint qualifications. Let us put

$$w = (u, x, \lambda) \in W = Z \times \Lambda ,$$

Further, we define subspaces $Z_0 \subset Z$ and $\hat{Z}_0 \subset \hat{Z}$ by

$$Z_0 = \{(u,x) \in Z \mid x(0) = 0\} , \quad \hat{Z}_0 = \{(u,x) \in \hat{Z} \mid x(0) = 0\} ,$$

and the mapping $\tilde{S} \in L(Z_0, Y) \cap L(\hat{Z}_0, \hat{Y})$ by

$$\tilde{S}(\sigma, \xi) = \dot{\xi} - A(\tilde{w})\xi - B(\tilde{w})\sigma .$$

It is well-known that the mapping $\tilde{S}$ is surjective (see e.g. Girsanov 1972). Hence, in particular Assumption (A3) of Section 2 is satisfied.

Let us define the Lagrangian associated with (O):

$$\begin{aligned} &\mathcal{L} : Z \times Y^* \to \mathbb{R} , \\ &\mathcal{L}(u, x, \lambda) = f(u, x, h) + (\lambda, \dot{x} - Ax - Bu) . \end{aligned}$$

It is also well-known (see e.g. Girsanov 1972) that there exists a unique Lagrange multiplier

$$\tilde{\lambda} \in W^{1,\infty}(0, T; \mathbb{R}^n) \tag{3.18}$$

associated with $(\tilde{u}, \tilde{x})$ such that the following stationarity conditions hold:

$$\dot{\tilde{\lambda}}(t) + A(t)^T \tilde{\lambda}(t) - f^0_x(\tilde{u}(t), \tilde{x}(t)) = 0 , \quad \tilde{\lambda}(T) = 0 ,$$

and for a.a. $t \in [0, T]$

$$\left(f^0_u(\tilde{u}(t), \tilde{x}(t)) - B(t)^T \tilde{\lambda}(t))\right)(u - \tilde{u}(t)) \geq 0 \quad \forall u \in U .$$

Note that (A4) is satisfied by (3.18).

For the sake of simplicity let us denote

$$\tilde{\mathcal{L}} := \mathcal{L}(\tilde{u}, \tilde{x}, \tilde{\lambda}) ,$$

and introduce for $w = (u, x, \lambda) \in W$ the Hamiltonian

$$\mathcal{H}(w(t)) = f^0(u(t), x(t)) - \langle \lambda(t), A(t)x(t) + B(t)u(t) \rangle .$$

According to Assumption (8) in Hager 1990, in addition to (B1)–(B2) we assume the following coercitivity condition:

(B3) There exists $\alpha_1 > 0$ such that

$$\left([u^T, x^T], \begin{bmatrix} \tilde{\mathcal{L}}_{uu} & \tilde{\mathcal{L}}_{ux} \\ \tilde{\mathcal{L}}_{xu} & \tilde{\mathcal{L}}_{xx} \end{bmatrix} \begin{bmatrix} u \\ x \end{bmatrix}\right)$$

$$= \int_0^T [u^T(t), x^T(t)] \begin{bmatrix} \mathcal{H}_{uu}(\tilde{w}(t)) & \mathcal{H}_{ux}(\tilde{w}(t)) \\ \mathcal{H}_{xu}(\tilde{w}(t)) & \mathcal{H}_{xx}(\tilde{w}(t)) \end{bmatrix} \begin{bmatrix} u(t) \\ x(t) \end{bmatrix} dt$$

$$\geq \alpha_1 \left(\|u\|_2^2 + \|x\|_{1,2}^2\right), \tag{3.19}$$

$$\tag{3.20}$$

for all pairs $(u, x) \in \hat{Z}$ satisfying

$$u = u_1 - u_2 \quad \text{for some } u_1, u_2 \in U,$$
$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \quad \forall' t \in [0, T],$$
$$x(0) = 0.$$

REMARK. Because of the linearity of the state equation we have

$$\begin{bmatrix} \mathcal{H}_{uu}(\tilde{w}(t)) & \mathcal{H}_{ux}(\tilde{w}(t)) \\ \mathcal{H}_{xu}(\tilde{w}(t)) & \mathcal{H}_{xx}(\tilde{w}(t)) \end{bmatrix} = \begin{bmatrix} f^0_{uu}(\tilde{w}(t)) & f^0_{ux}(\tilde{w}(t)) \\ f^0_{xu}(\tilde{w}(t)) & f^0_{xx}(\tilde{w}(t)) \end{bmatrix}$$

To apply the Lagrange–Newton method to (O) we construct the following linear–quadratic optimal control problems depending on the parameter $w = (u_w, x_w, \lambda_w)$ corresponding to problem $(\mathrm{QP})_w$ of Section 2:

$$(\mathrm{OQ})_w \qquad \mathrm{Min}_{(u,x)\in Z_1 \times Z_2} \quad \tilde{I}(u, x, w) = \int_0^T \tilde{f}^0(u(t), x(t), w)\, dt$$

subject to

$$\dot{x}(t) = A(t)(x(t) - x_w(t)) + B(t)(u(t) - u_w(t)) \quad \forall' t \in [0, T],$$
$$x(0) = x_0,$$
$$u(t) \in U \quad \forall' t \in [0, T],$$

where

$$\tilde{f}^0(x(t), u(t), w(t)) =$$
$$= \langle f^0_u(u_w(t), x_w(t)), u(t) - u_w(t)\rangle + \langle f^0_x(u_w(t), x_w(t)), x(t) - x_w(t)\rangle$$
$$+ \frac{1}{2} \begin{bmatrix} u(t) - u_w(t) \\ x(t) - x_w(t) \end{bmatrix}^T \begin{bmatrix} f^0_{uu}(w(t)) & f^0_{ux}(w(t)) \\ f^0_{xu}(w(t)) & f^0_{xx}(w(t)) \end{bmatrix} \begin{bmatrix} u(t) - u_w(t) \\ x(t) - x_w(t) \end{bmatrix}.$$

By $(\bar{u}_w, \bar{x}_w)$ we denote a solution to $(\mathrm{OQ})_w$. It is easy to see that by Assumption (B3) we have

$$(\bar{u}_{\tilde{w}}, \bar{x}_{\tilde{w}}) = (\tilde{u}, \tilde{x}).$$

In the Lagrange–Newton iterative procedure we put in $(\mathrm{OQ})_w$, $w := w_k$, compute the solution (u_{k+1}, x_{k+1}) of $(\mathrm{OQ})_{w_k}$ as well as the corresponding Lagrange multiplier λ_{k+1} and put $w_{k+1} = (u_{k+1}, x_{k+1}, \lambda_{k+1})$.

For a given value of the parameter $p = (a_u, a_x, b) \in X = V \times Y$ we define the auxiliary linear-quadratic optimal control problem $(\mathrm{OS})_p$ corresponding to $(\mathrm{QS})_p$ of Section 2:

$$
\begin{aligned}
(\mathrm{OS})_p \qquad \mathrm{Min}_{(u,x)\in Z_1\times Z_2} \quad \hat{I}(u,x,p) = \int_0^T & [\tilde{f}^0(u(t),x(t),\tilde{w}) - \\
& - \langle a_u(t), u(t) - \tilde{u}(t)\rangle - \langle a_x(t), x(t) - \tilde{x}(t)\rangle]\, dt
\end{aligned}
$$

subject to

$$
\begin{aligned}
&\dot{x}(t) = A(t)(x(t) - \tilde{x}(t)) + B(t)(u(t) - \tilde{u}(t)) - b(t) \ \forall' t \in [0,T]\,,\\
&x(0) = x_0\,,\\
&u(t) \in U \quad \forall' t \in [0,T]\,.
\end{aligned}
$$

It is easy to see that by Assumption (B3) Problem $(\mathrm{OS})_0$ has the unique solution $(\tilde{u}, \tilde{x})$.

Let $p_i = (a_u^i, a_x^i, b_i) \in X$, $i = 1,2$, be given, and denote by (u_i, x_i, λ_i), $i = 1,2$ the corresponding solutions and Lagrange multipliers. In order to estimate $\|u_1 - u_2\|_2$ we apply Lemma 1 of Hager 1990 regarding the state as an affine function of the control. Let $x = L(y)$ be the unique solution to

$$\dot{x} = Ax + y\,, \quad x(0) = 0\,,$$

and denote by ξ_0 the unique solution to

$$\dot{x} = Ax\,, \quad x(0) = x_0\,.$$

Then the unique solution to the state equation of Problem $(\mathrm{OS})_p$ for $p = p_i$ is

$$x = L(-A\tilde{x} + B(u_i - \tilde{u}) - b_i) + \xi_0\,.$$

The Problems $(\mathrm{OS})_p$ are special cases of Problem (13) in Hager 1990. Therefore, in the same way as in Hager 1990 we obtain

$$
\begin{aligned}
\alpha_1\|u_1 - u_2\|_2 \quad \leq \quad & \|\left(f^0_{ux}(\tilde{u},\tilde{x}) + B^T L^T f^0_{xx}(\tilde{u},\tilde{x})\right)(b_1 - b_2) \\
& + B^T L^T (a_x^1 - a_x^2) + a_u^1 - a_u^2\|_2\,. \qquad (3.21)
\end{aligned}
$$

From this, the state equation and the adjoint equation we then obtain by standard proof techniques the following result (see e.g. Hager 1990).

LEMMA 3.1 *Suppose Assumptions* (B1)–(B3) *are satisfied. Then there is a constant* $c_1 > 0$ *such that*

$$
\begin{aligned}
&\|u_2 - u_1\|_2 + \|x_2 - x_1\|_{1,2} + \|\lambda_2 - \lambda_1\|_{1,2} \\
&\qquad \leq \alpha_1^{-1} c_1 \left(\|a_u^1 - a_u^2\|_2 + \|a_x^1 - a_x^2\|_2 + \|b_2 - b_1\|_2\right) \qquad (3.22)
\end{aligned}
$$

for all $p_i = (a_u^i, a_x^i, b_i) \in X$, $i = 1,2$.

REMARK. By (3.21) the constant c_1 only depends on $f^0_{ux}(\tilde{u},\tilde{x})$, $f^0_{xx}(\tilde{u},\tilde{x})$, B, and L. More precisely,

$$c_1 = c_1\left(\|f^0_{ux}(\tilde{u},\tilde{x})\|_\infty, \|f^0_{xx}(\tilde{u},\tilde{x})\|_\infty, \|B\|_\infty, \|L\|_\infty\right) ,$$

where $\|\cdot\|_\infty$ denotes the operator norm for these mappings regarded as operators from L^2 to L^2, respectively to $W^{1,2}$. Moreover, by Assumption (B2) c_1 is a Lipschitz continuous function of its arguments.

Inequality (3.22) is too weak to obtain Assumption (A5) of Section 2. According to (15) in Hager 1990 we therefore assume in addition to (B3):

(B4) There exists $\alpha_2 > 0$ independent of t such that
$$\langle u, \mathcal{H}_{uu}(\tilde{w}(t))u\rangle = \langle u, f^0_{uu}(\tilde{u}(t),\tilde{x}(t))\, u\rangle \geq \alpha_2\, |u|^2$$
for all $u = u_1 - u_2$ with $u_1, u_2 \in U$, and all $t \in [0,T]$.

In the same way as in Hager 1990 we obtain from Assumption (B4)

LEMMA 3.2 *Suppose Assumptions* (B1)–(B4) *are satisfied. Then there is a constant* $c_2 > 0$ *such that*

$$|u_2(t) - u_1(t)| \leq \alpha_2^{-1} c_2 \left(|a^1_u(t) - a^2_u(t)| + |a^1_x(t) - a^2_x(t)| + |b_2(t) - b_1(t)|\right)$$

for all $p_i = (a^i_u, a^i_x, b_i) \in X,\ i = 1,2.$

REMARK. Again by (3.21) and Assumption (B2)

$$c_2 = c_2\left(\|f^0_{ux}(\tilde{u},\tilde{x})\|_\infty, \|f^0_{xx}(\tilde{u},\tilde{x})\|_\infty, \|B\|_\infty, \|L\|_\infty\right) ,$$

is a Lipschitz continuous function of its arguments.

Finally, from Lemmas 3.1 and 3.2 we obtain in the same way as in Hager 1990

LEMMA 3.3 *Suppose Assumptions* (B1)–(B4) *are satisfied. Then there is a constant* c,

$$c = c\left(\|f^0_{ux}(\tilde{u},\tilde{x})\|_\infty, \|f^0_{xx}(\tilde{u},\tilde{x})\|_\infty, \|B\|_\infty, \|L\|_\infty\right) ,$$

which is a Lipschitz continuous function of its arguments such that

$$\begin{aligned}&\|u_2 - u_1\|_\infty + \|x_2 - x_1\|_{1,\infty} + \|\lambda_2 - \lambda_1\|_{1,\infty}\\ &\qquad \leq \alpha_1^{-1}\alpha_2^{-1} c\left(\|a^1_u - a^2_u\|_\infty + \|a^1_x - a^2_x\|_\infty + \|b_2 - b_1\|_\infty\right)\end{aligned}$$

for all $p_i = (a^i_u, a^i_x, b_i) \in X,\ i = 1,2.$

By Lemma 3.3, Assumption (A5) of Section 2 is satisfied with

$$c_Q = \alpha_1^{-1}\alpha_2^{-1}c\,, \tag{3.23}$$

and $\sigma > 0$ arbitrary. Let c_ℓ be defined by (2.3), i.e.,

$$c_\ell = \frac{c_L}{2}(3 + \|\tilde{\lambda}\|_\Lambda)\,.$$

Then c_ℓ is a Lipschitz continuous function of $\|\tilde{\lambda}\|_\Lambda$. The radius r_2 is defined by (2.5). Since $\sigma > 0$ is arbitrary and $r_1 = 1$ we can choose $r_2 = 1$. Then by (2.7) r_3 is defined by

$$r_3 = \min\left\{1, \frac{2}{3(4 + \|\tilde{\lambda}\|_\Lambda)c_L c_Q}\right\}.$$

The radius r_4 defined by (2.9) is therefore given by

$$r_4 = \min\left\{1, \frac{2}{3(4 + \|\tilde{\lambda}\|_\Lambda)c_L c_Q}, \frac{1}{3c_\ell c_Q}\right\}.$$

In order to apply Theorem 2.6 we still have to verify Assumption (A6). But it follows from Lemma 5 in Hager 1990 (see also Lemma 5.3 in Malanowski 1993B) that the sufficient optimality condition (3.20) is satisfied for all w in some neighborhood of $\tilde{w}$. Hence, if (B1)–(B4) are satisfied, then all assumptions of Theorem 2.6 hold, and by this theorem we obtain the following result on local quadratic convergence of the Lagrange-Newton method.

THEOREM 3.4 *Suppose that Assumptions* (B1)–(B4) *are satisfied. Choose*

$$\rho < \min\{r_4, r_5\}\,,$$

and

$$\gamma < \min\{\rho, \frac{1}{3c_Q c_\ell}\}\,.$$

Then $\delta := 3c_Q c_\ell \gamma < 1$, *and for any starting point* $w_0 \in B_W(\tilde{w}, \gamma\delta)$ *the Lagrange-Newton method generates a unique sequence* $\{w_k\}$, $w_k = (u_k, x_k, \lambda_k)$ *convergent to* $\tilde{w}$. *Moreover,* (u_k, x_k) *is a global solution of Problem* $(\mathrm{QP})_k$, λ_k *is the unique associated Lagrange multiplier and*

$$\|w_k - \tilde{w}\|_W \le \gamma\,\delta^{2^k-1}\,,$$

for $k \ge 2$.

REMARK. Theorem 3.4 gives an explicit estimate of the radius of convergence for the Lagrange–Newton method. However, we want to emphasize that for concrete optimal control problems it is unlikely that this theoretical bound can be explicitly computed. Nevertheless, the result is very usefull if we replace the original control problem (O) by a discretized problem or by a perturbed problem. Since the constant c of Lemma 3.3 is a Lipschitz continuous function of its arguments, the constant c_Q defined by (3.23) is a Lipschitz continuous function of the same arguments. Since c_ℓ is a Lipschitz continuous function of $\|\tilde{\lambda}\|_\Lambda$, it follows that

$$r_4 = r_4\left(\|f^0_{ux}(\tilde{u},\tilde{x})\|_\infty, \|f^0_{xx}(\tilde{u},\tilde{x})\|_\infty, \|B\|_\infty, \|L\|_\infty, \|\tilde{\lambda}\|_\Lambda\right)$$

is a Lipschitz continuous function of its arguments. This fact can be used to show, that for suitable discretizations or sufficiently smooth perturbations of the original problem, the radius of convergence γ for the Lagrange–Newton method to compute the solution of the discretized or the perturbed problem changes Lipschitz continuously with the discretization or perturbation parameter.

References

ALT W. (1990) The Lagrange–Newton method for infinite–dimensional optimization problems, *Numerical Functional Analysis and Optimization* **11**, pp. 201–224.

ALT W. (1991) Parametric programming with applications to optimal control and sequential quadratic programming, *Bayreuther Mathematische Schriften* **35**, pp. 1–37.

ALT W. (1992) Sequential quadratic programming in Banach spaces. In W. Oettli and D. Pallaschke, editors, *Advances in Optimization*, vol. **382** of *Lecture Notes in Economics and Mathematical Systems*, pp. 281–301, Springer Verlag.

ALT W. AND MALANOWSKI K. (1993) The Lagrange–Newton method for nonlinear optimal control problems, *Computational Optimization and Applications*, **2**, pp. 77–100.

ALT W., SONTAG R. AND TRÖLTZSCH F. (1993) An SQP method for optimal control of weakly singular Hammerstein integral equations, preprint.

FIACCO A.V. (1983) Introduction to Sensitivity and Stability Analysis in Nonlinear Programming. Academic Press, New York.

FLETCHER R. (1987) Practical methods of optimization. John Wiley & Sons, New York, second edition.

GIRSANOV I.V. (1972) Lectures on Mathematical Theory of Extremum Problems, vol. **67** of *Lecture Notes in Economics and Mathematical Systems*. Springer Verlag, New York.

HAGER W.W. (1990) Multiplier methods for nonlinear optimal control, *SIAM Journal Numerical Analysis*, **17**, p. 1061–1080.

ITO K. AND KUNISCH K. (1992) Senistivity analysis of solutions to optimization problems in hilbert spaces with applications to optimal control and estimation, *Journal of Differential Equations*, **99**, pp. 1–40.

KELLEY C. AND WRIGHT S. (1990) rm Sequential quadratic programming for certain parameter identification problems, preprint.

LEVITIN E.S. AND POLYAK B.T. (1966) Constrained minimization methods, *USSR Journal on Computational Mathematics and Mathematical Physics*, **6(5)**, pp. 1–50.

MACHIELSEN K.C.P. (1987) Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space, *CWI Tract* **53**.

MALANOWSKI K. (1992) Second order conditions and constraint qualifications in stability and sensitivity analysis of solutions to optimization problems in Hilbert spaces, *Applied Mathematics and Optimization*, **25**, pp. 51–79.

MALANOWSKI K. (1993A) Stability and sensitivity analysis of solutions to infinite–dimensional optimization problems, preprint.

MALANOWSKI K. (1993B) Two–norm approach in stability and sensitivity analysis of optimization and optimal control problems, *Advances in Math. Sc. and Applications* **2**, pp. 397–443.

MALANOWSKI K. (1994) Stability and sensitivity analysis of solutions to nonlinear optimal control problems, *Applied Mathematics and Optimization*, to be published.

MAURER H. (1981) First- and second–order sufficient optimality conditions in mathematical programming and optimal control, *Mathematical Programming Study* **14**, pp. 163–177.

ROBINSON S.M. (1974) Perturbed Kuhn–Tucker points and rates of convergence for a class of nonlinear–programming algorithms, *Mathematical Programming* **7**, pp. 1–16.

ROBINSON S.M. (1976) Stability theory for systems of inequalities, part ii: Differentiable nonlinear systems, *SIAM Journal Numerical Analysis* **13**, pp. 497–513.

ROBINSON S.M. (1980) Strongly regular generalized equations, *Mathematics of Operations Research* **5**, pp. 43–62.

ROBINSON S.M. (1983) Generalized equations. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical Programming: The State of the Art*, p. 165–207, Berlin, Springer–Verlag.

ROBINSON S.M. (1987) Local structure of feasible sets in nonlinear programming, part iii: Stability and sensitivity, *Mathematical Programming Study* **30**, pp. 45–66.

STOER J. (1985) Principles of sequential quadratic programming methods for solving nonlinear programs. In K. Schittkowski, editor, *Computational Mathematical Programming*, vol. **F15**, pp. 165–207. Nato ASI Series.

WILSON R.B. (1963) A simplicial algorithm for concave programming. PhD thesis, Harvard University, Graduate School of Business Administration.