Int. J. Appl. Math. Comput. Sci., 2004, Vol. 14, No. 4, 531–540

CONVERGENCE OF THE LAGRANGE-NEWTON METHOD FOR OPTIMAL CONTROL PROBLEMS

KAZIMIERZ D. MALANOWSKI*

* Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01–447 Warszawa, Poland
e-mail: kmalan@ibspan.waw.pl

Convergence results for two Lagrange-Newton-type methods of solving optimal control problems are presented. It is shown how the methods can be applied to a class of optimal control problems for nonlinear ODEs, subject to mixed control-state constraints. The first method reduces to an SQP algorithm. It does not require any information on the structure of the optimal solution. The other one is the shooting method, where information on the structure of the optimal solution is exploited. In each case, conditions for well-posedness and local quadratic convergence are given. The scope of applicability is briefly discussed.

Keywords: optimal control, nonlinear ODEs, mixed constraints, Lagrange-Newton method

1. Introduction

In theoretical and numerical research, optimal control problems have been either treated as cone constrained optimization problems in functional spaces, or studied using some specialized tools. In the first approach, problems of optimal control are placed in a broader framework of optimization problems, and general techniques can be used to solve them, whereas the second approach allows us to take maximal advantage of the specific structure of the problems. Such a situation takes place also in applications of the Lagrange-Newton method for solving numerically optimal control problems.

The classical Lagrange-Newton method (see, e.g., Stoer and Bulirsch, 1980), one of the most efficient numerical methods of solving optimization problems, was developed for problems with equality-type constraints. In this method, the Newton procedure is applied to the first-order optimality system, which has the form of a system of equations. In the case of inequality-type constraints, the first-order optimality system cannot be expressed as an equation. However, it can be expressed as an inclusion, or the so-called *generalized equation* (Robinson, 1980). It was shown by S.M. Robinson (1980) that a Newton-type procedure applied to this general equation is locally quadratically convergent to the solution, provided that a property called *strong regularity* is satisfied. This approach has been successfully applied to a class of nonlinear cone-constrained optimization problems in infinite-dimensional spaces (Alt, 1990a; 1990b; 1990c) and optimal control problems subject to control and/or state constraints (see, e.g., Alt and Malanowski, 1993; 1995).

On the other hand, as early as at the beginning of the 1970s the so-called *shooting method* was proposed by R. Bulirsch (1971) (see Stoer and Bulirsch, 1980). This is a highly specialized method of numerically solving optimal control problems governed by ODEs. In the shoooting method for problems with inequality-type constraints, information on the *structure* of the optimal solution is crucial. Using this kind of information, the original optimization problem is reformulated as a problem with equality constraints. For the latter problem, the optimality system is expressed as a two- or multi-point boundary-value problem. This boundary-value problem is solved numerically, using the Newton method.

The literature devoted to Lagrange-Newton methods is enormous and this paper by no means pretends to give any survey of it. We just present, in a unified manner, the known covergence results for both of the above-mentioned approaches. The organization of the paper is the following: In Section 2 we briefly recall the Lagrange-Newton method for abstract optimization problems in Banach spaces, subject to equality and cone constraints, respectively. In Section 3 we introduce our model problem, which is an optimal control problem for nonlinear ODEs, subject to mixed control-state constraints. We present the application of the abstract approach to this problem and formulate assumptions under which the Lagrange-Newton method is locally quadratically convergent. In Section 4 we show how the additional information on the structure of the optimal control can be used to reformulate the problem as a problem with *equality*-type constraints. It is shown how the Lagrange-Newton procedure, applied to the latter problem, leads to the shooting method.

In the conclusion we give some comments on the scope of the applicability of each of the two presented methods.

We use the following notations: Capital letters $X, Y, Z, \Lambda, \dots$, sometimes with superscripts, denote Banach or Hilbert spaces. The norms are denoted by $\|\cdot\|$ with a subscript referring to the space. $\mathcal{O}_\rho^X(x_0) := \{x \in X \mid \|x - x_0\|_X < \rho\}$ is the open ball in X of radius ρ, centred at x_0. Asterisks denote dual spaces, as well as dual operators. Here (y, x), with $x \in X$ and $y \in X^*$, is a duality pairing between X and X^*.

For $f : X \times Y \to Z$, let $D_x f(x, y)$, $D_y f(x, y)$, $D^2_{xy} f(x, y), \dots$ denote the respective Fréchet derivatives in the corresponding arguments. $\mathbb{R}^n$ is the n-dimensional Euclidean space with the inner product denoted by $\langle x, y\rangle$ and the norm $|x| = \langle x, x\rangle^{\frac{1}{2}}$. Transposition is denoted by $*$.

$L^s(0, 1; \mathbb{R}^n)$, $s \in [1, \infty]$ are Banach spaces of measurable functions $f : [0, 1] \to \mathbb{R}^n$, supplied with the standard norms $\|\cdot\|_s$. $W^{1,s}(0, 1; \mathbb{R}^n)$ denotes Sobolev spaces of functions f which are absolutely continuous on $[0, 1]$ with the norms

$$\|f\|_{1,s} = \begin{cases} \big[|f(0)|^s + \|\dot{f}\|_s^s\big]^{1/s} & \text{for } s \in [1, \infty), \\ \max\big\{|f(0)|, \|\dot{f}\|_\infty\big\} & \text{for } s = \infty, \end{cases}$$

and c, l and ℓ denote generic constants, not necessarily the same in different places.

2. Lagrange–Newton Method for Abstract Optimization Problems in Banach Spaces

In this section we recall convergence results of the Lagrange-Newton method for abstract optimization problems subject to cone constraints, presented by Alt (1990c).

Let Z and Λ be Banach spaces of arguments and constraints, respectively. In the space Λ there is a closed convex cone K, which induces a partial order in that space. Further, let $F : Z \to \mathbb{R}$ and $\phi : Z \to \Lambda$. We consider the following optimization problem:

$$(\mathrm{P}) \quad \min F(z) \quad \text{subject to } \phi(z) \in K.$$

We make the following assumptions:

(A1) The mappings F and ϕ are twice Fréchet differentiable, with Lipschitz continuous second derivatives.

(A2) There exists a (local) solution $\widetilde{z}$ of (P).

Our purpose is to analyse the convergence of the Lagrange-Newton method, applied to (P), in a neighbourhood of $\widetilde{z}$. To formulate the Lagrange-Newton method, let us start with the problem subject to equality-type constraints, i.e., with the particular situation where $K = \{0\}$, and (P) reduces to

$$(\mathrm{P_e}) \quad \min F(z) \quad \text{subject to } \phi(z) = 0.$$

Let us introduce the following normal Lagrangian associated with (P_e):

$$\begin{aligned} \mathcal{L}_e : N &:= Z \times \Lambda^* \to \mathbb{R}, \\ \mathcal{L}_e(z, \lambda) &= F(z) + \big(\lambda, \phi(z)\big), \end{aligned} \tag{1}$$

and consider the first-order optimality system for (P_e):

$$\begin{aligned} &D_z\mathcal{L}_e(z, \lambda) := D_z F(z) + D_z\phi(z)^*\lambda = 0, \\ &\phi(z) = 0. \end{aligned} \tag{2}$$

We assume that there exists a Lagrange multiplier $\widetilde{\lambda} \in \Lambda^*$ such that (1) holds at $(\widetilde{z}, \widetilde{\lambda})$. Write $\eta := (z, \lambda) \in Z \times \Lambda^*$ and define

$$\begin{aligned} &\mathcal{F} : Z \times \Lambda^* \to Z^* \times \Lambda, \\ &\mathcal{F}(\eta) = \begin{pmatrix} D_z F(z) + D_z\phi(z)^*\lambda \\ \phi(z) \end{pmatrix}. \end{aligned} \tag{3}$$

In the Lagrange-Newton method, the Newton procedure is applied to the equation

$$\mathcal{F}(\eta) = 0, \tag{4}$$

i.e., starting with some initial element $\eta_1 := (z_1, \lambda_1)$, we construct the sequence $\{\eta_\alpha\}$, setting

$$D_\eta\mathcal{F}(\eta_\alpha)(\eta_{(\alpha+1)} - \eta_\alpha) + \mathcal{F}(\eta_\alpha) = 0. \tag{5}$$

Using the definition (3), we find that (5) amounts to

$$\begin{aligned} D^2_{zz}\mathcal{L}_e(z_\alpha, \lambda_\alpha)(z_{(\alpha+1)} - z_\alpha) + D_z\phi(z_\alpha)^*\lambda_{(\alpha+1)} \\ + D_z F(z_\alpha) = 0, \\ D_z\phi(z_\alpha)(z_{(\alpha+1)} - z_\alpha) + \phi(z_\alpha) = 0. \end{aligned} \tag{6}$$

Equations (6) can be interpreted as the optimality system for the following linear-quadratic optimization problem:

$$\begin{aligned} &(\mathrm{LP_e})_\alpha \\ &\min I_\alpha(z) := \frac{1}{2}\big((z - z_\alpha), D^2_{zz}\mathcal{L}_e(z_\alpha, \lambda_\alpha)(z - z_\alpha)\big) \\ &\qquad + \big(D_z F(z_\alpha), z\big), \\ &\text{subject to } D_z\phi(z_\alpha)(z - z_\alpha) + \phi(z_\alpha) = 0. \end{aligned}$$

Clearly, the Lagrange-Newton procedure is well defined in a neighbourhood $\mathcal{O}_\rho^N(\widetilde{\eta}) \subset N$ of the point

$\widetilde{\eta} := (\widetilde{x}, \widetilde{\lambda})$ if the Jacobian $D_\eta \mathcal{F}(\widetilde{\eta})$ is regular or, equivalently, if for any $\eta := (w, \nu) \in \mathcal{O}^N_\rho(\widetilde{\eta})$ the problem

$$\begin{aligned}&(\mathrm{QP_e})_\eta\\ &\min I_\eta(z) := \frac{1}{2}\big((z-w), D^2_{zz}\mathcal{L}_e(w,\nu)(z-w)\big)\\ &\qquad\qquad + \big(D_z F(w), z\big),\\ &\text{subject to } D_z\phi(w)(z-w) + \phi(w) = 0\end{aligned}$$

has a unique stationary point. Explicit conditions of regularity can be found, e.g., in Section 4.9.1 of (Bonnans and Shapiro, 2000).

Let us now pass to the cone-constrained problem (P). In the same way as in (1), we define the Lagrangian for (P):

$$\mathcal{L} : N \to \mathbb{R}, \quad \mathcal{L}(z,\lambda) = F(z) + \big(\lambda, \phi(z)\big). \tag{7}$$

The KKT (Karush-Kuhn-Tucker) optimality system for (P) has the form

$$\begin{aligned}&D_z F(z) + D_z\phi(z)^*\lambda = 0,\\ &\big(\lambda, \phi(z)\big) = 0, \quad \phi(z) \in K, \quad \lambda \in K^*.\end{aligned} \tag{8}$$

Define the following multivalued map, called the normal cone operator for K:

$$\begin{aligned}&\mathcal{N} : \Lambda^* \to 2^\Lambda,\\ &\mathcal{N}(\nu) = \begin{cases} \{y \in \Lambda \mid (\mu - \nu, y) \le 0 \quad \forall\, \mu \in K^*\} & \text{if } \nu \in K^*,\\ \emptyset & \text{if } \nu \notin K^*.\end{cases}\end{aligned} \tag{9}$$

In terms of $\mathcal{N}$, the three conditions in the second line of (8) can be written in the equivalent form $\phi(z) \in \mathcal{N}(\lambda)$. If we define the multivalued map

$$\mathcal{T} :\ N \to 2^\Delta, \tag{10}$$

where

$$\Delta := Z^* \times \Lambda,\ \mathcal{T}(\eta) = \begin{pmatrix} 0 \\ \mathcal{N}(\lambda)\end{pmatrix},$$

then, using (3), we can rewrite (8) in the form

$$\mathcal{F}(\eta) \in \mathcal{T}(\eta). \tag{11}$$

By analogy with (5) and (6), we define the Lagrange-Newton procedure for (11) by constructing the sequence $\{\eta_\alpha\}$, where

$$D_\eta\mathcal{F}(\eta_\alpha)(\eta_{(\alpha+1)} - \eta_\alpha) + \mathcal{F}(\eta_\alpha) \in \mathcal{T}(\eta_{(\alpha+1)}), \tag{12}$$

or, equivalently,

$$\begin{aligned}&D^2_{zz}\mathcal{L}(z_\alpha,\lambda_\alpha)(z_{(\alpha+1)} - z_\alpha)\\ &\qquad + D_z\phi(z_\alpha)^*\lambda_{(\alpha+1)} + D_z F(z_\alpha) = 0,\\ &D_z\phi(z_\alpha)(z_{(\alpha+1)} - z_\alpha) + \phi(z_\alpha) \in \mathcal{N}(\lambda).\end{aligned} \tag{13}$$

Just as in (6), we interpret (13) as the KKT optimality system for the following linear-quadratic optimal control problem:

$$\begin{aligned}&(\mathrm{LP})_\alpha && \min I_\alpha(z)\\ &\text{subject to} && D_z\phi(z_\alpha)(z - z_\alpha) + \phi(z_\alpha) \in K,\end{aligned}$$

where

$$\begin{aligned}I_\alpha(z) = {}& \frac{1}{2}\big((z - z_\alpha), D^2_{zz}\mathcal{L}_e(z_\alpha,\lambda_\alpha)(z - z_\alpha)\big)\\ & + \big(D_z F(z_\alpha), z\big).\end{aligned}$$

Thus, the Lagrange-Newton method reduces to an SQP-method (Alt, 1990a; 1990b; 1990c).

To analyse the convergence of the above Lagrange-Newton method, Robinson's implicit function theorem for strongly regular generalized equations is used (see, e.g., Alt, 1990a). We make the following assumption:

(A3) There exists $\widetilde{\lambda} \in K^*$ such that $(\widetilde{z}, \widetilde{\lambda})$ satisfies (8).

For any $\delta := (\delta^1, \delta^2) \in \Delta$, define the following accessory linear-quadratic problem:

$$\begin{aligned}&(\mathrm{QP})_\delta && \min I_{\widetilde{\eta}}(y) + (\delta^1, y),\\ &\text{subject to} && D_z\phi(\widetilde{z})(y - \widetilde{z}) + \phi(\widetilde{z}) + \delta^2 \in K,\end{aligned}$$

where

$$\begin{aligned}I_{\widetilde{\eta}}(y) := {}& \frac{1}{2}\Big((y - \widetilde{z}), D^2_{zz}\mathcal{L}(\widetilde{z},\widetilde{\lambda})(y - \widetilde{z})\Big)\\ & + (D_z F(\widetilde{z}), y).\end{aligned}$$

In addition to (A1)–(A3), we assume that

(A4) (*Strong regularity*) There exist constants $\rho_1 > 0, \rho_2 > 0$ and $l > 0$ such that, for each $\delta \in \mathcal{O}^\Delta_{\rho_1}(0)$, there is a unique stationary point $(y_\delta, \lambda_\delta) \in \mathcal{O}^N_{\rho_2}(\widetilde{\eta})$ of $(\mathrm{QP})_\delta$, and

$$\begin{aligned}&\|y_{\delta'} - y_{\delta''}\|_Z, \|\lambda_{\delta'} - \lambda_{\delta''}\|_\Lambda \le l\|\delta' - \delta''\|_\Delta,\\ &\qquad\qquad \forall \delta', \delta'' \in \mathcal{O}^\Delta_{\rho_1}(0).\end{aligned}$$

The following local convergence theorem for the Lagrange-Newton method holds (see Theorem 3.3 in (Alt, 1990a) or Lemma 7.2.3 in (Alt, 1990c)):

Theorem 1. *If Assumptions (A1)–(A4) are satisfied, then there exist constants* $\varrho > 0$, $c > 0$ *and* $h < 1$ *such that, for each initial point* $\eta_1 := (x_1, \lambda_1) \in \mathcal{O}_\varrho^N(\widetilde{\eta})$*, the Lagrange-Newton sequence* $\{\eta_\alpha\}$ *is well defined and*

$$\|\widetilde{\eta} - \eta_\alpha\|_N \leq c h^{2^\alpha - 1} \quad \textit{for} \ \alpha \geq 2.$$

Conditions of strong regularity for abstract cone constrained optimization problems can be found, e.g., in Section 5.1 of (Bonnans and Shapiro, 2000). Rather than to quote them, in the next section we proceed to a specific situation for optimal control problems.

3. SQP Method for Optimal Control Problems

In this section we introduce our model optimal control problem and apply to it the Lagrange-Newton procedure described in Section 2. We formulate conditions under which the assumptions of Theorem 1 are satisfied.

Consider the following optimal control problem:

(O)

$$\min_{(x,u)\in X^\infty} F(x,u) := \int_0^1 \varphi\big(x(t), u(t)\big)\,\mathrm{d}t + \psi\big(x(0), x(1)\big)$$

subject to

$$\begin{aligned} \dot{x}(t) - f\big(x(t), u(t)\big) &= 0 \quad \text{for a.a. } t \in [0,1],\\ \xi(x(0), x(1)) &= 0,\\ \theta\big(x(t), u(t)\big) &\leq 0 \quad \text{for a.a. } t \in [0,1], \end{aligned}$$

where

$$\begin{aligned} &X^\infty = W^{1,\infty}(0,1;\mathbb{R}^n) \times L^\infty(0,1;\mathbb{R}^m),\\ &\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, \qquad \psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R},\\ &f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n, \qquad \xi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^d,\\ &\theta : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k. \end{aligned}$$

We assume the following:

(B1) (*Data regularity*) The functions φ, ψ, f, ξ and θ are twice Fréchet differentiable in all their arguments and the derivatives are Lipschitz continuous.

(B2) (*Existence*) There exists a (local) solution $(\widetilde{x}, \widetilde{u})$ of (O).

By (B1) and (B2), conditions (A1) and (A2) are satisfied. To verify (A3), we need some constraint qualifications. To simplify notation, we set

$$\begin{aligned} &A(t) = D_x f\big(\widetilde{x}(t), \widetilde{u}(t)\big), \quad B(t) = D_u f\big(\widetilde{x}(t), \widetilde{u}(t)\big),\\ &\Xi_0 = D_{x(0)}\xi\big(\widetilde{x}(0), \widetilde{x}(1)\big), \quad \Xi_1 = D_{x(1)}\xi\big(\widetilde{x}(0), \widetilde{x}(1)\big),\\ &\Upsilon(t) = D_x\theta\big(\widetilde{x}(t), \widetilde{u}(t)\big), \quad \Theta(t) = D_u\theta\big(\widetilde{x}(t), \widetilde{u}(t)\big),\\ &I = \{1, \dots, k\}. \end{aligned} \tag{14}$$

For $\varepsilon \geq 0$, we introduce the sets of ε-active constraints

$$I^\varepsilon(t) = \big\{ i \in I \mid \theta^i\big(\widetilde{x}(t), \widetilde{u}(t)\big) \geq -\varepsilon \big\}, \tag{15}$$

and write

$$\begin{aligned} \Upsilon^\varepsilon(t) &= \big[D_x\theta^i\big(\widetilde{x}(t), \widetilde{u}(t)\big)\big]_{i\in I^\varepsilon(t)},\\ \Theta^\varepsilon(t) &= \big[D_u\theta^i\big(\widetilde{x}(t), \widetilde{u}(t)\big)\big]_{i\in I^\varepsilon(t)}. \end{aligned} \tag{16}$$

In addition to (B1) and (B2), we assume the following:

(B3) (*Linear independence*) There exist constants $\varepsilon, \beta > 0$ such that

$|\Theta^\varepsilon(t)^*\eta| \geq \beta|\eta|$ for all η of the appropriate dimensions and a.a. $t \in [0,1]$.

(B4) (*Controllability*) There is a $\varepsilon > 0$ such that, for each $e \in \mathbb{R}^d$, there exists $(y, v) \in X^\infty$, which satisfies the following equations:

$$\begin{aligned} \dot{y}(t) - A(t)y(t) - B(t)v(t) &= 0,\\ \Xi_0 y(0) + \Xi_1 y(1) &= e,\\ \Upsilon^\varepsilon(t)y(t) + \Theta^\varepsilon(t)v(t) &= 0. \end{aligned}$$

Introduce the space

$$Y^\infty := W^{1,\infty}(0,1;\mathbb{R}^n) \times \mathbb{R}^d \times L^\infty(0,1;\mathbb{R}^k),$$

and define the following Lagrangian and Hamiltonians:

$$\begin{aligned} &\mathcal{L} : X^\infty \times Y^\infty \to \mathbb{R}, \quad \mathcal{H} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R},\\ &\widehat{\mathcal{H}} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R},\\ &\mathcal{L}(x,u,p,\rho,\mu) = F(x,u) - \big(p, \dot{x} - f(x,u)\big)\\ &\qquad + \langle \rho, \xi(x(0), x(1))\rangle + \big(\mu, \theta(x,u)\big),\\ &\mathcal{H}(x,u,p) = \varphi(x,u) + \langle p, f(x,u)\rangle,\\ &\widehat{\mathcal{H}}(x,u,p,\mu) = \mathcal{H}(x,u,p) + \langle \mu, \theta(x,u)\rangle. \end{aligned} \tag{17}$$

It can be shown (see, e.g., Lemma 3.1 in (Malanowski, 2001)) that the following result holds:

Lemma 1. *If (B3) and (B4) hold, then there exists a unique Lagrange multiplier* $(\widetilde{p}, \widetilde{\rho}, \widetilde{\mu}) \in Y^\infty$ *such that the following KKT conditions are satisfied:*

$$\left.\begin{aligned} \dot{\widetilde{p}}(t) + D_x\widehat{\mathcal{H}}\big(\widetilde{x}(t), \widetilde{u}(t), \widetilde{p}(t), \widetilde{\mu}(t)\big) &= 0, \\ \widetilde{p}(0) + \Xi_0^*\widetilde{\rho} + D_{x(0)}\psi\big(\widetilde{x}(0), \widetilde{x}(1)\big) &= 0, \\ -\widetilde{p}(1) + \Xi_1^*\widetilde{\rho} + D_{x(1)}\psi\big(\widetilde{x}(0), \widetilde{x}(1)\big) &= 0, \end{aligned}\right\} \tag{18}$$

$$\left.\begin{aligned} &D_u\widehat{\mathcal{H}}\big(\widetilde{x}(t), \widetilde{u}(t), \widetilde{p}(t), \widetilde{\mu}(t)\big) = 0, \\ &\big\langle \widetilde{\mu}(t), \theta\big(\widetilde{x}(t), \widetilde{u}(t)\big)\big\rangle = 0, \quad \widetilde{\mu}(t) \geq 0. \end{aligned}\right\} \tag{19}$$

The above lemma shows that constraint qualifications ensure the existence of a normal Lagrange multiplier for (O), i.e., the abstract condition (A3) is satisfied. Moreover, the Lagrange multiplier is unique and more regular. In terms of the notation of Section 2, it belongs to Λ, rather than to Λ^*.

We define the following Lagrange-Newton procedure (LN1) for (O):

(1) Take $\eta_\alpha := (y_\alpha, v_\alpha, q_\alpha, \varrho_\alpha, \kappa_\alpha) \in X^\infty \times Y^\infty$.

(2) Find the stationary point

$$\eta_{(\alpha+1)} := \big(y_{(\alpha+1)}, v_{(\alpha+1)}, q_{(\alpha+1)}, \varrho_{(\alpha+1)}, \kappa_{(\alpha+1)}\big) \in X^\infty \times Y^\infty$$

of the following linear-quadratic optimal control problem:

$$(\mathrm{LO})_\alpha \qquad \min_{(y,v)\in X^\infty} \mathcal{I}_\alpha(y, v)$$

subject to

$$\begin{aligned} &\dot{y} - A_\alpha(y - y_\alpha) - B_\alpha(v - v_\alpha) - f(y_\alpha, v_\alpha) = 0, \\ &\Xi_{0\alpha}\big(y(0) - y_\alpha(0)\big) + \Xi_{1\alpha}\big(y(1) - y_\alpha(1)\big) \\ &\qquad\qquad + \xi\big(y_\alpha(0), y_\alpha(1)\big) = 0, \\ &\Upsilon_\alpha(y - y_\alpha) + \Theta_\alpha(v - v_\alpha) + \theta(y_\alpha, v_\alpha) \leq 0, \end{aligned}$$

where $A_\alpha, B_\alpha, \Xi_{0\alpha}, \Xi_{1\alpha}, \Upsilon_\alpha, \Theta_\alpha$ are defined as in (14), but evaluated at (y_α, v_α), while

$$\begin{aligned} \mathcal{I}_\alpha := \frac{1}{2}\big((&y - y_\alpha, v - v_\alpha) \\ &\times D^2\mathcal{L}(y_\alpha, v_\alpha, p_\alpha, \rho_\alpha, \mu_\alpha)(y - y_\alpha, v - v_\alpha)\big) \\ &+ (D_x\varphi(y_\alpha, v_\alpha), y) + (D_u\varphi(y_\alpha, v_\alpha), v) \\ &+ \langle D_{x(0)}\psi\big(y_\alpha(0), y_\alpha(1)\big), y(0)\rangle \\ &+ \langle D_{x(1)}\psi\big(y_\alpha(0), y_\alpha(1)\big), y(1)\rangle, \end{aligned}$$

with

$$\begin{aligned} \big((y, v), D^2\mathcal{L}(x, u, p, \rho, \mu)(y, v)\big) &:= \int_0^1 [y^*, v^*] \\ &\times \begin{bmatrix} D^2_{xx}\widehat{\mathcal{H}}(x, u, p, \mu) & D^2_{xu}\widehat{\mathcal{H}}(x, u, p, \mu) \\ D^2_{ux}\widehat{\mathcal{H}}(x, u, p, \mu) & D^2_{uu}\widehat{\mathcal{H}}(x, u, p, \mu) \end{bmatrix} \\ &\times \begin{bmatrix} y \\ v \end{bmatrix} \mathrm{d}t + [y(0)^*, y(1)^*] \\ &\times \begin{bmatrix} \mathcal{R}_{00}(x(0), x(1), \rho) & \mathcal{R}_{01}(x(0), x(1), \rho) \\ \mathcal{R}_{10}(x(0), x(1), \rho) & \mathcal{R}_{11}(x(0), x(1), \rho) \end{bmatrix} \\ &\times \begin{bmatrix} y(0) \\ y(1) \end{bmatrix}, \end{aligned} \tag{20}$$

where

$$\mathcal{R}_{rs} = D^2_{x(r)x(s)}\big(\xi(\widetilde{x}(0), \widetilde{\xi}(1))^*\widetilde{\rho} + \psi\big(\widetilde{x}(0), \widetilde{x}(1)\big)\big)$$
$$r = 0, 1, \; s = 0, 1.$$

(3) Increment α by 1 and go to (2).

In order for the Lagrange-Newton procedure to be well defined, problems $(\mathrm{LO})_\alpha$ must have unique stationary points. As in Section 2, to show the well-posedness and local convergence of the Lagrange-Newton procedure, we have to verify the strong regularity condition (A4).

Define the space of perturbations

$$\begin{aligned} \Delta := L^\infty(0, 1; \mathbb{R}^n) &\times L^\infty(0, 1; \mathbb{R}^m) \times \mathbb{R}^n \times \mathbb{R}^n \\ &\times L^\infty(0, 1; \mathbb{R}^n) \times \mathbb{R}^d \times L^\infty(0, 1; \mathbb{R}^k). \end{aligned} \tag{21}$$

For (O), the accessory problem analogous to $(\mathrm{QP})_\delta$ takes the form

$$(\mathrm{QO})_\delta \qquad \min_{(y,v)\in X^\infty} \mathcal{I}_\delta(y, v)$$

subject to

$$\begin{aligned} &\dot{y} - A(y - \widetilde{x}) - B(v - \widetilde{u}) - f(\widetilde{x}, \widetilde{u}) + \delta^5 = 0, \\ &\Xi_0\big(y(0) - \widetilde{x}(0)\big) + \Xi_1\big(y(1) - \widetilde{x}(1)\big) + \delta^6 = 0, \\ &\Upsilon(y - \widetilde{x}) + \Theta(v - \widetilde{u}) - \theta(\widetilde{x}, \widetilde{u}) + \delta^7 \leq 0, \end{aligned}$$

where $\delta := (\delta^1, \delta^2, \delta^3, \delta^4, \delta^5, \delta^6, \delta^7)$ and

$$\begin{aligned} \mathcal{I}_\delta(y, v) := \frac{1}{2}\big((&y - \widetilde{x}, v - \widetilde{u}), D^2\mathcal{L}(\widetilde{x}, \widetilde{u}, \widetilde{p}, \widetilde{\rho}, \widetilde{\mu}) \\ &\times (y - \widetilde{x}, v - \widetilde{u})\big) \\ &+ (D_x\varphi(\widetilde{x}, \widetilde{u}) + \delta^1, y) + (D_u\varphi(\widetilde{x}, \widetilde{u}) + \delta^2, v) \\ &+ \langle D_{x(0)}\psi\big(\widetilde{x}(0), \widetilde{x}(1)\big) + \delta^3, y(0)\rangle \\ &+ \langle D_{x(1)}\psi\big(\widetilde{x}(0), \widetilde{x}(1)\big) + \delta^4, y(1)\rangle. \end{aligned} \tag{22}$$

Just as in (15) and (16), for $\varepsilon \geq 0$ define

$$I_+^\varepsilon(t) = \{i \in I_0(t) \mid \widetilde{\mu}^i(t) > \varepsilon\},$$
$$\Upsilon_+^\varepsilon(t) = \left[D_x\theta^i(\widetilde{x}(t), \widetilde{u}(t)\right]_{i\in I_+^\varepsilon(t)}, \tag{23}$$
$$\Theta_+^\varepsilon(t) = \left[D_u\theta^i(\widetilde{x}(t), \widetilde{u}(t)\right]_{i\in I_+^\varepsilon(t)}.$$

In addition to (B1)–(B4), we assume the following:

(B5) (*Coercivity*) There exist $\varepsilon, \gamma > 0$ such that

$$\big((y,v), D^2\mathcal{L}\big(\widetilde{x},\widetilde{u},\widetilde{p},\widetilde{\rho},\widetilde{\mu}\big)\big)(y,v)\big) \geq \gamma(\|y\|_{1,2}^2 + \|v\|_2^2)$$

for all $(y,v) \in X^2$ such that

$$\dot{y}(t) - A(t)y(t) - B(t)v(t) = 0 \ \text{for a.a. } t \in [0,1],$$
$$\Xi_0 y(0) + \Xi_1 y(1) = 0,$$
$$\Upsilon_+^\varepsilon(t)y(t) + \Theta_+^\varepsilon(t)v(t) = 0 \ \text{for a.a. } t \in [0,1].$$

Remark 1. In the case when $\widetilde{u}(\cdot)$ and $\widetilde{\mu}(\cdot)$ are continuous functions and the conditions (B3)–(B5) are satisfied for $\varepsilon = 0$, they are also satisfied for $\varepsilon > 0$. Hence, in that case we can relax Assumptions (B3)–(B5) to $\varepsilon = 0$.

The following result can be found, e.g., in (Malanowski, 2001) (Proposition 5.4):

Lemma 2. *If Assumptions (B1)–(B5) are satisfied, then there exist constants* $\varsigma_1, \varsigma_2, \ell > 0$ *such that, for each* $\delta \in \mathcal{O}_{\varsigma_1}^{\Delta}(0)$, *there exists a unique stationary point* $(y_\delta, v_\delta, q_\delta, \varrho_\delta, \kappa_\delta) \in \mathcal{O}_{\varsigma_2}^{X^\infty\times Y^\infty}(\widetilde{\eta})$ *of* (QO)$_\delta$ *and*

$$\|y_{\delta'} - y_{\delta''}\|_{1,\infty}, \|v_{\delta'} - v_{\delta''}\|_\infty, \|q_{\delta'} - q_{\delta''}\|_{1,\infty},$$
$$|\varrho_{\delta'} - \varrho_{\delta''}|, \|\kappa_{\delta'} - \kappa_{\delta''}\|_\infty \leq \ell\|\delta' - \delta''\|_\Delta$$

for all $\delta', \delta'' \in \mathcal{O}_{\varsigma_2}^{X^\infty\times Y^\infty}(\widetilde{\eta})$.

Lemma 2 implies that Assumption (A4) is satisfied, and by Theorem 1 we obtain the following result:

Theorem 2. *If Assumptions (B1)–(B5) hold, then there exist constants* $\sigma > 0$, $c > 0$ *and* $h < 1$ *such that, for each initial point* $\eta_1 := (y_1, v_1, q_1, \varrho_1, \kappa_1) \in \mathcal{O}_\sigma^{X^\infty\times Y^\infty}(\widetilde{\eta})$, *the Lagrange-Newton procedure (LN1) is well defined and*

$$\|\widetilde{\eta} - \eta_\alpha\|_{X^\infty\times Y^\infty} \leq ch^{2^\alpha - 1} \quad \text{for } \alpha \geq 2.$$

4. Shooting Method

Theorem 2 was derived without any information on the form of the optimal solution. We were only assuming that some optimal control exists in the class of essentially bounded functions. Now, we will consider the situation where the optimal control is a continuous function of time and the number and order of active and nonactive constraints are known. This kind of information allows us to formulate our original optimal control problem as a problem with *equality* constraints. The Lagrange-Newton procedure for such problems leads to the well-known *shooting method* (see, e.g., (Bulirsch, 1971; Stoer and Bulirsch, 1980)).

Let us introduce the sets

$$\widetilde{\Omega}^i = \big\{t \in [0,1] \mid \theta^i\big(\widetilde{x}(t), \widetilde{u}(t)\big) = 0\big\}, \quad i \in I, \tag{24}$$

of those points at which the constraints are active for the optimal solution. Assume the following:

(C1) (*Solution structure*) The optimal control $\widetilde{u}$ is a continuous function. Each of the sets $\widetilde{\Omega}^i$, $i \in I$ consists of a finite number J^i of disjoint subintervals:

$$\widetilde{\Omega}^i = \cup_{j\in J^i}[\widetilde{\omega}_j^{i'}, \widetilde{\omega}_j^{i''}] \in (0,1).$$

There are no isolated touch points and none of the junction points $\widetilde{\omega}_j^{i'}$ or $\widetilde{\omega}_j^{i''}$ coincide with each other for any $i \in I$.

Set $\jmath = 2\sum_{i\in I} J^i$ and define the $(\jmath + 2)$-dimensional vector $\widetilde{\omega} := [0, \widetilde{\omega}_1, \ldots, \widetilde{\omega}_\jmath, 1]$, where the $\widetilde{\omega}_j$s are junction points for all constraints, arranged in an increasing order. Clearly, for each subinterval $(\widetilde{\omega}_j, \widetilde{\omega}_{j+1})$ a fixed set of constraints is active along $(\widetilde{x}, \widetilde{u})$. Write

$$\imath_j = \big\{i \in I \mid \theta^i\big(\widetilde{x}(t), \widetilde{u}(t)\big) = 0 \text{ for } t \in (\widetilde{\omega}_j, \widetilde{\omega}_{j+1})\big\}.$$

We can interpret $(\widetilde{x}, \widetilde{u})$ as a solution of the following optimal control problem ($\widehat{\text{O}}$) subject to *equality* constraints, active at a given number of subintervals, where the locations of these subintervals, i.e., of the corresponding entry and exit points become additional arguments of optimization. Namely,

$$(\widehat{\text{O}}) \qquad \min_{(x,u,\omega)} F(x,u)$$

subject to

$$\dot{x}(t) - f\big(x(t), u(t)\big) = 0 \text{ for a.a. } t \in [0,1],$$
$$\xi\big(x(0), x(1)\big) = 0,$$
$$\theta^i\big(x(t), u(t)\big) = 0$$

for all $t \in (\omega_j, \omega_{j+1}),\ i \in \imath_j,\ j = 1, \ldots, \jmath + 1,$

where the minimization is performed over the class of control functions which are piecewise C^1, with possible jumps at all junction points.

Setting $\mu^i(t) = 0$ for $t \notin (\omega_j^{i'}, \omega_j^{i''})$, we find that the Lagrangian and Hamiltonians for ($\widehat{\mathrm{O}}$) are given by (17). The constraints, together with the stationarity conditions of the Lagrangian, with respect to u and x, constitute the following system of equations:

$$\dot{x}(t) - f\big(x(t), u(t)\big) = 0, \tag{25}$$

$$\xi\big(x(0), x(1)\big) = 0, \tag{26}$$

$$\theta^i\big(x(t), u(t)\big) = 0 \quad \text{for } t \in (\omega_j, \omega_{j+1}), \quad i \in \imath_j,\ j = 1, \dots, \jmath + 1, \tag{27}$$

$$\dot{p}(t) + D_x\widehat{\mathcal{H}}\big(x(t), u(t), p(t), \mu(t)\big) = 0, \tag{28}$$

$$p(0) + D_{x(0)}\left[\xi_0^*\rho + \psi\big(x(0), x(1)\big)\right] = 0, \tag{29}$$

$$-p(1) + D_{x(1)}\left[\xi_1^*\rho + \psi\big(x(0), x(1)\big)\right] = 0, \tag{30}$$

$$D_u\widehat{\mathcal{H}}\big(x(t), u(t), p(t), \mu(t)\big) = 0. \tag{31}$$

Since in Problem ($\widehat{\mathrm{O}}$) optimization is performed also with respect to the vector ω of the junction points, we have to find stationarity conditions of the Lagrangian with respect to ω. These conditions yield

$$\varphi\big(x(t), u(t-)\big) = \varphi\big(x(t), u(t+)\big)$$

for all $t = \omega_j,\ j = 1, \dots, \jmath + 1$.

Clearly, the above conditions are satisfied if $u(\cdot)$ is a continuous function. In turn, the continuity of u implies, in particular,

$$\left.\begin{aligned} \theta^i\big(x(\omega_j^{i'}), u(\omega_j^{i'}-)\big) &= 0 \\ \theta^i\big(x(\omega_j^{i''}), u(\omega_j^{i''}+)\big) &= 0 \end{aligned}\right\} \ \forall j \in J^i,\ i \in I. \tag{32}$$

On the other hand, it can be shown (see Section 2 in (Malanowski and Maurer, 1996a)) that the conditions (B1)–(B3) and (B5), suplemented with (31), imply the continuity of u. Hence, we will treat (32) as stationarity conditions of $\mathcal{L}$ with respect to ω.

It will be convenient to eliminate u and μ from (25)–(32). To this end, note that, on each subinterval (ω_j, ω_{j+1}), the condition (31), together with (27), can be interpreted as stationarity conditions for the following parametric mathematical program, subject to *equality* constraints:

$$(\mathrm{MP})_j\big(x(t), p(t)\big) \qquad \min_{u \in \mathbb{R}^m} \mathcal{H}\big(x(t), u, p(t)\big)$$
subject to $\theta^i(x(t), u) = 0$ for $i \in \imath_j$.

This program depends on the vector parameter $(x(t), p(t)) \in \mathbb{R}^{2n}$. In view of (B1)–(B3) and (B5), there exist twice continuously differentiable functions

$$\eta_j : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m, \quad \chi_j : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^k$$

such that, for any $(x(t), p(t))$ in a neighbourhood of $(\widetilde{x}(t), \widetilde{p}(t))$,

$u(t) = \eta_j\big(x(t), p(t)\big)$ and $\mu(t) = \chi_j\big(x(t), p(t)\big)$ is a locally unique solution and a Lagrange multiplier of $(\mathrm{MP})_j\big(x(t), p(t)\big)$, i.e., (33)

$$\widetilde{u}(t) = \eta_j\big(\widetilde{x}(t), \widetilde{p}(t)\big),\ \widetilde{\mu}(t) = \chi_j\big(\widetilde{x}(t), \widetilde{p}(t)\big) \quad \text{for } t \in (\widetilde{\omega}_j, \widetilde{\omega}_{j+1}).$$

Using η_j and χ_j, we can rewrite the stationarity conditions (25)–(32) in the form of the following *multi-point boundary-value problem* for (x, p):

$$\left.\begin{aligned} \dot{x}(t) - f\left(x(t), \eta_j\big(x(t), p(t)\big)\right) &= 0, \\ \dot{p}(t) - D_x\widehat{\mathcal{H}}\Big(x(t), \eta_j\big(x(t), p(t)\big), p(t),& \\ \chi_j\big(x(t), p(t)\big)\Big) &= 0 \\ \text{for } t \in (\omega_j, \omega_{j+1}) \text{ and } j = 0, \dots, \jmath + 1.& \end{aligned}\right\} \tag{34}$$

$$\left.\begin{aligned} \xi\big(x(0), x(1)\big) &= 0, \\ p(0) + D_{x(0)}\Big[\xi\big(x(0), x(1)\big)^*\rho& \\ +\psi\big(x(0), x(1)\big)\Big] &= 0, \\ -p(1) + D_{x(1)}\Big[\xi\big(x(0), x(1)\big)^*\rho& \\ +\psi\big(x(0), x(1)\big)\Big] &= 0, \end{aligned}\right\} \tag{35}$$

$$\left.\begin{aligned} \theta^i(x(\omega_j^{i'}), \eta_{j-1}\big(x(\omega_j^{i'}), p(\omega_j^{i'})\big)) &= 0, \\ \theta^i(x(\omega_j^{i''}), \eta_{j+1}\big(x(\omega_j^{i''}), p(\omega_j^{i''})\big)) &= 0, \\ j \in J^i,\ i \in I.& \end{aligned}\right\} \tag{36}$$

Note that the solution to (34) is uniquely defined by the $2n$-dimensional vector $a = (x(0), p(0))$ of the initial conditions. Hence the system (34)–(36) can be expressed as the following equation in $\mathbb{R}^{2n+d+\jmath}$:

$$\mathcal{F}(a, \rho, \omega) = 0, \tag{37}$$

where

$$\mathcal{F}(a, \rho, \omega) = \begin{pmatrix} \mathcal{F}_1(a, \rho, \omega) \\ \mathcal{F}_2(a, \omega) \end{pmatrix},$$

with $\mathcal{F}_1$ and $\mathcal{F}_2$ given by the left-hand sides of (35) and (36), respectively. Clearly, $\mathcal{F}(\widetilde{a}, \widetilde{\rho}, \widetilde{\omega}) = 0$, where $\widetilde{a} = (\widetilde{x}(0), \widetilde{p}(0))$.

In the *shooting method* (LN2) the classical Newton procedure is applied to (37). This method is well defined and locally quadratically convergent if the Jacobian $D\mathcal{F}(\tilde{a},\tilde{\rho},\tilde{\omega})$ is regular, i.e., if the equation

$$\begin{bmatrix} D_a\mathcal{F}_1(\tilde{a},\tilde{\rho},\tilde{\omega}) & D_\rho\mathcal{F}_1(\tilde{a},\tilde{\rho},\tilde{\omega}) & D_\omega\mathcal{F}_1(\tilde{a},\tilde{\rho},\tilde{\omega}) \\ D_a\mathcal{F}_2(\tilde{a},\tilde{\rho},\tilde{\omega}) & 0 & D_\omega\mathcal{F}_2(\tilde{a},\tilde{\rho},\tilde{\omega}) \end{bmatrix} \times \begin{bmatrix} b \\ \varrho \\ \varpi \end{bmatrix} = \begin{bmatrix} r \\ s \end{bmatrix} \tag{38}$$

has a unique solution for any $r := (r^1, r^2, r^3) \in \mathbb{R}^{n+n+d}$ and $s \in \mathbb{R}^{\jmath}$. Note that

$$D_\omega\mathcal{F}_1(\tilde{a},\tilde{\rho},\tilde{\omega}) = 0. \tag{39}$$

This follows from the fact that, by the well-known properties of the solutions to ODEs and by the continuity of $\tilde{u}(\cdot)$, we have (see Maurer and Pesch, 1994):

$$\frac{\partial\tilde{x}}{\partial\omega}(t) = 0 \;\text{ and }\; \frac{\partial\tilde{p}}{\partial\omega}(t) = 0. \tag{40}$$

Thus (38) reduces to

$$\begin{bmatrix} D_a\mathcal{F}_1(\tilde{a},\tilde{\rho},\tilde{\omega}) & D_\rho\mathcal{F}_1(\tilde{a},\tilde{\rho},\tilde{\omega}) \end{bmatrix} \begin{bmatrix} b \\ \varrho \end{bmatrix} = r, \tag{41}$$

$$\begin{bmatrix} D_a\mathcal{F}_2(\tilde{a},\tilde{\omega}) & D_\omega\mathcal{F}_2(\tilde{a},\tilde{\omega}) \end{bmatrix} \begin{bmatrix} b \\ \varpi \end{bmatrix} = s. \tag{42}$$

In view of (33) and (40), $D_\omega\mathcal{F}(\tilde{a},\tilde{\rho},\tilde{\omega})$ is a diagonal matrix, with the diagonal elements given by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\theta^i(\tilde{x}(\tilde{\omega}_j^{i'}), \eta_{j-1}(x(\tilde{\omega}_j^{i'}), p(\tilde{\omega}_j^{i'}) &= \frac{\mathrm{d}}{\mathrm{d}t}\theta^i\big(\tilde{x}(t),\tilde{u}(t)\big)\big|_{t=\omega_j^{i'}-} \\ &= D_x\theta^i(\tilde{x}(\tilde{\omega}_j^{i'}), \tilde{u}(\tilde{\omega}_j^{i'}))\dot{\tilde{x}}(\tilde{\omega}_j^{i'}) \\ &\quad + D_u\theta^i(\tilde{x}(\tilde{\omega}_j^{i'}), \tilde{u}(\tilde{\omega}_j^{i'}))\dot{\tilde{u}}(\tilde{\omega}_j^{i'}-), \end{aligned}$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\theta^i(\tilde{x}(\tilde{\omega}_j^{i''}), \eta_{j+1}(x(\tilde{\omega}_j^{i''}), p(\tilde{\omega}_j^{i''}) &= \frac{\mathrm{d}}{\mathrm{d}t}\theta^i\big(\tilde{x}(t),\tilde{u}(t)\big)\big|_{t=\omega_j^{i''}+} \\ &= D_x\theta^i(\tilde{x}(\tilde{\omega}_j^{i''}), \tilde{u}(\tilde{\omega}_j^{i''}))\dot{\tilde{x}}(\tilde{\omega}_j^{i''}) \\ &\quad + D_u\theta^i(\tilde{x}(\tilde{\omega}_j^{i''}), \tilde{u}(\tilde{\omega}_j^{i''}))\dot{\tilde{u}}(\tilde{\omega}_j^{i''}+). \end{aligned}$$

This shows that, for any $b \in \mathbb{R}^{2n}$ and $s \in \mathbb{R}^{\jmath}$, (42) has a unique solution, if the following condition holds:

(C2) (*Nontangential junction*) At all junction points along the optimal trajectory, the following conditions are satisfied:

$$\left.\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\theta^i\big(\tilde{x}(t),\tilde{u}(t)\big)\big|_{t=\tilde{\omega}_j^{i'}-} \neq 0, \\ \frac{\mathrm{d}}{\mathrm{d}t}\theta^i\big(\tilde{x}(t),\tilde{u}(t)\big)\big|_{t=\tilde{\omega}_j^{i''}+} \neq 0 \end{aligned}\right\} \; j \in J^i,\ i \in I.$$

Thus, if (C2) holds, the Jacobian $D\mathcal{F}(\tilde{a},\tilde{\rho},\tilde{\omega})$ is regular provided that (41) has a unique solution for any $r := (r^1, r^2, r^3) \in \mathbb{R}^{n+n+d}$. Some calculations, similar to those in Section 2 of (Malanowski and Maurer, 1996a) and Section 5 in (Malanowski and Maurer, 1996b), show that any solution of (41) is equivalent to a stationary point of the following linear-quadratic accessory problem analogous to $(\mathrm{QO})_\delta$:

$$(\widehat{\mathrm{QO}})_r \qquad \min_{(y,v)\in X^\infty} \widehat{\mathcal{I}}_{\tilde{\eta}}(y,v,r)$$

subject to

$$\begin{aligned} \dot{y}(t) - A(t)y(t) - B(t)v(t) &= 0, \\ \Xi_0 y(0) + \Xi_1 y(1) + r^3 &= 0, \\ \langle \Upsilon^i(t), y(t)\rangle + \langle \Theta^i(t), v(t)\rangle &= 0 \end{aligned}$$

for all $t \in (\omega_j, \omega_{j+1}),\ i \in \imath_j,\ j = 1, \dots, \jmath + 1,$

where $\Upsilon^i(t)$ and $\Theta^i(t)$ are the i-th rows of $\Upsilon(t)$ and $\Theta(t)$, respectively, while

$$\widehat{\mathcal{I}}_{\tilde{\eta}}(y,v,r) := \frac{1}{2}(y, D^2\mathcal{L}(\tilde{x},\tilde{u},\tilde{p},\tilde{\rho},\tilde{\mu})y) + \langle r^1, y(0)\rangle + \langle r^2, y(1)\rangle.$$

In the same way as in the case of the accessory problem $(\mathrm{QO})_\delta$, we find that the conditions (B3)–(B5) imply that, for any $r \in R^{2n+d}$, Problem $(\widehat{\mathrm{QO}})_r$ has a unique stationary point. Thus, we have arrived at the following results:

Lemma 3. *If Assumptions (B1)–(B5) and (C1)–(C2) hold, then the Jacobian* $D\mathcal{F}(a,\rho,\omega)$ *is regular at* $(\tilde{a},\tilde{\rho},\tilde{\omega})$.

By Lemma 3 the shooting method (LN2) is locally quadratically convergent to the stationary point $(\tilde{a},\tilde{\rho},\tilde{\omega})$, i.e., for any $(b_1,\varrho_1,\varpi_1) \in \mathbb{R}^{2n+d+\jmath}$ the generated sequence $\{(b_\alpha,\varrho_\alpha,\varpi_\alpha)\}$ satisfies

$$\begin{aligned} &\big|(\tilde{a} - b_{(\alpha+1)}, \tilde{\rho} - \varrho_{(\alpha+1)}, \tilde{\omega} - \varpi_{(\alpha+1)})\big| \\ &\qquad \le c\,\big|(\tilde{a} - b_\alpha, \tilde{\rho} - \varrho_\alpha, \tilde{\omega} - \varpi_\alpha)\big|^2. \end{aligned}$$

Clearly, in a neighbourhood of $(\tilde{a},\tilde{\omega})$, there is a one-to-one correspondence between any vector $(b,\varpi) \in \mathbb{R}^{2n+\jmath}$ of the initial state and the junction points, and the solution $(x,p) \in W^{1,\infty}(0,1;\mathbb{R}^n) \times W^{1,\infty}(0,1;\mathbb{R}^n)$ of the state and adjoint equations (34). On the other hand, by (33), the corresponding control $u \in L^\infty(0,1;\mathbb{R}^m)$ and the Lagrange multiplier $\mu \in L^\infty(0,1;\mathbb{R}^m)$ depend continuously on (x,p) and ϖ. Hence we finally obtain the following convergence result analogous to Theorem 2:

Theorem 3. *If the assumptions (B1)–(B5) and (C1)–(C2) hold, then there exist constants* $\sigma > 0$, $c > 0$ *and* $h < 1$ *such that, for each initial point* $(a_1, \rho_1, \omega_1) \in \mathcal{O}_\sigma^{\mathbb{R}^{2n+d+\jmath}}(\widetilde{a}, \widetilde{\rho}, \widetilde{\omega})$, *the shooting method (LN2) is well defined. The sequence* $\{\eta_\alpha = (x_\alpha, u_\alpha, p_\alpha, \rho_\alpha, \mu_\alpha)\}$, *corresponding to* $\{(b_\alpha, \varrho_\alpha, \varpi_\alpha)\}$, *converges quadratically to* $\widetilde{\eta}$:

$$\|\widetilde{\eta} - \eta_\alpha\|_{X^\infty \times Y^\infty} \le ch^{2^\alpha - 1} \quad \textit{for } \alpha \ge 2.$$

In addition to that, the sequence of the junction points $\{\varpi_\alpha\}$ *converges quadratically to* $\widetilde{\omega}$.

5. Concluding Remarks

The results presented in Sections 3 and 4 show that the assumptions required for the well-posedness and local quadratic convergence of the SQP algorithm (LP1) are substentially weaker than those that ensure the same properties for the shooting method (LP2). The latter method requires additional assumptions: (C1) – on the structure of the optimal control, and (C2), which ensure that this structure is preserved in a neighbourhood of the reference solution. On the other hand, Algorithm (LP2) is more convenient from the numerical point of view, since it reduces to the Newton procedure for equations, while in (LP1) a linear-quadratic optimal control problem, subject to inequality-type constraints, has to be solved in each step.

Convergence results, similar to those presented here, occur for both algorithms applied to optimal control problems, where, in addition to mixed constraints, also pure state space constraints of *order one* are present (Alt and Malanowski, 1995; Malanowski and Maurer, 1996b). However, in this case the convergence analysis for the SQP method is much more complicated than that reported here, due to the presence of the so called *two norm discrepancy* (see e.g., Dontchev and Hager, 1998; Malanowski, 1994; 1995). To overcome this difficulty, some additional information on the regularity of the optimal solution is exploited (see Alt and Malanowski, 1995).

The scope of the applicability of the shooting method seems to be broader than that of (LP1). The point is that the latter is a general iterative algorithm for constrained optimization problems in functional spaces, whereas the shooting method is a technique specialized for optimal control problems governed by ODEs. In particular, it seems that for *higher-order* state constrained problems, one cannot avoid information on the structure of the optimal control. It is connected with the fact that higher-order state constraints can be viewed as cone constraints in spaces $W^{p,\infty}(0,1;\mathbb{R}^k)$, with $p > 1$. The analysis of projection onto such cones is difficult and requires more information on the projected element. At least in some cases of higher-order state constraints the shooting method was used and the local quadratic convergence ensured (see, e.g., Malanowski and Maurer, 2001).

Similarly, the shooting method can be extended to problems with free final time (Maurer and Oberle, 2002), whereas the algorithm (LP1) can be hardly applied there.

Throughout this paper we have assumed the coercivity of the Hessian of the Lagrangian. This assumption excludes an important class of optimal control, where the solution is of the bang-bang type. Clearly, for problems with the bang-bang solutions the local stability of the structure of the optimal solution is crucial for the convergence of Newton-type methods. Recent results concerning second-order optimality conditions and sensitivity analysis for this class of problems (Agrachev *et al.*, 2002; Felgenhauer, 2002; Kim and Maurer, 2003; Maurer and Osmolovskii, 2004) suggest that the local convergence results of the shooting method can be extended to some problems with bang-bang solutions.

Acknowledgment

The author would like to express his gratitude to the anonymous referees for their careful reading of the manuscript and the comments which allowed him to improve the presentation.

References

Agrachev A.A., Stefani G. and Zezza P.L. (2002): *Strong optimality for a bang-bang trajectory*. — SIAM J. Contr. Optim., Vol. 41, No. 4, pp. 991–1014.

Alt W. (1990a): *Lagrange-Newton method for infinite-dimensional optimization problems*. — Numer. Funct. Anal. Optim., Vol. 11, No. 3/4, pp. 201–224.

Alt W. (1990b): *Parametric programming with applications to optimal control and sequential quadratic programming*. — Bayreuther Math. Schriften, Vol. 34, No. 1, pp. 1–37.

Alt W. (1990c): *Stability of solutions and the Lagrange-Newton method for nonlinear optimization and optimal control problems*. — (Habilitationsschrift), Universität Bayreuth, Bayreuth.

Alt W. and Malanowski K. (1993): *The Lagrange-Newton method for nonlinear optimal control problems*. — Comput. Optim. Appl., Vol. 2, No. 1, pp. 77–100.

Alt W. and Malanowski K. (1995): *The Lagrange-Newton method for state constrained optimal control problems*. — Comput. Optim. Appl., Vol. 4, No. 3, pp. 217–239.

Bonnans J.F. and Shapiro A. (2000): *Perturbation Analysis of Optimization Problem*. — New York: Springer.

Bulirsch R. (1971): *Die Mehrzielmethode zur numerischen Lösung von nichtlinearen Randwertproblemen und Aufgaben der optimalen Steuerung*. — Report of the Carl-Cranz-Gesellschaft, Oberpfaffenhofen, 1971.

Dontchev A.L. and Hager W.W. (1998): *Lipschitz stability for state constrained nonlinear optimal control*. — SIAM J. Contr. Optim., Vol. 35, No. 2, pp. 696–718.

Felgenhauer U. (2002): *On stability of bang-bang type controls*. — SIAM J. Contr. Optim., Vol. 41, No. 6, pp. 1843–1867.

Kim J.-H.R. and Maurer H. (2003): *Sensitivity analysis of optimal control problems with bang-bang controls*. — Proc. 42nd IEEE Conf. *Decision and Control*, *CDC'2003*, Maui, Hawaii, USA, pp. 3281–3286.

Malanowski K. (1994): *Regularity of solutions in stability analysis of optimization and optimal control problems*. — Contr. Cybern., Vol. 23, No. 1/2, pp. 61–86.

Malanowski K. (1995): *Stability and sensitivity of solutions to nonlinear optimal control problems*. — Appl. Math. Optim., Vol. 32, No. 2, pp. 111–141.

Malanowski K. (2001): *Stability and sensitivity analysis for optimal control problems with control-state constraints*. — Dissertationes Mathematicae, Vol. CCCXCIV, pp. 1–51.

Malanowski K. and Maurer H. (1996a): *Sensitivity analysis for parametric optimal control problems with control-state constraints*. — Comput. Optim. Appl., Vol. 5, No. 3, pp. 253–283.

Malanowski K. and Maurer H. (1996b): *Sensitivity analysis for state-constrained optimal control problems*. — Discr. Cont. Dynam. Syst., Vol. 4, No. 2, pp. 241–272.

Malanowski K. and Maurer H. (2001): *Sensitivity analysis for optimal control problems subject to higher order state constraints*. — Ann. Oper. Res., Vol. 101, No. 2, pp. 43–73.

Maurer H. and Oberle J. (2002): *Second order sufficient conditions for optimal control problems with free final time: the Riccati approach*. — SIAM J. Contr. Optim., Vol. 41, No. 2, pp. 380–403.

Maurer H. and Osmolovskii N. (2004): *Second order optimality conditions for bang-bang control problems*. — Contr. Cybern., Vol. 32, No. 3. pp. 555–584.

Maurer H. and Pesch H.J. (1994): *Solution differentiability for parametric optimal control problems with control-state constraints*. — Contr. Cybern., Vol. 23, No. 1, pp. 201–227.

Robinson S.M. (1980): *Strongly regular generalized equations*. — Math. Oper. Res., Vol. 5, No. 1, pp. 43–62.

Stoer J. and Bulirsch R. (1980): *Introduction to Numerical Analysis*. — New York: Springer.