Dedicated to
Professor Czesław Olech

Control and Cybernetics

vol. **34** (2005) No. 3

Optimality properties of controls with bang-bang components in problems with semilinear state equation

by

Ursula Felgenhauer

Brandenburgische Technische Universität Cottbus, Institut für Mathematik
PF 101344, 03013 Cottbus, Germany
e-mail: felgenh@math.tu-cottbus.de

Abstract: In this paper we study optimal control problems with bang-bang solution behavior for a special class of semilinear dynamics. Generalizing a former result for linear systems, optimality conditions are derived by a duality based approach. The results apply for scalar as well as for vector control functions and, in particular, for the case of the so-called multiple switches, too.

Further, an iterative procedure for determining switching points is proposed, and convergence results are provided.

Keywords: optimality conditions, Riccati equation, bang-bang control, switching points optimization, sensitivity.

1. Introduction

The paper contributes to the recently widely discussed field of optimality conditions for bang-bang type optimal controls, see e.g. Sarychev (1997), Osmolovskii (2000), Milyutin and Osmolovskii (1998), Noble and Schaettler (2002), Agrachev, Stefani and Zezza (2002). Continuing former investigations (Felgenhauer, 2001a, b, 2003a) a duality based concept from Klötzler (1979), Maurer and Pickenhain (1995) is applied to derive sufficient optimality conditions. In contrast to Milyutin and Osmolovskii (1998) (or also Agrachev, Stefani and Zezza, 2002), the analysis is applicable for multiple switches of several control components, too. The proofs use variation estimates without control linearization or approximating cones (see Milyutin and Osmolovskii, 1998; Osmolovskii, 2000), and thus, directly yield strong local optimality results (see Theorem 3.2).

After introducing the problem in Section 2, the basic ideas of the concept are shortly described in Section 3. In the next two sections, particular types of conditions are proved: at first, the problem is considered in case of convex terminal functional under convexity assumptions on the Hamiltonian w.r.t. the state variable. In this special case, the test functions needed in the duality approach can be trivially found. Secondly, a certain Riccati type condition (Theorem 5.1, Section 5) is derived where the matrix solution is considered being piecewise continuous.

In Section 6, it is shown that the criteria obtained further guarantee strong optimality of the switching points position. We recall the finite-dimensional problem where minimization is performed over switching times as considered in Agrachev, Stefani and Zezza (2002) and find explicit representations for second variations of the objective functional. The result is compared to quadratic forms used in Milyutin and Osmolovskii (1998), Osmolovskii and Lempio (2002).

The final section is devoted to a primal-dual Newton method for iterating switching points, and a constructive optimality test for the auxiliary finite-dimensional problem is described.

Throughout the paper, the following notations are used:

The Euclidean vector space of dimension k is R^k with the norm $|\cdot|$ and scalar product $u^T v, \quad u, v \in R^k$. The superscript T herein and in general matrix calculations denotes the transposed matrix, or the (raw-)vector. Further, the Lebesgue space of measurable vector functions on $[0,1]$ with integrable $|\cdot|^p$ is written as $L_p(0,1;R^k)$, and W_p^m stands for the related Sobolev space of order m. The norm in L_p is given by $\|\cdot\|_p, \;\; 1 \le p \le \infty$. For continuously differentiable functions, we use spaces C^m. The (possibly partial) gradients and Hessian matrices are written as $\nabla_{(\cdot)}$ resp. $\nabla^2_{(\cdot)}$ where the subscripts refer to particular variables.

2. The problem. Regularity conditions

We consider the following optimal control problem with terminal functional and a *semilinear* state equation on the time interval $[0,1]$:

$$\textbf{(P)} \qquad \min \; J(x,u) \;=\; k(x(1))$$

$$s.t. \qquad \dot{x}(t) \;= f(t, x(t)) \;+\; B(t)\, u(t) \qquad a.e. \text{ in } [0,1], \tag{1}$$

$$x(0) \;=\; a, \tag{2}$$

$$|\, u_i(t)| \;\le\; 1, \quad i = 1, \ldots, m, \qquad a.e. \text{ in } [0,1]\,. \tag{3}$$

Notice that the control vector enters the state equation linearly, and the matrix $B = B(t)$ is independent of the state x.

The pair $(x,u) \in W^1_\infty(0,1;R^n) \times L_\infty(0,1;R^m)$ is called *admissible* for (P) if the state equation (1) together with the boundary condition (2) and the control constraints (3) are fulfilled. All data functions are assumed to be sufficiently smooth, e.g. k, f are supposed to be twice continuously differentiable functions with uniformly Lipschitz-continuous second derivatives on each compact $D \subset R^n$ respectively $K = [0,1] \times D$, and $B \in C^2([0,1])$.

An admissible pair (x^0, u^0) is called a (*global*) minimizer for (P), if $J(x^0, u^0) \leq J(x, u)$ for all admissible (x, u). If for some $\epsilon > 0$ the inequality holds for any admissible (x, u) with $\| x - x^0 \|_\infty < \epsilon$ then (x^0, u^0) is called a *strong local* minimizer for (P).

For the above problem, necessary optimality conditions are given by Pontryagin's *maximum principle.* Using the Hamiltonian function

$$H(t, x, u, p) \; = \; p^T f(t, x) \; + \; p^T B(t) u,$$

for the adjoint function p the system

$$\dot{p}(t) \; = \; -A(t)^T p(t) \, , \quad p(1) \; = \; \nabla_x k(x(1)) \tag{4}$$

with $A = \nabla_x f$ is obtained, and the optimal control u^0 satisfies

$$u^0(t) \; \in \; arg \max_{|v_i| \leq 1} \{ \, -H(t, x(t), v, p(t)) \, \}.$$

The function $\sigma = H_u = B^T p$ is called the switching function.

In case of problem (P) where the system is governed by a differential equation with given initial and free terminal state, the adjoint function is uniquely determined. Moreover, if B is independent of x then all coefficients in the adjoint equation are differentiable functions, and thus, the functions p and σ together with their first-order time derivatives are Lipschitz continuous.

If $\sigma \equiv 0$ on a certain interval then this part of the control trajectory is called a *singular arc.*

Assumption 2.1 (bang-bang regularity) *The pair (x^0, u^0) is a solution such that u^0 is piecewise constant and has no singular arcs. For every j, the set $\Sigma_j = \{ \, t \in [0,1] : \; \sigma_j(t) = 0 \, \}$ is finite, and $0, 1 \notin \Sigma_j$.*

Under the above assumption, almost everywhere the optimal control can componentwise be expressed by the formula

$$\sigma \; = \; B^T p, \quad u^0 \; = \; -sign(\sigma) \; . \tag{5}$$

We will further require that all points in $\Sigma_j, \; j = 1, \ldots, m$, be regular zeros of the respective σ-component:

Assumption 2.2 (strict bang-bang property) *For every $t_s \in \Sigma_j, \; j = 1, \ldots, m$: $\sigma_j(t_s) = 0 \quad \Rightarrow \quad \dot{\sigma}_j(t_s) \neq 0$.*

In the strict bang-bang case, the j-th control component switches at $t_s \in \Sigma_j$ in accordance with the jump condition

$$\left[\, u_j^0 \, \right]^s \; = \; u_j^0(t_s + 0) - u_j^0(t_s - 0) \; = \; -2 \, sign(\dot{\sigma}_j(t_s)) \, . \tag{6}$$

The set Σ of points where one or more components of σ vanish thus consists of the control switching points. Notice that a switching point is called *simple* if only one σ-component is zero. In order to allow for simultaneous (multiple) switches as well, introduce the notations

$$\begin{aligned}\Sigma_j &= \{t_{js} : \; s = 1, \dots, l(j)\} \\ \Sigma &= \{t_{js} : \; s = 1, \dots, l(j), \; j = 1, \dots, m\}, \qquad L = \sum_{j=1}^{m} l(j).\end{aligned} \tag{7}$$

It will be assumed that, in each Σ_j, the points are monotonically ordered so that, with the definitions $t_{j0} = 0$, $t_{j,l(j)+1} = 1$ for all j, we have

$$t_{js} < t_{j,s+1}, \quad s = 0, \dots, l(j), \quad j = 1, \dots, m. \tag{8}$$

3. Abstract sufficient optimality condition

In this section, we repeat some ideas from Klötzler (1979) and Maurer and Pickenhain (1995) for using an abstract duality concept in deriving sufficient optimality conditions for optimal control problems. The conditions are typically expressed in terms of certain Riccati type equations, respectively inequalities. As it was discussed in Felgenhauer (2001a), the main theorem in Maurer and Pickenhain (1995) may be well adapted to problems with discontinuous control solution. For the bang-bang situation, the scheme has been successfully applied to the *linear* system case in Felgenhauer (2003a). The generalization to nonlinear system dynamics requires relaxations for the class of dual feasible elements; in particular, it is useful to include piecewise continuous functions with jumps corresponding to the control discontinuities (see also Osmolovskii, 2000; Maurer and Osmolovskii, 2004).

Let us first reconsider the case when the state equation in (P) is *linear*, i.e $f(t,x) = A(t)x$. Suppose $(x^0, u^0) \in W^1_\infty \times L_\infty$ to be an extremal such that, with the costate p defined by (4), all conditions of the Pontryagin maximum principle are fulfilled.

Introduce the (dual) function $S : [0,1] \times R^n \to R$, and assume that S is continuously differentiable w.r.t. x, and at least piecewise continuously differentiable w.r.t. t. We will call S dual feasible if

$$\Psi(x,u,S) := \int_0^1 [H(t,x(t),u(t),\nabla_x S(t,x(t))) + S_t(t,x(t))] \, dt \geq 0 \tag{9}$$

for all $(x,u) \in W^1_\infty \times L_\infty$ such that $|u_i(t)| \leq 1$, $|x(t) - x^0(t)| \leq \epsilon$ for almost all $t \in [0,1]$. This condition consists in an *integrated form* of the Hamilton-Jacobi inequality for the constrained problem (P).

For given $\epsilon > 0$, further define

$$\Phi_\epsilon(S) = \inf_\xi \{k(\xi) + S(0,a) - S(1,\xi) : |\xi - x^0(1)| \leq \epsilon\}.$$

Then, for any admissible pair (x, u) such that $|x(t) - x^0(t)| \leq \epsilon$, and arbitrary dual feasible S, the following *duality relation* for $J = J(x, u)$ and $\Phi = \Phi_\epsilon(S)$ can be shown:

$$\begin{aligned} J(x,u) \;\; &= \;\; k(x(1)) \; - \int_0^1 \frac{d}{dt} S(t, x(t))\, dt \\ &\quad + \int_0^1 \left[H(t, x(t), u(t), \nabla_x S(t, x(t))) \; + \; S_t(t, x(t)) \right] dt \\ &\geq \;\; k(x(1)) \; + \; S(0, a) \; - \; S(1, x(1)) \;\; \geq \;\; \Phi_\epsilon(S) \, . \end{aligned} \tag{10}$$

In this sense, the problem (P) of minimizing J, and the problem of maximizing Φ_ϵ over all S with (9), may be considered as an abstract primal-dual problem pair, see Klötzler (1979).

In addition to Ψ from (9), introduce

$$\psi(\xi, S) \; := \; k(\xi) \; - \; k(x^0(1)) \; - \; S(1, \xi) \; + \; S(1, x^0(1)) \, . \tag{11}$$

Then one can characterize strict strong local minimizers by the following theorem (see Maurer and Pickenhain, 1995, and also Felgenhauer, 2003a):

THEOREM 3.1 *Let* (x^0, u^0) *be admissible for* (P). *Suppose that a function* $S : [0,1] \times R^n \to R$ *exists which is continuously differentiable w.r.t.* x *and piecewise continuously differentiable w.r.t.* t *such that for suitably chosen positive constants* c *and* ϵ *the following relations hold:*

(R1) $\Psi(x, u, S) \; \geq \; c \left(\|x - x^0\|_2^2 \; + \; \|u - u^0\|_1^2 \right), \qquad \Psi(x^0, u^0, S) \; = \; 0,$
for all admissible (x, u) *with* $\|x - x^0\|_\infty \leq \epsilon$ *a.e. in* $[0, 1]$*;*

(R2) $\psi(\xi, \, S) \; \geq \; 0 \qquad\qquad \forall \, \xi$ *with* $|\xi - x^0(1)| \; \leq \; \epsilon$.

Then (x^0, u^0) *is a strict strong local minimizer of* (P) *such that, for all admissible* (x, u) *with* $\|x - x^0\|_\infty \leq \epsilon$*, the objective functional suffices*

$$J(x, u) \; - \; J(x^0, u^0) \; \geq \; c \left(\|x - x^0\|_2^2 \; + \; \|u - u^0\|_1^2 \right) . \tag{12}$$

Proof. The definitions (9) and (11) for Ψ and ψ yield

$$\begin{aligned} J(x,u) \; - \; J(x^0, u^0) \;\; &= \;\; k(x(1)) \; + \int_0^1 \left[H(t, x, u, \nabla_x S(t, x)) \; + \; S_t(t, x) \right] dt \\ &\qquad + S(0, a) \; - \; S(1, x(1)) \; - \; k(x^0(1)) \\ &= \;\; \Psi(x, u, S) \; + \; \psi(x(1), S) \; + \; S(0, a) \; - \; S(1, x^0(1)). \end{aligned}$$

By the chain rule, we see that $S(0, a) \; - \; S(1, x^0(1)) \; = \; -\Psi(x^0, u^0, S)$. Thus, (12) is a direct consequence of (R1) together with (R2). ■

The above theorem may be generalized to the case of test functions S which are only piecewise continuous in time. It will be assumed that the discontinuity points do not depend on ξ, i.e. jump discontinuities may occur only for $t = \theta_k,\ k = 1, \ldots, l$. Setting $\theta_0 = 0$, and $\theta_\Sigma = (\theta_1, \ldots, \theta_l, \theta_{l+1})$ with $\theta_{l+1} = 1$, we assume that, for each ξ, $S(\cdot, \xi)$ is continuously differentiable on $(\theta_k, \theta_{k+1}),\ k = 0, \ldots, l$. Possible jump terms are written as

$$S(\theta_k + 0, \xi) - S(\theta_k - 0, \xi) \ = \ [\, S(\cdot, \xi)\,]^k\,, \qquad k = 1, \ldots, l.$$

These jump terms and their positions will be adapted later to the needs of the optimality proof, see Section 5. In general, their number will correspond to the overall number of (different) switching points of u^0, i.e. t_k from

$$\cup_{j=1}^m \Sigma_j \ = \ \{\, t_k : \ 1 \leq \ k \ \leq l\, \}, \qquad l \leq L, \tag{13}$$

with $0 = t_0 < t_1 < \cdots < t_l < t_{l+1} = 1$.

Let us redefine the auxiliary functionals Ψ resp. ψ by

$$\Psi(x, u, S) \ = \sum_{k=0}^{l} \int_{\theta_k}^{\theta_{k+1}} [H(t, x, u, \nabla_x S) + S_t]\ dt \tag{14}$$

and (see (11) and (9))

$$\begin{aligned} \psi(\xi_\Sigma, S) \ = \ \ & k(\xi_{l+1}) - k(x^0(1)) - S(1, \xi_{l+1}) + S(1, x^0(1)) \\ & + \sum_{k=1}^{l} [\, S(\cdot, \xi_k)\,]^k \ - \ \sum_{k=1}^{l} \left[\, S(\cdot, x^0(\theta_k))\, \right]^k \end{aligned}$$

where ξ_Σ denotes an arbitrary vector $\xi_\Sigma = (\xi_1, \ldots, \xi_l, \xi_{l+1}) \ \in \ R^{(l+1)n}$. Further, we abbreviate $(x(\theta_1), \ldots, x(\theta_{l+1}))$ by $x(\theta_\Sigma)$. Then in analogy to the proof of Theorem 3.1 we obtain

$$\begin{aligned} J(x, u) - J(x^0, u^0) \ \ &= \ \ \Psi(x, u, S) \ + \ \psi(x(\theta_\Sigma), S) \\ & \qquad + S(0, a) - S(1, x^0(1)) \ + \ \sum_{k=1}^{l} \left[S(\cdot, x^0(\theta_k)) \right]^k \\ &= \ \ \Psi(x, u, S) \ - \ \Psi(x^0, u^0, S) \ + \ \psi(x(\theta_\Sigma), S). \end{aligned} \tag{16}$$

For the modified situation, the result is summarized in the following theorem:

Theorem 3.2 *Let $S \ : \ \ [0, 1] \times R^n \ \rightarrow \ R$ be continuously differentiable w.r.t. x and piecewise continuous w.r.t. t. Further assume that the number of time-discontinuity points is not greater than L, their position is independent of x and, in all continuity points, S is continuously differentiable w.r.t. t.*
Suppose that for Ψ and ψ given by (14), (15) the relations (R1), (R2) *hold true. Then, for all (x, u) with $\|x - x^0\|_\infty \leq \epsilon$,*

$$J(x, u) \ - \ J(x^0, u^0) \ \geq \ c \left(\|x - x^0\|_2^2 \ + \ \|u - u^0\|_1^2 \right).$$

In order to find candidates for the function S with the above properties, as a rule, a quadratic ansatz is sufficient. For problem (P) it reduces to

$$S(t,x) \;=\; p(t)^T(x-x^0(t)) \;+\; 0.5\,(x-x^0(t))^T Q(t)(x-x^0(t)) \tag{17}$$

where p is the adjoint function. Thus, jump discontinuities in S may only occur if Q is discontinuous. In all continuity points we have

$$\begin{aligned} \nabla_x S(t,x) &= p(t) \;+\; Q(t)(x-x^0(t)), \\ S_t(t,x) &= \dot p(t)^T(x-x^0(t)) \;+\; 0.5\,(x-x^0(t))^T \dot Q(t)(x-x^0(t)) \\ &\qquad -\,(p(t)+Q(t)(x-x^0(t)))^T \dot x^0 . \end{aligned}$$

Using these expressions in $\Psi = \int_0^1 R[t]\,dt$ from (14), one can write the integrand in the following form:

$$\begin{aligned} R &= (p+Q(x-x^0))^T(f+Bu) \;+\; \dot p^T(x-x^0) \\ &\qquad -\,(p+Q(x-x^0))^T(f^0+Bu^0) \;+\; 0.5\,(x-x^0)^T\dot Q(x-x^0) \end{aligned}$$

(where f^0 stands for f evaluated along $x = x^0(t)$). Abbreviating further $x - x^0 = y$, $u-u^0 = v$, we get the representation $R = R_1 + R_2$ with

$$\begin{aligned} R_1 &= p^T(f-f^0) \;-\; p^T\nabla_x f^0 y \;+\; y^TQ(f-f^0) \;+\; 0.5\,y^T\dot Q y && (18) \\ &= 0.5\,y^T\left[\nabla_x^2 H^0 \;+\; Q\nabla_x f^0 \;+\; (\nabla_x f^0)^T Q \;+\; \dot Q\right] y \;+\; \mathrm{o}(|y|^2), && (19) \\ R_2 &= p^TBv \;+\; y^TQBv\,. && (20) \end{aligned}$$

The proof of the optimality conditions from Theorem 3.1, respectively. Theorem 3.2 thus reduces to the construction of an appropriate matrix function $Q = Q(t)$ such that, for the related S and admissible (x^0,u^0), (R1) and (R2) are locally satisfied.

4. Strong local optimality. Convex case

As a first application to the concept given in the previous section, consider problem (P) with a locally *convex* terminal functional $k = k(\xi)$. It will be shown that then the positive semi-definiteness of the Hessian of H w.r.t x together with the strict bang-bang property are sufficient for *strong* local optimality of a given extremal (x^0,u^0). The result is a modest generalization of Theorem 3.4, Felgenhauer (2003a), valid as well for *multiple* as for *simple* control switches. The proof uses only Theorem 3.1 without the extension to discontinuous dual test functions.

THEOREM 4.1 *Let* (x^0,u^0) *be an extremal of* (P) *and* p *a related adjoint function such that, with* $\sigma = B^Tp$, *Assumptions 2.1 and 2.2 are fulfilled. Suppose further that the function* $k = k(\xi)$ *is convex at* $\xi = x^0(1)$. *If the Hessian matrix* $\nabla_x^2 H^0[t]$ *(evaluated along the solution trajectory) is positive semi-definite on* $[0,1]$ *then*

(x^0, u^0) is a strict strong local minimizer. In particular, positive constants ϵ and c exist such that

$$J(x,u) - J(x^0,u^0) \;\geq\; c\left(\|x-x^0\|_2^2 \;+\; \|u-u^0\|_1^2 \right)$$

for all admissible (x,u) satisfying $\|x-x^0\|_\infty \leq \epsilon$.

For the proof, some auxiliary estimates are needed. The first preliminary result is a standard estimate for ordinary differential equations based on Gronwall's Lemma given here without proof.

Lemma 4.1 *Let (x,u) be an admissible pair and assume $\|x-x^0\|_\infty \leq M$. Then, a constant $c_1 = c(f,B,M)$ exists such that the functions $y = x-x^0$, $v = u-u^0$ satisfy the estimate*

$$\|\, y \,\|_\infty \leq \; c_1 \|\, v \,\|_1 \,. \tag{21}$$

In the next lemma, a matrix function Q for (17) is constructed.

Lemma 4.2 *Suppose that the Hessian matrix $\nabla_x^2 H^0[t] = \nabla_x^2 H(t, x^0(t), u^0(t), p(t))$ is positive semi-definite on $[0,1]$. Further, let (x,u) be admissible and denote $y = x-x^0$. Then, for every $\gamma \in (0,1)$, a constant $\epsilon = \epsilon(f,p,\gamma)$ and a matrix function Q with $\|Q\|_\infty = O(\gamma)$ exist such that R_1 from (18) satisfies*

$$\int_0^1 R_1[t]\, dt \;\geq\; \gamma \,\|\, y \,\|_2^2 \qquad \forall\, y :\; \|y\|_\infty \;\leq\; \epsilon. \tag{22}$$

Proof. For small y, the variation term R_1 has the expansion given by (19). If we denote by Q_1 the solution of the *linear* matrix differential equation

$$\dot{Q} \,+\, Q\nabla_x f + \nabla_x f^T Q \;=\; 2\,I, \qquad Q(1) \;=\; 0,$$

and set $Q \;=\; \gamma\, Q_1$, then

$$\|\, Q \,\|_\infty \;=\; \gamma \,\|\, Q_1 \|_\infty \;=:\; c_Q \gamma,$$

and $\exists\, \epsilon > 0$:

$$\begin{aligned} R_1[t] \;\; &\geq \;\; y(t)^T \nabla_x^2 H[t] y(t) \;+\; 2\,\gamma\, |y(t)|^2 + \,\mathrm{o}(|y(t)|^2) \\ &\geq \;\; \gamma\, |y(t)|^2 \qquad \forall\, y \text{ with } |y(t)| < \epsilon. \end{aligned}$$

For y such that $\|y\|_\infty < \epsilon$, after integrating over $[0,1]$ the last relation yields the desired estimate (22). ■

From Lemma 4.1 and Lemma 4.2 we directly conclude that, for y from a bounded set in L_∞,

$$\left| \int_0^1 y(t)^T Q(t) B(t) v(t)\, dt \right| \;\leq\; c_2 \gamma\, \|v\|_1^2 \tag{23}$$

with a positive constant c_2 depending on f and B but not on (x,u).

The next lemma consists of the crucial part in estimating R_2 from (20) under strict bang-bang assumptions. Up to minor changes, the proof is repeated from Felgenhauer (2003a), Section 3.

LEMMA 4.3 *For arbitrary admissible* (x, u) *and* $v = u - u^0$, *under Assumption 2.2 a constant* $c_3 = c(B, p, |\Sigma|)$ *exists such that*

$$\int_0^1 p(t)^T B(t) v(t)\, dt \;\geq\; c_3 \| v \|_1^2 \,. \tag{24}$$

Proof. First notice that, in case $v = u - u^0 = 0$, the relation (24) trivially holds. Thus it is supposed that $v \neq 0$ (i.e. $v(t)$ is not equal zero on some subset $I \subseteq [0, 1]$ of positive measure, and consequently, $\|v\|_p \neq 0 \quad \forall\, 1 \leq p < \infty$).

From Assumption 2.2, the following property of $B^T p = \sigma$ follows: For given $\delta > 0$ denote $\omega_\delta = \bigcup_{1 \leq s \leq l} (t_s - \delta, t_s + \delta)$. Then positive constants c_σ and $\bar{\delta}$ exist such that $\forall\, \delta \in (\,0, \bar{\delta}\,)$,

$$\min_i \,|\, \sigma_i(t) \,| \;\geq\; 0.5\, c_\sigma \delta \qquad \forall\, t \,\in\, [0, 1] \backslash \omega_\delta \,. \tag{25}$$

Using this estimate for the integral term from (24), we obtain

$$\begin{aligned} J = \int_0^1 p(t)^T B(t) v(t)\, dt \quad &= \quad \int_0^1 \sum_{i=1}^m |\sigma_i(t)|\, |v_i(t)|\, dt \\ &\geq \quad 0.5\, c_\sigma \delta \int_{[0,1]\backslash \omega_\delta} |\, v(t) \,| \;\, dt \,. \end{aligned} \tag{26}$$

But the functions $v \in L_\infty(0, 1; R^m)$ are uniformly bounded by a constant $M = 2\sqrt{m}$ due to the control box constraints so that, with $l = |\Sigma|$ denoting the number of switching points, from Assumption 2.1 we get

$$\begin{aligned} \|v\|_1 \quad &= \quad \int_0^1 |v(t)|\, dt = \int_{[0,1]\backslash \omega_\delta} |v(t)|\; dt \,+\, \int_{\omega_\delta} |v(t)|\; dt \\ &\leq \quad \int_{[0,1]\backslash \omega_\delta} |v(t)|\; dt \,+\, 2lM\, \delta \,. \end{aligned}$$

This last relation together with (26) yields

$$J \;\geq\; 0.5\, c_\sigma \delta \;(\, \|v\|_1 - 2lM\delta \,) \;\;. \tag{27}$$

Now one can choose an appropriate $\delta > 0$ depending on v via

$$\delta \;=\; M^{-1} \min\left\{ \frac{1}{4l}, \bar{\delta} \right\} \|v\|_1 \;=: \; c_\delta \|v\|_1 \;.$$

In particular, $\delta \,<\, \bar{\delta}$ so that from (27) it follows that

$$J \;\geq\; 0.5\, c_\sigma c_\delta (1 - 2lM\, c_\delta)\, \|v\|_1^2 \;\geq\; c_3 \|v\|_1^2,$$

where $c_3 > 0$ is independent of (x, u). ■

Proof of Theorem 4.1. Let γ be a positive number such that $\gamma < \min\{1, c_3/c_2\}$ for the constants $c_{2,3}$ from Lemma 4.3, respectively (23), and let $Q = \gamma\, Q_1$ be the related matrix function from Lemma 4.2.

Under the assumption that $x - x^0 = y$ is bounded by $\|y\|_\infty < \epsilon$ (see Lemma 4.2) consider first $\Psi(x, u, S)$ for S from (17), see (9):

$$\begin{aligned}\Psi(x,u,S) &= \int_0^1 R[t]\,dt \\ &= \int_0^1 R_1[t]\,dt \;+\; \int_0^1 \left(p^T Bv \;+\; y^T QBv\right)\,dt \\ &\geq \gamma\,\|y\|_2^2 \;+\; (c_3 -\, c_2\gamma)\,\|v\|_1^2\end{aligned}$$

as a consequence of Lemmas 4.2 and 4.3 together with (23). Taking into account that $\Psi(x^0, u^0, S) = S(1, x^0(1)) - S(0, a) = 0$ due to (17) and choosing $c = \min\{\gamma, c_3 -\, c_2\gamma\}$, we arrive at

$$\Psi(x,u,S) \,\geq\, c\left(\|x - x^0\|_2^2 \;+\; \|u - u^0\|_1^2\right),$$

i.e. the function S constructed from (17) with the given matrix function Q satisfies condition (R1) from Theorem 3.1.

Let us finally check the inequality (R2): recalling that $S(t, x^0(t)) \equiv 0, Q(1) = 0$ and $p(1) = \nabla_x k(x^0(1))$, the term ψ from (11) reduces to

$$\begin{aligned}\psi(\xi,S) &= k(\xi) - k(x^0(1)) \;-\; S(\xi,1) \\ &= k(\xi) - k(x^0(1)) \;-\; \nabla_x k(x^0(1))^T(\xi - x^0(1)).\end{aligned}$$

Due to the convexity assumption on k, the last expression is nonnegative, i.e. (R2) is satisfied. Applying Theorem 3.1, we end up with the assertion. ■

5. Strong local optimality. Riccati approach

So far, the duality approach from Section 3 was successfully applied to problem (P) without using the extensions made in Theorem 3.2. But for the case when either H, or the terminal functional k, do *not* satisfy convexity assumptions, next we will prove Riccati type optimality conditions including jump and multipoint boundary restrictions. The results show analogies to those presented in Section 1.9. of Osmolovskii (2000), but without assuming the switching points to be *simple*. The proof, however, is methodically independent of Osmolovskii and Lempio (2002), Milyutin and Osmolovskii (1998) (see also Maurer and Osmolovskii, 2004). Since, in particular, no structural restrictions on control variations are involved it supplements the former results in Felgenhauer (2003a).

In case of constrained control problems with continuous solutions, a modified weak Riccati approach was given in Maurer and Pickenhain (1995). Formally, in the bang-bang situation, these optimality conditions degenerate due to $\nabla_u^2 H = 0$. Thus, the estimation will be restarted from Theorem 3.2, where one has mainly to check Ψ and ψ from (15), (14) for their non-negativity. The related Riccati condition is induced by (19).

THEOREM 5.1 *Let* (x^0, u^0) *be an extremal of* (P) *and* p *a related adjoint function such that, with* $\sigma = B^T p$, *the Assumptions 2.1 and 2.2 hold. Further, let all switching points in* $\cup\Sigma_j$ *from (13) be enumerated in monotone order.*
On each interval $[t_k, t_{k+1}]$, *consider the matrix differential equation*

$$\dot{Q} + Q\nabla_x f + \nabla_x f^T Q + \nabla_x^2 H^0 = 0. \tag{28}$$

Suppose that the above system has solutions $Q = Q_k \in W^1_\infty(t_k, t_{k+1}; R^{n\times n})$ *satisfying the following conditions:*

(i) $Q_k(t_{k+1}) = 0, \quad k = 0, \ldots, l-1,$

(ii) $Q_k(t_k) \succ 0, \quad k = 1, \ldots, l,$

(iii) $\nabla_x^2 k(x^0(1)) - Q_l(1) \succ 0.$

Then there exist positive constants c, ϵ *such that*

$$J(x,u) - J(x^0,u^0) \geq c\,\|x - x^0\|_2^2 \tag{29}$$

for all admissible (x,u) *satisfying* $\|x - x^0\|_\infty \leq \epsilon$, *i.e.* (x^0, u^0) *is a strict strong local minimizer for* (P).

The proof of Theorem 5.1 consists of several parts. The first two parts deal with the estimation of Ψ consisting of the integral terms $\int R_1 dt$ and $\int R_2 dt$. The functional term ψ will be evaluated afterwards.
In a preliminary step, we consider properties of the system (28), (i)-(iii): First notice that, by continuity, a constant $\gamma > 0$ exists such that the differential inequality

$$\dot{Q} + Q\nabla_x f + \nabla_x f^T Q + \nabla_x^2 H^0 \succeq \gamma I. \tag{30}$$

has a solution $\tilde{Q}$ satisfying (i), a strengthened inequality (ii) with right-hand side γI, and terminal condition (iii). Moreover, for small $\delta > 0$, every solution arc $\tilde{Q}_k$ can be continued to a function in $W^1_\infty(t_k, t_{k+1} + \delta; R^{n\times n})$ satisfying (30) on this extended interval.
Denote $\theta_k = t_k + \delta,\ k = 1, \ldots, l$. If δ is taken sufficiently small then, by continuity,

$$\tilde{Q}_k(\theta_k) - \tilde{Q}_{k-1}(\theta_k) \succeq 0\,. \tag{31}$$

Further require that $\theta_k < t_{k+1} - \delta$ for all k. Notice that

$$\tilde{Q}_k(t_{k+1}) = Q_k(t_{k+1}) = 0, \qquad \nabla_x^2 k(x^0(1)) - \tilde{Q}_l(1) \succ 0$$

remain true. Patching the parts $\tilde{Q}_k \in W^1_\infty(\theta_k, \theta_{k+1}; R^{n\times n})$ together, we obtain functions $Q_\delta \in L_\infty(0,1; R^{n\times n})$ with the following properties: $\exists\delta_1 > 0$ such that for all $\delta < \delta_1$ the functions Q_δ are uniformly bounded in the sense

$$\|Q_\delta\|_\infty \leq M_0, \qquad \max_k \|\dot{\tilde{Q}}_k\|_\infty \leq M_1. \tag{32}$$

As a consequence, on each interval (θ_{k-1}, θ_k),

$$\|Q_\delta(t)\| \;\leq\; M_1 \,|\, t - t_k|\;. \tag{33}$$

The function Q_δ will be now inserted into S from (17) so that we obtain a piecewise differentiable function with jumps on θ_Σ (where $\theta_k = t_k + \delta, k = 1, \ldots, l$, and $\theta_{l+1} = t_{l+1} = 1$).

LEMMA 5.1 *Let the assumptions of Theorem 5.1 hold. Further, suppose $\delta < \delta_1$. Then there exists a constant $\epsilon_1 > 0$ independent of δ such that, for R_1 from (18) with $Q = Q_\delta$ and $y = x - x^0$, the following assertion holds:*

$$\int_0^1 R_1[t]\,dt \;\geq\; 0.25\,\gamma\,\|y\|_2^2 \qquad \forall\, y:\;\; \|y\|_\infty \leq \epsilon_1.$$

The proof is a direct consequence of the estimate (30) and the expansion (19) for R_1. Notice that, due to (32), the bound ϵ_1 does not depend on δ.

LEMMA 5.2 *Let $R_{2,i} = v_i B_i^T(p + Q_\delta y)$ be given in correspondence to (20). Under the assumptions of Lemma 5.1, there exist positive constants δ_2, ϵ_2 such that, for every $\delta < \min\{\delta_1, \delta_2\}$ and $i = 1, \ldots, m$,*

$$\int_0^1 R_{2,i}[t]\,dt \;\geq\; 0 \qquad \forall\, y:\;\; \|y\|_\infty \leq \epsilon_2\delta$$

and therefore, $\int_0^1 R_2[t]\,dt \;= \sum_{i=1}^m \int_0^1 R_{2,i}[t]\,dt$ is nonnegative.

Proof. As already noticed during the proof of Lemma 4.3 (see (26)), the control constraints together with the bang-bang nature of u^0 yield

$$R_{2,i} = p^T B_i v_i + y^T Q_\delta B_i v_i \;=\; |\sigma_i|\,|v_i| \,+\, y^T Q_\delta B_i v_i.$$

For the integral estimate, the interval $[0, 1]$ will be split into two sets $\omega_\delta = \omega_\delta(i)$ and $I_\delta = I_\delta(i) = [0, 1]\backslash\omega_\delta(i)$ defined by

$$\omega_\delta = \cup\{\omega_{\delta,k} :\, t_k \in \Sigma_i\} \qquad \text{with} \quad \omega_{\delta,k} = \,(t_k - \delta, t_k + \delta).$$

Then, by Assumption 2.2, choosing δ sufficiently small one can guarantee the following estimates for σ_i, $i = 1, \ldots, m$:

$$|\,\sigma_i(t)\,| \quad \geq \quad 0.5\,c_\sigma\delta \qquad \forall\, t\, \in\, I_\delta(i), \tag{34}$$

$$|\,\sigma_i(t)\,| \quad \geq \quad 0.5\,c_\sigma|t - t_k| \qquad \forall\, t\, \in \omega_{\delta,k}(i),\; t_k \in \Sigma_i \tag{35}$$

with a constant $c_\sigma > 0$ depending on the minimal slope of σ-components at their switching points (see (25)).

Denote $\max_i \|B_i\|_\infty = \beta$. For the integral over I_δ, from (34) and (32) we get

$$\begin{aligned}\int_{I_\delta} R_{2,i}[t]\,dt &\geq \int_{I_\delta} \big(\,|\sigma_i(t)| - |y(t)^T Q_\delta(t) B_i(t)\,|\big)\ |v_i(t)|\,dt \\ &\geq \ (\,0.5\,c_\sigma\delta \ -\ M_0\beta\|y\|_\infty) \int_{I_\delta} |v_i(t)|\,dt.\end{aligned}$$

As the last formula shows, the integral is nonnegative if e.g.

$$\|y\|_\infty \ \leq\ 0.5\,\frac{c_\sigma}{M_0\beta}\cdot\delta\ . \tag{36}$$

Next, consider the remaining integrals over ω_δ: estimates (35), (33) lead to

$$\begin{aligned}\int_{\omega_{\delta,k}} R_{2,i}[t]\,dt &\geq \int_{\omega_{\delta,k}} \big(\,|\sigma_i(t)| - |y(t)^T Q_\delta(t) B_i(t)\,|\big)\ |v_i(t)|\,dt \\ &\geq \int_{\omega_{\delta,k}} (\,0.5\,c_\sigma - M_1\beta\|y\|_\infty)\ |t-t_k|\cdot|v_i(t)|\,dt\ \geq\ 0\end{aligned}$$

if only $\|y\|_\infty$ does not exceed a certain bound, say

$$\|y\|_\infty \ \leq\ 0.5\,\frac{c_\sigma}{M_1\beta}. \tag{37}$$

Consequently, for some $\delta_2 \in (0,1)$ and $\epsilon_2 \leq 0.5c_\sigma/(\beta\max\{M_0, M_1\})$, from the above relations we deduce

$$\int_0^1 R_{2,i}[t]\,dt \ =\ \int_{I_\delta} R_{2,i}[t]\,dt\ +\sum_{k:\,t_k\in\Sigma_i}\int_{\omega_{\delta,k}} R_{2,i}[t]\,dt\ \geq\ 0$$

for all $\delta < \delta_2$ and y with $\|y\|_\infty \leq \epsilon_2\delta$, see (36), (37). Summation over i finally shows that $\int_0^1 R_2[t]\,dt\ \geq\ 0$. ■

Notice that in the proof of the last lemma the estimates are decoupled w.r.t. the control components. In particular, some of the sets $\Sigma_i \cap \Sigma_j$ may be nonempty, i.e multiple switches may be handled as well.

Proof of Theorem 5.1. Combining the last two lemmas we get the desired estimate for Ψ: first, determine $\bar\delta\ \leq\ \min\{1, \delta_1, \delta_2\}$ such that for $\delta = \bar\delta$ the conditions (31) – (33) and (34) – (35) are fulfilled. By setting $\epsilon' = \min\{\epsilon_1, \epsilon_2\bar\delta\}$, we obtain

$$\Psi(x,u,S)\ =\ \int_0^1 R[t]\,dt\ =\ \int_0^1 (R_1[t] + R_2[t])\,dt\ \geq\ 0.25\,\gamma\,\|\,x - x^0\|_2^2 \tag{38}$$

for all admissible (x,u) such that $\|\,x - x^0\|_\infty \leq \epsilon'$.

The proof is completed by checking the sign of $\psi = \psi(x(\theta_\Sigma), S)$: due to (17), we may write ψ as

$$\begin{aligned}
\psi(x(\theta_\Sigma), S) \; &= \; k(x(1)) - k(x^0(1)) \; - \; p(1)^T y(1) \; - \; 0.5\, y(1)^T Q_\delta(1) y(1) \\
&\qquad + 0.5 \sum_{k=1}^{l} y(\theta_k)^T \left[\, Q_\delta\right]^k y(\theta_k) \\
&= \; k(x^0(1) + y(1)) - k(x^0(1)) \; - \; \nabla_x k(x^0(1))^T y(1) \\
&\qquad - 0.5\, y(1)^T Q_\delta(1) y(1) \; + \; 0.5 \sum_{k=1}^{l} y(\theta_k)^T \left[\, Q_\delta\right]^k y(\theta_k) \\
&= \; 0.5\, y(1)^T \left(\nabla_x^2 k(x^0(1)) - Q_\delta(1) \right) y(1) \; + \; \mathrm{o}(|y(1)|^2) \\
&\qquad + 0.5 \sum_{k=1}^{l} y(\theta_k)^T \left[\, Q_\delta\right]^k y(\theta_k)
\end{aligned}$$

(where $\delta = \bar{\delta}$ is fixed). Conditions (iii) and (31) ensure that, for sufficiently small $|y(1)|$, this last term is nonnegative. Hence, by (16) it follows that

$$J(x,u) \; - \; J(x^0,u^0) \; = \; \Psi(x,u,S) \; + \; \psi(x(\theta_\Sigma), S) \; \geq \; c \, \| \, x - x^0 \|_2^2$$

for $c = 0.25\, \gamma$, and $\|y\|_\infty \leq \epsilon$ with an appropriately chosen $\epsilon \leq \epsilon'$. ■

REMARK 5.1 *As the above proof shows, assumption (i) may be replaced by the weaker condition*
(i') $Q_{k-1}(t_k) B_i(t_k) \; = \; 0$ for all i with $\sigma_i(t_k) = 0, \qquad k = 1, \dots, l.$

6. Strict optimality w.r.t. switching points

The conditions formulated in Theorem 4.1 (convex case) and Theorem 5.1 (general semilinear case) guarantee the strict strong local optimality of the reference solution (x^0, u^0). In particular, the optimality then holds true w.r.t. the subset of feasible local variations with fixed control structure and number of switching points. Considering only this particular type of variations, problem (P) can be related to a finite-dimensional mathematical program with switching points as main decision variables.

It will be shown that the switching set Σ^0 corresponding to u^0 provides a strict minimum for this auxiliary finite-dimensional problem such that a Strong Second-Order Optimality Condition (SSOC) is satisfied. To this aim, explicit formulas are derived for first- and second-order derivatives w.r.t. switching points. The formulas are new in that they allow to include *multiple* control switches. Their relation to quadratic forms used e.g. in Osmolovskii (2000) for deriving optimality conditions will be shortly discussed.

For a given strong local minimizer pair (x^0, u^0) satisfying Assumptions 2.1 and 2.2, we denote by $\Sigma^0 = \{\, t_{js}\}$ the set of ordered switching points according to (7) and (8). Let us consider vectors $\Sigma \in R^L$, $L = \sum_{j=1}^m l(j)$, in the neighborhood of Σ^0: if the distance $|\,\Sigma - \Sigma^0|$ is sufficiently small, then for the elements τ_{js} of Σ the monotonicity condition (8) is fulfilled. For simplicity, complete the switching points set by $\tau_{j0} = 0$, $\tau_{j,l(j)+1} = 1$ for all $j = 1, \dots, m$, and define

$$D_\Sigma \;=\; \{\, \Sigma = (\tau_{js}) :\; 0 < \tau_{js} < \tau_{j,s+1} < 1,\;\; s = 1, \dots, l(j) - 1,\; j = 1, \dots, m\,\}.$$

Then one can determine $u = u(t, \Sigma)$ and $x = x(t, \Sigma)$ by

$$\begin{aligned} u_j(t, \Sigma) &\;\equiv\; u_j^0(t_{js} + 0) \qquad\qquad \text{for}\;\; t \in (\tau_{js}, \tau_{j,s+1}), && (39)\\ \dot{x}(t) &\;=\; f(t, x(t)) \,+\, B(t)\, u(t, \Sigma), \qquad x(0) \,=\, a\,, && (40) \end{aligned}$$

and set $\phi(\Sigma) \;:=\; k(x(1, \Sigma))$. Obviously, $\Sigma = \Sigma^0$ solves the following finite-dimensional problem

$$\min \phi(\Sigma) \;=\; k\,(x(1, \Sigma)) \qquad \text{w.r.t.}\;\; \Sigma \in D_\Sigma. \tag{41}$$

Notice that the *Strong Second-Order Optimality Conditions* (SSOC) for (41) with its open feasible set $D_\Sigma \subset R^L$ have the form

$$\nabla_\Sigma\, \phi(\Sigma^0) \;=\; 0, \qquad \nabla^2_\Sigma \phi(\Sigma^0) \;\succ\; 0\,. \tag{42}$$

The derivatives of $\phi(\Sigma)$ given by (41) can be calculated from the chain rule using the functions $\eta_\alpha(t, \Sigma) = (\partial/\partial\tau_\alpha) x(t, \Sigma)$ and $\zeta_{\alpha\beta}(t, \Sigma) = (\partial^2/\partial\tau_\alpha\partial\tau_\beta) x(t, \Sigma)$. In order to possibly include the so-called *multiple* switching points where more than one control component may jump at moment $t = t_s$, we will use *multi* indices $\alpha = (i, r)$, $\beta = (j, s)$ for abbreviating e.g. $\tau_{ir} \in \Sigma_i$ by τ_α, τ_{js} by τ_β etc. Formally, one can write

$$\begin{aligned} \frac{\partial}{\partial \tau_\alpha}\, \phi(\Sigma) &\;=\; \nabla_x k(x(1, \Sigma))^T \eta_\alpha(1, \Sigma) \;=\; p(1)^T \eta_\alpha(1, \Sigma), && (43)\\ \frac{\partial^2}{\partial \tau_\alpha \partial \tau_\beta} \phi(\Sigma) &\;=\; \eta_\alpha(1, \Sigma)^T \nabla^2_x k(x(1, \Sigma)) \eta_\beta(1, \Sigma) && (44)\\ &\qquad\qquad + p(1)^T \frac{\partial}{\partial \tau_\beta} \eta_\alpha(1, \Sigma). \end{aligned}$$

Expressions for η_α are found from solving the differentiated state equation,

$$\dot{\eta}_\alpha(t, \Sigma) \;=\; A(t, \Sigma) \eta_\alpha(t, \Sigma) \quad \textit{a.e.}\,, \qquad \eta_\alpha(\tau_\alpha) = -b(\tau_\alpha), \tag{45}$$

with data

$$A(t, \Sigma) \;=\; \nabla_x f(t, x(t, \Sigma))\,, \quad b(\tau_\alpha) \;=\; B_i(\tau_{ir})\, \left[u_i^0 \right]^\alpha, \quad \alpha = (i, r),$$

and $\left[u^0\right]^\alpha = u^0(t_\alpha+0) - u^0(t_\alpha-0)$ (see (6)). Solutions can be represented by means of the fundamental matrix solutions $\Phi = \Phi(t,\Sigma)$, $\Psi = \Psi(t,\Sigma)$ determined from the systems

$$\dot{\Phi} + A^T\Phi = 0,\ \Phi(0) = I, \qquad \dot{\Psi} - A\Psi = 0,\ \Psi(0) = I, \tag{46}$$

and the Heaviside function χ in the following form:

$$\eta_\alpha(t,\Sigma) = -\chi(t,\tau_\alpha)\Psi(t,\Sigma)\Phi(\tau_\alpha,\Sigma)^T b(\tau_\alpha)\,. \tag{47}$$

Inserting (47) into (43), it follows from $p(t) = \Phi(t)\Psi(1)^T p(1)$ (see (4), (46)) that

$$\begin{aligned}\frac{\partial}{\partial\tau_\alpha}\phi(\Sigma) &= -\left[u_i^0\right]^\alpha B_i(\tau_\alpha)^T\Phi(\tau_\alpha,\Sigma)\Psi(1,\Sigma)^T\nabla_x k(x(1,\Sigma))\\ &= -\sigma_i(t_\alpha)\left[u_i^0\right]^\alpha = 0.\end{aligned} \tag{48}$$

Thus, Σ^0 is always a stationary solution of (41).

Repeating the differentiation of (47) w.r.t. τ_β one can also find appropriate representations for $\zeta_{\alpha\beta}$. To this aim, derivatives of the matrix functions $\Phi = \Phi(t,\Sigma)$ and $\Psi = \Psi(t,\Sigma)$ have to be provided. Denoting $M_\beta = \partial\Phi/\partial\tau_\beta$, $N_\beta = \partial\Psi/\partial\tau_\beta$ and $F_\beta = \partial A/\partial\tau_\beta$, from (46) we get

$$\dot{M}_\beta + A^T M_\beta = -F_\beta^T\Phi, \quad \dot{N}_\beta - A\,N_\beta = F_\beta\,\Psi$$

(with initial values equal to zero matrices). The matrix function $F_\beta(t,\Sigma) = (\partial/\partial\tau_\beta)A(t,\Sigma)$ is found by using a chain rule, and its k-th row (corresponding to the k-th component f_k of f) has the form $F_\beta^k = \eta_\beta^T\nabla_x^2 f_k$. Consequently, the functions M_β and N_β for $t \le \tau_\beta$ vanish, and for $t > \tau_\beta$ have integral representations of the form

$$\begin{aligned}M_\beta(t) &= -\Phi(t)\int_{\tau_\beta}^t \Psi(s)^T F_\beta(s)^T\Phi(s)\,ds,\\ N_\beta(t) &= \Psi(t)\int_{\tau_\beta}^t \Phi(s)^T F_\beta(s)\,\Psi(s)\,ds.\end{aligned}$$

In case $\alpha \ne \beta$, the derivatives of (47) at $t=1$ then may be written as

$$\frac{\partial}{\partial\tau_\beta}\eta_\alpha(1) = -\left[\,N_\beta(1)\Phi(\tau_\alpha)^T + \Psi(1)\,M_\beta(\tau_\alpha)^T\right]\,b(\tau_\alpha),$$

or, after some calculation,

$$\zeta_{\alpha\beta}(1,\Sigma) = -\Psi(1)\int_{t(\alpha,\beta)}^1 \Phi(s)^T F_\beta(s)\Psi(s)\,ds\ \cdot\ \Phi(\tau_\alpha)^T b(\tau_\alpha) \tag{49}$$

with $t(\alpha,\beta) = \max\{\tau_\alpha,\tau_\beta\}$.

Further, in case $\alpha = \beta$ one can see from (47) that the expression $\zeta_{\alpha\beta}$ has to be completed by an additive term,

$$\begin{aligned} \zeta_{\alpha\alpha}(1, \Sigma) &= D_\alpha + \int_{\tau_\alpha}^{1} \Psi(1)\Phi(t)^T F_\alpha(t)\eta_\alpha(t)\, dt, \\ D_\alpha &= -\Psi(1)\frac{d}{dt}\left(\Phi(t)^T B_i(t)\right)\Big|_{t=\tau_\alpha} \left[u_i^0\right]^\alpha . \end{aligned}$$

The derivatives are to be inserted into (44) and, after some simplifications, the Hessian may be given a symmetric formulation (see Felgenhauer, 2004)

$$\begin{aligned} \nabla^2_\Sigma \phi(\Sigma^0) = \eta(1)^T \nabla^2_x k(x(1, \Sigma^0))\, \eta(1) \; + \; diag_\alpha\{\, q_\alpha\} \\ + \int_0^1 \eta(t)^T \nabla^2_x H^0[t]\, \eta(t)\, dt \end{aligned} \tag{50}$$

with

$$q_\alpha \; = p(1)^T D_\alpha \; = \; -\dot\sigma_i(t_\alpha) \; \left[u_i^0\right]^\alpha \; > \; 0, \tag{51}$$

and the matrix function $\eta(t) \in R^{n\times L}$ assembled by columns $\eta_\alpha = \eta_{ir}$, $i = 1, \ldots, m$, $r = 1, \ldots, l(i)$, respectively.

The representation of $\nabla^2_\Sigma \phi$ in (50) allows for discussing some special cases where the SSOC (42) is fulfilled. As a first direct conclusion we deduce a coercivity result under the assumptions of Theorem 4.1:

Lemma 6.1 *Suppose that both Assumptions 2.1 and 2.2 hold. Further, assume that the matrices* $\nabla^2_x H^0[t]$ *(for all* $t \in [0, 1]$*) and* $\nabla^2_x k(x^0(1))$ *are positive semi-definite. Then* $\nabla^2_\Sigma \phi(\Sigma^0)$ *is positive definite, i.e. the second order condition (42) for problem (41) is satisfied at* Σ^0*.*

Secondly, formula (50) may be also compared to the quadratic form Ω used in Milyutin and Osmolovskii (1998), Osmolovskii (2000) as a main tool for testing optimality (see e.g. Milyutin, Osmolovskii, 1998, part II, par. 12.3 for details): if we introduce a vector $\bar\xi = (\bar\xi_\alpha) \in R^L$ corresponding to virtual shifts in the switching set $\Sigma^0 = (t_\alpha), \alpha = (i, r)$, then we obtain the second variation $\bar\xi^T \nabla^2_\Sigma k^0 \bar\xi$ for k (or equally: J) related to $\bar\xi$ resp. $\bar x = \eta \cdot \bar\xi$:

$$\Omega(p, \bar x) \; = \; \bar x(1)^T \nabla^2_x k^0 \bar x(1) \; + \; \sum_\alpha q_\alpha \bar\xi^2_\alpha \; + \; \int_0^1 \bar x(t)^T \nabla^2_x H^0[t] \bar x(t)\, dt. \tag{52}$$

For the problem class (P), this form coincides with Ω from Milyutin and Osmolovskii (1998) up to some natural extension needed for adapting the expansion to possibly *multiple* switches of several control components. (The above formula moreover suggests how the *critical cone* in Milyutin and Osmolovskii, 1998, Osmolovskii, 2000, should be modified for covering general switching set variations.) The positive definiteness of $\nabla^2_\Sigma \phi$ now can be checked by using (52).

In particular, one can apply the so-called Q-transformation from Osmolovskii and Lempio (2002) (see e.g. Proposition 2.1, or Theorem 2.3) and thus obtain the following result (where a detailed proof for shortness is omitted):

LEMMA 6.2 *Let for (x^0, u^0) and the adjoint function p the bang-bang regularity conditions from Assumptions 2.1 and 2.2 be fulfilled. Further suppose, that the Riccati equation (28) admits piecewise solutions satisfying the multi-point boundary restrictions (i)-(iii) from Theorem 5.1. Then $\nabla^2_\Sigma \phi(\Sigma^0)$ is positive definite.*

REMARK 6.1 *The boundary jump conditions formulated for Q in Theorem 5.1 differ from those in Osmolovskii and Lempio (2002) in that we prescribe one-sided limits for Q at the switching points rather than fixing the jump terms. Moreover, in case of simple switches, the conditions in Osmolovskii (2000) are certainly closer to the related necessary optimality conditions, and stronger for, e.g., problems with linear systems. The comparison for multiple switching remains to be an open question.*

7. Switching points iteration and optimality test

In case when the principal bang-bang structure of the optimal control is given, an iterative method for finding switching points can be derived by extending (39)-(40). As it was observed in Kim and Maurer (2003), the optimality properties of (41) can be further utilized for deriving sensitivity results w.r.t. switching points for parametric versions of the basic control problem (P). In Felgenhauer (2003b, 2004) certain shooting-type methods were used to this aim. In the present paper, we propose a primal-dual Newton type method for iterating switching points and analyze convergence conditions. As a by-product, an algorithm for testing definiteness of the Hessian in (41) is obtained which is suitable for numerical use in connection with so-called *indirect* methods.

Assume to be given a switching set approximation $\Sigma = (\tau_\alpha)$ (where $\tau_\alpha = \tau_{ir}$ corresponds to the r-th switching point of the i-th control component), and fix an initial guess $u^I \in \{-1, 1\}^m$ for $u^0(0)$. It is convenient to further denote $\tau_{i0} = 0$, $\tau_{i,l(i)+1} = 1$ for all i. In analogy to (39)-(40), determine primal variables $x(t, \Sigma)$, $u(t, \Sigma)$ by

$$u_i(t,\Sigma) \;=\; (-1)^r u_i^I \qquad \text{for } \; \tau_{ir} < t < \tau_{i,r+1}, \tag{53}$$

$$\dot{x}(t) \;=\; f(t, x(t)) \;+\; B(t)\, u(t,\Sigma), \quad x(0) \;=\; a.$$

Then, we will complete the mapping $\Sigma \;\rightarrow\; (x, u)$ by dual components $p = p(t, \Sigma)$, $\sigma = \sigma(t, \Sigma)$ solving

$$\dot{p}(t) \;=\; -A(t,\Sigma)^T p(t), \qquad p(1) = \nabla_x k(x(1,\Sigma)), \tag{54}$$

$$\sigma(t,\Sigma) \;=\; B(t)^T p(t,\Sigma). \tag{55}$$

If (x^0, u^0) is a solution of problem (P) with corresponding switching set Σ^0 then, for $u^I = u^0(0)$, we get $x^0(t) = x(t, \Sigma^0), u^0(t) = u(t, \Sigma^0)$, and together with $p(t) = p(t, \Sigma^0)$ and $\sigma(t) = \sigma(t, \Sigma^0)$ the pair (x^0, u^0) satisfies the maximum principle. If Assumption 2.1 holds true, we additionally have

$$W_\alpha(\Sigma^0) \;=\; \sigma_i(\tau_\alpha^0, \Sigma^0) \;=\; 0 \qquad \forall\, \tau_\alpha \in \Sigma_i,\; i = 1, \ldots, m \tag{56}$$

together with $\sigma_i(t, \Sigma^0) \neq 0$ for $t \notin \Sigma_i$. The above system consists of $L = dim\, \Sigma^0$ nonlinear equations. It can be used for improving the current switching points approximation $\Sigma = \Sigma^1$ by Newton's method:

$$\frac{\partial W(\Sigma^n)}{\partial \Sigma} \cdot \Delta\Sigma \;=\; -W(\Sigma^n), \qquad \Sigma^{n+1} \;=\; \Sigma^n \;+\; \Delta\Sigma. \tag{57}$$

The iteration (57) can be carried out if the Jacobian $\partial W/\partial\Sigma$ is regular. Notice that the matrix depends on the functions $\eta_\beta = (\partial/\partial\tau_\beta)x$ and $\rho_\beta = (\partial/\partial\tau_\beta)p$ which solve the following multi-point boundary value problem:

$$\begin{aligned}
\dot\eta_\beta(t) \;&=\; A(t, \Sigma)\, \eta_\beta(t), && (58)\\
&\qquad \eta_\beta(0) = 0,\;\; \eta_\beta(\tau_\beta) = -B_i(\tau_\beta)\, \left[\, u_i^I \right]^\beta, \\
\dot\rho_\beta(t) \;&=\; -A(t, \Sigma)^T \rho_\beta(t) \,-\, \nabla_x^2 \left(p(t, \Sigma)^T f(t, x(t, \Sigma))\right) \eta_\beta(t), && (59)\\
&\qquad \rho_\beta(1) = \nabla_x^2 k(x(1, \Sigma)) \cdot \eta_\beta(1).
\end{aligned}$$

This system is linear but coupled via terminal conditions.

The partial derivatives of W w.r.t. τ_β are given by

$$\frac{\partial W_\alpha}{\partial \tau_\beta} \;=\; \frac{\partial}{\partial \tau_\beta}\, \sigma_i(t)\big|_{t=\tau_\alpha} \;=\; B_i(\tau_\alpha)^T \rho_\beta(\tau_\alpha) \tag{60}$$

for $\alpha \neq \beta$ (see (56)), and

$$\frac{\partial W_\alpha}{\partial \tau_\alpha} \;=\; B_i(\tau_\alpha)^T \rho_\alpha(\tau_\alpha) \,+\, \dot\sigma_i(\tau_\alpha). \tag{61}$$

The following lemma provides important information for interpreting the Newton step (57) and assessing convergence properties of the resulting iteration.

Lemma 7.1 *The Newton update (57) for $\Sigma = (\tau_\alpha)$ is equivalent to a Newton step for minimizing $\phi(\Sigma) = k(x(1, \Sigma))$ in (41), i.e.*

$$\nabla_\Sigma^2 \phi(\Sigma^n) \cdot \Delta\Sigma \;=\; -\nabla_\Sigma \phi(\Sigma), \qquad \Sigma^{n+1} \;=\; \Sigma^n \,+\, \Delta\Sigma. \tag{62}$$

If, in particular, $\nabla_\Sigma^2 \phi(\Sigma)$ is positive definite then $\partial W/\partial\Sigma$ is regular at Σ.

Proof. The structure of the system (58), (59) allows for a solution representation in the form

$$\begin{aligned}
\eta_\beta(t, \Sigma) \;&=\; -\chi(t, \tau_\beta)\Psi(t, \Sigma)\Phi(\tau_\beta, \Sigma)^T B_j(\tau_\beta) \left[\, u_j^I \right]^\beta, \\
\rho_\beta(t, \Sigma) \;&=\; \Phi(t, \Sigma)\Psi(1, \Sigma)^T \nabla_x^2 k(x(1, \Sigma))\eta_\beta(1, \Sigma) \\
&\qquad + \Phi(t, \Sigma) \int_t^1 \Psi(s, \Sigma)^T \nabla_x^2 (p^T f)[s] \cdot \eta_\beta(s, \Sigma)\, ds\,,
\end{aligned}$$

see also (47). Inserting the last terms into (60) we obtain

$$\begin{aligned}\frac{\partial}{\partial \tau_\beta}\,\sigma_i(t,\Sigma)|_{t=\tau_\alpha} &= B_i(\tau_\alpha)^T \rho_\beta(\tau_\alpha)\\ &= B_i(\tau_\alpha)^T \Phi(\tau_\alpha)\Psi(1)^T \nabla_x^2 k(x(1,\Sigma))\,\eta_\beta(1,\Sigma)\\ &\quad +B_i(\tau_\alpha)^T\Phi(\tau_\alpha)\int_{\tau_\alpha}^1 \Psi(s)^T \nabla_x^2 (p^T f)[s]\cdot \eta_\beta(s)\,ds\\ &= -\frac{1}{\left[u_i^I\right]^\alpha}\,\eta_\alpha(1)^T \nabla_x^2 k(x(1,\Sigma))\eta_\beta(1)\\ &\quad -\frac{1}{\left[u_i^I\right]^\alpha}\int_{t(\alpha,\beta)}^1 \eta_\alpha(s)\nabla_x^2(p^T f)[s]\cdot\eta_\beta(s)\,ds.\end{aligned}$$

Remembering further that, due to (51),

$$-\left[u_i^I\right]^\alpha \dot\sigma_i(\tau_\alpha) \;=\; q_\alpha$$

and using the structure for $\nabla_\Sigma^2\phi$ given in (50), it follows that

$$\left[u_i^I\right]^\alpha \frac{\partial W_\alpha}{\partial \tau_\beta} \;=\; -\frac{\partial^2}{\partial\tau_\alpha\partial\tau_\beta}\phi(\Sigma) \qquad \forall\,\alpha,\beta. \tag{63}$$

The last relation shows that the Jacobian of W and the Hessian matrix of ϕ differ only by a diagonal matrix factor with nonzero entries so that $\partial W/\partial\Sigma$ is regular if and only if $\nabla^2\phi$ is a regular matrix.

Consider next the right hand side of the iteration (57): In analogy to (48), for W_α we obtain

$$W_\alpha(\Sigma) \;=\; \sigma_i(\tau_\alpha,\Sigma) \;=\; -\frac{1}{\left[u_i^I\right]^\alpha}\,\frac{\partial}{\partial\tau_\alpha}\phi(\Sigma). \tag{64}$$

Combining (64) and (63), the equivalence of iteration (57) and the Newton step for minimizing $\phi(\Sigma)$ follows. ■

COROLLARY 7.1 *Let $\Sigma^0 \in R^L$ corresponding to (x^0,u^0) be such that $\nabla_\Sigma^2\phi(\Sigma^0)$ is positive definite. If $u^I = u^0(0)$, and $\Sigma = \Sigma^1 \in R^L$ is sufficiently close to Σ^0 then the Newton sequence $\{\Sigma^n\}$ from (57) starting in Σ^1 converges quadratically in R^L, and the corresponding $(x^n,u^n) = (x(\cdot,\Sigma^n),u(\cdot,\Sigma^n))$ converge in $L_\infty \times L_1$ to (x^0,u^0): $\;\exists\, c>0$ such that*

$$\|x^{n+1}-x^0\|_\infty + \;\|u^{n+1}-u^0\|_1 \;\le\; c\,\|u^n-u^0\|_1^2\,. \tag{65}$$

Proof. Under the smoothness assumptions on the data of (P), $\nabla_\Sigma^2\phi$ depends Lipschitz continuously on Σ. Thus, locally the Newton methods (57), respectively (62) converge quadratically in R^L.

In the next step, we show the estimate (65). To this aim consider the norms

$$\|v\|_1 \;=\; \int_0^1 |\,v(t)\,|\,dt, \quad \|v\|_{(1)} \;=\; \int_0^1 |\,v(t)\,|_1\,dt$$

where $|\cdot|$ stands for the Euclidean norm, and $|\cdot|_1$ for the l_1 norm in R^m.
For functions $v \in L_\infty$ these two norms are obviously equivalent with $\|v\|_1 \le \|v\|_{(1)} \le \sqrt{m}\|v\|_1$.
Let Σ^n be sufficiently close to Σ^0. Then,

$$\begin{array}{rcl}\|u^n - u^0\|_{(1)} & = & \displaystyle\int_0^1 |\,u^n(t) - u^0(t)\,|_1\,dt \;=\; \sum_{i=1}^m \int_0^1 |u_i^n(t) - u_i^0(t)\,|\,dt \\ & = & \displaystyle 2\sum_{i=1}^m \sum_{r=1}^{l(i)} \left|\,\tau_{ir}^n - \tau_{ir}^0\right| \;=\; 2\left|\,\Sigma^n - \Sigma^0\,\right|_1 .\end{array}$$

The quadratic convergence property of $\{\Sigma^n\}$ thus yields the relation

$$\|u^{n+1} - u^0\|_1 \;\le\; c\,\left(\|u^n - u^0\|_{(1)}\right)^2 \;\le\; m\,c\,\left(\|u^n - u^0\|_1\right)^2$$

for some $c > 0$ independent of n, so that the convergence of $x^n = x(\cdot, \Sigma^n)$ in L_∞ and the estimate (65) follow directly from Lemma 4.1. ■

The close relation between the matrices $\nabla^2\phi$ and $\partial W/\partial\Sigma$ on the one hand and the sensitivity differentials $\eta_\Sigma = \partial x/\partial\Sigma$, $\rho_\Sigma = \partial p/\partial\Sigma$ on the other hand may be further utilized for an optimality test for stationary solutions Σ of (41). The test procedure consists of the following steps:
Suppose $\Sigma \in R^L$ be given and enumerated in a way that the elements in $\Sigma = (\tau_{\alpha_k})_{k=1,\dots,L}$ are monotonically ordered in the whole, i.e.

$$(\tau_{\alpha_0} =)\,0 \;<\; \tau_{\alpha_1} \;\le\; \tau_{\alpha_2} \;\le\; \dots \;\le\; \tau_{\alpha_L} \;<1\,(= \tau_{\alpha_{L+1}}).$$

In a forward-process, for $u = u(t, \Sigma)$ defined by (53) solve the system

$$\dot{x}(t) \;=\; f(t, x(t)) \;+\; B(t)\,u(t, \Sigma), \qquad x(0) \;=\; a,$$

and, successively for $\beta = \alpha_1, \dots, \alpha_L$, $t \ge \tau_\beta$,

$$\dot{\eta}_\beta(t) \;=\; -\nabla_x f(t, x(t))\eta_\beta(t), \quad \eta_\beta(\tau_\beta) = \;-B_j(\tau_\beta)\;\left[\,u_i^I\right]^\beta .$$

In the second stage, for $k = L, \dots, 0$ solve backwards on each $[\tau_{\alpha_k}, \tau_{\alpha_{k+1}}]$

$$\begin{array}{rcl}\dot{p}(t) & = & -\nabla_x f(t, x(t))^T p(t), \quad p(1) = \nabla_x k(x(1)), \\ \dot{\rho}_\beta(t) & = & -\nabla_x f(t, \Sigma)^T \rho_\beta(t) \;-\; \nabla_x^2\left(p(t, \Sigma)^T f(t, x(t, \Sigma))\right)\eta_\beta(t), \\ & & \rho_\beta(1) = \nabla_x^2 k(x(1, \Sigma)) \cdot \eta_\beta(1), \qquad \beta = \alpha_L, \dots, \alpha_{k+1}.\end{array}$$

Due to (63), successively the second derivatives of ϕ are obtained and may be assembled to principal minors of $\nabla^2\phi$ taken in the reverse order

$$\begin{aligned}\left(\nabla^2\phi\right)^{(k)} &= \left((\nabla^2\phi)_{\alpha_i\alpha_j}\right)_{i,j\geq k}, \\ (\nabla^2\phi)_{\alpha\beta} &= -\left[u_i^I\right]^\alpha \left[B_i(\tau_\alpha)^T\rho_\beta(\tau_\alpha) + \delta_{\alpha\beta}\dot{\sigma}_i(\tau_\alpha)\right].\end{aligned}$$

Here $\delta_{\alpha\beta}$ stands for the Kronecker symbol, and the slope $\dot{\sigma}$ of the switching function has the representation

$$\dot{\sigma}_i(t) = \dot{B}_i(t)^T p(t) + B_i(t)^T \dot{p}(t) = \left(\dot{B}_i(t) - \nabla_x f(t, x(t)) B_i(t)\right)^T p(t).$$

If the determinants of all principal minors above are positive, then the Hessian $\nabla^2_\Sigma\phi$ is positive definite, i.e. Σ^0 is a strict minimizer for (41).

Conclusion: As it was shown in Agrachev, Stefani and Zezza (2002) for the case of *simple* switches, the Strong Second-Order Optimality Condition for the auxiliary problem (41) combined with the bang-bang regularity Assumptions 2.1 and 2.2, are sufficient conditions for the strong local optimality of (x^0, u^0). It should be noticed that the technique used in Osmolovskii (2000), Osmolovskii and Lempio (2002) (and also Milyutin and Osmolovskii, 1998) to our knowledge has led to widely equivalent statements. Thus, the considerations of the last two sections give reason to the hypotheses that the mentioned optimality criteria should also apply in case when the optimal control has *multiple* switches.

References

Agrachev, A., Stefani, G. and Zezza, P.L. (2002) Strong optimality for a bang-bang trajectory. *SIAM J. Control Optim.* **41**, 991-1014.

Felgenhauer, U. (2001a) Weak and strong optimality in a problem with discontinuous control behavior. *J. Optim. Theor. Appl.* **110**, 361-387.

Felgenhauer, U. (2001b) Stability and local growth near bounded-strong local optimal controls. In: E. Sachs and R. Tichatschke, eds., *System Modelling and Optimization XX*, 20th IFIP TC7 Conference Trier 2001; Kluwer Academic Publ., Dordrecht, The Netherlands, 2003, 213-227.

Felgenhauer, U. (2003a) On stability of bang-bang type controls. *SIAM J. Control Optim.* **41** (6), 1843-1867.

Felgenhauer, U. (2003b) Optimality and sensitivity properties of bang-bang controls for linear systems. In: J. Cagnol, J.-P. Zolesio, eds., *Information Processing: Recent Mathematical Advances in Optimization and Control*, 21st IFIP TC7 Conference Sophia Antipolis 2003; Presses de l'Ecole des Mines de Paris, 2004, 87-99.

Felgenhauer, U. (2004) Optimality and sensitivity for semilinear bang-bang type optimal control problems. *Internat. J. Appl. Math. Computer Sc.* **14** (4), 447-454.

Kim, J.R. and Maurer, H. (2003) Sensitivity analysis of optimal control problems with bang-bang controls. In: *Proc. IEEE-Conference on Decision and Control*, Hawaii 2003, **4**, 3281-3286.

Klötzler, R. (1979) On a general conception of duality in optimal control. *Lect. Notes Math.* **703**, 189-196, Springer, New York.

Malanowski, K. (2001) Stability and sensitivity analysis for optimal control problems with control-state constraints. *Dissertationes Mathematicae*, Polska Akad. Nauk, Inst. Matemat., Warszawa.

Maurer, H. and Osmolovskii, N.P. (2004) Second order sufficient conditions for time-optimal bang-bang control. *SIAM J. Control Optim.* **42** (6), 2239–2263.

Maurer, H. and Pickenhain, S. (1995) Second order sufficient conditions for optimal control problems with mixed control-state constraints. *J. Optim. Theor. Appl.* **86**, 649–667.

Milyutin, A.A. and Osmolovskii, N.P. (1998) *Calculus of Variations and Optimal Control*. Amer. Mathem. Soc., Providence, Rhode Island.

Noble, J. and Schaettler, H. (2002) Sufficient conditions for relative minima of broken extremals in optimal control theory. *J. Math. Anal. Appl.* **269**, 98-128.

Osmolovskii, N.P. (2000) Second-order conditions for broken extremals. In: A. Ioffe et al., eds., *Calculus of Variations and Optimal Control*, Chapman & Hall/CRC Res. Notes Math. **411**, Boca Raton, FL, 198-216.

Osmolovskii N.P. and Lempio, F. (2002) Transformation of quadratic forms to perfect squares for broken extremals. *Set-Valued Analysis* **10**, 209 – 232.

Sarychev, A.V. (1997) First- and second-order sufficient optimality conditions for bang-bang controls. *SIAM J. Control Optim.* **35**, 315-340.