Control and Cybernetics

vol. **38** (2009) No. 4B

Optimality and stability result for bang–bang optimal controls with simple and double switch behaviour*†

by

Ursula Felgenhauer[1], **Laura Poggiolini**[2], **Gianna Stefani**[2]

1 Brandenburgische Technische Universität Cottbus, Institut für Angewandte Mathematik und Wissenschaftliches Rechnen, Germany

[2] Università di Firenze, Dip. di Matematica Applicata "G. Sansone", Italy

e-mail: felgenh@tu-cottbus.de, laura.poggiolini@math.unifi.it, gianna.stefani@unifi.it

Abstract: The paper considers parametric optimal control problems with bang–bang control vector function. For this problem we give regularity and second–order optimality conditions at the nominal solution which are sufficient to: *(i)* existence and local uniqueness of extremals, *(ii)* local structure stability, *(iii)* strong local optimality, under parameter perturbations. Here "local" means in a L_∞–neighbourhood of the nominal trajectory, regardless of the control values.

Stability results were obtained by the first author using the shooting approach, while optimality results were obtained by the other authors, using the Hamiltonian approach. The paper, combining both approaches, allows to unify the assumptions and to close some gaps between optimality and stability results.

Keywords: bang–bang control, parametric control problems, Hamiltonian flows.

1. Introduction

From optimal control theory in case of continuous control functions we know that, in analogy to the mathematical programming situation, a successful sensitivity analysis of the solutions requires certain strong second–order optimality conditions and possibly additional constraint qualifications. In recent years, essential progress has been made in deriving sufficient optimality conditions for

*This paper was prepared during first author's stay at the Department of Applied Mathematics, University of Florence, which was supported by CNR, INDAM grant No. 43169224, and by Florence University.

†Submitted: February 2009; Accepted: September 2009.

the so–called bang–bang controls in control–affine systems. Important contributions on second–order sufficient optimality conditions in case of bang–bang controls are given in Osmolovskii (1995, 2004), Milyutin and Osmolovskii (1998), Sarychev (1997), Agrachev et al. (2002), Poggiolini and Stefani (2004), Maurer and Osmolovskii (2004), Noble and Schättler (2002).

A standard assumption in the above mentioned papers, except for Sarychev (1997), consists in the fact that all control switches are *simple* (i.e. control components do not switch simultaneously). In the case of simultaneous (or *multiple*) switches, strong local optimality results have been given in Poggiolini and Stefani (2006), Poggiolini and Spadini (2008 and 2009). Prior to this, up to the authors' knowledge, only Sarychev (1997) faced the multiple switches case, considering L_1–local optimality, under stronger regularity assumptions. An example where a simultaneous switch of two control components was detected numerically was given in Oberle (1987) and personal communication.

For stability and sensitivity investigations on bang–bang extremals, results are given in Kim and Maurer (2003) and in Felgenhauer (2003, 2004 and 2008b). In the first paper the authors exploit the finite–dimensional sub–problem of minimising over switching times. In the other papers the author starts from the state–adjoint system given in the form of Pontryagin's Maximum Principle: this method requires to analyse parameter dependencies of the canonical system.

For linear or semi–linear dynamics, differentiability of the switching times with respect to parameters was shown in Felgenhauer (2005) without the simple switches restriction. For a general control–affine system this property holds true only under the simple switches assumption, see Kim and Maurer (2003), while only Lipschitz continuity holds if multiple switches occur, see Felgenhauer (2008a) for a proof and Felgenhauer (2008b) for an example where the multiple switching time is not differentiable.

In the present paper, the authors have brought together and compared different methods:

1. the so–called shooting approach in analysing bang–bang extremals with simultaneous control switches, already exploited by the first author to obtain stability and sensitivity results,
2. the Hamiltonian formalism and analysis of maximised Hamiltonians, previously used by the other authors to obtain strong local optimality results.

The common feature of the two approaches is to consider the second–order conditions associated to the optimisation problem over switching times positions. As a result, the assumptions of the two approaches are proven to be equivalent. Moreover, the shooting procedure can be seen as an embedded part of the construction of the Hamiltonian flow. This observation made it possible to find a new proof for the local stability of the switching structure and to obtain a unique optimiser in an L_∞ neighbourhood of the reference trajectory, independently of the related control values. In the result, the structural stability properties as obtained in Felgenhauer (2008a), that is in a $L_\infty \times L_1$ neighbourhood of the reference trajectory–control pair, could be strengthened.

The plan of the paper is the following: in Section 2 we define the notation; in Section 3 we state the problem and the regularity assumptions on the bang–bang structure of a reference extremal; in Section 4 we discuss the finite dimensional sub–problem, provide second order conditions and state the main theorem of the paper. In Section 5 we explain the Hamiltonian approach and derive an equivalent Hamiltonian formulation of the main assumptions. Finally, in Section 6 we prove the result stated in Section 4.

Hopefully, the result of this paper will be extended to general Mayer problems.

2. Notation

Let $\mathbb{R}^n$ be the Euclidean space of column vectors with norm $|\cdot|$. We shall denote the space of row vectors as $(\mathbb{R}^n)^*$. If $p \in (\mathbb{R}^n)^*$ and $q \in \mathbb{R}^n$, then $\langle p, q\rangle$ will denote their duality product, that is $\langle p, q\rangle = pq$.

The Lebesgue space of order r of vector–valued functions on $[0,1]$ is denoted by $L_r([0,1];\mathbb{R}^n)$. $W^k_r([0,1];\mathbb{R}^n)$ is the corresponding Sobolev space, and norms are given as $\|\cdot\|_r$ and $\|\cdot\|_{k,r}$, $(1 \leq r \leq \infty,\ k \geq 1)$, respectively. For the space of k–times continuously differentiable functions we will write C^k. The subspace of functions with Lipschitz continuous derivatives of order k is denoted $C^{k,1}$.

Let Ω be an open interval of $\mathbb{R}$ containing the origin. For any smooth parameter dependent function $\alpha\colon (q,h) \in \mathbb{R}^n \times \Omega \mapsto \alpha^h(q) \in \mathbb{R}$, the symbol $\mathrm{D}_q\, \alpha^h(q)$ denotes the gradient row vector, with respect to the q variable. The symbol ∂_q is used for (partial) generalised derivative in the sense of Clarke.

For any smooth parameter dependent vector field in $\mathbb{R}^n$, $f\colon (q,h) \in \mathbb{R}^n \times \Omega \mapsto f^h(q) \in \mathbb{R}^n$ we denote by $f^h\colon q \in \mathbb{R}^n \mapsto f^h(q) \in \mathbb{R}^n$ the vector field obtained by fixing h. Therefore, the Jacobian matrix of f with respect to the q variable, evaluated at (q,h) is denoted as $\mathrm{D}_q\, f^h(q)$. By "smooth parameter dependent vector field" we mean that f is at least C^2.

The Lie bracket between two vector fields, f_1 and f_2, is denoted as $[f_1, f_2]$:

$$[f_1, f_2] := (\mathrm{D}_q\, f_2)\, f_1 \,-\, (\mathrm{D}_q\, f_1)\, f_2.$$

The symbol $\exp tf(q)$ denotes the solution of the Cauchy problem

$$\dot{x}(t) = f(x(t)), \qquad x(0) = q.$$

The directional derivative of a smooth function α with respect to a vector field f in a point q is denoted as $f \cdot \alpha(q)$, i.e.

$$f \cdot \alpha(q) := \frac{\mathrm{d}}{\mathrm{d}\, t}\alpha(\exp tf(q))\bigg|_{t=0}.$$

Further, by $\operatorname{conv} M$ we denote the convex hull of a set M. For characterising discontinuities, jump terms are denoted as $[v]^s = v(t_s+) - v(t_s-)$ where the index s will become clear from the context.

We shall identify the tangent space to $\mathbb{R}^n$ in a point x with $\mathbb{R}^n$ itself, the cotangent space with $(\mathbb{R}^n)^*$. The tangent bundle and the cotangent bundle will be identified with $\mathbb{R}^n \times \mathbb{R}^n$ and $(\mathbb{R}^n)^* \times \mathbb{R}^n$, respectively. The elements of the cotangent bundle will be denoted as $\ell = (p, q)$, with $p \in (\mathbb{R}^n)^*$, $q \in \mathbb{R}^n$.

The projection of the cotangent bundle onto the state space is denoted as π:

$$\pi\colon (p, q) \in (\mathbb{R}^n)^* \times \mathbb{R}^n \mapsto q \in \mathbb{R}^n.$$

The symbols F^h and $\overrightarrow{F}^h$ will denote the Hamiltonian function and the corresponding Hamiltonian field associated to f^h, respectively, namely:

$$\begin{aligned}
&F\colon (p, q, h) \in (\mathbb{R}^n)^* \times \mathbb{R}^n \times \Omega \mapsto F^h(p, q) := \langle p, f^h(q)\rangle \in \mathbb{R},\\
&\overrightarrow{F}\colon (\mathbb{R}^n)^* \times \mathbb{R}^n \times \Omega \to (\mathbb{R}^n)^* \times \mathbb{R}^n\\
&\overrightarrow{F}\colon (p, q, h) \mapsto \overrightarrow{F}^h(p, q) := \big(-\mathrm{D}_q F^h(p, q), \mathrm{D}_p F^h(p, q)\big)\,.
\end{aligned}$$

With the symbol $\mathcal{F}^h$ we denote the flow of the parameter–dependent Hamiltonian field $\overrightarrow{F}$, emanating at time $t = 1$, i.e. $\mathcal{F}^h_t(\ell) = \exp(t-1)\overrightarrow{F}^h(\ell)$ denotes the solution $\lambda^h(t) = (p^h(t), q^h(t)) \in (\mathbb{R}^n)^* \times \mathbb{R}^n$, evaluated at time t, of the Hamiltonian system

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\lambda^h(t) = \overrightarrow{F}^h(\lambda(t))\,, \qquad \lambda^h(1) = \ell.$$

Finally, we denote as $Sign\colon \mathbb{R} \multimap \mathbb{R}$ the set–valued sign function, i.e $Sign(x) = sign(x)$ for $x \neq 0$, and $Sign(0) = [-1, 1]$.

3. The problem and the regularity assumptions

We consider a family of optimal control problems. To be more precise, we consider the following one–parameter family of Mayer problems on the fixed time interval $[0, 1]$ and with vector–valued bounded control $u\colon [0, 1] \to [-1, 1]^m$ entering the system dynamics linearly:

$(\mathbf{P}_{\boldsymbol{h}})$ minimise $\beta^h(x(1))$ subject to

$$\dot{x}(t) = f^h(x(t)) + \sum_{i=1}^m u_i(t) g_i^h(x(t)) \text{ a.e. in } [0, 1], \tag{1}$$

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

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

We suppose that the initial state value, the dynamics and the cost depend on a real parameter $h \in \Omega$, where Ω is an open interval containing the nominal value of the parameter $h_0 = 0$. The unknowns of the problem are the state of the system $x : [0,1] \mapsto \mathbb{R}^n$ and the control $u : [0,1] \mapsto \mathbb{R}^m$. The control set will be given by

$$\mathcal{U} := \{v \in L_\infty([0,1];\mathbb{R}^m) \colon \|v\|_\infty \leq 1\}.$$

Further, we denote by $g^h(x)$ the $n \times m$ matrix whose columns are the parameter dependent controlled vector fields g_i^h and assume that all the data of the problem are at least C^2.

We assume that, at $h = 0$, a bang–bang admissible state–control pair (x^0, u^0) in $W^1_\infty([0,1];\mathbb{R}^n) \times L_\infty([0,1];\mathbb{R}^m)$, which contains at most one double switch, is given. We want to pursue optimality and stability analysis for problems $(\mathbf{P}_{\boldsymbol{h}})$ near $h_0 = 0$.

In this paper we are concerned with *strong local optimality* where by (x^h, u^h) being a strong local optimiser of $(\mathbf{P}_{\boldsymbol{h}})$ we mean that (x^h, u^h) minimises the cost functional $\beta^h(x(1))$ among all the admissible couples (x, u) such that $\|x - x^h\|_\infty$ is small enough, regardless of any distance between u and u^h.

We recall that a necessary condition for an admissible couple (x^h, u^h) to be an optimiser of $(\mathbf{P}_{\boldsymbol{h}})$ is Pontryagin's Maximum Principle (PMP), which ensures the existence of an adjoint covector function $\mu^h \colon [0,1] \mapsto (\mathbb{R}^n)^*$. In particular, (μ^h, x^h, u^h) can be characterised as a solution of the following shooting problem for the canonical system:

Find z such that

$$\begin{aligned}
&\dot{x}(t) = f^h(x(t)) + g^h(x(t))u(t),\\
&\dot{\mu}(t) = -\langle \mu(t), \mathrm{D}_q(f^h + g^h u(t))(x(t))\rangle,\\
&u_j(t) \in Sign\langle \mu(t), g_j^h(x(t))\rangle, \quad j = 1, \dots, m
\end{aligned} \tag{4}$$

$$x(1) = z \tag{5}$$

$$\mu(1) = -\mathrm{D}_q\, \beta^h(z), \tag{6}$$

with target (see (2))

$$T(z,h) = x(0,z,h) = a(h). \tag{7}$$

We call a couple $\lambda := (\mu, x)$ solving (4) a *Pontryagin extremal of control system* (1). The solutions λ of (4)–(7) are called *Pontryagin extremals of problem* $(\mathbf{P}_{\boldsymbol{h}})$

For any fixed h, system (4)–(6) describes a backward parametrised family of Pontryagin extremals for the problem obtained from $(\mathbf{P}_{\boldsymbol{h}})$ by removing the initial constraint. Thus, problem (7) can be interpreted as the backward shooting system for determining Pontryagin extremals of $(\mathbf{P}_{\boldsymbol{h}})$

In what follows we state the regularity assumptions made on the control structure of the reference extremal (μ^0, x^0, u^0). We define

$$\begin{aligned}
&\sigma_j^0(t) := \langle \mu^0(t), g_j^0(x^0(t))\rangle,\\
&\Sigma_j^0 := \{t \in [0,1] \colon \sigma_j^0(t) = 0\}.
\end{aligned}$$

We point out that such conditions are the strengthening of the necessary conditions coming from Pontryagin's maximum principle.

Assumption 1 (bang–bang regularity) *The triple* (μ^0, x^0, u^0) *is a Pontryagin extremal of* $(\mathbf{P_0})$ *such that* u^0 *is piecewise constant,* $u_j^0(t) \in \{-1, 1\}$ *for almost any* $t \in [0,1]$ *and for any* $j = 1, \ldots, m$*. For each* $j = 1, \ldots, m$*, the set* $\Sigma_j^0 := \{\, t \in [0,1] : \ \sigma_j^0(t) = 0 \,\}$ *is finite and coincides with the set of discontinuity points of* u_j^0*. Moreover,* $0, 1 \notin \Sigma_j^0$*.*

Assumption 2 (strict bang–bang property) *For every* $j = 1, \ldots, m$ *and for any* $t_s \in \Sigma_j^0$*,* $\dot{\sigma}_j^0(t_s+) \cdot \dot{\sigma}_j^0(t_s-) > 0$*.*

For the purpose of this paper, we add

Assumption 3 (double switch restriction) *Any switching time of* u^0*,* $t_s \in \bigcup_{j=1}^m \Sigma_j^0$*, is the switching time of at most two control components.*

4. The finite–dimensional sub–problem and the main result

In order to formulate second–order optimality conditions for the control problem, consider the *induced* finite–dimensional problem obtained from $(\mathbf{P}_{\boldsymbol{h}})$ when the control structure is fixed but switching times positions are allowed to vary, see Agrachev et al. (2002). For each control component, the switching times in general can move independently. In particular, the double switching time may bifurcate to two simple switching times for the related u–components. According to the order of the switches, the control will take a different intermediate value on the new continuity interval.

Denote the double switching time of u^0 by τ^0 and assume that the control components switching there are the first two, i.e. $\{\tau^0\} = \Sigma_1^0 \cap \Sigma_2^0$. Without loss of generality, the sign of u_1, u_2 may be adjusted in such a way that

$$\big(u_i^0(\tau^0+) - u_i^0(\tau^0-)\big) \,=\, 2, \quad i = 1, 2. \tag{8}$$

The jump of u^0 at τ^0 is then accomplished by $u^0(\tau^0+) - u^0(\tau^0-) = 2(e_1 + e_2)$, where e_i stands for the i–th unit vector in $\mathbb{R}^m$.

Let $\Sigma^0 := (\Sigma_1^0, \ldots, \Sigma_m^0)$ be the vector of all switching times of u^0. We will enumerate them in decreasing order and assume that there are $R - 1$ simple switching points between τ^0 and 1, and other $R' - 1$ simple switches between 0 and τ^0. If we set $\theta_0^0 := 1$, $\theta_R^0 = \theta_{R+1}^0 := \tau^0$ and $\theta_{R+R'+1}^0 = 0$, we have

$$\begin{aligned} 1 = \theta_0^0 > \theta_1^0 > \ldots > \theta_{R-1}^0 > \theta_R^0 = \tau^0 = \theta_{R+1}^0 > \\ > \theta_{R+2}^0 > \ldots > \theta_{R+R'}^0 > \theta_{R+R'+1}^0 = 0. \end{aligned} \tag{9}$$

The length of Σ^0 is $L = R + R'$.
Further,

$$I_r^0 = \left(\theta_{r+1}^0, \theta_r^0\right), \qquad r = 0, \ldots, R + R',\ r \neq R, \tag{10}$$

defines pairwise disjoint intervals where the control takes constant value $v_r^0 := u^0(\theta_r^0-)$.

Any given vector $(\delta, \varepsilon) \in \mathbb{R}^{R+R'-2} \times \mathbb{R}^2$ of sufficiently small norm defines perturbed switching times

$$\theta_r = \theta_r^0 + \delta_r, \qquad r \neq R, R+1, \tag{11}$$

$$\tau_i = \tau^0 + \varepsilon_i, \qquad i = 1, 2, \tag{12}$$

where τ_1 and τ_2 stand for the new switching times of u_1, u_2 respectively.
If (δ, ε) is sufficiently close to zero then the new switching vector remains partially ordered in the following sense:

$$1 > \theta_r^0 > \theta_s^0 > 0 \quad \Rightarrow \quad 1 > \theta_r > \theta_s > 0. \tag{13}$$

The times

$$\theta_R := \max\{\tau_1,\ \tau_2\}, \qquad \theta_{R+1} := \min\{\tau_1,\ \tau_2\}, \tag{14}$$

define an interval $I_R = (\theta_{R+1}, \theta_R)$ where the control will take one of the values

$$u^\nu = u(\tau^0+) - 2\, e_\nu = v_{R-1}^0 - 2\, e_\nu, \quad \nu = 1, 2,$$

depending of the order of control switches: $\nu = 1$ if $\varepsilon_1 \geq \varepsilon_2$, and $\nu = 2$ otherwise. On $I_r, r \neq R$, use $u = v_r^0 = u^0(\theta_r^0-)$.
Now, the auxiliary parametric optimisation problem w.r.t. switching times positions is given by

$$(\mathbf{OP}_{\boldsymbol{h}}) \quad \text{minimise} \quad J^h(\delta, \varepsilon) = \beta^h(x(1)) \quad \text{subject to}$$

$$\dot{x}(t) = \begin{cases} f^h(x(t)) + g^h(x(t))\, v_r^0 & t \in I_r,\ r = 0, \ldots, R + R', r \neq R \\ f^h(x(t)) + g^h(x(t))\, u^\nu, & t \in I_R \end{cases} \tag{15}$$

$$x(0) = a(h) \tag{16}$$

$$\delta_R := \max\{\varepsilon_1,\ \varepsilon_2\}, \delta_{R+1} := \min\{\varepsilon_1,\ \varepsilon_2\} \tag{17}$$

$$\left.\begin{aligned} &\nu := 1 \text{ if } \varepsilon_1 \geq \varepsilon_2, \quad \nu := 2 \text{ if } \varepsilon_1 < \varepsilon_2, \\ &\theta_r := \theta_r^0 + \delta_r, \quad r = 0, \ldots R + R' + 1, \\ &\delta_0 = 0,\ \delta_{R+R'+1} = 0, \\ &I_r := (\theta_{r+1}, \theta_r), \quad r = 0, \ldots, R + R'. \end{aligned}\right. \tag{18}$$

The feasible set for $(\mathbf{OP}_{\boldsymbol{h}})$ (say: M) will be further restricted to those (δ, ε) that ensure the monotonicity property (13).

Proposition 4.1 *For each sufficiently small h, the functional J^h is a $C^{1,1}$ function of (δ, ε) near zero. Define M_ν, $\nu = 1, 2$ as*

$$M_1 = \{(\delta, \varepsilon) : \varepsilon_1 \geq \varepsilon_2\}, \quad M_2 = \{(\delta, \varepsilon) : \varepsilon_1 \leq \varepsilon_2\}.$$

Then the restriction J^h_ν of J^h to M_ν is C^2 for either $\nu = 1$ or $\nu = 2$. On $M_1 \cap M_2$, the generalised Hessian in the sense of Clarke is defined by

$$\partial_{(\delta,\varepsilon)} \left(\nabla_{(\delta,\varepsilon)} J^h\right) = \operatorname{conv}\{\nabla^2_{(\delta,\varepsilon)} J^h_1, \nabla^2_{(\delta,\varepsilon)} J^h_2\}.$$

At $h = 0$, the first variation $\nabla_{(\delta,\varepsilon)} J^0(0)$ equals zero.

The proof was given in Felgenhauer (2008a), where $(\mathbf{OP}_{\boldsymbol{h}})$ was formulated as a minimisation problem over the vector $\Sigma = (\Sigma_1, \dots, \Sigma_m) \in R^L$ which assembles component–wise the switching times of u. The data correspond one–to–one to (δ, ε) since, for each $\theta^0_r \in \Sigma^0_j$, the perturbed $\theta_r = \theta^0_r + \delta_r$ belongs to Σ_j if $r \neq R, R+1$, and $\tau_i = \theta^0_R + \varepsilon_i \in \Sigma_i$. In Felgenhauer (2008a), section 4, the piecewise C^2 behaviour of the objective functional has been proven under the Assumptions 1–3. Moreover, explicit formulas for ∇J^h and for the matrices spanning Clarke's generalised Hessian had been provided (Felgenhauer, 2008a, (23) respectively (24)–(25)). For problem $(\mathbf{OP}_{\boldsymbol{h}})$, now the following generalised strong second–order optimality condition is assumed to hold at $h = 0$:

Assumption 4 *Each matrix $Q \in \partial_{(\delta,\varepsilon)} \left(\nabla_{(\delta,\varepsilon)} J^0\right)(0) = \operatorname{conv}\{Q_1, Q_2\}$ with $Q_\nu = \nabla^2_{(\delta,\varepsilon)} J^0_\nu(0)$, $\nu = 1, 2$, is positive definite on R^L.*

The assumption is obviously equivalent to requiring that both Q_1 and Q_2 be positive definite.

Remark 1 *Assumption 4 is a strong second–order optimality condition for the non-smooth problem $(\mathbf{OP_0})$ since, together with the first–order condition $\nabla_{(\delta,\varepsilon)} J^0(0) = 0$, it ensures strict local optimality of the solution $(\delta, \varepsilon) = 0$, see Corollary 6.21 from Klatte and Kummer (2002), or Theorem 13.24, Rockafellar and Wets (1998).*

Remark 2 *The coercivity Assumption 4, together with Assumptions 1-3, yields, moreover, the existence, local uniqueness and Lipschitz regularity with respect to h of the solution $(\delta, \varepsilon) = (\delta(h), \varepsilon(h))$ of $(\mathbf{OP}_{\boldsymbol{h}})$ (see Felgenhauer, 2008a, Theorem 2). In Felgenhauer (2008b), Theorem 3, it was proven that the switching times can equally be found from the shooting system (4)–(7) where the latter was restricted to an appropriate neighbourhood of $(x^0, u^0, 0)$ in $C^0([0, 1]; \mathbb{R}^n) \times L_1([0, 1]; \mathbb{R}^m) \times \Omega$. As it will be shown in the following (see Remark 4 in Section 6.2), one can get rid of the L_1 constraint for u by revising the proof of the structural stability result from Felgenhauer (2008a), Theorem 1.*

For the parametric problem $(\mathbf{P}_{\boldsymbol{h}})$ the following main result is obtained:

Theorem 1 *Suppose Assumptions 1–4 are fulfilled for the reference solution* (x^0, u^0) *of* $(\mathbf{P_0})$ *at* $h = 0$*. Then there exist a constant* $c > 0$ *and a neighbourhood* $\mathcal{V}$ *of* x^0 *in* $C^0([0,1];\mathbb{R}^n)$ *such that, for each* h *with* $|h| \le c$*, there exists a strong local minimiser* (x^h, u^h) *of* $(\mathbf{P}_{\boldsymbol{h}})$ *with* $x^h \in \mathcal{V}$*. Furthermore, such minimiser is unique in* $\mathcal{V} \times \mathcal{U}$*. The corresponding control* u^h *is bang–bang with the same number and type of switches of each component of* u^0*, and the associated vector of switching times positions* Σ^h *is Lipschitz continuous with respect to* h*.*

5. Hamiltonian formulation of the Assumptions

In this Section we reformulate the assumptions in Hamiltonian formalism so that we can compare them with the assumptions in Poggiolini and Spadini (2009) (see also Poggiolini and Spadini, 2008, and Poggiolini and Stefani, 2006) where strong local optimality sufficient conditions were proven.

In Subsection 5.1 we reformulate the regularity Assumptions 1–3, while in Subsection 5.2 we reformulate the coercivity Assumption 4 on the second variation of the finite dimensional sub–problem $(\mathbf{OP_0})$.

Since all the assumptions can be given by means of strict inequalities involving Hamiltonians, they are preserved under parameter perturbations.

5.1. Regularity assumptions

We denote the couple (μ^0, x^0) as

$$\lambda^0 \colon t \in [0,1] \mapsto \lambda^0(t) \in (\mathbb{R}^n)^* \times \mathbb{R}^n.$$

According to (9), we also define

$$\ell^0_r = (p^0_r, q^0_r) := \lambda^0(\theta^0_r) \quad r = 0, \dots, R + R' + 1.$$

Notice that $\ell^0_R = \ell^0_{R+1} = \lambda^0(\tau^0)$.

On each interval I^0_r from (10), we define $k^h_r(q) := f^h(q) + g^h(q)v^0_r$, so that $k^0_r(q)$ is the restriction of the time–dependent vector field defined by the reference control $u^0(t)$ to the time–interval I^0_r:

$$k^0_r(q) = f^0(q) + g^0(q)v^0_r, \quad r \ne R.$$

Also, we denote as $K^h_r(p,q)$ the parameter dependent Hamiltonian function obtained by lifting $k^h_r(q)$, i.e.

$$K^h_r(p,q) := \langle p, k^h_r(q)\rangle, \quad r \ne R.$$

At time τ^0, where two components of the reference control u^0, u^0_1 and u^0_2 switch simultaneously, we have

$$\begin{aligned} k^0_{R-1} &= k^0_{R+1}(q) + 2g^0_1(q) + 2\, g^0_2(q) \\ K^0_{R-1}(p,q) &= K^0_{R+1}(p,q) + 2\,\langle p, g^0_1(q) + g^0_2(q)\rangle. \end{aligned}$$

according to the switching terms given by (8).

In the newly appearing interval I_R, according to which control component switches first, one of the two following vector fields (and Hamiltonian function) drives the control system:

$$\begin{aligned} k^h_{R\nu}(q) &:= k^h_{R-1}(q) - 2\, g^h_\nu(q) \\ K^h_{R\nu}(p,q) &:= K^h_{R-1}(p,q) - 2\,\langle p, g^h_\nu(q)\rangle \end{aligned} \qquad \nu = 1, 2. \tag{19}$$

With such formalism, Assumptions 1 and 2 can be easily restated in the following way

1. *Bang–bang regularity.*
   For any $t \in I^0_r$, $r \neq R$ and for any $u \in [-1,1]^m \setminus \{v^0_r\}$ we have

$$\langle \mu^0(t), f^0(x^0(t)) + g^0(x^0(t))u\rangle < K^0_r(\lambda^0(t)). \tag{20}$$

2. *Strict bang–bang property at switching times.*
   For each choice $K^0_R \in \{K^0_{R1}, K^0_{R2}\}$,

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left(K^0_r - K^0_{r-1}\right)(\lambda^0(t))\Big|_{t=\theta^0_r} < 0, \quad r = 1,\dots,R+R', \tag{21}$$

   equivalently: for each choice $k^0_R \in \{k^0_{R1}, k^0_{R2}\}$,

$$\langle p^0_r, [k^0_{r-1}, k^0_r](q^0_r)\rangle < 0, \quad r = 1,\dots,R+R'. \tag{22}$$

REMARK 1 *The strict inequalities "$<$" in equations* (20)–(22) *can be replaced by "$\neq$" since mild inequalities "$\leq$" hold true by PMP.*

As a direct consequence of the above reformulation and by continuity with respect to $h \in \Omega$ we get

LEMMA 5.1 *There exists $c > 0$ such that for any (p,q,h) with $\left\|(p,q) - \ell^0_r\right\| + |h| < c$ and any choice $k^h_R \in \{k^h_{R1}, k^h_{R2}\}$,*

$$\langle p, [k^h_{r-1}, k^h_r](q)\rangle < 0, \quad r = 1,\dots,R+R'. \tag{23}$$

5.2. The second order conditions

In Poggiolini and Spadini (2009) the second variation for the finite dimensional sub–problem ($\mathbf{OP_0}$) has been determined in terms of the vector fields defined by the nominal parameter $h = 0$, see also Poggiolini and Stefani (2006) for the case when only one (double) switch occurs. In those papers it is also proven that ($\mathbf{OP_0}$) is C^1.

Here we give the ideas of the construction to demonstrate the approach. To start with, we denote by $S_t(q,\delta,\varepsilon)$ the solution of (15), for the nominal value of the parameter $h = 0$, in case of $x(1) = q$:

$$S_0(q,\delta,\varepsilon) = \exp(-\theta_{R+R'})k^0_{R+R'} \circ \exp(\theta_{R+R'} - \theta_{R+R'-1})k^0_{R+R'-1} \circ \dots \\ \circ \exp(\theta_{R+2} - \theta_{R+1})k^0_{R+1} \circ \exp(\theta_{R+1} - \theta_R)k^0_{R\nu(\varepsilon)} \circ \\ \circ \exp(\theta_R - \theta_{R-1})k^0_{R-1} \circ \dots \circ \exp(\theta_2 - \theta_1)k^0_1 \circ \exp(\theta_1 - 1)k^0_0(q). \quad (24)$$

Here and in the sequel the "$\circ$" operation denotes the composition of maps. In particular, the flow $S_t(q,0,0)$, associated to the reference control, will be shortly denoted as $\hat{S}_t(q)$.

By the properties of flows, formula (24) can be written by means of the push–forward of the involved vector fields at time $t = 1$.

Namely, in a neighbourhood of $x^0(1)$ and for any choice $k^0_R \in \{k^0_{R1}, k^0_{R2}\}$ define

$$\widetilde{k}^0_r(q) := (\mathrm{D}_q\, \hat{S}_{\theta_r}(q))^{-1} k^0_r(\hat{S}_{\theta_r}(q)) \qquad r = 0, \dots, R + R'. \quad (25)$$

In order to write more compact formulas we introduce as variables the variations of the lengths of the intervals I_r:

$$\omega_r := \ -\delta_{r+1} + \delta_r \quad r \neq R, \qquad \omega_R := |\varepsilon_1 - \varepsilon_2| = -\delta_{R+1} + \delta_R. \quad (26)$$

Is is easy to see that for each fixed ν, the map $(\delta,\varepsilon) \mapsto \omega$ is one–to–one.

For the push-forward of the flow we thus have (see e.g. Poggiolini and Stefani, 2004)

$$(\hat{S}_0)^{-1} S_0(q,\delta,\varepsilon) = \exp \omega_{R+R'} \widetilde{k}^0_{R+R'} \circ \dots \circ \exp \omega_{R+1} \widetilde{k}^0_{R+1} \circ \\ \circ \exp \omega_R \widetilde{k}^0_{R\nu} \circ \exp \omega_{R-1} \widetilde{k}^0_{R-1} \circ \dots \circ \exp \omega_0 \widetilde{k}^0_0(q).$$

Recalling definitions (19)–(25) and using Campbell–Hausdorff formula (see e.g. Goodman, 1976) it is easy to see that the first–order approximation $L(q,\delta,\varepsilon)$ at $(x^0(1), 0, 0)$ of the map $(\hat{S}_0)^{-1} S_0$, is well defined and is given by:

$$L(q,\delta,\varepsilon) := q + \left(\sum_{r=R+2}^{R+R'} \omega_r \widetilde{k}^0_r + \delta_{R+1}\, \widetilde{k}^0_{R+1} \right) (x^0(1)) - \\ - 2 \sum_{i=1}^{2} \varepsilon_i \widetilde{g}^0_i(x^0(1)) + \left(-\delta_{R-1}\, \widetilde{k}^0_{R-1} + \sum_{r=0}^{R-2} \omega_r \widetilde{k}^0_r \right) (x^0(1)). \quad (27)$$

Remark 2 *The first order approximation formula* (27) *proves that problem* $(\mathbf{OP_0})$ *is* C^1.

From formula (27), applying Pontryagin Maximum Principle, it is easy to see that the first variation of $(\mathbf{OP_0})$ is zero on the whole space of admissible tangent directions (q,δ,ε), that is on those (q,δ,ε) such that $L(q,\delta,\varepsilon) = 0$.

Applying once again the Campbell–Hausdorff formula it is now a lengthy but straightforward calculation to obtain that the second variation of ($\mathbf{OP_0}$) on M_ν, $\nu = 1, 2$ is given by

$$\mathcal{Q}_\nu[\delta, \varepsilon]^2 := \frac{1}{2}\Big\{ \Big(\sum_{r=0}^{R+R'} \omega_r \widetilde{k}_r^0 \Big)^2 \cdot \beta^0(x^0(1)) + \\ + \sum_{r=0}^{R+R'} \omega_r \Big[\widetilde{k}_r^0, \sum_{s=0}^{r-1} \omega_s \widetilde{k}_s^0 \Big] \cdot \beta^0(x^0(1)) \Big\} \quad (28)$$

where $\widetilde{k}_R^0 := \widetilde{k}_{R\nu}^0$.

Recalling (19), (25), (26) one can check that the analytic expression of the second order variation of ($\mathbf{OP_0}$) changes according to the sign of $\varepsilon_1 - \varepsilon_2$ while they coincide on $M_1 \cap M_2$. Namely, the second order variation is given by a quadratic form $\mathcal{Q}_1$ on M_1 and by another quadratic form $\mathcal{Q}_2$ on M_2 that coincide on $M_1 \cap M_2$, i.e. when $\varepsilon_1 = \varepsilon_2$.

Requiring the coercivity of such second variations is equivalent to requiring the coercivity of both $\mathcal{Q}_1$ and $\mathcal{Q}_2$ on the whole tangent space.

To clarify the non–smoothness of the second variations, we give explicit formulas for the case when only the double switch occurs. In this case we have $\delta_1 = \max\{\varepsilon_1, \varepsilon_2\}$, $\delta_2 = \min\{\varepsilon_1, \varepsilon_2\}$, $\delta_0 = \delta_3 = 0$. Define

$$\widetilde{g}_\nu^0(q) := (\mathrm{D}_q \, \hat{S}_{\tau^0}(q))^{-1} g_\nu^0(\hat{S}_\tau(q)). \quad (29)$$

For the two cases $\nu = 1, 2$ the second variation is given by the two quadratic forms:

$$\begin{aligned} \mathcal{Q}_1[\epsilon_1, \epsilon_2]^2 = & \, 2 \left(\varepsilon_1 \widetilde{g}_1^0 + \varepsilon_2 \widetilde{g}_2^0 \right) \cdot \left(\varepsilon_1 \widetilde{g}_1^0 + \varepsilon_2 \widetilde{g}_2^0 \right) \cdot \beta^0(x^0(1)) + \\ & + \varepsilon_1^2 \, [\widetilde{g}_1^0, \widetilde{k}_0^0] \cdot \beta^0(x^0(1)) + \varepsilon_2^2 \, [\widetilde{g}_2^0, \widetilde{k}_0^0] \cdot \beta^0(x^0(1)) + \\ & + 2 \, \varepsilon_2 (\varepsilon_2 - \varepsilon_1) \, [\widetilde{g}_1^0, \widetilde{g}_2^0] \cdot \beta^0(x^0(1)) \end{aligned} \quad (30)$$

and

$$\begin{aligned} \mathcal{Q}_2[\epsilon_1, \epsilon_2]^2 = & \, 2 \left(\varepsilon_1 \widetilde{g}_1^0 + \varepsilon_2 \widetilde{g}_2^0 \right) \cdot \left(\varepsilon_1 \widetilde{g}_1^0 + \varepsilon_2 \widetilde{g}_2^0 \right) \cdot \beta^0(x^0(1)) + \\ & + \varepsilon_1^2 \, [\widetilde{g}_1^0, \widetilde{k}_0^0] \cdot \beta^0(x^0(1)) + \varepsilon_2^2 \, [\widetilde{g}_2^0, \widetilde{k}_0^0] \cdot \beta^0(x^0(1)) + \\ & + 2 \, \varepsilon_1 (\varepsilon_2 - \varepsilon_1) \, [\widetilde{g}_1^0, \widetilde{g}_2^0] \cdot \beta^0(x^0(1)). \end{aligned} \quad (31)$$

REMARK 3 *Notice that $\mathcal{Q}_1 = \mathcal{Q}_2$ and the problem is C^2 if and only if the duality product $\langle \mu^0(\tau^0), [g_1^0, g_2^0](x^0(\tau^0)) \rangle = [\widetilde{g}_1^0, \widetilde{g}_2^0] \cdot \beta^0(x^0(1))$ is null.*

REMARK 4 *Under Assumptions 1–4 the optimality of the Pontryagin extremal for the nominal problem ($\mathbf{P_0}$) is shown in Poggiolini and Spadini (2009), see also Poggiolini and Stefani (2006) and Poggiolini and Spadini (2008) for some preliminary results.*

6. Proof of the result

In this Section we give the proof of Theorem 1.

To start with we give the main steps explaining the underlying ideas coming from the Hamiltonian approach to optimal control.

From the Hamiltonian point of view, each solution of the shooting system (4)–(7) is a solution (μ^h, x^h) of the Hamiltonian system associated to the maximised Hamiltonian $H^{\max}$

$$H^{\max}(p, q, h) := \max\{\langle p, f^h(q) + g^h(q)u\rangle,\ u \in [-1, 1]^m\}, \tag{32}$$

with boundary conditions $\mu^h(1) = -\,\mathrm{D}_q\,\beta^h(x^h(1))$ and $x^h(0) = a(h)$.

In our case $H^{\max}$ turns out to be non–smooth but the regularity Assumptions 1–3 ensure that the associated Hamiltonian system admits a unique flow which is Lipschitz continuous with respect to the initial point.

The construction of such flow clearly shows that the regularity assumptions yield the structural stability of Pontryagin extremals of control system (1) with boundary conditions $(\mu^h(1), x^h(1))$ in a neighbourhood of $(\mu^0(1), x^0(1))$ in the cotangent bundle and small enough $|h|$. This construction is described in Section 6.1.

Using this construction and Assumption 4, in Section 6.2, we prove that the shooting system (4)–(7) has a unique solution for small enough $|h|$ and z sufficiently close to $x^0(1)$.

Finally, in Section 6.3, by continuity, we observe that the sufficient optimality conditions stated in Poggiolini and Spadini (2009) are fulfilled by the extremal (μ^h, x^h) of $(\mathbf{P}_{\boldsymbol{h}})$.

6.1. The maximised flow

Let $t \in [0, 1] \mapsto \mathcal{H}^h_t(p, q)$ be the solution (if it exists) of the Hamiltonian system associated to the maximised Hamiltonian with boundary condition $\mathcal{H}^h_1(p, q) = (p, q)$. Equivalently, $\mathcal{H}^h_t(p, q) = (\mu^h(t), x^h(t))$ where (μ^h, x^h) is the solution of (4) with boundary conditions $\mu^h(1) = p$, $x^h(1) = q$.

In what follows we explain how our regularity conditions give the existence of $\mathcal{H}^h_t(p, q)$. The construction will be pursued in a similar way to what is done in Agrachev et al. (2002), Poggiolini and Stefani (2006), Poggiolini and Spadini (2009) using the Implicit Function Theorem.

Since, by Assumption 1, the duality product $\langle \mu^0(1), g^0_j(x^0(1))\rangle = \sigma^0_j(1)$, $j = 1, \ldots, m$ is not zero, there exists $c > 0$ such that $\langle p, g^h_j(q)\rangle \neq 0$ for any (p, q, h) such that $\|(p, q) - \ell^0_0\| + |h| < c$. That is, there exists a neighbourhood $\mathcal{W}$ of ℓ^0_0 and $c > 0$ such that for any $(p, q, h) \in \mathcal{W} \times (-c, c)$ the maximised Hamiltonian is given by

$$H^{\max}(p, q, h) = K^h_0(p, q) = \langle p, f^h(q) + g^h(q)v^0_0\rangle.$$

Therefore, $\mathcal{H}^h_t(p,q)$ coincides, for t close to 1, with the flow $\exp(t-1)\overrightarrow{K}^h_0(p,q)$ of $\overrightarrow{K}^h_0$ emanating from (p,q) at time $t=1$. By continuity and because of regularity Assumption 1, this flow coincides with the maximised flow until it intersects the hyper–surface $K^h_0 - K^h_1 = 0$. By possibly restricting $\mathcal{W}$ and for $c > 0$ we get that such intersection is transversal, since

$$\frac{\partial}{\partial t}\left(K^h_0 - K^h_1\right) \circ \exp(t-1)\overrightarrow{K}^h_0(p,q)\Big|_{(\theta^0_1,\ell^0_1,0)} = -\langle p^0_1, [k^0_0, k^0_1](q^0_1)\rangle$$

which is positive by (22). Hence, we may apply the implicit function theorem and define a smooth function $\theta_1(\ell,h)$ such that

$$\begin{cases}\theta_1(\ell^0_0,0) = \theta^0_1\\ \left(K^h_0 - K^h_1\right)\circ\exp(\theta_1(\ell,h)-1)\overrightarrow{K}^h_0(\ell) = 0.\end{cases}$$

We iterate this procedure by defining

$$\theta_0(\ell,h) := 1 \qquad\qquad \varphi_0(\ell,h) := \ell$$

and, $\theta_r(\ell,h)$, $\varphi_r(\ell,h)$ as

$$\begin{cases}\theta_r(\ell^0_0,0) = \theta^0_r\\ \left(K^h_{r-1} - K^h_r\right)\circ\exp(\theta_r(\ell,h)-1)\overrightarrow{K}^h_{r-1}\circ\varphi_{r-1}(\ell,h) = 0.\end{cases}$$

$$\varphi_r(\ell,h) = \exp\left(1-\theta_r(\ell,h)\right)\overrightarrow{K}^h_r \circ \exp\left(\theta_r(\ell,h)-1\right)\overrightarrow{K}^h_{r-1}\circ\varphi_{r-1}(\ell,h)$$

for $r = 1,\ldots,R-1$. Thus, for any $r = 0,\ldots,R-2$ the maximised flow is thus given by

$$\mathcal{H}^h_t(\ell) := \exp(t-1)\overrightarrow{K}^h_r \circ \varphi_r(\ell,h) \qquad t\in[\theta_{r+1}(\ell,h),\theta_r(\ell,h)].$$

Notice that, possibly restricting $\mathcal{W}$, such flow is well defined because the number of switches is finite. Moreover the procedure coincides with the one introduced in Agrachev et al. (2002) in the case of simple switches.

We now define the decoupling $\tau_1(\ell,h)$ and $\tau_2(\ell,h)$ of the double switching time τ^0 by

$$\begin{cases}\tau_\nu(\ell^0_0,0) = \tau^0\\ \left(K^h_{R\nu} - K^h_{R-1}\right)\circ\exp(\tau_\nu(\ell,h)-1)\overrightarrow{K}^h_{R-1}\circ\varphi_{R-1}(\ell,h) = 0\end{cases}$$

and

$$\varphi^\nu_R(\ell,h) = \exp(1-\tau_\nu(\ell,h))\overrightarrow{K}^h_{R\nu}\circ\exp(\tau_\nu(\ell,h)-1)\overrightarrow{K}^h_{R-1}\circ\varphi_{R-1}(\ell,h).$$

Also we define $\widetilde{\tau}_1(\ell,h)$, $\widetilde{\tau}_2(\ell,h)$ by

$$\begin{cases}\widetilde{\tau}_1(\ell^0_0,0) = \tau^0\\ \left(K^h_{R+1} - K^h_{R2}\right)\circ\exp(\widetilde{\tau}_1(\ell,h)-1)\overrightarrow{K}^h_{R2}\circ\varphi^2_R(\ell,h) = 0\end{cases}$$

and

$$\begin{cases} \widetilde{\tau}_2(\ell_0^0, 0) = \tau^0 \\ \left(K_{R+1}^h - K_{R1}^h\right) \circ \exp(\widetilde{\tau}_2(\ell, h) - 1)\overrightarrow{K}_{R1}^h \circ \varphi_R^1(\ell, h) = 0. \end{cases}$$

That is, the index 1 always refers to the switching time of the first component of the control, which is $\tau_1(\ell, h)$ if $\tau_1(\ell, h) \geq \tau_2(\ell, h)$, $\widetilde{\tau}_1(\ell, h)$ otherwise. The index 2 always refers to the switching time of the second component, which is $\tau_2(\ell, h)$ if $\tau_2(\ell, h) \geq \tau_1(\ell, h)$, $\widetilde{\tau}_2(\ell, h)$ otherwise.

The decoupled switching times are then defined as

$$\theta_R(\ell, h) = \max\{\tau_1(\ell, h), \tau_2(\ell, h)\}$$

$$\theta_{R+1}(\ell, h) = \begin{cases} \widetilde{\tau}_2(\ell, h) & \text{if } \tau_1(\ell, h) \geq \tau_2(\ell, h) \\ \widetilde{\tau}_1(\ell, h) & \text{if } \tau_1(\ell, h) \leq \tau_2(\ell, h). \end{cases}$$

REMARK 1 $\theta_{R+1}(\ell, h) \leq \theta_R(\ell, h)$. *Moreover,* $\tau_1(\ell, h) = \tau_2(\ell, h)$ *if and only if* $\theta_{R+1}(\ell, h) = \theta_R(\ell, h)$. *For a short proof see Poggiolini and Spadini (2009).*

The maximised flow is thus given by

$$\mathcal{H}_t^h(\ell) = \begin{cases} \exp(t-1)\overrightarrow{K}_{R-1}^h \circ \varphi_{0r}(\ell, h) & \text{if } t \in [\theta_R(\ell, h), \theta_{R-1}(\ell, h)] \\ \exp(t-1)\overrightarrow{K}_{R\nu}^h \circ \varphi_R^\nu(\ell, h) & \begin{array}{c}\text{if } \theta_R(\ell, h) = \tau_\nu(\ell, h) \\ \text{and } t \in [\theta_{R+1}(\ell, h), \theta_R(\ell, h)].\end{array} \end{cases}$$

Now we repeat the procedure given before for the simple switches for each of the two possible paths. Define

$$\varphi_{R+1}^\nu(\ell, h) = \exp(1 - \theta_{R+1}(\ell, h))\overrightarrow{K}_{R+1}^h \circ \exp(\theta_{R+1}(\ell, h) - 1)\overrightarrow{K}_{R\nu}^h \circ \varphi_R^\nu(\ell, h).$$

We can define θ_r, $r = R+2, \ldots, R+R'$ as follows: first define $\theta_r^\nu(\ell, h)$, $\varphi_r^\nu(\ell, h)$ by

$$\begin{cases} \theta_r^\nu(\ell_0^0, 0) = \theta_r^0 \\ \left(K_{r-1}^h - K_r^h\right) \circ \exp(\theta_r^\nu(\ell, h) - 1)\overrightarrow{K}_{r-1}^h \circ \varphi_{r-1}^\nu(\ell, h) = 0 \end{cases}$$

$$\varphi_r^\nu(\ell, h) = \exp(1 - \theta_r^\nu(\ell, h))\overrightarrow{K}_r^h \circ \exp(\theta_r^\nu(\ell, h) - 1)\overrightarrow{K}_{r-1}^h \circ \varphi_{r-1}^\nu(\ell, h),$$

then choose $\theta_r(\ell, h)$ according to the choice made at the decoupling of the double switching time τ^0:

$$\theta_r(\ell, h) = \begin{cases} \theta_r^1(\ell, h) & \text{if } \tau_1(\ell, h) \geq \tau_2(\ell, h) \\ \theta_r^2(\ell, h) & \text{if } \tau_1(\ell, h) \leq \tau_2(\ell, h). \end{cases}$$

The maximised flow is thus defined as

$$\mathcal{H}_t^h(\ell) = \exp(t-1)\overrightarrow{K}_r^h \circ \varphi_r^\nu(\ell, h) \qquad \begin{array}{c} t \in [\theta_{r+1}(\ell, h), \theta_r(\ell, h)] \\ r = R+2, \ldots, R+R'+1 \end{array}$$

where

$$\begin{cases} \nu = 1 & \text{if } \theta_R(\ell,h) = \tau_1(\ell,h), \\ \nu = 2 & \text{if } \theta_R(\ell,h) = \tau_2(\ell,h). \end{cases}$$

By construction, the maximised flow is Lipschitz continuous and piecewise smooth. Also we point out that $t \mapsto \mathcal{H}^h_t(\ell,h)$ is a bang–bang Pontryagin extremal of control system (1), hence the vector of switching times $\Sigma(\ell,h)$ is well defined and has the same length of Σ^0.

The previous procedure can be summarised in the following Lemma:

LEMMA 6.1 *There exists a neighbourhood $\mathcal{W}$ of ℓ^0_0 in $(\mathbb{R}^n)^* \times \mathbb{R}^n$ and $c > 0$ such that the maximised flow*

$$\mathcal{H}^h \colon (t,p,q) \in [0,1] \times \mathcal{W} \mapsto \mathcal{H}^h_t \colon (p,q) \in (\mathbb{R}^n)^* \times \mathbb{R}^n$$

is well defined for any $h \in (-c,c)$. Moreover, for any $(p,q,h) \in \mathcal{W} \times (-c,c)$, the extremal (x^h,u^h) of $(\mathbf{P}_{\boldsymbol{h}})$ associated to

$$t \mapsto \mathcal{H}^h_t(p,q)$$

preserves the structure of (x^0,u^0), that is: u^h is a strict bang–bang control, each component u^h_i switches $|\Sigma^0_i|$ times and the double switch restriction also holds true.

REMARK 2 *The only difference between the qualitative behaviour of u^0 and u^h that may occur is that the double switching time $\tau^0 = \theta^0_R = \theta^0_{R+1}$ splits in two different switching times for the components u^h_1 and u^h_2 of u^h.*

6.2. Proof of the stability result

Starting from Lemma 6.1 we consider the map

$$\Psi = (\psi, \mathrm{id}) \colon \pi\mathcal{W} \times (-c,c) \to \mathbb{R}^n \times (-c,c)$$
$$\Psi = (\psi, \mathrm{id}) \colon (q,h) \mapsto (\psi(q,h),h) := (\pi\mathcal{H}^h_0(-\operatorname{D}_q \beta^h(q),q),h).$$

The first step is to prove that if Clarke's generalised Jacobian of $\psi^0 \colon q \mapsto \psi(q,0)$ at $q = x^0(1)$ is maximal rank then Ψ is locally Lipschitz one–to–one and this will yield the existence of one and only one local solution to the shooting system. Namely we prove the following Lemma:

LEMMA 6.2 *If $\partial_q\psi^0(x^0(1))$ is maximal rank, then there exist a neighbourhood $\mathcal{V}$ of $x^0(1)$ in $\mathbb{R}^n$ and $c > 0$ such that for any $h \in (-c,c)$ the shooting system (4)–(7) in $\mathcal{V}$ admits a unique solution $z(h)$. Moreover the map*

$$h \in (-c,c) \mapsto z(h) \in \mathcal{V}$$

is Lipschitz continuous.

Proof. The map Ψ is Lipschitz continuous, hence its Lipschitz invertibility can be proven using Clarke's Implicit Function Theorem (theorem 7.1.1 in Clarke, 1983). An easy computation shows that if $\partial_q \psi^0(x^0(1))$ is of maximal rank, then also $\partial_{(q,h)}\Psi(x^0(1),0)$ is of maximal rank.

Possibly restricting $c > 0$, there exists a neighbourhood $\mathcal{V}_0$ of $a(0) = x^0(0)$ such that the inverse mapping Ψ^{-1} is defined in $\mathcal{V}_0 \times (-c, c)$:

$$\Psi^{-1} \colon \mathcal{V}_0 \times (-c, c) \to \mathcal{V} \times (-c, c).$$

Since a depends continuously on h, again possibly restricting c, we may assume that $(a(h), h) \in \mathcal{V}_0 \times (-c, c)$.

Thus, the point $z(h)$ defined by $(z(h), h) = \Psi^{-1}(a(h), h)$, solves the shooting system (4)–(7), and the structure of (x^0, u^0) is preserved by construction. ■

LEMMA 6.3 *Under Assumption 4 the set $\partial_q \psi^0(x^0(1))$ is of maximal rank.*

Proof. A complete proof will appear in Poggiolini and Spadini (2009), here we give the proof only in the study case when no simple switch occurs (so that $\tau^0 = \theta_1^0 = \theta_2^0$ is the double switching time and $R + R' + 1 = 3$), and the reference vector field is given by

$$\begin{cases} k_2^0 = k_0^0 - 2(g_1^0 + g_2^0) & t \in [0, \tau^0], \\ k_0^0 & t \in [\tau^0, 1]. \end{cases}$$

In this case

$$\begin{aligned} \langle \mathrm{d}\,\widetilde{\tau}_2(\ell_0^0, 0), (\delta\ell, 0)\rangle &= \langle \mathrm{d}\tau_1(\ell_0^0, 0), (\delta\ell, 0)\rangle - \\ &\quad - \langle \mathrm{d}(\tau_1 - \tau_2)(\ell_0^0, 0), (\delta\ell, 0)\rangle \frac{\langle p_1^0, [k_0^0, k_{12}^0](q_1^0)\rangle}{\langle p_1^0, [k_{11}^0, k_2^0](q_1^0)\rangle} \\ \langle \mathrm{d}\,\widetilde{\tau}_1(\ell_0^0, 0), (\delta\ell, 0)\rangle &= \langle \mathrm{d}\tau_2(\ell_0^0, 0), (\delta\ell, 0)\rangle + \\ &\quad + \langle \mathrm{d}(\tau_1 - \tau_2)(\ell_0^0, 0), (\delta\ell, 0)\rangle \frac{\langle p_1^0, [k_0^0, k_{11}^0](q_1^0)\rangle}{\langle p_1^0, [k_{12}^0, k_2^0](q_1^0)\rangle} \end{aligned} \tag{33}$$

and

$$\begin{aligned} \mathcal{H}_0^0(\ell) &= \exp(-\widetilde{\tau}_2(\ell))\overrightarrow{K}_2^0 \circ \exp(\widetilde{\tau}_2 - \tau_1)(\ell)(\overrightarrow{K}_0^0 - 2\overrightarrow{G}_1^0) \circ \\ &\quad \circ \exp(\tau_1(\ell) - 1)\overrightarrow{K}_0^0(\ell) \qquad \text{if } \tau_1(\ell) \geq \tau_2(\ell), \\ \mathcal{H}_0^0(\ell) &= \exp(-\widetilde{\tau}_1(\ell))\overrightarrow{K}_2^0 \circ \exp(\widetilde{\tau}_1 - \tau_2)(\ell)(\overrightarrow{K}_0^0 - 2\overrightarrow{G}_2^0) \circ \\ &\quad \circ \exp(\tau_2(\ell) - 1)\overrightarrow{K}_0^0(\ell) \qquad \text{if } \tau_2(\ell) \geq \tau_1(\ell). \end{aligned}$$

The differentials at ℓ_0^0 are given by

$$\begin{aligned} &\mathrm{D}_\ell\, \mathcal{H}_0^0(\ell_0^0)\delta\ell = \mathrm{D}_\ell(\exp(-\tau^0\overrightarrow{K}_2^0))(\ell_1^0)\big\{\langle \mathrm{d}\,\widetilde{\tau}_2(\ell_0^0), \delta\ell\rangle 2\overrightarrow{G}_2^0(\ell_1^0) + \\ &+ \langle \mathrm{d}\tau_1(\ell_0^0), \delta\ell\rangle 2\overrightarrow{G}_1^0(\ell_1^0) + \mathrm{D}_\ell(\exp(\tau^0 - 1)\overrightarrow{K}_0^0))(\ell_0^0)\delta\ell\big\} \\ &\qquad\qquad\qquad\qquad \text{if } \langle \mathrm{d}\tau_1(\ell_0^0), \delta\ell\rangle \geq \langle \mathrm{d}\tau_2(\ell_0^0), \delta\ell\rangle, \end{aligned}$$

$$\begin{aligned}
&\mathrm{D}_\ell\, \mathcal{H}_0^0(\ell_0^0)\delta\ell = \mathrm{D}_\ell(\exp(-\tau^0\overrightarrow{K}_2^0))(\ell_1^0)\big\{\langle \mathrm{d}\,\widetilde{\tau}_1(\ell_0^0), \delta\ell\rangle 2\overrightarrow{G}_1^0(\ell_1^0)+\\
&+ \langle \mathrm{d}\tau_2(\ell_0^0), \delta\ell\rangle 2\overrightarrow{G}_2^0(\ell_1^0) + \mathrm{D}_\ell(\exp(\tau^0-1)\overrightarrow{K}_0^0))(\ell_0^0)\delta\ell\big\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{if } \langle \mathrm{d}\tau_2(\ell_0^0), \delta\ell\rangle \geq \langle \mathrm{d}\tau_1(\ell_0^0), \delta\ell\rangle.
\end{aligned}$$

For any $\delta x \in \mathbb{R}^n$ let $\delta\ell := (-\,\mathrm{D}_q\,\beta^0(q_0^0)(\delta x, \cdot), \delta x)$. The differentials at q_0^0 of ψ^0 are thus given by

$$\begin{aligned}
&\mathrm{D}_q\,\psi^0(q_0^0)\delta x = \mathcal{A}_1\delta x\\
&:= \mathrm{D}_q\,\hat{S}_0(q_0^0)\big\{2\langle \mathrm{d}\,\widetilde{\tau}_2(\ell_0^0), \delta\ell\rangle \widetilde{g}_2^0(q_0^0) + 2\langle \mathrm{d}\tau_1(\ell_0^0), \delta\ell\rangle \widetilde{g}_1^0(q_0^0) + \delta x\big\}\\
&= \mathrm{D}_q\,\hat{S}_0(q_0^0)\big\{2\langle \mathrm{d}\tau_1(\ell_0^0), \delta\ell\rangle (\widetilde{g}_1^0+\widetilde{g}_2^0)(q_0^0) + 2\langle \mathrm{d}(\widetilde{\tau}_2-\tau_1)(\ell_0^0), \delta\ell\rangle \widetilde{g}_2^0(q_0^0) + \delta x\big\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{if } \langle \mathrm{d}\tau_1(\ell_0^0), \delta\ell\rangle \geq \langle \mathrm{d}\tau_2(\ell_0^0), \delta\ell\rangle,
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{D}_q\,\psi^0(q_0^0)\delta x = \mathcal{A}_2\delta x\\
&:= \mathrm{D}_q\,\hat{S}_0(q_0^0)\big\{2\langle \mathrm{d}\,\widetilde{\tau}_1(\ell_0^0), \delta\ell\rangle \widetilde{g}_1^0(q_0^0) + 2\langle \mathrm{d}\tau_2(\ell_0^0), \delta\ell\rangle \widetilde{g}_2^0(q_0^0) + \delta x\big\}\\
&= \mathrm{D}_q\,\hat{S}_0(q_0^0)\big\{2\langle \mathrm{d}\tau_2(\ell_0^0), \delta\ell\rangle (\widetilde{g}_1^0+\widetilde{g}_2^0)(q_0^0) + 2\langle \mathrm{d}(\widetilde{\tau}_1-\tau_2)(\ell_0^0), \delta\ell\rangle \widetilde{g}_1^0(q_0^0) + \delta x\big\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{if } \langle \mathrm{d}\tau_2(\ell_0^0), \delta\ell\rangle \geq \langle \mathrm{d}\tau_1(\ell_0^0), \delta\ell\rangle.
\end{aligned}$$

We have to show that every map in Clarke's Jacobian of ψ^0 at $q_0^0 = x^0(1)$ is maximal rank, i.e. we must show that for any $\gamma \in [0, 1]$ the map $\gamma\mathcal{A}_1 + (1-\gamma)\mathcal{A}_2$ is invertible. Taking into account that the map $\mathrm{D}_q\,\hat{S}_0(q_0^0)$ is an isomorphism, this is equivalent to proving that the map

$$\mathcal{B}_\gamma := (\mathrm{D}_q\,\hat{S}_0(q_0^0))^{-1}\,(\gamma\mathcal{A}_1 + (1-\gamma)\mathcal{A}_2)$$

is an isomorphism for any $\gamma \in [0, 1]$. Since

$$\begin{aligned}
\mathcal{B}_\gamma\delta x = \delta x &+ 2\gamma\left\{\langle \mathrm{d}\tau_1(\ell_0^0), \delta\ell\rangle(\widetilde{g}_1^0+\widetilde{g}_2^0)(q_0^0) + \langle \mathrm{d}(\widetilde{\tau}_2-\tau_1)(\ell_0^0), \delta\ell\rangle\widetilde{g}_2^0(q_0^0)\right\} +\\
&+ 2(1-\gamma)\left\{\langle \mathrm{d}\tau_2(\ell_0^0), \delta\ell\rangle(\widetilde{g}_1^0+\widetilde{g}_2^0)(q_0^0) + \langle \mathrm{d}(\widetilde{\tau}_1-\tau_2)(\ell_0^0), \delta\ell\rangle\widetilde{g}_1^0(q_0^0)\right\},
\end{aligned}$$

the kernel of $\mathcal{B}_\gamma$ is a linear sub-space of

$$V := \mathrm{span}\{\widetilde{g}_1^0(q_0^0), \widetilde{g}_2^0(q_0^0)\}.$$

Also, if $V \subset \ker \mathrm{d}\tau_1(\ell_0^0) \cap \ker \mathrm{d}\tau_2(\ell_0^0)$, then the restriction of $\mathcal{B}_\gamma$ to V is the identity map, hence $\mathcal{B}_\gamma\delta x \neq 0$ for any $\delta x \in \mathbb{R}^n \setminus \{0\}$ and for any $\gamma \in [0, 1]$.

Let us assume that $V \not\subset \ker \mathrm{d}\tau_1(\ell_0^0) \cap \ker \mathrm{d}\tau_2(\ell_0^0)$. In Poggiolini and Spadini (2009) it is shown that both $\mathcal{B}_0$ and $\mathcal{B}_1$ have the same orientation of the identity mapping, that is: if we choose bases of $\mathbb{R}^n$, and the matrices B_0, B_1 represent $\mathcal{B}_0$ and $\mathcal{B}_1$ in such bases, then $\det B_0 \det B_1 > 0$

If $\dim V = 1$ then $\det B_\gamma$ is a linear function of γ, thus $\mathcal{B}_\gamma$ is an isomorphism for any $\gamma \in [0, 1]$ and we are done. If $\dim V = 2$, then we distinguish between two cases:

A) $V \subset \ker \mathrm{d}(\tau_1 - \tau_2)(\ell_0^0)$
In this case, for any $\delta x \in V$ we have

$$\mathcal{B}_\gamma \delta x = \mathcal{B}_0 \delta x = \mathcal{B}_1 \delta x \quad \forall \gamma \in [0,1].$$

B) $V \not\subset \ker \mathrm{d}(\tau_1 - \tau_2)(\ell_0^0)$
In this case $V \cap \ker \mathrm{d}(\tau_1 - \tau_2)(\ell_0^0)$ is a one–dimensional linear space. Let us fix $\delta y \in V \cap \ker \mathrm{d}(\tau_1 - \tau_2)(\ell_0^0)$ and $\delta z \in V \cap \ker \mathrm{d}\tau_1(\ell_0^0) \setminus \ker \mathrm{d}\tau_2(\ell_0^0)$ so that $\mathrm{span}\{\delta y, \delta z\} = V$.
Denote by B_γ the matrix associated to $\mathcal{B}_\gamma$ in the bases $\{\delta y, \delta z\}$ and $\{\widetilde{g}_1^0(q_0^0), \widetilde{g}_2^0(q_0^0)\}$. Since

$$\mathcal{B}_\gamma \delta y = \delta y + 2\, \langle \mathrm{d}\tau_1(\ell_0^0), \delta y \rangle (\widetilde{g}_1^0(q_0^0) + \widetilde{g}_2^0(q_0^0)),$$
$$\mathcal{B}_\gamma \delta z = \gamma \mathcal{A}_1 \delta z + (1-\gamma)\mathcal{A}_0 \delta z,$$

we get

$$\det B_\gamma = \gamma \det B_1 + (1-\gamma) \det B_0.$$

In Poggiolini and Spadini (2009) the authors prove that the determinants of both matrixes B_0 and B_1 have the same sign of the matrix representing the identity map, thus our claim is proven, both in case A) and in case B). ■

Remark 3 *$\psi^h(q)$ is the target mapping T defined in* (7)*. Moreover, the extremal (x^h, u^h) preserves the structure of (x^0, u^0), in the sense of Lemma 6.1: u^h is a strict bang–bang control, each component u_i^h switches $|\Sigma_i^0|$ times and the double switch restriction also holds true. By construction of the maximised flow it is clear that the switching times of u^h depend smoothly on the initial point ℓ. Since $\ell = (-\mathrm{D}_q\, \beta^h(z(h)), z(h))$ is Lipschitz continuous with respect to h, then also Σ^h has the same regularity.*

Remark 4 *The statements in Lemmata 6.2, 6.3 strengthen the stability results from Kim and Maurer (2003) and Felgenhauer (2008b) specifying them on state–parameter neighbourhoods (i.e. w.r.t.* strong *topology) for simple as well as for double switching points. The restriction to an L_1 neighbourhood of the reference control as formerly used in Felgenhauer (2008b) is no longer required.*

6.3. Proof of the optimality

Lemmata 6.2 and 6.3 yield that there exists $c > 0$ such that for any $h \in \Omega$, such that $|h| < c$, there is a unique solution $z(h) \in \mathcal{V}$ of the shooting system (4)–(7). Therefore, we obtain a Pontryagin extremal given by

$$t \mapsto (\mu^h(t), x^h(t)) := \mathcal{H}_t(-\mathrm{D}_q\, \beta^h(z(h)), z(h))$$

with associated bang–bang control u^h. Hence, for sufficiently small h, such extremal is a regular bang–bang extremal with the strict bang–bang property and it has at most one double switch.

Thus, we can define the quadratic forms $\mathcal{Q}_1(h)$, $\mathcal{Q}_2(h)$ along the extremal (μ^h, x^h) by means of $\widetilde{k}_0^h, \ldots, \widetilde{k}_{R\nu}^h, \ldots, \widetilde{k}_{R+R'}^h$. It is clear that, possibly restricting c, they are coercive by continuity.

Applying the results in Poggiolini and Spadini (2008 and 2009) we have that (x^h, u^h) is a strong local optimiser of $(\mathbf{P}_{\boldsymbol{h}})$.

This completes the proof of the theorem.

References

AGRACHEV, A. and SACHKOV, Y. (2004) *Control theory from the geometric viewpoint*. Encyclopaedia of Mathematical Sciences, 87. Control Theory and Optimization, II. Springer, Berlin.

AGRACHEV, A., STEFANI, G. and ZEZZA, P. L. (2002) Strong optimality for a bang–bang trajectory. *SIAM J. Control Optim.* **41** (4), 991–1014.

ARUTYUNOV, A.V., AVAKOV, E.R. and IZMAILOV, A.F. (2007) Directional regularity and metric regularity. *SIAM J. Optim.* **18** (3), 810–833.

CLARKE, F.H. (1983) *Optimization and Nonsmooth Analysis*. Wiley Inc., New York.

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

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

FELGENHAUER, U. (2005) Optimality properties of controls with bang–bang components in problems with semilinear state equation. *Control & Cybernetics* **34** (3), 763–785.

FELGENHAUER, U. (2008a) Lipschitz stability of broken extremals in bang–bang control problems. In: I. Lirkov et al., eds., *Proc. th Internat. Conf. Large–Scale Scientific Computing, Sozopol 2007, Lecture Notes Comp. Sci.* **4818**, Springer, 306–314.

FELGENHAUER, U. (2008b) The shooting approach in analyzing bang–bang extremals with simultaneous control switches. *Control & Cybernetics* **37** (2), 307–327.

GOODMAN, R.W. (1976) Nilpotent Lie groups: structure and applications to analysis, *Lecture Notes in Mathematics* **562**, Springer

KIM, J. R. and MAURER, H. (2003) Sensitivity analysis of optimal control problems with bang–bang controls. In: *Proc. 42nd IEEE Conference on Decision and Control, Hawaii 2003*, **4**, 3281–3286.

KLATTE, D. and KUMMER, B. (2002) *Nonsmooth Equations in Optimization*. Kluwer Acad. Publ, Dordrecht.

LEDZEWICZ, U., NOWAKOWSKI, A. and SCHÄTTLER, H. (2004) Stratifiable families of extremals and sufficient conditions for optimality in optimal control problems. *J. Optim. Theory Appl.* **122** (2), 345–370.

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 OSMOLOVSKII, N.P. (2005) Equivalence of second–order optimality conditions for bang–bang control problems. *Control & Cybernetics* **34** (3), 927–950.

MAURER, H. and OSMOLOVSKII, N.P. (2007) Equivalence of second order optimality conditions for bang–bang control problems. II. Proofs, variational derivatives and representations. *Control & Cybernetics* **36** (1), 5–45.

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

MORDUKHOVICH, B.S. (2006) *Variational Analysis and Generalized Differentiation I,II.* Springer, Berlin.

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

OBERLE H.J. (1987) Numerical computation of singular control functions for a two-link robot arm. In: R. Burlisch et al., eds., *Optimal Control.* **LNCIS 95**, Springer, 244-253.

OSMOLOVSKII, N.P. (1995) Quadratic conditions for nonsingular extremals in optimal control (A theoretical treatment). *Russian J. of Mathem. Physics* **2**, 487–512.

OSMOLOVSKII, N.P. (2004) Quadratic extremality conditions for broken extremals in the general problem of the calculus of variations. Optimal control and dynamical systems. *J. of Mathem. Sciences* **123** (3), 3987–4122.

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

POGGIOLINI, L. and SPADINI, M. (2008) Sufficient optimality conditions for a bang–bang trajectory in a Bolza Problem. In: A. Sarychev et al., eds., *Mathematical Control Theory and Finance*, Springer, 337–357.

POGGIOLINI, L. and SPADINI, M. (2009) Strong local optimality for a bang–bang trajectory in a Mayer problem. Submitted.

POGGIOLINI, L. and STEFANI, G. (2004) State–local optimality of a bang–bang trajectory: a Hamiltonian approach. *Systems Control Lett.* **53**, 269–279.

POGGIOLINI, L. and STEFANI, G. (2006) Sufficient optimality conditions for a bang–bang trajectory. In: *Proc. 45th IEEE Conference on Decision and Control, San Diego (USA).*

ROCKAFELLAR, R.T. and WETS, R.J. (1998) *Variational Analysis.* Springer, Berlin.

SCHÄTTLER, H. (2006) Local fields of extremals for optimal control problems with state constraints of relative degree 1. *J. Dyn. Control Syst.* **12** (4), 563–599.

SARYCHEV, A.V. (1997) First– and second–order sufficient optimality conditions for bang–bang controls. *SIAM J. Control Optim.* **35** (1), 315–340.