This paper was presented at the 38th IEEE Conference on Decision and Control, held in Phoenix, Arizona, December 7-10, 1999, and has appeared in the conference proceedings. The date of this version is August 12, 1999.

A maximum principle for hybrid optimal control problems *

Héctor J. Sussmann
Department of Mathematics; Rutgers, the State University of New Jersey
Hill Center—Busch Campus, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA
E-mail: sussmann@hamilton.rutgers.edu Web site: http://www.math.rutgers.edu/~sussmann

1. Introduction

We present a version of the maximum principle for hybrid optimal control problems, under weak regularity conditions. The class of hybrid problems to be considered is defined in §2. (In particular, we only consider autonomous systems, in which the dynamical behavior and the cost are invariant under time translations.) The maximum principle is stated in §3 as a general assertion involving terms that are not yet precisely defined, and without a detailed specification of technical assumptions. One version of the principle, where the terms are precisely defined and the appropriate technical requirements are completely specified, is stated in §4 for problems where all the basic objects—the dynamics, the Lagrangian, and the cost functions for the switchings and the endpoint constraints—are differentiable along the reference arc. Another version, involving nonsmooth maps, is stated in §5, and some brief remarks on even more general versions are given in §6. To illustrate the use of the maximum principle. two very simple examples, involving problems that can easily be solved directly, are shown in §7. More interesting examples will be discussed in the paper by B. Piccoli in these proceedings.

Our results are stronger than the usual versions of the finite-dimensional maximum principle. For example, even the theorem for classical differentials applies to situations where the maps are not of class C^1, and can fail to be Lipschitz continuous. The "nonsmooth" result applies to maps that are neither Lipschitz continuous nor differentiable in the classical sense. This has been shown for the non-hybrid case in several of our earlier papers on the maximum principle (cf. Sussmann [9, 10]), and in each case it would be trivial to construct hybrid examples of a similar nature.

On the other hand, the results presented here are much weaker than what can actually be proved by our methods. More general versions, involving nonautonomous systems of differential inclusions, will be discussed in subsequent papers.

2. Hybrid optimal control problems

Definition 1 A *hybrid control system* is a 7-tuple $\Sigma = (\mathcal{Q}, M, U, f, \mathcal{U}, J, \mathcal{S})$ such that

H1. $\mathcal{Q}$ is a finite set;

H2. $M = \{M_q\}_{q\in\mathcal{Q}}$ is a family of smooth manifolds, indexed by $\mathcal{Q}$;

H3. $U = \{U_q\}_{q\in\mathcal{Q}}$ is a family of sets;

H4. $f = \{f_q\}_{q\in\mathcal{Q}}$ is a family such that each f_q is a map from the product $M_q \times U_q$ to the tangent bundle TM_q of M_q, such that $f_q(x,u) \in T_xM_q$ for every $(x,u) \in M_q \times U_q$;

H5. $\mathcal{U} = \{\mathcal{U}_q\}_{q\in\mathcal{Q}}$ is a family consisting, for each q, of a set $\mathcal{U}_q$ each of whose members is a map $\eta : I_\eta \to U_q$, defined on some interval $I_\eta \subset \mathbb{R}$;

H6. $\mathcal{S}$ is a subset of $\hat{\mathcal{M}}^2(\Sigma)$, where

$$\hat{\mathcal{M}}^2(\Sigma) \stackrel{\text{def}}{=} \{(q,x,q',x') : q,q' \in \mathcal{Q}, x \in M_q, x' \in M_{q'}\}\,; \quad (1)$$

H7. $J = \{J_q\}_{q\in\mathcal{Q}}$ is a family of subintervals of $\mathbb{R}$, that are allowed to be empty. ♢

The members of $\mathcal{Q}$ are called *locations*. The families M, U, are, respectively, the *family of state spaces* and the *family of control spaces* of Σ. For each q, the manifold M_q, the set U_q, the map f_q, and the set $\mathcal{U}_q$ are, respectively, the *state space*, the *control space*, the *dynamical law*, and the *class of admissible controls* at location q. Usually, $\mathcal{Q}$ will be the set of states of some finite automaton.

For q, q' in $\mathcal{Q}$, we write

$$\mathcal{S}_{q,q'} \stackrel{\text{def}}{=} \{(x,x') \in M_q \times M_{q'} : (q,x,q',x') \in \mathcal{S}\}\,.$$

The sets $\mathcal{S}_{q,q'}$ are the *switching sets* of Σ.

Definition 2 A *control* for a hybrid system Σ as above is a triple $\zeta = (\mathbf{q}, \mathbf{t}, \boldsymbol{\eta})$ such that

- $\mathbf{q} = (q_1, \ldots, q_\nu)$ is a finite sequence of locations;
- $\mathbf{t} = (t_0, t_1, \ldots, t_\nu)$, is a finite sequence of real numbers such that $t_0 < t_1 < \cdots < t_\nu$;
- $t_j - t_{j-1} \in J_{q_j}$ for all $j \in \{1, \ldots, \nu - 1\}$;
- $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_\nu)$ is a finite sequence such that η_j belongs to $\mathcal{U}_{q_j}$ and $I_{\eta_j} = [t_{j-1}, t_j]$ for $j = 1, \ldots, \nu$.

If $\zeta(t) = (\mathbf{q}, \mathbf{t}, \boldsymbol{\eta})$ is a control, and $\mathbf{t} = (t_0, t_1, \ldots, t_\nu)$, we use $\mathbf{q}(\zeta)$, $\mathbf{t}(\zeta)$, $\boldsymbol{\eta}(\zeta)$, $\nu(\zeta)$, a_ζ, b_ζ, I_ζ, to denote, respectively, the finite sequences $\mathbf{q}$, $\mathbf{t}$, $\boldsymbol{\eta}$, the natural number ν, the real numbers t_0, t_ν, and the interval $[a_\zeta, b_\zeta]$. Then I_ζ is the *domain* of ζ, $\nu(\zeta) - 1$ is the *number of switchings*, the t_j are the *switching times*, and $\mathbf{q}(\zeta)$ is the *switching strategy*. ♢

Definition 3 If Σ is a hybrid system as above, ζ is a control for Σ, and $\nu = \nu(\zeta)$, then a *trajectory* for ζ is a ν-tuple $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_\nu)$ with the property that, if $\mathbf{t}(\zeta) = (t_0, \ldots, t_\nu)$, $\mathbf{q}(\zeta) = (q_1, \ldots, q_\nu)$, $\boldsymbol{\eta}(\zeta) = (\eta_1, \ldots, \eta_\nu)$, then, for each $j \in \{1, \ldots, \nu\}$,

- ξ_j is an absolutely continuous map from $[t_{j-1}, t_j]$ to M_{q_j} such that $\dot{\xi}_j(t) = f_{q_j}(\xi_j(t), \eta_j(t))$ for almost all $t \in [t_{j-1}, t_j]$;

*Research supported in part by NSF Grant DMS98-03411-00798 and AFOSR Grant 0923.

- the "switching condition" $(\xi_j(t_j),\xi_{j+1}(t_j))\in\mathcal{S}_{q_j,q_{j+1}}$ holds if $j<\nu$. $\diamondsuit$

Definition 4 If Σ is a hybrid system as above, a *trajectory-control pair* for Σ is a pair $(\boldsymbol{\xi},\zeta)$ such that ζ is a control for Σ and $\boldsymbol{\xi}$ is a trajectory of Σ for ζ. $\diamondsuit$

We use $TCP(\Sigma)$ to denote the set of all trajectory-control pairs of Σ.

Definition 5 If Σ is a hybrid system as above, and $\Xi=(\boldsymbol{\xi},\zeta)\in TCP(\Sigma)$, then the *endpoint condition* $\partial\Xi$ of Ξ is the 4-tuple $(q_0,\xi_1(a_\zeta),q_\nu,\xi_\nu(b_\zeta))$, where $\nu=\nu(\zeta)$, $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_\nu)$, $\mathbf{q}(\zeta)=(q_1,\ldots,q_\nu)$. $\diamondsuit$

It is clear that $\partial\Xi\in\hat{\mathcal{M}}^2(\Sigma)$ whenever $\Xi\in TCP(\Sigma)$. An *endpoint constraint* for Σ is a subset of $\hat{\mathcal{M}}^2(\Sigma)$. If $\mathcal{E}$ is an endpoint constraint for Σ, and $q,q'\in\mathcal{Q}$, we write $\mathcal{E}_{q,q'}=\{(x,x'):(q,x,q',x')\in\mathcal{E}\}$. We use $TCP(\Sigma;\mathcal{E})$ to denote the set of all $\Xi\in TCP(\Sigma)$ such that $\partial\Xi\in\mathcal{E}$. If, in addition, $[a,b]$ is a compact subinterval of $\mathbb{R}$, then $TCP(\Sigma;\mathcal{E};a,b)$ will denote the set of all $\Xi=(\boldsymbol{\xi},\zeta)$ in $TCP(\Sigma;\mathcal{E})$ such that $a_\zeta=a$ and $b_\zeta=b$.

Definition 6 If Σ is a hybrid system as above, then a *Lagrangian* for Σ is a family $L=\{L_q\}_{q\in\mathcal{Q}}$ such that

- each L_q is a real-valued function on $M_q\times U_q$,
- whenever $q\in\mathcal{Q}$, $\eta\in\mathcal{U}_q$ has domain $[\alpha,\beta]$, and $\xi:[\alpha,\beta]\to M_q$ is an absolutely continuous solution of $\dot\xi(t)=f_q(\xi(t),\eta(t))$ a.e., then the function $[\alpha,\beta]\ni t\to L(\xi(t),\eta(t))$ is integrable.

A *switching cost function* for Σ is an extended real-valued function on $\mathcal{S}$ that never takes the value $-\infty$.

An *endpoint cost function* for Σ is an extended real-valued function on $\hat{\mathcal{M}}^2(\Sigma)$ that never takes the value $-\infty$. $\diamondsuit$

If $L=\{L_q\}_{q\in\mathcal{Q}}$ is a Lagrangian for the hybrid control system Σ, then we can define the corresponding *Lagrangian cost functional* $C_L:TCP(\Sigma)\to\mathbb{R}$, by letting

$$C_L(\boldsymbol{\xi},\zeta)=\sum_{j=1}^{\nu}\int_{t_{j-1}}^{t_j}L_{q_j}(\xi_j(t),\eta_j(t))\,dt\,,\qquad(2)$$

where $\nu=\nu(\zeta)$, $\mathbf{t}(\zeta)=(t_0,\ldots,t_\nu)$, $\mathbf{q}(\zeta)=(q_1,\ldots,q_\nu)$, $\boldsymbol{\eta}(\zeta)=(\eta_1,\ldots,\eta_\nu)$, and $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_\nu)$.

If Φ is a switching cost function for Σ, and φ is an endpoint cost function, then we associate with Φ and φ the functional $\hat C_{\Phi,\varphi}:TCP(\Sigma)\to\mathbb{R}\cup\{+\infty\}$ that assigns to each $\Xi=(\boldsymbol{\xi},\zeta)\in TCP(\Sigma)$ the number

$$\hat C_{\Phi,\varphi}(\boldsymbol{\xi},\zeta)=\varphi(\partial\Xi)+\sum_{j=1}^{\nu-1}\Phi(q_j,\xi_j(t_j),q_{j+1},\xi_{j+1}(t_j)),$$

where $\nu=\nu(\zeta)$, $(q_1,\ldots,q_\nu)$ is the switching strategy of ζ, and $t_0,\ldots,t_\nu$ are the switching times.

Definition 7 A *hybrid Bolza cost functional* for Σ is a functional $\mathcal{C}:TCP(\Sigma)\to\mathbb{R}\cup\{+\infty\}$ such that $\mathcal{C}=C_L+\hat C_{\Phi,\varphi}$ for some L,Φ,φ that are, respectively, a Lagrangian, a switching cost function, and an endpoint cost function for Σ. $\diamondsuit$

Given a hybrid control system Σ, a Bolza cost functional $\mathcal{C}$ for Σ, and an endpoint constraint $\mathcal{E}\subseteq\hat{\mathcal{M}}^2(\Sigma)$, we will consider the *variable time interval optimal control problem* $\mathcal{P}(\Sigma,\mathcal{C},\mathcal{E})$, whose objective is to minimize $\mathcal{C}(\boldsymbol{\xi},\zeta)$ in the class $TCP(\Sigma;\mathcal{E})$, and also, for each compact subinterval $[a,b]$ of $\mathbb{R}$, the *fixed time interval optimal control problem* $\mathcal{P}(\Sigma,\mathcal{C},\mathcal{E},a,b)$, whose objective is to minimize $\mathcal{C}(\boldsymbol{\xi},\zeta)$ in $TCP(\Sigma;\mathcal{E};a,b)$.

3. The maximum principle as a principle

We assume throughout this section that

A1. $\Sigma=(\mathcal{Q},M,U,f,\mathcal{U},J,\mathcal{S})$ *is a hybrid control system;*

A2. $\mathcal{C}=C_L+\hat C_{\Phi,\varphi}$ *is a hybrid Bolza cost functional;*

A3. $\mathcal{E}\subseteq\hat{\mathcal{M}}^2(\Sigma)$ *is an endpoint constraint;*

A4. $\Xi^\#=(\boldsymbol{\xi}^\#,\zeta^\#)\in TCP(\Sigma;\mathcal{E};a,b)$.

The *maximum principle* gives a necessary condition for $\Xi^\#$ to be a solution of $\mathcal{P}(\Sigma,\mathcal{C},\mathcal{E},a,b)$ or $\mathcal{P}(\Sigma,\mathcal{C},\mathcal{E})$. The result only depends on comparing trajectories with the same switching strategy, and does not require the candidate arc $\Xi^\#$ to be a true solution. Moreover, even within the class of arcs corresponding to a fixed switching strategy, only arcs that are close to $\Xi^\#$ are compared with $\Xi^\#$. So we define a *local solution* of a problem $\mathcal{P}(\Sigma,\mathcal{C},\mathcal{E},a,b)$ or $\mathcal{P}(\Sigma,\mathcal{C},\mathcal{E})$ to be a trajectory-control pair $\Xi^\#=(\boldsymbol{\xi}^\#,\zeta^\#)=(\xi_1^\#,\ldots,\xi_{\nu^\#}^\#,\zeta^\#)$ such that there exist neighborhoods $\mathcal{N}_1,\ldots,\mathcal{N}_{\nu^\#}$ of the graphs of $\xi_1^\#,\ldots,\xi_{\nu^\#}^\#$ with the property that $\Xi^\#$ minimizes the cost $\mathcal{C}(\Xi)$ in the class of all the trajectory-control pairs $\Xi=(\boldsymbol{\xi},\zeta)=(\xi_1,\ldots,\xi_\nu,\zeta)$ such that $\partial\Xi\in\mathcal{E}$, $\mathbf{q}(\zeta)=\mathbf{q}(\zeta^\#)$ (so that, in particular, $\nu=\nu^\#$), and the graph of ξ_j is contained in $\mathcal{N}_j$ for $j=1,\ldots,\nu^\#$.

We first present the maximum principle as a true "principle," that is, a not very precise mathematical statement that can be rendered precise in various ways, giving rise to different "versions" of the principle. Two such versions—both completely precise and rigorous—will be stated in the next two sections.

The maximum principle. *Assume that A1-A4 hold, and $\mathcal{P}$ is one of the problems $\mathcal{P}(\Sigma,\mathcal{C},\mathcal{E},a,b)$, $\mathcal{P}(\Sigma,\mathcal{C},\mathcal{E})$. Let $\Xi^\#$ be a local solution of $\mathcal{P}$. Then there exists an adjoint pair $(\boldsymbol{\psi},\psi_0)$ along $\Xi^\#$ that satisfies the Hamiltonian maximization, nontriviality, transversality, and Hamiltonian value conditions for $\mathcal{P}$.* $\diamondsuit$

4. A version involving classical differentials

We now make the maximum principle precise in the setting of maps having classical differentials, by giving formal mathematical definitions of the notions of "adjoint pair," "Hamiltonian maximization," "nontriviality," and "transversality," and of the "Hamiltonian value conditions" for a problem $\mathcal{P}$, and by stating technical conditions that make the result true.

For each pair $(q,q')\in\mathcal{Q}\times\mathcal{Q}$, we define two functions $\Phi_{q,q'}:\mathcal{S}_{q,q'}\to\mathbb{R}\cup\{+\infty\}$, $\varphi_{q,q'}:\mathcal{E}_{q,q'}\to\mathbb{R}\cup\{+\infty\}$, by letting

$$\begin{aligned}\Phi_{q,q'}(x,x')&=\Phi(q,x,q',x')\ \text{ for }(x,x')\in\mathcal{S}_{q,q'}\,,\\ \varphi_{q,q'}(x,x')&=\varphi(q,x,q',x')\ \text{ for }(x,x')\in\mathcal{E}_{q,q'}\,.\end{aligned}$$

Also, we define $\gamma_j(\Xi^\#)$—the "j-th jump" of $\Xi^\#$— for $j=1,\ldots,\nu^\#-1$, to be the pair $(\xi_j^\#(t_j),\xi_{j+1}^\#(t_j))$, and we let $\gamma_e(\Xi^\#)$ be the pair $(\xi_1^\#(a),\xi_{\nu^\#}^\#(b))$.

In order to state the transversality condition and to define the notion of adjoint pair we need a concept of "tangent cone" to a set. There are many nonequivalent definitions of tangent cone, and we choose for our purposes the notion of a "Boltyanskii approximating cone":

Definition 8 Let S be a subset of a smooth manifold X, and let $\bar{x} \in X$. A *Boltyanskii approximating cone* to S at $\bar{x}$ is a closed convex cone K in the tangent space $T_{\bar{x}}X$ to X at $\bar{x}$ such that there exist a neighborhood V of 0 in $T_{\bar{x}}X$ and a continuous map $\mu : V \cap K \to M$ with the property that $\mu(V \cap K) \subseteq S$, $\mu(0) = \bar{x}$, and $\mu(v) = \bar{x} + v + o(\|v\|)$ as $v \to 0$ via values in $V \cap K$. $\diamondsuit$

(The condition "$\mu(v) = \bar{x}+v+o(\|v\|)$ as $v \to 0$" appears to depend, in principle, on a choice of local coordinates about $\bar{x}$, but is in fact independent of that choice.)

We now define the notion of an "adjoint pair" along $\Xi^{\#}$, and what it means for such a pair to be "Hamiltonian-maximizing." For that purpose, we first stipulate that

S1. *"tangent cone" means "Boltyanskii approximating cone."*

We then let

$$\zeta^{\#} = (\mathbf{q}^{\#}, \mathbf{t}^{\#}, \boldsymbol{\eta}^{\#}), \qquad \mathbf{t}^{\#} = (t_0^{\#}, t_1^{\#}, \ldots, t_{\nu^{\#}}^{\#}),$$
$$\mathbf{q}^{\#} = (q_1^{\#}, \ldots, q_{\nu^{\#}}^{\#}), \qquad \boldsymbol{\eta}^{\#} = (\eta_1^{\#}, \ldots, \eta_{\nu^{\#}}^{\#}),$$
$$\boldsymbol{\xi}^{\#} = (\xi_1^{\#}, \ldots, \xi_{\nu^{\#}}^{\#}),\ I_j^{\#} = [t_{j-1}^{\#}, t_j^{\#}] \text{ for } j = 1, \ldots, \nu^{\#}.$$

and make the following assumptionas.

A5. *For $j \in \{1, \ldots, \nu - 1\}$, K_j is a tangent cone to the set $\mathcal{S}_{q_j^{\#}, q_{j+1}^{\#}}$ at $\gamma_j(\Xi^{\#})$.*

A6. *K_e is a tangent cone to the set $\mathcal{E}_{q_1^{\#}, q_\nu^{\#}}$ at $\gamma_e(\Xi^{\#})$.*

We recall that if V is a finite-dimensional real linear space and $S \subseteq V$, then the *polar* of S is the set

$$S^{\perp} = \{w \in V^{\dagger} : w \cdot v \leq 1 \text{ for all } v \in S\},$$

where $V^{\dagger}$ is the dual of V. If S is a cone (that is, S is nonempty and such that $r \cdot v \in S$ whenever $v \in S$ and $r \geq 0$), then $S^{\perp} = \{w \in V^{\dagger} : w \cdot v \leq 0 \text{ for all } v \in S\}$.

It is well known that if S is a cone then $S^{\perp}$ is a closed convex cone, and $S^{\perp\perp}$ is the closed convex hull of S (using the standard identification of $V^{\dagger\dagger}$ with V), so $S^{\perp\perp} = S$ if and only if S is closed and convex.

Next, we assume

A7. *If $j \in \{1, \ldots, \nu\}$ then for almost every $t \in I_j^{\#}$ the maps $x \to f_{q_j^{\#}}(x, \eta_j^{\#}(t))$ and $x \to L_{q_j^{\#}}(x, \eta_j^{\#}(t))$ are differentiable at $x = \xi_j^{\#}(t)$.*

A8. *If $j \in \{1, \ldots, \nu - 1\}$, then $\Phi_{q_j^{\#}, q_{j+1}^{\#}}$ is differentiable at $\gamma_j(\Xi^{\#})$ along $\mathcal{S}_{q_j^{\#}, q_{j+1}^{\#}}$.*

A9. *$\varphi_{q_1^{\#}, q_\nu^{\#}}$ is differentiable at $\gamma_e(\Xi^{\#})$ along $\mathcal{E}_{q_1^{\#}, q_\nu^{\#}}$.*

The precise meaning of "differentiability along a set" in A8 and A9 is as follows: a partially defined function h on a smooth manifold X is *differentiable along* S at a point $\bar{x}$ of a subset S of X if (a) for some neighborhood V of $\bar{x}$, the inclusion $V \cap S \subseteq \text{Domain}(h)$ holds, and (b) there exists a linear functional λ on the tangent space $T_{\bar{x}}X$ such that

$$\lim_{x \to \bar{x}, x \in S} \frac{h(x) - h(\bar{x}) - \lambda \cdot (x - \bar{x})}{\|x - \bar{x}\|} = 0. \qquad (3)$$

Naturally, (3) makes sense relative to a choice of local coordinates about $\bar{x}$, but it is easy to verify that the validity of (3) does not depend on that choice. The linear functional λ need not be unique but, if K is a tangent cone to S at $\bar{x}$, then the restriction of λ to K is unique, since $\lambda(v) = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1}(h(\mu(\varepsilon v)) - h(\bar{x}))$, if μ is any map that satisfies the conditions of Definition 8. In particular, the condition "$\nabla h(\bar{x}) \in K^{\perp}$" makes sense intrinsically, if it is interpreted to mean "$\lambda \in K^{\perp}$ for any λ such that (3) holds."

For each $q \in \mathcal{Q}$, we define a real-valued function H_q on $T^*M_q \times U_q \times \mathbb{R}$—where T^*M_q is the cotangent bundle of M_q—by letting

$$H_q(x, p, u, p_0) = p \cdot f_q(x, u) - p_0 L_q(x, u) \qquad (4)$$

for $x \in M_q$, $u \in U_q$, $p \in T_x^*M_q$, $p_0 \in \mathbb{R}$.

Definition 9 If S1 and A1-A9 hold, then an *adjoint pair along* $\Xi^{\#}$ is a pair $(\boldsymbol{\psi}, \psi_0)$ with the property that:

- $\boldsymbol{\psi}$ is a $\nu^{\#}$-tuple $(\psi_1, \ldots, \psi_{\nu^{\#}})$ such that each ψ_j is a field of covectors along $\xi_j^{\#}$ (that is, ψ_j is a map from the interval $I_j^{\#}$ to the cotangent bundle $T^*M_{q_j^{\#}}$, with the property that $\psi_j(t)$ belongs to the cotangent space $T^*_{\xi_j^{\#}(t)}M_{q_j^{\#}}$ for every $t \in I_j^{\#}$);
- each ψ_j is absolutely continuous;
- $\psi_0 \in \mathbb{R}$ and $\psi_0 \geq 0$;
- each ψ_j satisfies the *adjoint equation*

$$\dot{\psi}_j(t) = -\frac{\partial H_{q_j^{\#}}}{\partial x}(\xi_j^{\#}(t), \psi_j(t), \eta_j^{\#}(t), \psi_0) \text{ for a.e. } t \in I_j^{\#};$$

- for each $j \in \{1, \ldots, \nu^{\#}-1\}$, the *switching condition*

$$(-\psi_j(t_j^{\#}), \psi_{j+1}(t_j^{\#})) - \psi_0 \omega_j \in K_j^{\perp} \qquad (5)$$

holds, where $\omega_j = \nabla \Phi_{q_j^{\#}, q_{j+1}^{\#}}(\gamma_j(\Xi^{\#}))$. $\diamondsuit$

Definition 10 If S1 and A1-A9 hold, and $(\boldsymbol{\psi}, \psi_0)$ is an adjoint pair along $\Xi^{\#}$, we say that $(\boldsymbol{\psi}, \psi_0)$ satisfies the *Hamiltonian maximization condition* if there exist real numbers $h_1, \ldots, h_{\nu^{\#}}$ such that for each $j \in \{1, \ldots, \nu^{\#}\}$

$$\begin{aligned} h_j &= H_{q_j^{\#}}(\xi_j^{\#}(t), \psi_j(t), \eta_j^{\#}(t), \psi_0) \\ &= \max\{H_{q_j^{\#}}(\xi_j^{\#}(t), \psi_j(t), u, \psi_0) : u \in U_{q_j^{\#}}\} \end{aligned} \qquad (6)$$

for almost all $t \in I_j^{\#}$. $\diamondsuit$

Definition 11 If S1 and A1-A9 hold, and $(\boldsymbol{\psi}, \psi_0)$ is an adjoint pair along $\Xi^{\#}$, we say that $(\boldsymbol{\psi}, \psi_0)$ satisfies the *transversality condition* for $\mathcal{E}$ and φ if

$$(\psi_1(a), -\psi_\nu(b)) - \psi_0 \omega_e \in C_e^{\perp}, \qquad (7)$$

where $\omega_e = \nabla \varphi_{q_1^{\#}, q_{\nu^{\#}}^{\#}}(\gamma_e(\Xi^{\#}))$. $\diamondsuit$

Definition 12 If $(\boldsymbol{\psi}, \psi_0)$ is an adjoint pair along $\Xi^{\#}$, we say that $(\boldsymbol{\psi}, \psi_0)$ satisfies the *nontriviality condition* if either $\psi_0 \neq 0$ or at least one of the functions ψ_j is not identically zero. $\diamondsuit$

Definition 13 If A1-A9 hold, $(\boldsymbol{\psi}, \psi_0)$ is an adjoint pair along $\Xi^{\#}$, and $h_1, \ldots, h_{\nu^{\#}}$ are the constants for which (6) holds, we say that $(\boldsymbol{\psi}, \psi_0)$ satisfies the *Hamiltonian value condition for the fixed time interval problem* $\mathcal{P}(\Sigma, \mathcal{C}, \mathcal{E}, a, b)$ if, for every $j \in \{1, \ldots, \nu^{\#}-1\}$,

- if $t_j^{\#} - t_{j-1}^{\#} \in \mathrm{Interior}(J_{q_j^{\#}})$, then $h_j = h_{\nu^{\#}}$;
- if $t_j^{\#} - t_{j-1}^{\#}$ is the left endpoint of $J_{q_j^{\#}}$, but $J_{q_j^{\#}}$ is nontrivial, then $h_j \leq h_{\nu^{\#}}$;
- if $t_j^{\#} - t_{j-1}^{\#}$ is the right endpoint of $J_{q_j^{\#}}$, but $J_{q_j^{\#}}$ is nontrivial, then $h_j \geq h_{\nu^{\#}}$;

We say that $(\boldsymbol{\psi}, \psi_0)$ satisfies the *Hamiltonian value condition for the variable time interval problem* $\mathcal{P}(\Sigma, \mathcal{C}, \mathcal{E})$ if it satisfies the Hamiltonian value condition for $\mathcal{P}(\Sigma, \mathcal{C}, \mathcal{E}, a, b)$ and in addition $h_{\nu^{\#}} = 0$. $\diamondsuit$

We have now completed the list of definitions needed to make the statement of the maximum principle meaningful, but in order to make the result true we need the following additional conditions.

A10 *$f_q(x, u)$ and $L_q(x, u)$ are continuous with respect to x for each $q \in \mathcal{Q}$ and each $u \in U_q$.*

A11 *The maps $I_j^{\#} \ni t \to (f_{q_j^{\#}}(x, \eta_j^{\#}(t)), L_{q_j^{\#}}(x, \eta_j^{\#}(t)))$ are measurable for every $q \in \mathcal{Q}$, $x \in M_q$.*

A12 *$\Phi_{q,q'}$ and $\varphi_{q,q'}$ are finite-valued and continuous on the sets $\mathcal{S}_{q,q'}$, $\mathcal{E}_{q,q'}$, for every q, q'.*

A13 *The differentiability of the maps*

$$M_q \ni x \to \tilde{f}_j(x, t) \stackrel{\text{def}}{=} (f_{q_j^{\#}}(x, \eta_j^{\#}(t)), -L_{q_j^{\#}}(x, \eta_j^{\#}(t)))$$

at $x = \xi_j^{\#}(t)$ has the following Carathéodory-type uniformity: there exists integrable functions $k_\delta : [a, b] \to \mathbb{R} \cup \{+\infty\}$ such that $\lim_{\delta \downarrow 0} \int_a^b k_\delta(t)\, dt = 0$, having the property that, for almost all $t \in I_j^{\#}$, the inequality

$$\|\hat{f}_j(x, t) - w_j(t) \cdot (x - \xi^{\#}(t))\| \leq k_\delta(t) \|x - \xi_j^{\#}(t)\| \qquad (8)$$

holds whenever $\|x - \xi_j^{\#}(t)\| \leq \delta$, where we define

$$\hat{f}_j(x, t) \stackrel{\text{def}}{=} \tilde{f}_j(x, t) - \tilde{f}_j(\xi_j^{\#}(t), t),\ w_j(t) = \frac{\partial \tilde{f}_j}{\partial x}(\xi_j^{\#}(t), t).$$

A14. *Each set $\mathcal{U}_q$ is invariant under time translations, restrictions and concatenations—that is, if a control $\eta : [a, b] \to U_q$ belongs to $\mathcal{U}_q$, $a \leq c \leq d \leq b$, and $\tau \in \mathbb{R}$, then the maps $[c, d] \ni t \to \eta(t) \in U_q$ and $[a - \tau, b - \tau] \ni t \to \eta(t + \tau) \in U_q$ also belong to $\mathcal{U}_q$; if, in addition, $\eta' : [b, b'] \to U_q$ also belongs to $\mathcal{U}_q$ and, we define $\eta''(t) = \eta(t)$ for $a \leq t \leq b$, $\eta''(t) = \eta'(t)$ for $b < t \leq b'$, then $\eta'' \in \mathcal{U}_q$—and contains all the constant U_q-valued maps defined on compact intervals.*

Theorem 1 *If S1 and A1-A14 hold, then the maximum principle is true.* $\diamondsuit$

Theorem 1 is stronger than the classical versions of the maximum principle given, e.g., in Pontryagin *et al.* [6] and Berkovitz [1], because it does not require the maps to be of class C^1, and only assumes that they are differentiable along the reference trajectory.

5. A nonsmooth version

The simplest way to get a nonsmooth analogue of Theorem 1 is to allow the map $t \to w(t)$ of A13 to be set-valued, and to allow "generalized differentials" instead of ordinary differentials in A8 and A9.

For simplicity, we assume now that

A15 *each M_q is an open subset of a Euclidean space X_q.*

Then a vector field on M_q is just a map from M_q to X_q, the classical differential at a point of a vector field is a member of the space $L(X_q, X_q)$ of linear maps from X_q to X_q, and the classical differentials at a point of the maps $\tilde{f}_j$ defined in A14 are members of the space $\Lambda_j \stackrel{\text{def}}{=} L(X_{q_j^{\#}}, X_{q_j^{\#}} \times \mathbb{R})$ of linear maps from $X_{q_j^{\#}}$ to $X_{q_j^{\#}} \times \mathbb{R}$.

In our generalization, we substitute the following for A8, A9 and A13.

A8' - A9' *There exist a function $\delta :]0, +\infty[\to]0, +\infty]$, compact convex nonempty subsets Ω_j of the product $X_{q_j^{\#}}^{\dagger} \times X_{q_{j+1}^{\#}}^{\dagger}$ for $j = 1, \ldots, \nu - 1$, and a compact convex nonempty subset Ω_e of $X_{q_1^{\#}}^{\dagger} \times X_{q_\nu^{\#}}^{\dagger}$, such that $\theta(\delta) \to 0$ as $\delta \downarrow 0$,*

$$\hat{\Phi}_{q_j^{\#}, q_{j+1}^{\#}}(x_1, x_2) - \lambda_j(x_1, x_2) \leq \delta \theta(\delta)$$

whenever $\|x_1 - \xi_j^{\#}(t_j^{\#})\| + \|x_2 - \xi_{j+1}^{\#}(t_j^{\#})\| \leq \delta$ and $(x_1, x_2) \in \mathcal{S}_{q_j^{\#} q_{j+1}^{\#}}$, and

$$\hat{\varphi}_{q_1^{\#}, q_{\nu^{\#}}^{\#}}(x_1, x_2) - \lambda_e(x_1, x_2) \leq \delta \theta(\delta)$$

whenever $\|x_1 - \xi_1^{\#}(a)\| + \|x_2 - \xi_{\nu^{\#}}^{\#}(b)\| \leq \delta$ and $(x_1, x_2) \in \mathcal{E}_{q_1^{\#}, q_{\nu^{\#}}^{\#}}$, where

$$\hat{\Phi}_{q_j^{\#}, q_{j+1}^{\#}}(x_1, x_2) = \Phi_{q_j^{\#}, q_{j+1}^{\#}}(x_1, x_2) - \Phi_{q_j^{\#}, q_{j+1}^{\#}}(\gamma_j(\Xi^{\#})),$$

$$\lambda_j(x_1, x_2) = \max\Big\{\omega \cdot \Big(x_1 - \xi_j^{\#}(t_j^{\#}), x_2 - \xi_{j+1}^{\#}(t_j^{\#})\Big) : \omega \in \Omega_j\Big\},$$

$$\hat{\varphi}_{q_1^{\#}, q_{\nu^{\#}}^{\#}}(x_1, x_2) = \varphi_{q_1^{\#}, q_{\nu^{\#}}^{\#}}(x_1, x_2) - \varphi_{q_1^{\#}, q_{\nu^{\#}}^{\#}}(\gamma_e(\Xi^{\#})),$$

$$\lambda_e(x_1, x_2) = \max\Big\{\omega \cdot \Big(x_1 - \xi_1^{\#}(a), x_2 - \xi_{\nu^{\#}}^{\#}(b)\Big) : \omega \in \Omega_e\Big\}.$$

A13' *There exist*

- *integrable functions $k_\delta : [a, b] \to \mathbb{R} \cup \{+\infty\}$, for $\delta > 0$,*
- *an integrable function $k : [a, b] \to \mathbb{R} \cup \{+\infty\}$,*
- *measurable set-valued maps $W_j : I_j^{\#} \to 2^{\Lambda_j}$ with nonempty compact convex values,*

such that

- $\min\{\|\hat{f}_j(x, t) - w \cdot (x - \xi_j^{\#}(t))\| : w \in W_j(t)\} \leq k_\delta(t) \|x - \xi_j^{\#}(t)\| \qquad (9)$

  (where the $\hat{f}_j$ are the maps defined in A13) whenever $\|x - \xi_j^{\#}(t)\| \leq \delta$,
- $\lim_{\delta \downarrow 0} \int_a^b k_\delta(t)\, dt = 0$,
- $\max\{\|w\| : w \in W_j(t)\} \leq k(t)$ *for a.e.* $t \in I_j^{\#}$.

We then observe that in the classical case of §4 the adjoint equation is exactly equivalent to the statement that

$$\dot{\psi}_j(t) = -(\psi_j(t), -\psi_0) \cdot w_j(t) \text{ for a.e. } t \in I_j^\#, \quad (10)$$

and modify Definitions 9, 15 as follows.

Definition 14 If S1, A1-A7, A8'-A9', A13' and A15 hold, then an *adjoint pair along* $\Xi^\#$ is a pair $(\boldsymbol{\psi}, \psi_0)$ with the property that:

- $\boldsymbol{\psi}$ is a $\nu^\#$-tuple $(\psi_1, \ldots, \psi_{\nu^\#})$ such that each ψ_j is a field of covectors along $\xi_j^\#$
- each ψ_j is absolutely continuous;
- each ψ_j satisfies the *adjoint equation* (10) for some measurable selection $t \to w_j(t)$ of W_j,
- for each $j \in \{1, \ldots, \nu^\# - 1\}$,

$$(\exists \omega_j \in \Omega_j)\Big((-\psi_j(t_j^\#), \psi_{j+1}(t_j^\#)) - \psi_0\omega_j \in K_j^\perp\Big); \quad (11)$$

- $\psi_0 \in \mathbb{R}$ and $\psi_0 \geq 0$. ◇

Definition 15 If S1, A1-A7, A8'-A9', A13' and A15 hold, and $(\boldsymbol{\psi}, \psi_0)$ is an adjoint pair along $\Xi^\#$, we say that $(\boldsymbol{\psi}, \psi_0)$ satisfies the *transversality condition* for $(\mathcal{E}, \varphi)$ if $(\psi_1(a), -\psi_\nu(b)) - \psi_0\omega_e \in K_e^\perp$ for some $\omega_e \in \Omega_e$. ◇

The nontriviality, Hamiltonian maximization and Hamiltonian value conditions remain unchanged, and the following nonsmooth analogue of Theorem 1 holds.

Theorem 2 *If S1, A1-A7, A8'A9', A10-A12, A13', A14 and A15 hold, then the maximum principle is true provided that the concept of "adjoint pair" and the transversality condition are interpreted according to Definitions 14 and 15.* ◇

6. Other versions

The nonsmooth version stated in Theorem 2 applies, for example, if the maps $x \to \tilde{f}_j(x,t)$ are Lipschitz continuous for fixed t—with an integrable bound on the Lipschitz constant—and the maps $\Phi_{q_j^\#, q_{j+1}^\#}$, $\varphi_{q_1^\#, q_n^\#}$ are Lipschitz continuous near $\gamma_j(\Xi^\#)$, $\gamma_e(\Xi^\#)$. In this case, one can simple let $W_j(t)$ to be the Clarke generalized Jacobian (cf. Clarke [3]) of $\tilde{f}_j$ with respect to x at $x = \xi_j^\#$, and let Ω_j, Ω_e be the Clarke generalized Jacobians of $\Phi_{q_j^\#, q_{j+1}^\#}$ and $\varphi_{q_1^\#, q_\nu^\#}$ at $\gamma_j(\Xi^\#)$, $\gamma_e(\Xi^\#)$. This special case of Theorem 2 is already more general than the usual nonsmooth versions of the maximum principle (cf., e. g., Clarke [2]), because it allows the "Lojasiewicz improvement," consisting of not requiring the maps $x \to \tilde{f}_j(x,u)$ to be Lipschitz continuous for values of u other than those that occur in the reference control, and imposing only the weaker condition A10.

In addition, Theorem 2 applies even in the case when each map $x \to \tilde{f}_j(x,t)$ is just *semidifferentiable* at $\xi_j^\#(t)$—with some integral bounds, as usual— and the maps $\Phi_{q_j^\#, q_{j+1}^\#}$, $\varphi_{q_1^\#, q_n^\#}$ are semidifferentiable at $\gamma_j(\Xi^\#)$, $\gamma_e(\Xi^\#)$. (Semidifferentiability and semidifferentials are defined, for example, in [9].) In that case, one can take the $W_j(t)$, Ω_j and Ω_e to be convex semidifferentials of the maps.

Even more general versions can be proved, using other theories of generalized differentials such as the "multidifferentials" introduced in [8], as generalizations of J. Warga's theory of "derivate containers" (cf. Warga [11, 12]) and H. Halkin's theory of "screens" (cf. Halkin [4, 5]). This leads to a more general hybrid maximum principle for systems of differential inclusions, along the lines of the results of [10].

Assumption A15, that the systems under consideration are defined on open subsets of Euclidean spaces, was made in order to simplify the statements. The appropriate setting for these results is that of systems on manifolds. All the results carry over with no change, provided only that the objects of interest are required to belong to the appropriate spaces or bundles. For example, the generalized differentials Ω_j, Ω_e must be subsets of the products $T^*_{\xi_j^\#(t_j^\#)} M_{q_j^\#} \times T^*_{\xi_{j+1}^\#(t_j^\#)} M_{q_{j+1}^\#}$ and $T^*_{\xi_1^\#(a)} M_{q_1^\#} \times T^*_{\xi_{\nu^\#}^\#(b)} M_{q_{\nu^\#}^\#}$, respectively. Defining in an intrinsic way the generalized differential $W_j(t)$ of the map $x \to \tilde{f}_j(x,t)$ at $\xi_j^\#(t)$ is possible but somewhat more delicate. The correct answer turns out to be that $W_j(t) \subseteq \mathcal{JVF}^1(\xi_j^\#(t), f_j(\xi_j^\#(t), \eta_j^\#(t))) \times T^*_{\xi_j^\#(t)} M_{q_j^\#}$. (If X is a manifold, $x \in X$, and $v \in T_xX$, then $\mathcal{JVF}^1(x,v)$ is the space of all 1-jets at x of smooth vector fields V such that $V(x) = v$, so $\mathcal{JVF}^1(x,v)$ has a canonical affine space structure[†]. A detailed discussion of these differential-geometric issues is given in Sussmann [7], where the members of $\mathcal{JVF}^1(x) \stackrel{\text{def}}{=} \bigcup_{v \in T_xX} \mathcal{JVF}^1(x,v)$ are called "connections at x.")

Finally, we point out that all the results stated here for autonomous problems carry over to the time-varying case, provided that some extra hypotheses are made to deal with several technical complications. In particular, one has to assume that, on some tube about the reference trajectory, $\|f(x,u,t)\|$ and $|L(x,u,t)|$ are bounded by integrable functions of t. (These conditions can be dispensed with, but then one must assume instead that the classes $\mathcal{U}_q$ of admissible controls are closed under the operation of replacement on arbitrary measurable sets rather than just intervals.) Also, the Hamiltonian maximization has to be reformulated in the "weak" form (as explained in Sussmann [7]), in which the inequality $H_{q_j^\#}(\xi_j^\#(t), \psi_j(t), u, \psi_0) \leq H_{q_j^\#}(\xi_j^\#(t), \psi_j(t), \eta_j^\#(t), \psi_0)$ is required to hold almost everywhere for each u—that is, with the possible exception of a "bad" null set B_u that may depend on u. (Or one could assume instead that the U_q are separable metric spaces—rather than general sets with no structure—and the maps $u \to \tilde{f}_j(\xi_j^\#(t), u, t)$ are continuous for a. e. t[‡].)

7. Examples

We present two simple examples to illustrate the use of our results. More interesting examples of harder problems appear in B. Piccoli's paper in these proceedings.

[†]but not a linear space structure!

[‡]This assumption is not needed for autonomous problems because in that case one can always fix compact neighborhoods N_j of the sets $\xi_j^\#(I_j^\#)$ and substitute for each control set $U_{q_j^\#}$ the space U_j^{new} of all pairs (V, σ), where V is a continuous vector field on N_j and σ is a continuous real-valued function on N_j. With these new spaces instead of the original ones, $(*)$ automatically holds.

Example 1 *(Reflection)* We study the light rays—i.e., minimum time paths— in a medium in $\mathbb{R}^n$—with coordinates $x^1, \ldots, x^n$—where the speed of light is a smooth function $x \to c(x)$ with positive values, subject to the extra constraint that at least one point along the ray must belong to the hyperplane $S = \{x : x^n = 0\}$. The control law is $\dot{x} = c(x)u$, $u \in U$, where the control set U is the closed unit ball of $\mathbb{R}^n$. This is a standard—non-hybrid—control problem, but the extra constraint that the rays must intersect S makes it convenient to represent it as a "hybrid problem" in our sense, with just one location q, $J_q = \mathbb{R}$, and a switching set $\mathcal{S}_{qq} = \{(x, x') : x = x' \in S\}$.

Suppose that a path $\xi : [a, b] \to \mathbb{R}^n$, going from $A = \xi(a)$ to $B = \xi(b)$ is a solution of the original problem, corresponding to a control $[a, b] \ni t \to \eta(t) \in U$. Assume, moreover, that $A \notin S$, $B \notin S$, and A and B are on the same side of S. (In the other cases the extra constraint is automatically satisfied for all paths going from A to B, so our "hybrid" problem is exactly equivalent to the non-hybrid one.) Then ξ arises in an obvious way from a trajectory-control pair Ξ of the hybrid problem, with switching strategy (q, q), and this new pair is a local solution of the hybrid optimal control problem vith variable time interval. (Notice that Ξ need not be a *global* solution since, for example, a path going directly from A to B without going through S might be faster.)

Theorem 1 implies the existence of an adjoint pair (ψ, ψ_0) of the usual kind, except that now ψ can have a jump at the switching time $\bar{t}$. Clearly, $a < \bar{t} < b$.

Since the hybrid optimal control problem has a variable time interval and the set J_q of possible switching times is $\mathbb{R}$, the Hamiltonian value condition implies that the Hamiltonian $h(t) = c(\xi(t))\psi(t)\cdot\eta(t) - \psi_0$ vanishes on $[a, b]$. The Hamiltonian maximization condition implies that $h(t) = c(\xi(t))\|\psi(t)\| - \psi_0$. It then follows from the nontriviality condition that $\psi_0 \neq 0$ and $\psi(t) \neq 0$. Then $\eta(t) = \frac{\psi(t)}{\|\psi(t)\|}$. After dividing by ψ_0, we may assume that $\psi_0 = 1$, so $\eta(t) \equiv c(\xi(t))\psi(t)$. Clearly, then, η is continuous except possibly at the switching time $\bar{t}$.

At any point of S, the set S is its own tangent cone. The switching condition says that $(-\eta(\bar{t}-), \eta(\bar{t}+))$ belongs to $S^\perp$, and this easily implies the identity $(\eta(\bar{t}+) - \eta(\bar{t}-)) \cdot v = 0$ for every $v \in S$. This says that $\eta^1(t), \ldots, \eta^{n-1}(t)$—that is, the components of $\eta(t)$ tangential to S—are continuous at $\bar{t}$. The normal component $\eta^n(t)$ is allowed to jump, but the facts that $\|\eta(t)\|$ is constant and the other components are continuous imply that the only possible jump is for η^n to change sign, i.e. to be such that $\eta^n(t+) = -\eta^n(t-)$. Thus we recover the well known law about reflection. ♢

Example 2 Consider the same problem as in the previous example, except that now the set S is the half-space $\{(x^1, \ldots, x^n) : x^n \leq 0\}$. In this case, if the "switching" occurs at an interior point of S, the tangent cone K to $\mathcal{S}_{qq}$ is $\{(v, v) : v \in \mathbb{R}^n\}$. Hence in this case the switching condition simply says that $\eta(\bar{t}+) = \eta(\bar{t}-)$, so that there is no discontinuity of ψ and η at $\bar{t}$, and ξ satisfies the necessary conditions for optimality for the unconstrained problem. (If $A \notin S$ and $B \notin S$, then this possibility is excluded if, for example, c is constant, because in that case the adjoint equation says that $\dot{\psi} = 0$, so η will actually have to be constant, and then ξ is a straight-line segment, and cannot go through S. But, naturally, if c is not constant it may very well happen that the fastest path from A to B goes through S.)

If, on the other hand, the switching occurs on the boundary of S, then K is the set $\{(v, v) : v \in S\}$. Then the switching condition $(-\eta(\bar{t}-), \eta(\bar{t}+)) \in K^\perp$ implies, as before, that $\eta^1, \ldots, \eta^{n-1}$ are continuous and η^n can only change sign at $\bar{t}$. Moreover, the switching condition also implies that $\eta^n(\bar{t}+) \geq \eta^n(\bar{t}-)$, consistent with the obvious fact that η^n can only jump at $\bar{t}$ from a negative value to a positive one. ♢

References

[1] Berkovitz, L. D., *Optimal Control Theory.* Springer-Verlag, New York, 1974.

[2] Clarke, F. H., *The Maximum Principle under minimal hypotheses.* SIAM J. Control Optim. **14**, 1976, pp. 1078–1091.

[3] Clarke, F. H., *Optimization and Nonsmooth Analysis.* Wiley Interscience, New York, 1983.

[4] Halkin, H., *Necessary conditions for optimal control problems with differentiable or nondifferentiable data.* In *Mathematical Control Theory,* Lect. Notes in Math. **680**, Springer-Verlag, Berlin, 1978, pp. 77–118.

[5] Halkin, H., *Necessary conditions in mathematical programming and optimal control theory.* In *Optimal Control Theory and its Applications, Part I,* Lect. Notes in Econom. and Math. Systems **105**, Springer-Verlag, Berlin, 1974, pp. 113–165.

[6] Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mischenko, *The Mathematical Theory of Optimal Processes.* Wiley, New York, 1962.

[7] Sussmann, H. J., *An introduction to the coordinate-free maximum principle.* In *Geometry of Feedback and Optimal Control*, B. Jakubczyk and W. Respondek Eds., Marcel Dekker, New York, 1997, pp. 463–557.

[8] Sussmann, H. J., *Multidifferential calculus: chain rule, open mapping and transversal intersection theorems.* In *Optimal Control: Theory, Algorithms, and Applications*, W. W. Hager and P. M. Pardalos Eds., Kluwer, 1998, pp. 436–487.

[9] Sussmann, H. J., *Geometry and optimal control.* In *Mathematical Control Theory*, J. Baillieul and J. C. Willems Eds., Springer-Verlag, New York, 1998, pp. 140–198.

[10] Sussmann, H. J., *Transversality conditions and a strong maximum principle for systems of differential inclusions.* In *Proc. 37th IEEE Conference on Decision and Control, Tampa, FL, Dec. 1998,* IEEE publications, New York, 1998, pp. 1–6.

[11] Warga, J., *Controllability, extremality and abnormality in nonsmooth optimal control.* J. Optim. Theory Applic. **41**, 1983, pp. 239–260.

[12] Warga, J., *Optimization and controllability without differentiability assumptions.* SIAM J. Control and Optimization **21**, 1983, pp. 837–855.