1. A nonsmooth hybrid maximum principle

Héctor J. Sussmann[1,2]

Department of Mathematics
Rutgers, the State University of New Jersey
Hill Center—Busch Campus
110 Frelinghuysen Road
Piscataway, NJ 08854-8019, USA
sussmann@hilbert.rutgers.edu

Summary.

We present two versions of the maximum principle for nonsmooth hybrid optimal control problems, the first one of which requires differentiability along the reference trajectory and yields an adjoint equation of the usual kind, while the second one only requires approximability to first order by Lipschitz maps, and yields an adjoint differential inclusion involving a generalized gradient of the approximating Hamiltonian.

1.1 Introduction

In this paper we present a version of the maximum principle for nonsmooth hybrid optimal control problems, under weak regularity conditions. The class of hybrid problems to be considered is defined in §1.2. The maximum principle is stated in §1.3 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 §1.4 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, but a considerable amount of nonsmoothness is allowed away from the reference control. A fairly detailed outline of the proof of the result of §1.4 is presented in §1.5. Another version

[1] Research supported in part by NSF Grant DMS-9803411 and AFOSR Grant 0923.
[2] Most of this work was done in the Netherlands, during a three-month visit at the University of Groningen, to which the author is immensely grateful for its generous hospitality and exciting intellectual atmosphere.

of the hybrid maximum principle, allowing nondifferentiability along the reference control—and requiring only "first-order approximability by Lipschitz maps"—is stated without proof in §1.6.

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 [5, 6]), 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 systems of differential inclusions, and "flows" that do not arise from vector fields, will be discussed in subsequent papers.

A simpler version of our results, dealing only with autonomous systems, was announced—without proof—in the conference paper [7].

1.2 Hybrid optimal control problems

Throughout this paper, the expression "smooth manifold"—or, simply, the word "manifold"—means "finite-dimensional Hausdorff manifold of class C^1 without boundary." If M is a manifold, and $x \in M$, then T_xM, T_x^*M, TM, T^*M denote, respectively, the tangent and cotangent spaces of M at x, and the tangent and cotangent bundles of M.

Definition 1.2.1. *A* ***finite family of state spaces*** *is a pair* $(\mathcal{Q}, \mathcal{M})$ *such that*

FFSS1. $\mathcal{Q}$ *is a finite set;*

FFSS2. $\mathcal{M} = \{M_q\}_{q \in \mathcal{Q}}$ *is a family of smooth manifolds, indexed by* $\mathcal{Q}$. ♢

If $(\mathcal{Q}, \mathcal{M})$ is a finite family of state spaces, then for each pair $(q, q') \in \mathcal{Q} \times \mathcal{Q}$ we use $\mathcal{M}_{q,q'}$ to denote the product $M_q \times M_{q'} \times \mathbb{R} \times \mathbb{R}$.

Definition 1.2.2. *A* ***switching constraint*** *for a finite family of state spaces* $(\mathcal{Q}, \mathcal{M})$ *is a family* $\mathcal{S} = \{S_{q,q'}\}_{(q,q') \in \mathcal{Q} \times \mathcal{Q}}$ *such that* $S_{q,q'}$ *is a subset of* $\mathcal{M}_{q,q'}$ *for every pair* $(q, q') \in \mathcal{Q} \times \mathcal{Q}$. ♢

The following is the definition of "hybrid control system" that will be adopted for the purposes of this paper.

Definition 1.2.3. *A **hybrid control system** is a 6-tuple*

$$\Sigma = (\mathcal{Q}, \mathcal{M}, \mathcal{U}, f, \boldsymbol{\mathcal{U}}, \mathcal{S})$$

such that

HCS1. $(\mathcal{Q}, \mathcal{M})$ *is a finite family of state spaces;*

HCS2. $\mathcal{U} = \{U_q\}_{q \in \mathcal{Q}}$ *is a family of sets;*

HCS3. $f = \{f_q\}_{q \in \mathcal{Q}}$ *is a family such that* f_q *is, for each* q*, a partially defined map from* $M_q \times U_q \times \mathbb{R}$ *to* TM_q*, having the property that* $f_q(x, u, t)$ *belongs to* $T_x M_q$ *for every* $(x, u, t) \in M_q \times U_q \times \mathbb{R}$ *for which* $f_q(x, u, t)$ *is defined;*

HCS4. $\boldsymbol{\mathcal{U}} = \{\boldsymbol{U}_q\}_{q \in \mathcal{Q}}$ *is a family consisting, for each* q*, of a set* $\boldsymbol{U}_q$*, each of whose members is a map* $\eta : I_\eta \to U_q$ *defined on some subinterval* I_η *of* $\mathbb{R}$*;*

HCS5. $\mathcal{S} = \{S_{q,q'}\}_{(q,q') \in \mathcal{Q} \times \mathcal{Q}}$ *is a switching constraint for* $(\mathcal{Q}, \mathcal{M})$*.* ♢

The sets $S_{q,q'}$ are the *switching sets* of Σ, and are allowed to be empty. One should think of $S_{q,q'}$ as the set of all 4-tuples (x, x', t, t') such that $x \in M_q$, $x' \in M_{q'}$, and a switching (or "jump") from state $x \in M_q$ to state $x' \in M_{q'}$ is permitted at time t, with a resetting of the clock to time t'. Usually, one does not want to permit clock resetting, but for mathematical reasons it is better to allow it in principle, and exclude it, when desired, by just taking the switching sets $S_{q,q'}$ to consist only of points of the form (x, x', t, t).

The members of $\mathcal{Q}$ are called *locations*. The families $\mathcal{M}$, $\mathcal{U}$, 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 $\boldsymbol{U}_q$ 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.

Definition 1.2.4. *A **control** for a hybrid system* Σ *as above is a triple* $\zeta = (\mathbf{q}, \mathbf{I}, \boldsymbol{\eta})$ *such that*

- $\mathbf{q} = (q_1, \ldots, q_\nu)$ *is a finite sequence of locations;*
- $\mathbf{I} = (I_1, \ldots, I_\nu)$ *is a finite sequence of compact intervals;*
- $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_\nu)$ *is a finite sequence such that* η_j *belongs to* $\boldsymbol{U}_{q_j}$ *and* $I_{\eta_j} = I_j$ *for* $j = 1, \ldots, \nu$*.*

If $\zeta = (\mathbf{q}, \mathbf{I}, \boldsymbol{\eta})$ *is a control, and* $\mathbf{I} = (I_1, \ldots, I_\nu)$ *for* $j = 1, \ldots, \nu$*, we use* $\mathbf{q}(\zeta)$*,* $\mathbf{I}(\zeta)$*,* $\boldsymbol{\eta}(\zeta)$*,* $\nu(\zeta)$*, to denote, respectively, the finite sequences* $\mathbf{q}$*,* $\mathbf{I}$*,* $\boldsymbol{\eta}$*, and the natural number* ν*. If* $I_j = [t_j, \tau_j]$*, we use* $\mathbf{t}(\zeta)$*,* $\boldsymbol{\tau}(\zeta)$ *to denote the* *sequences* $(t_1,\ldots,t_\nu)$ *and* $(\tau_1,\ldots,\tau_\nu)$, *and we let* $a_\zeta = t_1$, $b_\zeta = \tau_\nu$. *Then* a_ζ, b_ζ, $\nu(\zeta)-1$, *and* $\mathbf{q}(\zeta)$ *are, respectively, the* ***initial time****, the* ***terminal time****, the* ***number of switchings****, and the* ***switching strategy*** *of* ζ. ♢

Definition 1.2.5. *If* $\Sigma = (\mathcal{Q},\mathcal{M},\mathcal{U},f,\boldsymbol{\mathcal{U}},\mathcal{S})$ *is a hybrid system as above,* ζ *is a control for* Σ, *and* $\nu = \nu(\zeta)$, *then a* ***pretrajectory*** *for* ζ *is a* ν*-tuple* $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_\nu)$ *such that, if*

$$\mathbf{I}(\zeta) = (I_1,\ldots,I_\nu)\,,\ I_j = [t_j,\tau_j],\ \mathbf{q}(\zeta) = (q_1,\ldots,q_\nu)\,,\ \boldsymbol{\eta}(\zeta) = (\eta_1,\ldots,\eta_\nu)\,,$$

then, for each $j \in \{1,\ldots,\nu\}$, ξ_j *is an absolutely continuous map from* I_j *to the manifold* M_{q_j}, *having the property that* $f_{q_j}(\xi_j(t),\eta_j(t),t)$ *is defined and* $\dot{\xi}_j(t) = f_{q_j}(\xi_j(t),\eta_j(t),t)$ *for almost all* $t \in I_j$. ♢

Definition 1.2.6. *If* Σ *is a hybrid system as above, a* ***pretrajectory-control pair*** *for* Σ *is a pair* $(\boldsymbol{\xi},\zeta)$ *such that* ζ *is a control for* Σ *and* $\boldsymbol{\xi}$ *is a pretrajectory of* Σ *for* ζ.

We use $PTCP(\Sigma)$ *to denote the set of all pretrajectory-control pairs of the system* Σ. ♢

Definition 1.2.7. *An* ***endpoint constraint*** *for a finite family of state spaces* $(\mathcal{Q},\mathcal{M})$ *is a family* $\mathcal{E} = \{E_{q,q'}\}_{(q,q')\in\mathcal{Q}\times\mathcal{Q}}$ *of sets such that* $E_{q,q'}$ *is, for each* $(q,q') \in \mathcal{Q}\times\mathcal{Q}$, *a subset of* $\mathcal{M}_{q,q'}$. ♢

Notice that, mathematically, an endpoint constraint is exactly the same kind of object as a switching condition. This is why the part of the maximum principle that has to do with the switchings will have the same form as the transversality condition.

Definition 1.2.8. *Let* $\Sigma = (\mathcal{Q},\mathcal{M},\mathcal{U},f,\boldsymbol{\mathcal{U}},\mathcal{S})$ *be a hybrid control system as in the previous definitions, and let* $\Xi = (\boldsymbol{\xi},\zeta)$ *belong to* $PTCP(\Sigma)$. *Let* $\nu = \nu(\zeta)$, $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_\nu)$, $\mathbf{q}(\zeta) = (q_1,\ldots,q_\nu)$, $\mathbf{t}(\zeta) = (t_1,\ldots,t_\nu)$, $\boldsymbol{\tau}(\zeta) = (\tau_1,\ldots,\tau_\nu)$, $\mathbf{I}(\zeta) = (I_1,\ldots,I_\nu)$, $\mathcal{S} = \{S_{q,q'}\}_{(q,q')\in\mathcal{Q}\times\mathcal{Q}}$. *Then*

- *The* ***endpoint condition*** *of* $\boldsymbol{\xi}$ *(or of* Ξ*) is the 4-tuple*

$$\partial\boldsymbol{\xi} \stackrel{\text{def}}{=} \partial\Xi \stackrel{\text{def}}{=} (\xi_\nu(b_\zeta),\xi_1(a_\zeta),b_\zeta,a_\zeta) \in \mathcal{M}_{q_\nu,q_1}\,. \tag{1.1}$$

- *If* $1 \le j < \nu$, *the "j-th jump" of* $\boldsymbol{\xi}$ *(or of* Ξ*) is the 4-tuple*

$$\partial_j\boldsymbol{\xi} \stackrel{\text{def}}{=} \partial_j\Xi \stackrel{\text{def}}{=} (\xi_j(\tau_j),\xi_{j+1}(t_{j+1}),\tau_j,t_{j+1}) \in \mathcal{M}_{q_j,q_{j+1}}\,. \tag{1.2}$$

- *If* $\mathcal{E} = \{E_{q,q'}\}_{(q,q')\in\mathcal{Q}\times\mathcal{Q}}$ *is an endpoint constraint for* $(\mathcal{Q},\mathcal{M})$, *we say that* Ξ ***satisfies the constraint*** $\mathcal{E}$ *if* $\partial\Xi$ *belongs to* E_{q_ν,q_1}.

- *We say that* $\boldsymbol{\xi}$ *(or* Ξ*)* ***satisfies the switching conditions for*** Σ *if* $\partial_j \Xi$ *belongs to* $S_{q_j,q_{j+1}}$ *whenever* $j \in \{1, \dots, \nu - 1\}$. ♢

Definition 1.2.9. *If* $\Sigma = (\mathcal{Q}, \mathcal{M}, \mathcal{U}, f, \boldsymbol{\mathcal{U}}, \mathcal{S})$ *is a hybrid system as above, then*

- *we say that a pretrajectory* $\boldsymbol{\xi}$ *of* Σ *is a* ***trajectory*** *of* Σ *if* $\boldsymbol{\xi}$ *satisfies the switching conditions for* Σ*;*
- *we use* $TCP(\Sigma)$ *to denote the set of all trajectory-control pairs of* Σ *(i.e., the set of all* $\Xi = (\boldsymbol{\xi}, \zeta) \in PTCP(\Sigma)$ *such that* $\boldsymbol{\xi}$ *is a trajectory of* Σ*), and* $TCP(\Sigma; \mathcal{E})$ *to denote the set of all* $\Xi \in TCP(\Sigma)$ *that satisfy the endpoint constraint* $\mathcal{E}$. ♢

Definition 1.2.10. *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*

- L_q *is, for each* $q \in \mathcal{Q}$*, a partially defined real-valued function on the product* $M_q \times U_q \times \mathbb{R}$*,*
- *whenever* $q \in \mathcal{Q}$*,* $\eta \in \boldsymbol{U}_q$ *has domain* $[\alpha, \beta]$*, and* $\xi : [\alpha, \beta] \to M_q$ *is an absolutely continuous solution of* $\dot{\xi}(t) = f_q(\xi(t), \eta(t), t)$ *a.e., it follows that the function* $[\alpha, \beta] \ni t \to L(\xi(t), \eta(t), t)$ *is defined for almost every* t*, and is integrable.*

A ***switching cost function*** *for* Σ *is a family* $\Phi = \{\Phi_{q,q'}\}_{(q,q') \in \mathcal{Q} \times \mathcal{Q}}$ *such that each* $\Phi_{q,q'}$ *is an extended real-valued function on* $S_{q,q'}$ *that never takes the value* $-\infty$.

An ***endpoint cost function*** *for* Σ *is a family* $\varphi = \{\varphi_{q,q'}\}_{(q,q') \in \mathcal{Q} \times \mathcal{Q}}$ *such that each* $\varphi_{q,q'}$ *is an extended real-valued function on* $\mathcal{M}_{q,q'}$ *that never takes the value* $-\infty$. ♢

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

$$C_L(\boldsymbol{\xi}, \zeta) = \sum_{j=1}^{\nu} \int_{I_j} L_{q_j}(\xi_j(t), \eta_j(t), t)\, dt\,, \tag{1.3}$$

where $\nu = \nu(\zeta)$, $\mathbf{I}(\zeta) = (I_1, \dots, I_\nu)$, $\mathbf{q}(\zeta) = (q_1, \dots, q_\nu)$, $\boldsymbol{\eta}(\zeta) = (\eta_1, \dots, \eta_\nu)$, and $\boldsymbol{\xi} = (\xi_1, \dots, \xi_\nu)$.

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

$$\hat{C}_{\Phi,\varphi}(\boldsymbol{\xi},\zeta) = \varphi_{q_\nu,q_1}(\partial \Xi) + \sum_{j=1}^{\nu-1} \Phi_{q_j,q_{j+1}}(\partial_j \Xi)\,, \tag{1.4}$$

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

Definition 1.2.11. *A* ***hybrid Bolza cost functional*** *for Σ is an extended real-valued 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.* ♢

Given a hybrid control system Σ, a Bolza cost functional $\mathcal{C}$ for Σ, and an endpoint constraint $\mathcal{E}$, we will consider the *optimal control problem* $\mathcal{P}(\Sigma,\mathcal{C},\mathcal{E})$, whose objective is to minimize $\mathcal{C}(\boldsymbol{\xi},\zeta)$ in the class $TCP(\Sigma;\mathcal{E})$. We observe that the endpoint constraint sets $E_{q,q'}$ could all be of the special form $E^0_{q,q'} \times \{b\} \times \{a\}$, where a, b are fixed real numbers, independent of q, q', and each $E^0_{q,q'}$ is a subset of $M_q \times M_{q'}$. In that special case, all the members $\Xi = (\boldsymbol{\xi},\zeta)$ of $TCP(\Sigma;\mathcal{E})$ satisfy $a_\zeta = a$, $b_\zeta = b$, so we have a *problem with fixed initial and terminal times.* In addition, the switching sets $S_{q,q'}$ could be of the form $S_{q,q'} = S^0_{q,q'} \times \{\bar{t}_{q,q'}\} \times \{\bar{t}_{q,q'}\}$, where $S^0_{q,q'} \subseteq M_q \times M_{q'}$ and the $\bar{t}_{q,q'}$ are fixed real numbers, in which case we would be dealing with a *problem with fixed switching times and no clock resetting.*

1.3 The general form of the maximum principle

Let us assume that

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

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

A3. $\mathcal{E}$ is an endpoint constraint for $(\mathcal{Q},\mathcal{M})$;

A4. $\Xi^\#$ (the "reference trajectory-control pair") belongs to $TCP(\Sigma;\mathcal{E})$, and

$$\begin{aligned}
\Xi^\# &= (\boldsymbol{\xi}^\#,\zeta^\#)\,, & \boldsymbol{\xi}^\# &= (\xi^\#_1,\ldots,\xi^\#_{\nu^\#})\,,\\
\zeta^\# &= (\mathbf{q}^\#,\mathbf{I}^\#,\boldsymbol{\eta}^\#)\,, & \mathbf{q}^\# &= (q^\#_1,\ldots,q^\#_{\nu^\#})\,,\\
\mathbf{I}^\# &= (I^\#_1,\ldots,I^\#_{\nu^\#})\,, & \boldsymbol{\eta}^\# &= (\eta^\#_1,\ldots,\eta^\#_{\nu^\#})\,.
\end{aligned}$$

The *maximum principle* gives a necessary condition for $\Xi^\#$ to be a solution of $\mathcal{P}(\Sigma,\mathcal{C},\mathcal{E})$. 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 introduce the following definition.

Definition 1.3.1. *A **local solution** of a problem $\mathcal{P}(\Sigma, \mathcal{C}, \mathcal{E})$ is 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 $\xi_1^{\#}, \ldots, \xi_{\nu^{\#}}^{\#}$ in $M_{q_1} \times \mathbb{R}, \ldots, M_{q_\nu^{\#}} \times \mathbb{R}$ having the property that $\Xi^{\#}$ minimizes the cost $\mathcal{C}(\Xi)$ in the class of all the trajectory-control pairs $\Xi = (\boldsymbol{\xi}, \zeta) = (\xi_1, \ldots, \xi_\nu, \zeta) \in TCP(\Sigma, \mathcal{E})$ such that $\mathbf{q}(\zeta) = \mathbf{q}(\zeta^{\#})$ (so that, in particular, $\nu = \nu^{\#}$) and the graph $G(\xi_j)$ of ξ_j is contained in $\mathcal{N}_j$ for $j = 1, \ldots, \nu^{\#}$. (Here the "graph" of ξ_j is the set*

$$G(\xi_j) \stackrel{\text{def}}{=} \{(\xi_j(t), t) : t \in \text{Domain}(\xi_j)\}\,, \tag{1.5}$$

so $G(\xi_j) \subseteq M_{q_j} \times \mathbb{R}$.) ♢

We now present the maximum principle for hybrid systems 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 subsequent sections of the paper.

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

To turn the above statement into a theorem, we have to specify technical assumptions on the 12-tuple of data $(\mathcal{Q}, \mathcal{M}, \mathcal{U}, f, \boldsymbol{\mathcal{U}}, \mathcal{S}, L, \Phi, \varphi, \mathcal{E}, \boldsymbol{\xi}^{\#}, \zeta^{\#})$, and assign a precise meaning to the notions of "adjoint pair," "weak Hamiltonian maximization," "nontriviality," and "transversality." This will be done in detail in the following section for problems whose reference vector fields and Lagrangians are differentiable along the reference trajectory. The changes needed for the nondifferentiable case will be sketched in §1.6.

1.4 A version involving classical differentials

We now make the maximum principle precise in the setting of maps having classical differentials.

First, we let

$$I_0^{\#} = I_{\nu^{\#}}^{\#}\,, \qquad I_{\nu^{\#}+1}^{\#} = I_1^{\#}\,, \qquad q_0^{\#} = q_{\nu^{\#}}^{\#}\,, \qquad q_{\nu^{\#}+1}^{\#} = q_1^{\#}\,,$$

and write

$$
\begin{array}{lll}
I_j^\# & = [t_j^\#, \tau_j^\#] & \text{for} \quad j = 0, 1, \ldots, \nu^\# + 1\,, \\
a^\# & = t_1^\#\,, & \\
b^\# & = \tau_{\nu^\#}^\#\,, & \quad , \\
M_j^\# & = M_{q_j^\#} & \text{for} \quad j = 0, 1, \ldots, \nu^\# + 1\,, \\
\mathcal{M}_j^\# & = \mathcal{M}_{q_j^\#, q_{j+1}^\#} & \text{for} \quad j = 1, \ldots, \nu^\#\,, \\
S_j^\# & = S_{q_j^\#, q_{j+1}^\#} & \text{for} \quad j = 1, \ldots, \nu^\# - 1\,, \\
S_j^\# & = E_{q_{\nu^\#}^\#, q_1^\#} & \text{for} \quad j = \nu^\#\,, \\
S_0^\# & = S_{\nu^\#}^\# \quad , & \\
\Phi_j^\# & = \Phi_{q_j^\#, q_{j+1}^\#} & \text{for} \quad j = 1, \ldots, \nu^\# - 1\,, \\
\Phi_j^\# & = \varphi_{q_{\nu^\#}^\#, q_1^\#} & \text{for} \quad j = \nu^\#\,.
\end{array}
$$

Next, we drop the superscript $^\#$ in our discussion (except in the statement of our hypotheses A5-A15, where the notation of the previous sections will be maintained), and write ν, Ξ, $\boldsymbol{\xi}$, ξ_j, $\mathbf{I}$, I_j, ζ, $\mathbf{q}$, q_j, $\boldsymbol{\eta}$, η_j, t_j, τ_j, a, b, M_j, $\mathcal{M}_j$, S_j, Φ_j instead of $\nu^\#$, $\Xi^\#$, $\boldsymbol{\xi}^\#$, $\xi_j^\#$, $\mathbf{I}^\#$, $I_j^\#$, $\zeta^\#$, $\mathbf{q}^\#$, $q_j^\#$, $\boldsymbol{\eta}^\#$, $\eta_j^\#$, $t_j^\#$, $\tau_j^\#$, $a^\#$, $b^\#$, $M_j^\#$, $\mathcal{M}_j^\#$, $S_j^\#$, $\Phi_j^\#$. Then

$$
\begin{array}{lll}
\Xi = (\boldsymbol{\xi}, \zeta)\,, & \boldsymbol{\xi} = (\xi_1, \ldots, \xi_\nu)\,, & \zeta = (\mathbf{q}, \mathbf{I}, \boldsymbol{\eta})\,, \\
\mathbf{I} = (I_1, \ldots, I_\nu)\,, & \mathbf{q} = (q_1, \ldots, q_\nu)\,, & \boldsymbol{\eta} = (\eta_1, \ldots, \eta_\nu)\,.
\end{array}
$$

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

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

We now define the notion of an "adjoint pair" along Ξ, 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 make the following assumptions.

A5. For $j \in \{1, \dots, \nu^{\#} - 1\}$, K_j is a tangent cone to the set $S_{q_j^{\#}, q_{j+1}^{\#}}$ at $\partial_j \Xi^{\#}$.

A6. K_e is a tangent cone to the set $E_{q_1^{\#}, q_{\nu^{\#}}^{\#}}$ at $\partial \Xi^{\#}$.

(In particular, K_j is a subset of the tangent space $T_{\partial_j \Xi} \mathcal{M}_{q_j, q_{j+1}}$, which is canonically identified with the product $T_{\xi_j(\tau_j)} M_{q_j} \times T_{\xi_{j+1}(t_{j+1})} M_{q_{j+1}} \times \mathbb{R} \times \mathbb{R}$. Similarly, the cone K_e is a subset of $T_{\partial \Xi} \mathcal{M}_{q_\nu, q_1}$, which is canonically identified with $T_{\xi_\nu(b)} M_{q_\nu} \times T_{\xi_1(a)} M_{q_1} \times \mathbb{R} \times \mathbb{R}$.)

We will also write K_ν for K_e and $\partial_\nu \Xi$ for $\partial \Xi$, so *K_j is a tangent cone to S_j at $\partial_j \Xi$ for $j = 1, \dots, \nu$.*

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 $j \in \{1, \dots, \nu^{\#}\}$ then for almost every time $t \in I_j^{\#}$ the maps $x \to f_{q_j^{\#}}(x, \eta_j^{\#}(t), t)$ and $x \to L_{q_j^{\#}}(x, \eta_j^{\#}(t), t)$ are defined on a neighborhood of $\xi_j^{\#}(t)$ and are differentiable at $\xi_j^{\#}(t)$.

A8. If $j \in \{1, \dots, \nu^{\#} - 1\}$, then $\Phi_{q_j^{\#}, q_{j+1}^{\#}}$ is differentiable at $\partial_j \Xi^{\#}$ along $S_{q_j^{\#}, q_{j+1}^{\#}}$.

A9. $\varphi_{q_{\nu^{\#}}^{\#}, q_1^{\#}}$ is differentiable at $\partial \Xi^{\#}$ along $E_{q_{\nu^{\#}}^{\#}, q_1^{\#}}$.

Remark 1.4.1. 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 $\bar{s}$ of a subset S of X if (a) for some neighborhood V of $\bar{s}$, the inclusion $V \cap S \subseteq \text{Domain}(h)$ holds, and (b) there exists a linear functional λ on the tangent space $T_{\bar{s}} X$ such that

$$\lim_{s \to \bar{s}, s \in S} \frac{h(s) - h(\bar{s}) - \lambda \cdot (s - \bar{s})}{\| s - \bar{s} \|} = 0 . \qquad (1.6)$$

Naturally, (1.6) makes sense relative to a choice of local coordinates about $\bar{s}$, but it is easy to verify that the validity of (1.6) does not depend on that choice. The linear functional λ need not be unique but, if K is a tangent cone to S at $\bar{s}$ (in the sense of S1), then the restriction of λ to K is unique, since $\lambda(v) = \lim_{\varepsilon\downarrow 0} \varepsilon^{-1}(h(\mu(\varepsilon v)) - h(\bar{s}))$ whenever $v \in K$, if μ is any map that satisfies the conditions of Definition 1.4.1. In particular, a condition such as "$\theta + \rho\nabla h(\bar{s}) \in K^{\perp}$" makes sense intrinsically, if θ is a given covector at $\bar{s}$ and ρ is a number. Indeed, it suffices to interpret the condition to mean "$\theta + \rho\lambda \in K^{\perp}$ for any λ such that (1.6) holds." ♢

For each $q \in \mathcal{Q}$, we introduce a partially defined real-valued function H_q on $T^*M_q \times U_q \times \mathbb{R} \times \mathbb{R}$ by letting

$$H_q(x, p, u, p_0, t) = p \cdot f_q(x, u, t) - p_0 L_q(x, u, t) \tag{1.7}$$

for $(x, p, u, p_0, t) \in T^*M_q \times U_q \times \mathbb{R} \times \mathbb{R}$ such that $f_q(x, u, t)$ and $L_q(x, u, t)$ are defined.

Definition 1.4.2. *If S1 and A1-A9 hold, then an* ***adjoint pair along*** Ξ *is a pair* $(\boldsymbol{\psi}, \psi_0)$ *with the property that:*

- $\boldsymbol{\psi}$ *is a* ν*-tuple* $(\psi_1, \dots, \psi_\nu)$ *such that each* ψ_j *is a field of covectors along* ξ_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* $\psi_j(t) \in T^*_{\xi_j(t)}M_{q_j}$ *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***

$$\dot{\psi}_j(t) = -\frac{\partial H_{q_j}}{\partial x}(\xi_j(t), \psi_j(t), \eta_j(t), \psi_0, t) \text{ for a.e. } t \in I_j; \tag{1.8}$$

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

$$(-\psi_j(\tau_j), \psi_{j+1}(t_{j+1}), h_j^+, -h_{j+1}^-) - \psi_0\omega_j \in K_j^{\perp} \tag{1.9}$$

holds, where $\omega_j = \nabla\Phi_{q_j,q_{j+1}}(\partial_j\Xi)$*, and*

$$h_j^+ = \begin{cases} \lim_{s\downarrow 0} \frac{1}{s}\int_{\tau_j - s}^{\tau_j} H_{q_j}(\xi_j(t), \psi_j(t), \eta_j(t), \psi_0, t)\, dt & \text{if the limit exists}, \\ 0 & \text{otherwise}, \end{cases}$$

$$h_j^- = \begin{cases} \lim_{s\downarrow 0} \frac{1}{s}\int_{t_j}^{t_j + s} H_{q_j}(\xi_j(t), \psi_j(t), \eta_j(t), \psi_0, t)\, dt & \text{if the limit exists}, \\ 0 & \text{otherwise}. \end{cases}$$

Definition 1.4.3. *If S1 and A1-A9 hold,* $(\boldsymbol{\psi}, \psi_0)$ *is an adjoint pair along* Ξ*, and* $j \in \{1, \ldots, \nu\}$*, we say that* $(\boldsymbol{\psi}, \psi_0)$ *satisfies the* ***weak Hamiltonian maximization condition on the*** j***-th interval*** *if for every* $u \in U_{q_j}$ *the inequality*

$$H_{q_j}(\xi_j(t), \psi_j(t), u, \psi_0, t) \leq H_{q_j}(\xi_j(t), \psi_j(t), \eta_j(t), \psi_0, t) \tag{1.10}$$

holds for almost all $t \in I_j$*. We say that* $(\boldsymbol{\psi}, \psi_0)$ *satisfies the* ***strong Hamiltonian maximization condition on the*** j***-th interval*** *if the identity*

$$H_{q_j}(\xi_j(t), \psi_j(t), \eta_j(t), \psi_0, t) = \max\{H_{q_j}(\xi_j(t), \psi_j(t), u, \psi_0, t) : u \in U_{q_j}\} \tag{1.11}$$

holds for almost all $t \in I_j$*.* $\diamondsuit$

Notice that the only difference between "weak" and "strong" Hamiltonian maximization is that weak maximization says that (1.10) holds, for each $u \in U_{q_j}$, at all points in the complement of a "bad" null set $B_j(u)$ that could depend on u, whereas strong maximization says that $B_j(u)$ can be chosen independently of u.

Weak maximization implies strong maximization in several important cases, such as, for example, (a) when the control sets are separable topological spaces and the dynamics and Lagrangian are continuous with respect to the control, and (b) when the dynamics and Lagrangian do not depend explicitly on time.

Proposition 1.4.1. *Let* $j \in \{1, \ldots, \nu\}$*. Assume that either*

(1) U_{q_j} *is a separable topological space and the maps*

$$U_{q_j} \ni u \to (f_{q_j}(\xi_j(t), u, t), L_{q_j}(\xi_j(t), u, t)) \in TM_{q_j} \times \mathbb{R}$$

are continuous for almost every $t \in I_j$*,*

or

(2) there exists a compact subset X_j *of the set space* M_{q_j}*, containing the set* $\xi_j(I_j) \stackrel{\text{def}}{=} \{\xi_j(t) : t \in I_j\}$ *and such that* $f_{q_j}(x, u, t)$ *and* $L_{q_j}(x, u, t)$ *are defined for all* $(x, u, t) \in X_j \times U_{q_j} \times I_j$ *and do not depend on* t*.*

Then weak Hamiltonian maximization on the j*-th interval implies strong Hamiltonian maximization on the* j*-th interval.*

Proof. If V is a countable dense subset of U_{q_j}, then weak maximization implies that there exists a null subset B of I_j such that (1.10) holds whenever

$t \in I_j \backslash B$ and $u \in V$. If the continuity hypotheses of (1) hold, then (1.10) holds whenever $t \in I_j \backslash B$, and $u \in U_{q_j}$, so (1.11) holds for all $t \in I_j \backslash B$.

If (2) holds, then for each $u \in U_{q_j}$ the map

$$X_j \ni x \to f_{q_j}(x, u, t) \stackrel{\text{def}}{=} F^u(x) \in TM_{q_j}$$

is a continuous vector field on X_j, and the scalar function

$$X_j \ni x \to L_{q_j}(x, u, t) \stackrel{\text{def}}{=} L^u(x) \in \mathbb{R}$$

is continuous. Let $VF(X_j)$ denote the space of all continuous vector fields on X_j, endowed with the topology of uniform convergence. Then $VF(X_j)$ is metrizable and separable. Clearly, the space $C^0(X_j, \mathbb{R})$ of continuous real-valued functions on X_j is metric and separable. It follows that the set $\mathcal{W} = \{(F^u, L^u) : u \in U_{q_j}\}$ is a subset of the separable metrizable space $VF(X_j) \times C^0(X_j, \mathbb{R})$. Hence $\mathcal{W}$ is a metrizable separable space. So there exists a countable subset V of U_{q_j} such that the set $\mathcal{W}^V \stackrel{\text{def}}{=} \{(F^u, L^u) : u \in V\}$ is dense in $\mathcal{W}$. For each $u \in V$, let B^u be a null subset of I_j such that (1.10) holds for all $t \in I_j \backslash B^u$. Let $B = \cup_{u \in V} B^u$. Then B is a null set, and (1.10) holds whenever $u \in V$ and $t \in I_j \backslash B$. If $u \in U_{q_j}$ is arbitrary, then there exists a sequence $\{u_k\}_{k=1}^{\infty}$ in U_{q_j} such that $F^{u_k} \to F^u$ and $L^{u_k} \to L^u$ uniformly on X_j as $k \to \infty$. Then (1.11) holds whenever $t \in I_j \backslash B$. ♢

Definition 1.4.4. *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*

$$(-\psi_\nu(b), \psi_1(a), h_\nu^+, -h_1^-) - \psi_0 \omega_e \in K_e^\perp , \tag{1.12}$$

where $\omega_e = \nabla \varphi_{q_\nu, q_1}(\partial \Xi)$, and the numbers $h_j^\pm$ are those that were introduced in Definition 1.4.2. ♢

Definition 1.4.5. *If $(\boldsymbol{\psi}, \psi_0)$ is an adjoint pair along Ξ, 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.* ♢

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 additional conditions. In order to state these conditions, we first give some preliminary definitions. For this purpose, we first define the "augmented dynamical laws"

$$\tilde{f}_q(x, u, t) \stackrel{\text{def}}{=} (f_q(x, u, t), L_q(x, u, t)) , \tag{1.13}$$

so $\tilde{f}_q(x, u, t) \in T_x M_q \times \mathbb{R}$ whenever $x \in M_q$, $u \in U_q$, $t \in \mathbb{R}$ are such that both $f_q(x, u, t)$ and $L_q(x, u, t)$ are defined.

Definition 1.4.6. *We say that Σ is* ***autonomous at location*** q *if the set of those $(x,u) \in M_q \times U_q$ for which $\tilde{f}_q(x,u,t)$ is defined is independent of t and, for (x,u) in that set, the value of $\tilde{f}_q(x,u,t)$ is independent of t.* $\diamondsuit$

Next, we define various types of "control variations." In the following discussion, "$\lceil$" denotes "restriction." If U is a set, a, b, c are real numbers such that $a \leq b \leq c$, and $\eta_1 : [a,b] \to U$, $\eta_2 : [b,c] \to U$ are maps, then the ***concatenation of*** η_1 ***and*** η_2 is the map $\eta_2 * \eta_1 : [a,c] \to U$ given by

$$(\eta_2 * \eta_1)(t) = \begin{cases} \eta_1(t) & \text{if} \quad a \leq t \leq b\,, \\ \eta_2(t) & \text{if} \quad b < t \leq c\,. \end{cases}$$

So $\eta_2 * \eta_1$ is defined only when the domains I_{η_i} of η_i are compact intervals such that $\max(I_{\eta_1}) = \min(I_{\eta_2})$. Clearly, concatenation is an associative operation, in the sense that $(\eta_3 * \eta_2) * \eta_1$ is defined if and only if $\eta_3 * (\eta_2 * \eta_1)$ is defined, and in that case $(\eta_3 * \eta_2) * \eta_1 = \eta_3 * (\eta_2 * \eta_1)$.

Definition 1.4.7. *Let U be a set, let a, b be real numbers such that $a \leq b$, and let $\eta : [a,b] \to U$ be a map. Then*

1. *if $r \geq 0$, the* ***right*** r***-translation*** *and* ***left*** r***-translation of*** η ***by*** r *are the maps $T_r^+(\eta) : [a+r, b+r] \to U$, $T_r^-(\eta) : [a-r, b-r] \to U$, given by*

$$T_r^+(\eta)(t) = \eta(t-r) \quad \text{for} \quad t \in [a+r, b+r]\,,$$
$$T_r^-(\eta)(t) = \eta(t+r) \quad \text{for} \quad t \in [a-r, b-r]\,;$$

2. *if $a \leq c \leq b$, $r \geq 0$, and η' is a U-valued map whose domain contains the interval $[c, c+r]$, then the* ***right*** r***-expansion of the map*** η ***by*** η' ***at time*** c *is the map $E^+_{r,c,\eta'}(\eta) : [a, b+r] \to U$ given by*

$$E^+_{r,c,\eta'}(\eta) = T_r^+(\eta\lceil[c,b]) * \eta'\lceil[c, c+r] * \eta\lceil[a,c]\,;$$

   if $c = b$, then we omit the subscript c, so $E^+_{r,\eta'}(\eta) \stackrel{\text{def}}{=} \eta'\lceil[b, b+r] * \eta\,;$

3. *if $a \leq c \leq b$, $r \geq 0$, and η' is a U-valued map whose domain contains the interval $[c-r, c]$, then the* ***left*** r***-expansion of the map*** η ***by*** η' ***at time*** c *is the map $E^-_{r,c,\eta'}(\eta) : [a-r, b] \to U$ given by*

$$E^-_{r,c,\eta'}(\eta) = \eta\lceil[c,b] * \eta'\lceil[c-r, c] * T_r^-(\eta\lceil[a,c])\,;$$

   if $c = a$, then we omit the subscript c, so $E^-_{r,\eta'}(\eta) \stackrel{\text{def}}{=} \eta * \eta'\lceil[a-r, a] * \eta\lceil[a,c]\,;$

3. *if $a \leq c_1 \leq c_2 \leq b$, then*

   a. *the* ***right shortening of*** η ***at*** $[c_1, c_2]$ *is the map*

$$S^+_{c_1,c_2}(\eta) = \eta\lceil[c_2, b] * T^+_{c_2-c_1}(\eta\lceil[a, c_1])\,;$$

*b. the **left shortening of** η **at** $[c_1, c_2]$ is the map*

$$S^-_{c_1,c_2}(\eta) = T^-_{c_2-c_1}(\eta\lceil[c_2, b]) * \eta\lceil[a, c_1]\,;$$

*4. if A is a subset of $[a, b]$, and θ is a U-valued map whose domain contains A, then the **replacement of** η **by** θ **on** A is the map $R_{A,\theta}(\eta) : [a, b] \to U$ given by*

$$R_{A,\theta}(\eta) = \begin{cases} \eta(t) & \text{if} \quad t \in [a, b]\backslash A\,, \\ \theta(t) & \text{if} \quad t \in A\,. \end{cases} \qquad \diamondsuit$$

We let $\mu(U)$ denote the set of all U-valued maps defined on compact intervals, so

$$\mu(U) \stackrel{\text{def}}{=} \bigcup_{-\infty<a\leq b<\infty} U^{[a,b]}\,, \tag{1.14}$$

where $U^{[a,b]}$ is the set of all mappings from $[a, b]$ to U. Then the transformations T^+_r, T^-_r, $R_{A,\theta}$, $S^+_{c_1,c_2}$, $S^-_{c_1,c_2}$, $E^+_{r,c,\eta'}$, $E^-_{r,c,\eta'}$, $E^+_{r,\eta'}$, $E^-_{r,\eta'}$, are partially defined maps from $\mu(U)$ to $\mu(U)$. (For example, $R_{A,\theta}(\eta)$ is defined if and only if $A \subset \text{Domain}(\eta) \cap \text{Domain}(\theta)$.) So we can define *iterates* of these maps. For example, if $\mathbf{A} = (A_1, \ldots, A_m)$ is an m-tuple of subsets of $\mathbb{R}$, and $\mathbf{u} = (u_1, \ldots, u_m)$ is an m-tuple of members of U (regarded as constant U-valued maps) then $R_{\mathbf{A},\mathbf{u}}$ is the composite map

$$R_{\mathbf{A},\mathbf{u}} \stackrel{\text{def}}{=} R_{A_1,u_1} \circ R_{A_2,u_2} \circ \ldots \circ R_{A_m,u_m}\,. \tag{1.15}$$

Similarly, $E^\sigma_{\mathbf{r},\mathbf{c},\theta}$ and $E^\sigma_{\mathbf{r},\theta}$ are the composite maps

$$\begin{aligned} E^\sigma_{\mathbf{r},\mathbf{c},\theta} &\stackrel{\text{def}}{=} E^{\sigma_1}_{r_1,c_1,\theta_1} \circ E^{\sigma_2}_{r_2,c_2,\theta_2} \circ \ldots \circ E^{\sigma_m}_{r_m,c_m,\theta_m}\,, \\ E^\sigma_{\mathbf{r},\theta} &\stackrel{\text{def}}{=} E^{\sigma_1}_{r_1,\theta_1} \circ E^{\sigma_2}_{r_2,\theta_2} \circ \ldots \circ E^{\sigma_m}_{r_m,\theta_m}\,, \end{aligned}$$

if $\mathbf{r} = (r_1, \ldots, r_m)$ and $\mathbf{c} = (c_1, \ldots, c_m)$ are m-tuples of nonnegative real numbers, $\theta = (\theta_1, \ldots, \theta_m)$ is an m-tuple of members of $\mu(U)$, and $\sigma = (\sigma_1, \ldots, \sigma_m)$ is a sequence of members of $\{-, +\}$.

Definition 1.4.8. *We say that $\boldsymbol{U}_{q_j}$ is a **fixed-time measurable weak variational neighborhood of the control** η_j if*

(FTMVN) For every positive integer m and every m-tuple $\mathbf{u} = (u_1, \ldots, u_m)$ of members of U_{q_j} there exists a positive number δ such that, whenever $\mathbf{A} = (A_1, \ldots, A_m)$ is an m-tuple of pairwise disjoint measurable subsets of I_j for which $\text{meas}(A_1 \cup \ldots \cup A_m) \leq \delta$, it follows that the function $R_{\mathbf{A},\mathbf{u}}(\eta)$ belongs to $\boldsymbol{U}_{q_j}$. $\diamondsuit$

Definition 1.4.9. *We say that* $\boldsymbol{U}_{q_j}$ *is a* ***weak variable-time measurable variational neighborhood of the control*** η_j *if*

(WVTMVN) *For every positive integer* m *and every triple* $(\mathbf{u},\mathbf{v},\mathbf{w})$ *of* m*-tuples* $\mathbf{u}=(u_1,\ldots,u_m)$, $\mathbf{v}=(v_1,\ldots,v_m)$, $\mathbf{w}=(w_1,\ldots,w_m)$ *of members of* U_{q_j} *there exists a positive number* δ *with the property that, if* $\alpha,\beta,r_1,\ldots,r_m,s_1,\ldots,s_m$ *are nonnegative real numbers whose sum does not exceed* δ*, and* $\mathbf{A}=(A_1,\ldots,A_m)$ *is an arbitrary* m*-tuple of pairwise disjoint measurable subsets of* $[t_j+\alpha,\tau_j-\beta]$ *such that* $\mathrm{meas}(A_1\cup\ldots\cup A_m)\le\delta$*, then the function*

$$V_{\mathbf{r},\mathbf{v},\alpha,\mathbf{s},\mathbf{w},\beta,\mathbf{A},\mathbf{u}}(\eta)\stackrel{\mathrm{def}}{=}E^-_{\mathbf{r},\mathbf{v}}(E^+_{\mathbf{s},\mathbf{w}}(S^+_{t_j,t_j+\alpha}(S^-_{\tau_j-\beta,\tau_j}(R_{\mathbf{A},\mathbf{u}}(\eta)))))\tag{1.16}$$

(where $\mathbf{r}=(r_1,\ldots,r_m)$ *and* $\mathbf{s}=(s_1,\ldots,s_m)$*) belongs to* $\boldsymbol{U}_{q_j}$*.* $\diamondsuit$

Definition 1.4.10. *We say that* $\boldsymbol{U}_{q_j}$ *is a* ***strong variable-time measurable variational neighborhood of the control*** η_j *if*

(SVTMVN) *For every positive integer* m *and every* m*-tuple* $\mathbf{u}=(u_1,\ldots,u_m)$ *of members of* U_{q_j} *there exists a positive number* δ *with the property that, if* $\mathbf{c}=(c_1,\ldots,c_m)$, $\mathbf{d}=(d_1,\ldots,d_m)$ *are* m*-tuples of members of* I_j *such that* $t_j<c_1<\ldots<c_m<\tau_j$, $t_j<d_1<\ldots<d_m<\tau_j$*, and* $c_i\neq d_k$ *for all* i,k*,* (σ,σ') *is a pair of sequences* $\sigma=(\sigma_1,\ldots,\sigma_m)$, $\sigma'=(\sigma'_1,\ldots,\sigma'_m)$ *of members of* $\{-,+\}$, $\alpha,\beta,r_1,\ldots,r_m,s_1,\ldots,s_m$ *are nonnegative real numbers such that*

$$\alpha+\beta+r_1+\cdots+r_m+s_1+\cdots+s_m\le\delta\,,$$

and $\mathbf{A}=(A_1,\ldots,A_m)$ *is an arbitrary* m*-tuple of pairwise disjoint measurable subsets of the interval* $[t_j+\alpha,\tau_j-\beta]$ *satisfying* $\mathrm{meas}(A_1\cup\ldots\cup A_m)\le\delta$*, then the function*

$$W_{\mathbf{r},\mathbf{d},\sigma',\mathbf{s},\mathbf{c},\mathbf{r},\sigma,\mathbf{A},\mathbf{u}}(\eta)\stackrel{\mathrm{def}}{=}E^{\sigma'}_{\mathbf{s},\mathbf{d},\eta}(S^{\sigma}_{\mathbf{c},\mathbf{r}}(R_{\mathbf{A},\mathbf{u}}(\eta)))\tag{1.17}$$

(where $\mathbf{r}=(r_1,\ldots,r_m)$ *and* $\mathbf{s}=(s_1,\ldots,s_m)$*) belongs to* $\boldsymbol{U}_{q_j}$ *.* $\diamondsuit$

Definition 1.4.11. *For each of the three types (fixed-time, weak variable-time, and strong variable-time) of "measurable variational neighborhoods" introduced above, the corresponding notion of an* ***interval variational neighborhood*** *is defined in exactly the same way, except that the sets* A_k *are required to be intervals rather than general measurable subsets.* $\diamondsuit$

Thus, we have a total of six different types of "variational neighborhoods" of a control η.

Remark 1.4.2. Our definitions of the various types of variational neighborhoods say, in all six cases, that $\boldsymbol{U}_{q_j}$ is a "variational neighborhood" of η_j if it contains all the controls obtained from η_j by making "sufficiently small" variations in an appropriate class $\mathcal{V}$. Where the six definitions differ is, naturally, in the choice of $\mathcal{V}$. The fixed-time variational neighborhood corresponds to choosing $\mathcal{V}$ to consist of *substitutions of finitely many constant controls for the control η, without changing the total time interval.* Weak variable-time variational neighborhoods are obtained when, in addition, we allow *shortenings at the endpoints and expansions by constant controls at the endpoints.* Finally, strong variable-time variational neighborhoods arise when we allow *shortenings and expansions by constant controls not only at the endpoints but at all points of I_j.*

Those readers who prefer neighborhoods to be derived from a topology can easily verify that ***for each of our six definitions of "neighborhood," if we define an "open subset" of $\mu(U)$ to be a subset which is a neighborhood of each of its points, then the set of open subsets constitutes a topology, and the corresponding notion of neighborhood is exactly the one of the original definition.*** ♢

Definition 1.4.12. *If $q \in \mathcal{Q}$, J is an interval, $\eta : J \to \boldsymbol{U}_q$ is a function, and $x \in M_q$, we say that (x,t) is a* ***forward regular point for*** *η if*

(1) there exists a positive $\bar{\delta}$ such that $\tilde{f}_q(x', \eta(s), s)$ is defined whenever x' belongs to M_q, $\mathrm{dist}(x', x) \le \bar{\delta}$, and $t \le s \le t+\bar{\delta}$, and depends continuously on x' and measurably on s,

and

(2) if $\sigma^{\eta,q}_{x,t,\delta}$ is the modulus of continuity defined by

$$\sigma^{\eta,q}_{x,t,\delta}(s) \stackrel{\text{def}}{=} \sup\{\|\tilde{f}_q(x', \eta(s), s) - \tilde{f}_q(x, \eta(t), t)\| : x' \in M_q, d(x', x) \le \delta\}, \quad (1.18)$$

then the number

$$\nu(\delta) \stackrel{\text{def}}{=} \frac{1}{\delta} \int_t^{t+\delta} \sigma^{\eta,q}_{x,t,\delta}(s)\, ds \quad (1.19)$$

goes to zero as $\delta \downarrow 0$. ♢

Remark 1.4.3. In Definition 1.4.12, "dist" is the distance function corresponding to any Riemannian metric g on a neighborhood of x in M_q. It is easy to see that the forward regularity property does not depend on the choice of g. ♢

"Backward regularity" is defined in the same way, except that in (1) the condition "$t \leq s \leq t + \bar{\delta}$" is replaced by "$t - \bar{\delta} \leq s \leq t$," and in (2) the integral of (1.19) is taken over the interval $[t-\delta, t]$ rather than over $[t, t+\delta]$.

We are now ready to state our technical conditions. In the statements, we will use

$$\tilde{Y}_j(t) \stackrel{\text{def}}{=} \frac{\partial \tilde{f}_{q_j^\#}}{\partial x}(\xi_j^\#(t), \eta_j^\#(t), t)\,. \tag{1.20}$$

We recall that the existence of the right-hand side of (1.20) for almost every $t \in I_j$ is guaranteed by A7.

A10. For each $j \in \{1, \ldots, \nu^\#\}$ and each $u \in U_{q_j^\#}$

1. $\tilde{f}_{q_j^\#}(x, u, t)$ is defined for all (x, t) belonging to some neighborhood $\mathcal{N}_j(u)$ in $M_{q_j^\#} \times \mathbb{R}$ of the graph $G(\xi_j^\#)$ (cf. (1.5);
2. for each $t \in \mathbb{R}$, the map $x \to \tilde{f}_{q_j^\#}(x, u, t)$ is continuous on the set $\{x : (x,t) \in \mathcal{N}_j(u)\}$,
3. for each $x \in M_{q_j^\#}$, the map $t \to \tilde{f}_{q_j^\#}(x, u, t)$ is Lebesgue measurable on the set $\{t : (x,t) \in \mathcal{N}_j(u)\}$.

A11. The maps $I_j^\# \ni t \to \tilde{f}_{q_j^\#}(x, \eta_j^\#(t), t)$ are Lebesgue measurable for every $j \in \{1, \ldots \nu^\#\}$, $x \in M_{q_j^\#}$.

A12. For every $j \in \{1, \ldots, \nu^\#\}$, the map $\tilde{Y}_j$ is Lebesgue integrable on $I_j^\#$.

A13. $\Phi_{q,q'}$ and $\varphi_{q,q'}$ are finite-valued and continuous on the sets $S_{q,q'}$, $E_{q,q'}$, for all $(q, q') \in \mathcal{Q} \times \mathcal{Q}$.

A14. The differentiability at the point $x = \xi_j^\#(t)$ of the augmented vector fields $M_{q_j^\#} \ni x \to \tilde{f}_{q_j^\#}(x, \eta_j^\#(t), t)$ has the following Carathéodory-type uniformity: there exist measurable functions $k_j^\delta : I_j^\# \to [0, +\infty]$ such that

$$\lim_{\delta \downarrow 0} \int_{I_j^\#} k_j^\delta(t)\, dt = 0\,,$$

having the property that, for almost all $t \in I_j^\#$, the inequality

$$\left\| \tilde{f}_{q_j^\#}(x, \eta_j^\#(t), t) - \tilde{f}_{q_j^\#}(\xi_j^\#(t), \eta_j^\#(t), t) - \tilde{Y}_j(t) \cdot (x - \xi^\#(t)) \right\| \\ \leq k_j^\delta(t) \| x - \xi_j^\#(t) \| \tag{1.21}$$

holds whenever $\|x - \xi_j^{\#}(t)\| \leq \delta$.

A15. For each $j \in \{1, \ldots, \nu^{\#}\}$, either

I. one of the following two conditions holds:

I.1. $\boldsymbol{U}_{q_j^{\#}}$ is a weak variable-time measurable variational neighborhood of $\eta_j^{\#}$, $(\xi_j^{\#}(t_j^{\#}), t_j^{\#})$ is a backward regular point for all constant controls and a forward regular point for $\eta_j^{\#}$, and $(\xi_j^{\#}(\tau_j^{\#}), \tau_j^{\#})$ is forward regular for all constant controls and backward regular for $\eta_j^{\#}$.

I.2. $\boldsymbol{U}_{q_j^{\#}}$ is a fixed-time measurable variational neighborhood of $\eta_j^{\#}$, the switching set $S_j^{\#}$ is a subset of the Cartesian product $M_j^{\#} \times M_{j+1}^{\#} \times \{\tau_j^{\#}\} \times \mathbb{R}$, and the switching set $S_{j-1}^{\#}$ is a subset of $M_{j-1}^{\#} \times M_j^{\#} \times \mathbb{R} \times \{t_j^{\#}\}$.

or

II. the following two conditions hold:

a. the "constant control augmented vector fields" satisfy integral bounds

$$\|\tilde{f}_{q_j^{\#}}(x, u, t)\| \leq k_{u,j}(t) \quad \text{for all} \quad (x, u, t) \in \mathcal{N}_j(u)\,, \tag{1.22}$$

where each $k_{u,j}$ is a nonnegative Lebesgue integrable function on $\mathbb{R}$,

b. one of conditions I.1, I.2 holds, with "measurable" replaced by "interval,"

or

III. Σ is autonomous at location $q_j^{\#}$, and $\boldsymbol{U}_{q_j^{\#}}$ is a strong variable-time interval variational neighborhood of $\eta_j^{\#}$.

Theorem 1.4.1. *If S1 and A5-A15 hold, then the maximum principle is true. Moreover, the adjoint vector can be chosen in such a way that, in addition, the Hamiltonian function* $t \to H_{q_j^{\#}}(\xi_j^{\#}(t), \psi_j(t), \eta_j^{\#}(t), \psi_0, t)$ *is almost everywhere constant on* $I_j^{\#}$ *for every* j *for which A15.III holds.* $\diamondsuit$

Theorem 1.4.1 is stronger than the classical versions of the maximum principle given, e.g., in Pontryagin *et al.* [3] and Berkovitz [1], because it does not require the maps to be of class C^1, or even Lipschitz continuous, and only assumes that they are differentiable along the reference trajectory.

1.5 Outline of the proof

For $n \in \mathbb{N}$, we use $\mathbb{R}^n$, $\mathbb{R}_n$, $\mathbb{R}^{n\times n}$ to denote, respectively, the spaces of n-dimensional real column vectors, n-dimensional real row vectors, and $n \times n$ real matrices.

For simplicity, we will assume that the state spaces M_q are open subsets of Euclidean spaces $\mathbb{R}^{n_q}$.

Remark 1.5.1. The proof for the general manifold case is identical, except that more care has to be exercised to define intrinsically various objects such as $\frac{\partial f_q}{\partial x}$. Alternatively, one could prove the theorem on manifolds by an extra "hybridization" of our problem, in which each curve ξ_j is covered by domains Ω_j^k of coordinate patches (Ω_j^k, κ_j^k), and these coordinate patches are treated as new locations, with state spaces $\kappa_j^k(\Omega_j^k) \subseteq \mathbb{R}^{\dim M_j}$, while the changes of coordinates are treated as switchings. $\diamondsuit$

Then the vector fields f_q are $\mathbb{R}^{n_q}$-valued maps, and the differentials $\frac{\partial f_q}{\partial x}$ are $\mathbb{R}^{n_q \times n_q}$-valued. We let $n_j = n_{q_j}$, and define

$$Y_j(t) = \frac{\partial f_{q_j}}{\partial x}(\xi_j(t), \eta_j(t), t)\,, \quad y_j(t) = \frac{\partial L_{q_j}}{\partial x}(\xi_j(t), \eta_j(t), t)\,, \tag{1.23}$$

so Y_j and y_j are integrable functions, with values in the spaces $\mathbb{R}^{n_j \times n_j}$ and $\mathbb{R}_{n_j}$, respectively.

The proof of Theorem 1.4.1 is carried out by means of "needle variations." Our variations will be *set-valued maps*, because they are constructed by associating to each value of the variation parameter "the solution" of a certain initial value problem with a continuous but non-Lipschitz right-hand side. Since the solution need not be unique, the object associated to the parameter value will in fact be a *set* of curves rather than a single curve. We introduce the following notation: "$F : A \longrightarrow\!\!\!\!\rightarrow B$" means "$F$ is a set-valued map from A to B." If $F : A \longrightarrow\!\!\!\!\rightarrow B$, then the *graph* of F is the set $\{(a, b) : a \in A, b \in F(a)\}$. (Often, one defines a set-valued map from A to B to be a subset of $A \times B$. In that case, the graph of F is just F.)

We first remark that for every $j \in \{1, \ldots, \nu\}$ there exists an integrable function $\hat{k}_j$ on I_j such that the integral bound

$$\|\tilde{f}_{q_j}(x, \eta_j(t), t)\| \leq \hat{k}_j(t) \quad \text{for all} \quad (x, t) \in \mathcal{N}_j \tag{1.24}$$

holds for some neighborhood $\mathcal{N}_j$ of the graph of ξ_j in $M_{q_j} \times I_j$. (This follows from A12, A14, the absolute continuity of ξ_j, and the integrability of the function $t \to L_{q_j}(\xi_j(t), \eta_j(t), t)$.)

We will begin by doing the proof for the case corresponding to Part II of condition A15, assuming the the "interval" analogue of I.2 holds. We will then discuss how to modify the proof to take care of the other cases.

So we assume that the integral bounds of A15.II.a hold for all j, so that for every j and every $u \in U_{q_j}$ there exists an integrable function $k_{u,j}$ on $\mathbb{R}$ such that (1.22) holds. We also assume that every U_{q_j} is a fixed-time interval variational neighborhood of η_j, and that each switching set S_j is contained in the product $M_j \times M_{j+1} \times \{\tau_j\} \times \{t_{j+1}\}$. (As explained before, this includes in particular the set S_ν, which is defined to be E_{q_ν,q_1}. Recall that $M_{\nu+1} = M_1$ and $t_{\nu+1} = t_1$.) It follows that $S_j = S_j^0 \times \{\tau_j\} \times \{t_{j+1}\}$ for some subset S_j^0 of $M_j \times M_{j+1}$.

Fix a positive integer m and then choose, for each index $j \in \{1, \dots, \nu\}$, an m-tuple $\mathbf{u}_j = (u_j^1, u_j^2, \dots, u_j^m) \in U_{q_j}^m$ of control values. Let

$$\mathcal{N}_j(\mathbf{u}) = \mathcal{N}_j(u_j^1) \cap \ldots \cap \mathcal{N}_j(u_j^m)\,.$$

Then the Scorza-Dragoni theorem together with the integral bounds implies that for each j there exists a null subset $B_j(\mathbf{u})$ of $\mathbb{R}$ such that if $(x,t) \in \mathcal{N}_j(\mathbf{u})$ and $t \notin B_j(\mathbf{u})$ then (x,t) is a forward and backward regular point for η_j and for each of the constant controls $u_j^1, \dots, u_j^m$.

Pick, for each j, m distinct points $s_j^1, \dots, s_j^m$ of I_j, not belonging to $B_j(\mathbf{u})$, and such that $s_j^1 < s_j^2 < \ldots < s_j^m < \tau_j$. Fix a small positive real number $\bar{\varepsilon}$. Let $\mathbf{E} = ([0,\bar{\varepsilon}]^m)^\nu$, so $\mathbf{E}$ is the set of all ν-tuples $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_\nu)$ of m-tuples $\boldsymbol{\varepsilon}_j = (\varepsilon_j^1, \dots, \varepsilon_j^m)$ of numbers ε_j^k such that $0 \le \varepsilon_j^k \le \bar{\varepsilon}$ for all j,k. Then for each $\boldsymbol{\varepsilon} \in \mathbf{E}$ we can construct the modified control

$$\zeta^{\mathbf{u},\mathbf{s},\boldsymbol{\varepsilon}} \stackrel{\text{def}}{=} (\mathbf{q}, \mathbf{I}, \boldsymbol{\eta}^{\mathbf{u},\mathbf{s},\boldsymbol{\varepsilon}})\,,$$

having the same switching strategy $\mathbf{q}$ and the same sequence $\mathbf{I}$ of intervals as ζ, such that

$$\begin{aligned} \boldsymbol{\eta}^{\mathbf{u},\mathbf{s},\boldsymbol{\varepsilon}} &= (\eta_1^{\mathbf{u},\mathbf{s},\boldsymbol{\varepsilon}}, \dots, \eta_\nu^{\mathbf{u},\mathbf{s},\boldsymbol{\varepsilon}})\,, \\ \eta_j^{\mathbf{u},\mathbf{s},\boldsymbol{\varepsilon}}(t) &= \begin{cases} \eta_j(t) & \text{if} \quad t \in I_j \backslash (\cup_{k=1}^m [s_j^k, s_j^k + \varepsilon_j^k])\,, \\ u_j^k & \text{if} \quad t \in [s_j^k, s_j^k + \varepsilon_j^k]\,. \end{cases} \end{aligned}$$

Fix small balls $D_1, \dots, D_m$ in the tangent spaces $T_{\xi_1(t_1)}M_1, \dots, T_{\xi_\nu(t_\nu)}M_\nu$, centered at 0, and smooth maps $\Delta_j : D_j \to M_j$ such that $\Delta_j(0) = \xi_j(t_j)$ and the differential of Δ_j at 0 is the identity map of $T_{\xi_j(t_j)}M_j$.

Let $\mathbf{P} = \mathbf{E} \times \mathbf{D}$. For each $(\boldsymbol{\varepsilon}, \mathbf{v}) \in \mathbf{P}$ let $\xi^{\mathbf{u},\mathbf{s}}(\boldsymbol{\varepsilon}, \mathbf{v})$ be the set of all ν-tuples $\tilde{\xi} = (\tilde{\xi}_1, \dots, \tilde{\xi}_\nu)$ such that each $\tilde{\xi}_j$ is an absolutely continuous curve in M_j, defined on I_j, which is a solution of the initial value problem

$$\begin{cases} \dot{x} &= f_{q_j}(x, \eta_j^{\mathbf{u},\mathbf{s},\boldsymbol{\varepsilon}}(t), t) \quad \text{for} \quad \text{a.e.} \quad t \in I_j\,, \\ x(t_j) &= \Delta_j(v_j)\,. \end{cases} \tag{1.25}$$

Notice that in general $\xi^{\mathbf{u},\mathbf{s}}(\boldsymbol{\varepsilon},\mathbf{v})$ will not consist of a single element, because the right-hand of the differential equation of (1.25) is just continuous with respect to x, so uniqueness of solutions of (1.25) is not guaranteed. On the other hand, global existence of solutions on I_j follows— if $\bar{\varepsilon}$ and the balls D_j are small enough—because (a) the differential equation of (1.25) has local solutions —thanks to the continuity of the right-hand side with respect to x, the measurability with respect to t, and the integral bounds (1.22) and (1.24), and (b) a solution of (1.25), as long as it exists, will stay in any prespecified neighborhood of the graph of ξ_j —provided, once again, that $\bar{\varepsilon}$ and the D_j are small enough—due to A12 and A14 (which imply bounds

$$\|\tilde{f}_{q_j}(x,\eta_j(t),t)-\tilde{f}_{q_j}(\xi_j(t),\eta_j(t),t)\|\le\kappa_j(t)\|x-\xi_j(t)\|\,,$$

with κ_j integrable, valid whenever $\|x-\xi_j(t)\|\le\delta$ and δ is small enough), and Gronwall's inequality.

We let $\mathcal{V}^{\mathbf{u},\mathbf{s}}$ be the set-valued map $\mathbf{P}\ni(\boldsymbol{\varepsilon},\mathbf{v})\to\xi^{\mathbf{u},\mathbf{s}}(\boldsymbol{\varepsilon},\mathbf{v})$. It then follows by a standard application of the Ascoli-Arzelà theorem that the graph of $\mathcal{V}^{\mathbf{u},\mathbf{s}}$— i.e., the set of all pairs $((\boldsymbol{\varepsilon},\mathbf{v}),\tilde{\xi})$ such that $\tilde{\xi}\in\xi^{\mathbf{u},\mathbf{s}}(\boldsymbol{\varepsilon},\mathbf{v})$—is a compact subset of the product $\mathbf{P}\times\mathbf{C}^0(\mathbf{I},\mathbf{M})$, where $\mathbf{C}^0(\mathbf{I},\mathbf{M})\stackrel{\mathrm{def}}{=}C^0(I_1,M_1)\times\ldots\times C^0(I_\nu,M_\nu)$. So $\mathcal{V}^{\mathbf{u},\mathbf{s}}$ *is an upper semicontinuous set-valued map from* $\mathbf{P}$ *to* $\mathbf{C}^0(\mathbf{I},\mathbf{M})$*, with nonempty compact values.*

To the variation $\mathcal{V}^{\mathbf{u},\mathbf{s}}$ we associate the set-valued "endpoint-cost map"

$$\mathcal{EC}^{\mathbf{u},\mathbf{s}}:\mathbf{P}\longrightarrow\!\!\!\!\rightarrow\mathcal{M}_1^0\times\ldots\times\mathcal{M}_\nu^0\times\mathbb{R}$$

given by

$$\mathcal{EC}^{\mathbf{u},\mathbf{s}}(\boldsymbol{\varepsilon},\mathbf{v})=\left\{\left(\boldsymbol{\partial}^0\tilde{\xi},C_L(\tilde{\xi},\zeta^{\mathbf{u},\mathbf{s},\boldsymbol{\varepsilon}})\right):\tilde{\xi}\in\xi^{\mathbf{u},\mathbf{s}}(\boldsymbol{\varepsilon},\mathbf{v})\right\},\tag{1.26}$$

where

$$\begin{aligned}\boldsymbol{\partial}^0\tilde{\xi}&\stackrel{\mathrm{def}}{=}(\partial_1^0\tilde{\xi},\partial_2^0\tilde{\xi},\ldots,\partial_{\nu-1}^0\tilde{\xi},\partial_\nu^0\tilde{\xi})\,,\\ \mathcal{M}_j^0&\stackrel{\mathrm{def}}{=}M_j\times M_{j+1}\,,\\ \partial_j^0\tilde{\xi}&\stackrel{\mathrm{def}}{=}(\tilde{\xi}_j(\tau_j),\tilde{\xi}_{j+1}(t_j))\,.\end{aligned}$$

Then $\mathcal{EC}^{\mathbf{u},\mathbf{s}}$ *is an upper semicontinuous set-valued map from* $\mathbf{P}$ *to the product* $\mathcal{M}_1^0\times\ldots\mathcal{M}_\nu^0\times\mathbb{R}$*, with nonempty compact values.*

Moreover, the maps $\mathcal{V}^{\mathbf{u},\mathbf{s}}$ and $\mathcal{EC}^{\mathbf{u},\mathbf{s}}$ are "regular" in the sense of [4], [5], or [6]. (This follows because the vector fields that occur in (1.25) can be regularized, i.e., approximated by smooth vector fields, for which the corresponding initial value problems have unique solutions. The variation maps $\mathcal{V}_k^{\mathbf{u},\mathbf{s}}$ and endpoint-cost maps $\mathcal{EC}_k^{\mathbf{u},\mathbf{s}}$ arising from these regularized vector fields are then single-valued and continuous, and approximate the set-valued maps $\mathcal{V}^{\mathbf{u},\mathbf{s}}$, $\mathcal{EC}^{\mathbf{u},\mathbf{s}}$ in the "graph convergence" sense used in [4] to define regularity.)

The map $\mathcal{EC}^{\mathbf{u},\mathbf{s}}$ is differentiable at 0, and its differential $D\mathcal{EC}^{\mathbf{u},\mathbf{s}}(0)$ can be explicitly computed. The result is, of course, a linear map

$$D\mathcal{EC}^{\mathbf{u},\mathbf{s}}(0) : (\mathbb{R}^m)^\nu \times T_{\xi_1(t_1)}M_1 \times \ldots \times T_{\xi_\nu(t_\nu)}M_\nu \to T_{\partial_1^0 \Xi}\mathcal{M}_1^0 \times \ldots \times T_{\partial_\nu^0 \Xi}\mathcal{M}_\nu^0 \times \mathbb{R}\,,$$

and is given by the formula

$$D\mathcal{EC}^{\mathbf{u},\mathbf{s}}(0)(\boldsymbol{\varepsilon},\mathbf{v}) = ((w_1,v_2),(w_2,v_3),\ldots,(w_\nu,v_1),\alpha(\boldsymbol{\varepsilon},\mathbf{v}))\,, \tag{1.27}$$

where

$$\begin{aligned}\alpha(\boldsymbol{\varepsilon},\mathbf{v}) = &\sum_{j=1}^{\nu}\int_{t_j}^{\tau_j} y_j(t)\cdot D\Theta^j_{t,t_j}(\xi_j(t_j))\cdot v_j\,dt \\ &+\sum_{j=1}^{\nu}\sum_{k=1}^{m}\varepsilon_j^k\Big(\lambda_j^k + \int_{s_j^k}^{\tau_j} y_j(t)\cdot D\Theta^j_{t,s_j^k}(\xi_j(s_j^k))\cdot \tilde{w}_j^k\,dt\Big)\,,\end{aligned}$$

$$w_j = D\Theta^j_{\tau_j,t_j}(\xi_j(t_j))\cdot v_j + \sum_{k=1}^{m}\varepsilon_j^k D\Theta^j_{\tau_j,s_j^k}(\xi_j(s_j^k))\cdot\tilde{w}_j^k\,, \tag{1.28}$$

$$\tilde{w}_j^k = f_{q_j}(\xi_j(s_j^k),u_j^k,s_j^k) - f_{q_j}(\xi_j(s_j^k),\eta_j(s_j^k),s_j^k)\,, \tag{1.29}$$

$$\lambda_j^k = L_{q_j}(\xi_j(s_j^k),u_j^k,s_j^k) - L_{q_j}(\xi_j(s_j^k),\eta_j(s_j^k),s_j^k)\,, \tag{1.30}$$

and $\{\Theta^j_{t,s}\}_{s,t\in I_j}$ is the family of flow maps corresponding to the reference vector field $(x,t)\to f_{q_j}(x,\eta_j(t),t)$. (Each differential $D\Theta^j_{t,s}(\xi_j(s))$ is then a linear map from $T_{\xi_j(s)}M_j$ to $T_{\xi_j(t)}M_j$. The flow maps $\Theta^j_{t,s}$ are of course set-valued, but our technical hypotheses imply that $\Theta^j_{t,s}$ is differentiable at $\xi_j(s)$ in the ordinary sense.)

Let $\Omega_j(\cdot,\cdot)$ denote the fundamental matrix solution of the linear time-varying differential equation $\dot{X} = Y_j(t)\cdot X$, so $I_j\times I_j \ni (t,s)\to\Omega_j(t,s)\in\mathbb{R}^{n_j\times n_j}$ is a continuous map having the property that

$$\Omega_j(t,s) = \mathrm{identity}_{\mathbb{R}^n_j} + \int_s^t Y_j(r)\cdot\Omega_j(r,s)\,dr\,. \tag{1.31}$$

Then

$$\begin{aligned}\alpha(\boldsymbol{\varepsilon},\mathbf{v}) = &\sum_{j=1}^{\nu}\int_{t_j}^{\tau_j} y_j(t)\cdot\Omega_j(t,t_j)\cdot v_j\,dt \\ &+\sum_{j=1}^{\nu}\sum_{k=1}^{m}\varepsilon_j^k\Big(\lambda_j^k + \int_{s_j^k}^{\tau_j} y_j(t)\cdot\Omega_j(t,s_j^k)\cdot\tilde{w}_j^k\,dt\Big)\,,\end{aligned}$$

$$w_j = \Omega_j(\tau_j, t_j) \cdot v_j + \sum_{k=1}^{m} \varepsilon_j^k \Omega_j(\tau_j, s_j^k) \cdot \tilde{w}_j^k \,.$$

Every pair $(\tilde{\xi}, \zeta^{\mathbf{u},\mathbf{s},\boldsymbol{\mathcal{E}}})$, for $\tilde{\xi} \in \xi^{\mathbf{u},\mathbf{s}}(\boldsymbol{\varepsilon}, \mathbf{v})$, is a pretrajectory-control pair of Σ. Moreover, $(\tilde{\xi}, \zeta^{\mathbf{u},\mathbf{s},\boldsymbol{\mathcal{E}}})$ belongs to $TCP(\Sigma; \mathcal{E})$ if and only if

$$\partial_j^0 \tilde{\xi} \in S_j^0 \quad \text{for} \quad j = 1, \ldots, \nu \,. \tag{1.32}$$

If (1.32) holds, then it is clear that the total cost $\mathcal{C}(\tilde{\xi}, \zeta^{\mathbf{u},\mathbf{s},\boldsymbol{\mathcal{E}}})$ is equal to the sum $C_L(\tilde{\xi}, \zeta^{\mathbf{u},\mathbf{s},\boldsymbol{\mathcal{E}}}) + c(\tilde{\xi}, \zeta^{\mathbf{u},\mathbf{s},\boldsymbol{\mathcal{E}}})$, where

$$c(\tilde{\xi}, \zeta^{\mathbf{u},\mathbf{s},\boldsymbol{\mathcal{E}}}) = \sum_{j=1}^{\nu} \Phi_j^0(\partial_j^0 \tilde{\xi}) \,, \tag{1.33}$$

and Φ_j^0 is the function $S_j^0 \ni (x, x') \to \Phi_j(x, x', \tau_j, t_{j+1})$.

The assumption that Ξ is optimal implies that

$$\mathcal{C}(\tilde{\xi}, \zeta^{\mathbf{u},\mathbf{s},\boldsymbol{\mathcal{E}}}) \geq \mathcal{C}(\Xi) \quad \text{whenever} \quad (1.32) \quad \text{holds}. \tag{1.34}$$

Now, fix smooth functions $\sigma_j : \mathcal{M}_j^0 \to \mathbb{R}$ such that $\sigma_j(\partial_j^0 \Xi) = 0$ and $\sigma_j(x, x') > 0$ whenever $(x, x') \neq \partial_j^0 \Xi$. Then let $\mathcal{G}$ be the set of all points $((x_1, x_1'), \ldots, (x_\nu, x_\nu'), r)$ of $\mathcal{M}_1^0 \times \ldots \times \mathcal{M}_\nu^0 \times \mathbb{R}$ having the property that $(x_j, x_j') \in S_j^0$ for $j = 1, \ldots, \nu$ and

$$r \leq \mathcal{C}(\Xi) - \sum_{j=1}^{\nu} \Phi_j^0(x_j, x_j') - \sum_{j=1}^{\nu} \sigma_j(x_j, x_j') \,.$$

Then

$$\mathcal{E}\mathcal{C}^{\mathbf{u},\mathbf{s}}(\mathbf{P}) \cap \mathcal{G} = \{ G_* \} \,, \tag{1.35}$$

where

$$G_* \stackrel{\text{def}}{=} (\, \partial_1^0 \Xi \,, \, \ldots \,, \, \partial_\nu^0 \Xi \,, \, C_L(\Xi) \,) \,. \tag{1.36}$$

We now determine a Boltyanskii approximating cone $\mathcal{K}$ for the set $\mathcal{G}$ at G_*. For this purpose, we first observe that each cone K_j must be equal to the product $K_j^0 \times \{0\} \times \{0\}$, where K_j^0 is a closed convex cone in $T_{\partial_j^0 \Xi} \mathcal{M}_j^0$ which is a Boltyanskii approximating cone to S_j^0 at $\partial_j^0 \Xi$. (Proof: let V_j be a neighborhood of 0 in the tangent space $T_{\partial_j \Xi} \mathcal{M}_j$, and let $\alpha_j : V_j \cap K_j \to S_j$ be a continuous map such that

$$\alpha_j(z) = \partial_j \Xi + z + o(\|z\|) \quad \text{as} \quad z \to_{K_j} 0 \,.$$

Write $\alpha_j(z) = (\alpha_j^1(z), \alpha_j^2(z), \alpha_j^3(z), \alpha_j^4(z))$, so $\alpha_j^1(z) \in M_j$, $\alpha_j^2(z) \in M_{j+1}$, $\alpha_j^3(z) \in \mathbb{R}$, and $\alpha_j^4(z) \in \mathbb{R}$. Then $\alpha_j^3(z) \equiv \tau_j$ and $\alpha_j^4(z) \equiv t_{j+1}$. If we let $z = (z_1, z_2, z_3, z_4)$, we see that $\alpha_j^3(z) = \alpha_j^3(0)$, while on the other hand $\alpha_j^3(z) = \alpha_j^3(0) + z_3 + o(\|z\|)$. Therefore $\alpha_j^3(\rho z) = \alpha_j^3(0) + \rho z_3 + o(\rho)$ as $\rho \downarrow 0$, and then $\rho z_3 = o(\rho)$, so $z_3 = 0$. A similar argument proves that $z_4 = 0$. So $K_j = K_j^0 \times \{0\} \times \{0\}$, as stated. Moreover, if we define a map β_j by letting $\beta_j(w) = (\alpha_j^1(w,0,0), \alpha_j^2(w,0,0))$, then β is a continuous map from a neighborhood of 0 in K_j^0 to S_j^0, and $\beta_j(w) = \partial_j^0 \Xi + w + o(\|w\|)$ as $w \to_{K_j^0} 0$.) We then define $\mathcal{K}$ to be the set of all $((z_1, z_1'), \ldots, (z_\nu, z_\nu'), r)$ such that $(z_j, z_j') \in K_j^0$ for $j = 1, \ldots, \nu$, and

$$r \le -\sum_{j=1}^{\nu} \nabla \Phi_j^0(\partial_j^0 \Xi) \cdot (z_j, z_j') \,. \tag{1.37}$$

Then $\mathcal{K}$ is a closed convex cone in $T_{G_*}(\mathcal{M}_1^0 \times \ldots \times \mathcal{M}_\nu^0 \times \mathbb{R})$, and it is easy to see that $\mathcal{K}$ is a Boltyanskii approximating cone to $\mathcal{G}$ at G_*. Moreover, $\mathcal{K}$ is not a linear subspace of $T_{G_*}(\mathcal{M}_1^0 \times \ldots \times \mathcal{M}_\nu^0 \times \mathbb{R})$, because $(0,-1) \in \mathcal{K}$ but $(0,1) \notin \mathcal{K}$.

It then follows from the general separation theorem of [4] that there exists a nonzero linear functional Ψ on $T_{G_*}(\mathcal{M}_1^0 \times \ldots \times \mathcal{M}_\nu^0 \times \mathbb{R})$ which is nonnegative on $\mathcal{K}$ and nonpositive on $\hat{\mathcal{K}}$, where

$$\hat{\mathcal{K}} \stackrel{\text{def}}{=} D\mathcal{E}\mathcal{C}^{\mathbf{u},\mathbf{s}}(0)\Big(([0,\infty[^{\,m})^\nu \times T_{\xi_1(t_1)}M_1 \times \ldots \times T_{\xi_\nu(t_\nu)}M_\nu\Big) \,. \tag{1.38}$$

Write

$$\Psi = (\psi_1^+, \psi_2^-, \psi_2^+, \psi_3^-, \ldots, \psi_\nu^+, \psi_1^-, -\psi_0) \,, \tag{1.39}$$

where each ψ_j^+ is a linear functional on $T_{\xi_j(\tau_j)}M_j$, each ψ_j^- is a linear functional on $T_{\xi_j(t_j)}M_j$, and $\psi_0 \in \mathbb{R}$. Since the vector $(0,-1)$ belongs to $\mathcal{K}$, the fact that Ψ is nonnegative on $\mathcal{K}$ implies that $\psi_0 \ge 0$. Next, fix a j and a vector $v \in T_{\xi_j(t_j)}M_j$, and let $\boldsymbol{\varepsilon} = 0$, $\mathbf{v} = (v_1, \ldots, v_\nu)$, where $v_\ell = 0$ if $\ell \ne j$, $v_j = v$. With this choice of $\boldsymbol{\varepsilon}$ and $\mathbf{v}$, the vector $D\mathcal{E}\mathcal{C}^{\mathbf{u},\mathbf{s}}(0)(\boldsymbol{\varepsilon}, \mathbf{v})$ is equal to $((w_1, v_2), \ldots, (w_\nu, v_1), r)$, where the v_ℓ and w_ℓ vanish when $\ell \ne j$, $v_j = v$, $w_j = \Omega_j(\tau_j, t_j) \cdot v$, and $r = \int_{t_j}^{\tau_j} y_j(t) \cdot \Omega_j(t, t_j) \cdot v \, dt$. The fact that Ψ is nonpositive on $\hat{\mathcal{K}}$ implies that

$$\psi_j^- \cdot v + \psi_j^+ \cdot \Omega_j(\tau_j, t_j) \cdot v - \psi_0 \int_{t_j}^{\tau_j} y_j(t) \cdot \Omega_j(t, t_j) \cdot v \, dt \le 0 \,. \tag{1.40}$$

Since this is true for all v in the linear space $T_{\xi_j(t_j)}M_j$, it follows that

$$\psi_j^- + \psi_j^+ \cdot \Omega_j(\tau_j, t_j) - \psi_0 \int_{t_j}^{\tau_j} y_j(t) \cdot \Omega_j(t, t_j) \, dt = 0 \,. \tag{1.41}$$

Let ψ_j be the unique solution of the adjoint equation (1.8) that satisfies the terminal condition $\psi_j(\tau_j) = \psi_j^+$. Then $\dot{\psi}_j(t) = -\psi_j(t) \cdot Y_j(t) + \psi_0 \cdot y_j(t)$, from which it follows that

$$\psi_j(t) = \psi_j^+ \cdot \Omega_j(\tau_j, t) - \psi_0 \int_t^{\tau_j} y_j(s) \cdot \Omega_j(s,t)\, ds\,. \tag{1.42}$$

(Proof: $\Omega_j(t,s) \cdot \Omega_j(s,t) =$ identity, so

$$\frac{\partial \Omega_j(t,s)}{\partial s} \cdot \Omega_j(s,t) + \Omega_j(t,s) \cdot \frac{\partial \Omega_j(s,t)}{\partial s} = 0\,,$$

and then

$$\frac{\partial \Omega_j(t,s)}{\partial s} \cdot \Omega_j(s,t) + \Omega_j(t,s) \cdot Y_j(s) \cdot \Omega_j(s,t) = 0\,,$$

so

$$\frac{\partial \Omega_j(t,s)}{\partial s} = -\Omega_j(t,s) \cdot Y_j(s)\,.$$

Using this, if we define $\psi_j^*(t)$ to be the right-hand side of (1.42), we see that

$$\dot{\psi}_j^*(t) = -\psi_j^*(t) \cdot Y_j(t) + \psi_0 \cdot y_j(t)\,.$$

Since $\psi_j^*(\tau_j) = \psi_j^+ = \psi_j(\tau_j)$, the conclusion follows.)

The identities (1.41) and (1.42) imply that

$$\psi_j^- = -\psi_j(t_j)\,. \tag{1.43}$$

So we have shown that (1.43) holds for all j.

Now fix j, pick a point (z,z') in K_j^0, and define $\mathbf{z} = ((z_1,z_1'),\ldots,(z_\nu,z_\nu'),r)$, where $z_\ell = 0$ and $z_\ell' = 0$ if $\ell \neq j$, $z_j = z$, $z_j' = z'$, and $r = -\nabla \Phi_j^0(\partial_j^0 \Xi)\cdot(z,z')$. Then $\mathbf{z}$ clearly belongs to $\mathcal{K}$, so the nonnegativity of Ψ on $\mathcal{K}$ implies that $\psi_j^+ \cdot z + \psi_{j+1}^- \cdot z' - \psi_0 r \geq 0$, that is,

$$\Big((\psi_j(\tau_j), -\psi_{j+1}(t_{j+1})) + \psi_0 \nabla \Phi_j^0(\partial_j^0 \Xi)\Big) \cdot (z,z') \geq 0\,,$$

or, equivalently,

$$\Big((-\psi_j(\tau_j), \psi_{j+1}(t_{j+1})) - \psi_0 \nabla \Phi_j^0(\partial_j^0 \Xi)\Big) \cdot (z,z') \leq 0\,.$$

Since this is true for all $(z,z') \in K_j^0$, and $K_j = K_j^0 \times \{0\} \times \{0\}$, the conclusion that

$$(-\psi_j(\tau_j), \psi_{j+1}(t_{j+1}), h_j^+, -h_{j+1}^-) - \psi_0 \nabla \Phi_j(\partial_j \Xi) \in K_j^\perp$$

follows trivially. For $j = 1,\ldots,\nu-1$, this yields the switching conditions, and for $j = \nu$ this is the transversality condition.

Next, let us fix j, and also fix $k \in \{1, \ldots, m\}$. Choose $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_\nu)$, where we let $\boldsymbol{\varepsilon}_\ell = 0$ if $\ell \neq j$, and $\boldsymbol{\varepsilon}_j = (\varepsilon_j^1, \ldots, \varepsilon_j^m)$, where $\varepsilon_j^i = 0$ if $i \neq k$, and $\varepsilon_j^k = 1$. Also, choose $\mathbf{v} = 0$. Then $D\mathcal{E}\mathcal{C}^{\mathbf{u},\mathbf{s}}(0)(\boldsymbol{\varepsilon}, \mathbf{v})$ is equal to $((w_1, 0), (w_2, 0), \ldots, (w_\nu, 0), \alpha(\boldsymbol{\varepsilon}, 0))$, where

$$\begin{aligned}
w_i &= 0 \qquad\qquad \text{if} \qquad i \neq j\,, \\
w_j &= \Omega_j(\tau_j, s_j^k) \cdot \tilde{w}_j^k\,, \\
\alpha(\boldsymbol{\varepsilon}, 0) &= \lambda_j^k + \int_{s_j^k}^{\tau_j} y_j(t) \cdot \Omega_j(t, s_j^k) \cdot \tilde{w}_j^k \, dt\,.
\end{aligned}$$

The nonpositivity of Ψ on $\hat{\mathcal{K}}$ then implies that $\psi_j^+ \cdot w_j - \psi_0 \alpha(\boldsymbol{\varepsilon}, 0) \leq 0$, i.e., that

$$\psi_j^+ \cdot \Omega_j(\tau_j, s_j^k) \cdot \tilde{w}_j^k - \psi_0 \int_{s_j^k}^{\tau_j} y_j(t) \cdot \Omega_j(t, s_j^k) \cdot \tilde{w}_j^k \, dt - \psi_0 \lambda_j^k \leq 0\,,$$

that is

$$\psi_j(s_j^k) \cdot \tilde{w}_j^k - \psi_0 \lambda_j^k \leq 0\,. \tag{1.44}$$

In view of (1.29) and (1.30), this says that

$$\begin{aligned}
&\psi_j(s_j^k) \cdot f_{q_j}(\xi_j(s_j^k), u_j^k, s_j^k) - \psi_0 L_{q_j}(\xi_j(s_j^k), u_j^k, s_j^k) \\
&\leq \psi_j(s_j^k) \cdot f_{q_j}(\xi_j(s_j^k), \eta_j(s_j^k), s_j^k)\,, -\psi_0 L_{q_j}(\xi_j(s_j^k), \eta_j(s_j^k), s_j^k)\,,
\end{aligned}$$

i.e., that

$$H_{q_j}(\xi_j(s_j^k), \psi_j(s_j^k), u_j^k, \psi_0, s_j^k) \leq H_{q_j}(\xi_j(s_j^k), \psi_j(s_j^k), \eta_j(s_j^k), \psi_0, s_j^k)\,. \tag{1.45}$$

The inequality (1.45) is precisely the Hamiltonian maximization condition, except that we have only established it for finitely many times and finitely many control values. The complete Hamiltonian maximization condition then follows by a standard compactness argument that we omit.

This completes the proof under the special assumptions described earlier in this section. We now outline how to prove the result in the general case, still assuming that A15.II.a holds. Assume that, for a particular j, $\boldsymbol{U}_{q_j}$ is a weak variable-time interval variational neighborhood of η_j, and the regularity conditions of A15.I.1 are satisfied. This implies, to begin with, that the limits that occur in the right-hand sides of the definitions of $h_j^\pm$ (cf. Def. 1.4.2) exist, and are given by

$$\begin{aligned}
h_j^+ &= H_{q_j}(\xi_j(\tau_j), \psi_j(\tau_j), \eta_j(\tau_j), \psi_0, \tau_j)\,, \\
h_j^- &= H_{q_j}(\xi_j(t_j), \psi_j(t_j), \eta_j(t_j), \psi_0, t_j)\,.
\end{aligned}$$

In this case, we can make some extra variations in addition to those considered in the previous argument. Specifically, we can do "shortenings" at the endpoints t_j, τ_j (that is, we can restrict η_j to a subinterval $[t_j+\alpha, \tau_j-\beta]$, with $\alpha \geq 0$ and $\beta \geq 0$. In the definition of the endpoint-cost map given in (1.26), we substitute $\partial_j\tilde{\xi}$ for $\partial_j^0\tilde{\xi}$, that is, we add two extra time variables, corresponding to variations of t_j and of τ_j. Corresponding to these variables, the functional Ψ will now have two new components $\hat{\psi}_j^-$, $\hat{\psi}_j^+$, and the cone $\hat{\mathcal{K}}$ will have two new coordinates δt_j, $\delta\tau_j$. The linear part of the effect on $\mathcal{EC}^{\mathbf{u},\mathbf{s}}$ of a "shortening at the right endpoint" will involve $-\beta f_{q_j}(\xi_j(\tau_j), \eta_j(\tau_j), \tau_j)$ (variation of $\xi_j(t_j)$), $-\beta$ (variation of τ_j), and $-\beta L_{q_j}(\xi_j(\tau_j), \eta_j(\tau_j), \tau_j)$ (variation of the Lagrangian cost). This will give the inequality

$$-\psi_j^+ \cdot f_{q_j}(\xi_j(\tau_j), \eta_j(\tau_j), \tau_j) - \hat{\psi}_j^+ + \psi_0 L_{q_j}(\xi_j(\tau_j), \eta_j(\tau_j), \tau_j) \leq 0\,,$$

that is,

$$H_{q_j}(\xi_j(\tau_j), \psi_j(\tau_j), \eta_j(\tau_j), \psi_0, \tau_j) \geq -\hat{\psi}_j^+\,.$$

On the other hand, the linear part of the effect on $\mathcal{EC}^{\mathbf{u},\mathbf{s}}$ of an "expansion at the right endpoint" using the control value u will involve $\beta f_{q_j}(\xi_j(\tau_j), u, \tau_j)$ (variation of $\xi_j(t_j)$), β (variation of τ_j), and $\beta L_{q_j}(\xi_j(\tau_j), u, \tau_j)$ (variation of the Lagrangian cost). This will give the inequality

$$\psi_j^+ \cdot f_{q_j}(\xi_j(\tau_j), u, \tau_j) + \hat{\psi}_j^+ - \psi_0 L_{q_j}(\xi_j(\tau_j), u, \tau_j) \leq 0\,,$$

that is,

$$H_{q_j}(\xi_j(\tau_j), \psi_j(\tau_j), u, \psi_0, \tau_j) \leq -\hat{\psi}_j^+\,.$$

Since this happens for all values of u, we can specialize to $u = \eta_j(\tau_j)$ and conclude that

$$\hat{\psi}_j^+ = -H_{q_j}(\xi_j(\tau_j), \psi_j(\tau_j), \eta_j(\tau_j), \psi_0, \tau_j) = -h_j^+\,.$$

A similar argument yields the identity

$$\hat{\psi}_j^- = H_{q_j}(\xi_j(t_j), \psi_j(t_j), \eta_j(t_j), \psi_0, t_j) = h_j^-\,.$$

Now, if the same conditions are also satisfied for $j+1$, then we use the cone K_j instead of K_j^0. If $(z, z', \delta\tau_j, \delta t_{j+1})$ belongs to K_j, we will get a member $(z, z', \delta\tau_j, \delta t_{j+1}, r)$ of $\mathcal{K}$ by choosing $r = -\nabla\Phi_j(\partial_j\Xi).(z, z', \delta\tau_j, \delta t_{j+1})$. This will then yield the inequality

$$\psi_j^+ \cdot z + \psi_{j+1}^- \cdot z' + \hat{\psi}_j^+\delta\tau_j + \hat{\psi}_{j+1}^-\delta t_{j+1} - \psi_0 r \geq 0\,,$$

that is,

$$\begin{aligned} 0 \;\leq\; & \psi_j(\tau_j)\cdot z - \psi_{j+1}(t_{j+1})\cdot z' - h_j^+\delta\tau_j + h_{j+1}^-\delta t_{j+1} \\ & +\psi_0\nabla\Phi_j(\partial_j\Xi).(z, z', \delta\tau_j, \delta t_{j+1}) \end{aligned}$$

which implies

$$(-\psi_j(\tau_j), \psi_{j+1}(t_j), h_j^+, -h_{j+1}^-) - \psi_0 \nabla \Phi_j(\partial_j \Xi) \in K_j^\perp \,, \tag{1.46}$$

as desired.

The intermediate case, when the interval analogue of A15.I.2 holds for $j+1$, is easily handled, and we omit the argument.

Finally, in the autonomous case we observe, first of all, that the integral bounds are automatically satisfied. In this case, one can make shortening and expansion variations at almost all times. Once again, these variations change the times t_j and τ_j, so Ψ has corresponding components $\hat{\psi}_j^-$, $\hat{\psi}_j^+$. The identities

$$-\hat{\psi}_j^+ = H_{q_j}(\xi_j(t), \psi_j(t), \eta_j(t), \psi_0) = \hat{\psi}_j^-$$

(where we have explicitly spelled out the fact that in H_{q_j} there is no direct dependence on t) now follow for almost all $t \in I_j$. This clearly implies that the function $I_j \ni t \to H_{q_j}(\xi_j(t), \psi_j(t), \eta_j(t), \psi_0)$ is a. e. constant. It then follows, once again, that the limits occurring in the definition of $h_j^\pm$ exist, and in addition

$$H_{q_j}(\xi_j(t), \psi_j(t), \eta_j(t), \psi_0) = h_j^+ = -h_j^-$$

for almost all t. The switching condition (1.46) is now established exactly as before.

To conclude, we sketch how to remove the assumption on the integral bounds, if the "measurable" conditions A15.I.1 or A15.I.2 hold for a particular j, instead of their weaker "interval" counterparts A15.II.1, A15.II.2.

In this case, we replace the vector field system f_{q_j} and the corresponding Lagrangian by a new system $\hat{f}_{q_j}$ and a new Lagrangian $\hat{L}_{q_j}$, with a different control space $\hat{U}_{q_j}$. We take $\hat{U}_{q_j}$ to be $U_{q_j} \times \mathbb{N}$, so a control value for the new system is now a pair (u, N) where $u \in U_{q_j}$ and $N \in \mathbb{N}$. We fix a compact neighborhood $\mathcal{N}$ of the graph of ξ_j and define, for each $u \in U_{q_j}$, the upper bound function

$$\theta_u(t) = \sup\{\|f_{q_j}(x, u, t)\| + |L_{q_j}(x, u, t)| : x \in \mathcal{N}(t)\}\,,$$

where $\mathcal{N}(t) = \{x : (x, t) \in \mathcal{N}\}$. Then θ_u is measurable and almost everywhere finite. We then define

$$\hat{f}_{q_j}(x, (u, N), t) = \begin{cases} f_{q_j}(x, u, t) & \text{if} \quad \sigma_u(t) \le N\,, \\ f_{q_j}(x, \eta_j(t), t) & \text{if} \quad \sigma_u(t) > N\,, \end{cases}$$

$$\hat{L}_{q_j}(x, (u, N), t) = \begin{cases} L_{q_j}(x, u, t) & \text{if} \quad \sigma_u(t) \le N\,, \\ L_{q_j}(x, \eta_j(t), t) & \text{if} \quad \sigma_u(t) > N\,. \end{cases}$$

It is then easy to see that the integral bounds are automatically satisfied for the new system. Moreover, the substitution of a constant control (u, N) for the reference control on a subinterval J of I_j is equivalent to the substitution of the control value u for $\eta_j(t)$ for t in the measurable set $\{s \in J : \sigma_u(s) \leq N\}$. In view of conditions A15.I or A15.II, this results in an admissible control for the original system. So the theorem under condition A15.II applies.

1.6 The case of a nondifferentiable reference vector field and Lagrangian

We now outline—without proof—how to remove the differentiability assumptions. For simplicity, we will only consider the removal of the differentiability requirement on the reference vector field and Lagrangian, even though it is also possible to weaken the differentiability requirement on the functions Φ_j.

Instead of A7, A12 and A14, we should now assume

A'. For each j belonging to the index set $\{1, \ldots, \nu^{\#}\}$ there exist maps $(x,t) \to \hat{f}_j(x,t)$ and $(x,t) \to \hat{L}_j(x,t)$ such that

- A'.1 for almost every $t \in I_j^{\#}$ the maps $x \to \hat{f}_j(x,t)$ and $x \to \hat{L}_j(x,t)$ are Lipschitz with a Lipschitz constant $c_j(t)$ such that $\int_{t_j}^{\tau_j} c_j(t)\,dt < \infty$,
- A'.2. for every x the maps $t \to \hat{f}_j(x,t)$ and $t \to \hat{L}_j(x,t)$ are measurable,
- A'.3. there exist measurable functions $k_j^{\delta} : I_j^{\#} \to [0, +\infty]$ such that

$$\lim_{\delta \downarrow 0} \int_{I_j^{\#}} k_j^{\delta}(t)\,dt = 0\,,$$

having the property that, for almost all $t \in I_j^{\#}$, the inequality

$$\begin{aligned} &\|f_{q_j^{\#}}(x, \eta_j^{\#}(t), t) - \hat{f}_j(x,t)\| + |L_{q_j^{\#}}(x, \eta_j^{\#}(t), t) - \hat{L}_j(x,t)| \\ &\leq k_j^{\delta}(t)\|x - \xi_j^{\#}(t)\| \end{aligned} \tag{1.47}$$

holds whenever $\|x - \xi_j^{\#}(t)\| \leq \delta$.

Under these new conditions, all the terms that occur in the statement of the maximum principle are well defined, except only for the adjoint equation (1.8). We now reinterpret the adjoint equation to mean

$$-\dot{\psi}_j(t) \in \partial_x \hat{H}_j(\xi_j(t), \psi_j(t), \psi_0, t) \text{ for a.e. } t \in I_j, \tag{1.48}$$

where $\partial_x \hat{H}_j(x, p, p_0, t)$ is the generalized gradient (in the sense of Clarke [2]) of the function $x \to \hat{H}_j(x, p, p_0, t)$, where

$$\hat{H}_j(x, p, p_0, t) = p \cdot \hat{f}_j(x, t) - p_0 \hat{L}_j(x, t) .$$

Theorem 1.6.1. *If S1, A5-6, A8-11, A13, A15 and A' hold, and in the definition of the adjoint equation (1.48) is substituted for (1.8), then the maximum principle is true.* ♢

References

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

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

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

[4] 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, Editors, Kluwer Academic Publishers, 1998, pp. 436-487.

[5] 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.

[6] 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, 1998, pp. 1-6.

[7] Sussmann, H. J., "A Maximum Principle for hybrid optimal control problems," to appear in *Proc. 38th IEEE Conference on Decision and Control*, Phoenix, AZ, Dec. 1999. IEEE publications, 1999.