REGULAR SYNTHESIS AND SUFFICIENCY CONDITIONS FOR OPTIMALITY

Benedetto Piccoli

SISSA–ISAS
via Beirut 2-4
34014 Trieste, Italy

piccoli@shannon.sissa.it

Héctor J. Sussmann*

Department of Mathematics, Rutgers University
110 Frelinghuysen Rd.
Piscataway, NJ 08854, USA

sussmann@hilbert.rutgers.edu
http://www.math.rutgers.edu/~sussmann

October 1998

ABSTRACT

We propose a definition of "regular synthesis," more general than those suggested by other authors such as Boltyanskii and Brunovský, and an even more general notion of "regular presynthesis." We give a complete proof of the corresponding sufficiency theorem, a slightly weaker version of which had been stated in an earlier article, with only a rough outline of the proof. We illustrate the strength of our result by showing that the optimal synthesis for the famous Fuller problem satisfies our hypotheses. We also compare our concept of synthesis with the simpler notion of a "family of solutions of the closed-loop equation arising from an optimal feedback law," and show by means of examples why the latter is inadequate, and why the difficulty cannot be resolved by using other concepts of solution—such as Filippov solutions, or the limits of sample-and-hold solutions recently proposed as feedback solutions by Clarke, Ledyaev, Subbotin and Sontag —for equations with a non-Lipschitz and possibly discontinuous right-hand side.

Keywords. Optimal control, regular synthesis, sufficient conditions.

Ref. SISSA 68/96/M

* Supported in part by NSF Grant DMS95-00798 and AFOSR Grant 0923. The final part of the work was carried out between January and June 1998, while this author was a visitor at the Institut Henri Poincaré (IHP), in Paris. He is grateful to the IHP for its kind hospitality during this period.

§1. Introduction

The purpose of the theory of "regular synthesis" is to turn into a rigorous, precise theorem the vague assertion that "if a collection Γ of extremals covers the whole space in a smooth enough way, then all the members of Γ are optimal." Naturally, a good definition should at the same time be stringent enough to lead to a correct proof of the sufficiency assertion, and broad enough to cover as large as possible a class of optimal control problems. In particular, it would be desirable for the theory to apply to well-known examples such as the famous "Fuller's problem" (cf. Fuller [12], Marchal [13], Zelikin-Borisov [28]), where the optimal controls have infinitely many switchings.

The three main goals of this paper are (a) to propose a new definition of regular synthesis, more general than those suggested by other authors (e. g. Boltyanskii [2], Brunovský [5], [6]), as well as an even more general notion of "regular presynthesis," (b) to give a detailed statement and proof of the corresponding sufficiency theorem, a slightly weaker version of which had been announced by one of us in an earlier article (Sussmann [27]), with only a rough sketch of the proof, and (c) to compare our definition with other concepts of synthesis such as those of Boltyanskii and Brunovský, and with the simpler notion of a "family of solutions of the closed-loop equation arising from an optimal feedback law."

Our main sufficiency result is Theorem 2.13. The key to the proof turns out to be a technical result—Theorem 2.14—on the relation between the adjoint vector on the Pontryagin Maximum Principle and the gradient of the cost functional. This in turn depends on results about differentiability of trajectories with respect to a parameter under weak differentiability assumptions on the dependence of the data on the parameter. These results are of interest in themselves and are presented separately in Appendix A.

Our definitions of regular presynthesis and regular synthesis are explained in §2, cf. Definition 2.3 for the notion of "presynthesis," 2.4 for that of "synthesis," and 2.12 for "regularity." The strength of Theorem 2.13 is illustrated in §3 by showing that it applies to the synthesis for Fuller's problem, where other sufficiency theorems such as those of [2] and [6] cannot be used because of the technical problem arising from the occurrence of an infinite number of switchings. (Not surprisingly, the optimality of Fuller's synthesis was originally proved without resorting to regular synthesis arguments, and using instead special symmetry properties of the problem, cf., e. g., Fuller [12], Marchal [13], Zelikin-Borisov [28].) Other definitions of synthesis—including those of Boltyanskii and Brunovský—are reviewed and compared with ours in §4, where we also discuss in detail why the concept of "optimal synthesis" cannot just be defined by simply identifying it with that of "optimal feedback law." In particular, we explain in §4 that, if $x \to v(x)$ is an optimal feedback, then in general the concept of "solution" of the closed-loop equation $\dot{x} = f(x, v(x))$ is problematic (because $x \to f(x, v(x))$ may fail to be Lipschitz-continuous) and compare the classical definition of solution with two other notions: (i) Filippov solutions, and (ii) the limits of sample-and-hold solutions recently proposed as feedback solutions by Clarke, Ledyaev, Subbotin and Sontag in [8], which will be referred to here as "CLSS solutions." We argue that the classical notion of solution fares better than the other two concepts, since there are situations where the optimal trajectories are not Filippov solutions of the optimal closed-loop equation, as well as cases where they are not CLSS solutions. The analysis of §4 is supplemented, in §5, with several examples that clarify the relationship between the various definitions. Finally, a number of technical points, including some arguments pertaining to the proof of Theorem 2.13, are relegated to two appendices.

As explained in §4, there are essentially two opposite starting points for defining the notion of a "regular" optimal synthesis. One approach, following Boltyanskii [2], Brunovský [5]-[6], Cesari [7], Fleming and Soner [10], and Sussmann [16], is to take as the primary object a smooth or "piecewise smooth" feedback control v, and then let the synthesis be the collection $Traj(v)$ of all trajectories generated by v. If this collection is too large (for example, if the closed-loop equation determined by v does not have unique solutions), then a synthesis would be a suitably defined subclass Γ of $Traj(v)$. The other strategy, proposed in Piccoli [14] and Sussmann [27], is to take a "synthesis" to be just a collection of trajectories—or, more precisely, trajectory-control pairs—not necessarily arising from a feedback control. To carry out the first stategy, Boltyanskii and Brunovský were forced to make precise sense of the notion of piecewise smoothness, which they did by requiring that the set of interest—that is, the set of initial points of optimal trajectories—be partitioned into submanifolds of various dimensions on which a smooth feedback control is specified. It turns out, however, that with this definition the feedback control by itself is not sufficient for the synthesis to be completely determined, since a discontinuous feedback control may generate too many trajectories, some of which could fail to be optimal, as shown in Example 5.3 below. So, in any case, the final outcome of the effort to implement the first strategy is a result more consistent with the second one, since one ends up with a notion of synthesis as a pair consisting of a feedback control v together with a selection Γ of a family of trajectories generated by v, suggesting that the specification of Γ is needed but that of v might not be. This leads naturally to the approach of this paper, in which v is dispensed with altogether.

An important point that ought to be further investigated but will not be touched upon here is the relation between the geometrical approaches of [5], [6], [16], [17], [18], [20], [21], [24], [27], and the theory of viscosity solutions (cf. Bardi and Capuzzo Dolcetta [1], Fleming and Soner [11]). We point out that the existing results of this theory for deterministic optimal control do not appear to cover cases such as Fuller's problem. (Cf. for example [1], where the main result is for Lagrangians that are bounded away from zero.)

§2. Regular Presynthesis and Synthesis, and the Sufficiency Theorem

We start with some definitions that will enable us to state and prove a mild generalization of the theorem whose proof was outlined in [27]. Our assumptions will be minimal, so as to achieve maximum generality.

We consider a control system:

$$\dot{x} = f(x,u)\,, \qquad x \in \Omega\,, \qquad u \in U\,, \tag{2.1}$$

and a family $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{P}}(\Omega, U, f, L, \mathcal{T}, \tau) = \{\mathcal{P}(\Omega, U, f, L, \mathcal{T}, \tau, x)\}_{x\in\Omega}$ of minimization problems, where $\mathcal{P}(\Omega, U, f, L, \mathcal{T}, \tau, x)$ is the problem:

$$\text{minimize} \quad \tau(\gamma(b)) + \int_a^b L(\gamma(t), \eta(t))dt \qquad \text{subject to } (2.1)\,,\ \gamma(a) = x\,,\ \text{ and } \gamma(b) \in \mathcal{T}\,. \tag{2.2}$$

(A more precise formulation of the problems $\mathcal{P}(\Omega, U, f, L, \mathcal{T}, \tau, x)$, carefully specifying the space of of curves to be considered, will be given below. Since both f and L have no explicit dependence on t, we can always take $b = 0$.) We assume that

(A1) *n is a positive integer and Ω—the "state space"—is an open subset of $\mathbb{R}^n$;*
(A2) *U—the "control set"—is a nonempty set;*
(A3) *$f : \Omega \times U \to \mathbb{R}^n$—the "dynamics"—and $L : \Omega \times U \to \mathbb{R}$—the "Lagrangian"—are maps;*
(A4) *$f(\cdot, u)$ and $L(\cdot, u)$ are maps of class C^1 for every fixed $u \in U$;*
(A5) *$\mathcal{T}$—the "target"—is a nonempty subset of Ω;*
(A6) *$\tau : \Omega \to \mathbb{R}$ is a function of class C^1.*

We use $T^c_x S$ to denote the *contingent tangent cone* to a set S at a point $x \in S$, defined as follows: a vector v is in $T^c_x S$ iff there exist a sequence $\{x_j\}$ of points of S and a sequence $\{h_j\}$ of strictly positive numbers such that $h_j \to 0$ and $x_j = x + h_j v + o(h_j)$ as $j \to \infty$. We write

$$\tilde{T}_x S = T^c_x S \cap (-T^c_x S) = \{v : v \in T^c_x S \quad \text{and} \quad -v \in T^c_x S\}\,. \tag{2.3}$$

As in Sussmann [27], $\tilde{f}$ will denote the map $\Omega \times U \ni (x,u) \to \tilde{f}(x,u) \stackrel{\text{def}}{=} (f(x,u), L(x,u)) \in I\!R^{n+1}$.

If $\Omega \times I\!R \ni (x,t) \to g(x,t) \in I\!R^m$ is a time-varying vector-valued map on Ω, then $Dg(x,t)$ denotes, for each $(x,t) \in \Omega$, the Jacobian matrix at x of the map $y \to g(y,t)$, so $Dg(x,t) \in I\!R^{m\times n}$.

If $\mu : A \to B$ is a map, we use $\mathrm{Dom}(\mu)$ to indicate the domain of μ, i.e. the set A.

A *control* is a U-valued map η such that $\mathrm{Dom}(\eta)$ is a compact subinterval $[a,b]$ of $I\!R$. If η is a control, we define maps $f_\eta : \Omega \times \mathbb{R} \to I\!R^n$, $L_\eta : \Omega \times \mathbb{R} \to I\!R$, $\tilde{f}_\eta : \Omega \times \mathbb{R} \to I\!R^{n+1}$ by letting

$$\begin{array}{llll} f_\eta(x,t) = f(x,\eta(t)), & L_\eta(x,t) = L(x,\eta(t)), & \tilde{f}_\eta(x,t) = \tilde{f}(x,\eta(t)), & \text{if } t \in \mathrm{Dom}(\eta)\,; \\ f_\eta(x,t) = 0, & L_\eta(x,t) = 0, & \tilde{f}_\eta(x,t) = 0, & \text{if } t \notin \mathrm{Dom}(\eta)\,. \end{array}$$

A *trajectory* for a control η is an absolutely continuous map $\gamma : \mathrm{Dom}(\eta) \to \Omega$ with the property that $\dot\gamma(t) = f(\gamma(t),\eta(t))$ for almost every $t \in \mathrm{Dom}(\eta)$. If $\mathrm{Dom}(\eta) = [a,b]$ (so that $\mathrm{Dom}(\gamma) = [a,b]$ as well), and $x = \gamma(a)$, $y = \gamma(b)$, we say that γ (or the pair (γ,η)) *goes from x to y*, and that η *steers x to y*, and we write $\gamma^- = x$, $\gamma^+ = y$.

A control η is $\mathcal{C}^1$-*admissible* for $\tilde{f}$ if $\tilde{f}_\eta$ satisfies the following $\mathcal{C}^1$ Carathéodory conditions:

(C.i) *the map* $\mathrm{Dom}(\eta) \ni t \to \tilde{f}_\eta(x,t) \in I\!R^{n+1}$ *is measurable for each* $x \in \Omega$;

(C.ii) *for every compact subset K of Ω there exists an integrable function $\varphi_K : I\!R \to I\!R$ such that for every $x \in K$ and $t \in \mathrm{Dom}(\eta)$:*

$$\|\tilde{f}_\eta(x,t)\| + \|D\tilde{f}_\eta(x,t)\| \le \varphi_K(t)\,. \tag{2.4}$$

If η is a $\mathcal{C}^1$-admissible control for $\tilde{f}$, and γ is a trajectory corresponding to η, then we say that (γ,η) is a $\mathcal{C}^1$-*admissible pair* for $\tilde{f}$. We use $\mathrm{Adm}^{\mathcal{C}^1}(\tilde{f})$ to denote the set of all $\mathcal{C}^1$-admissible pairs for $\tilde{f}$, and write $\mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f})$, $\mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f},\mathcal{T})$, $\mathrm{Adm}^{\mathcal{C}^1,0}_x(\tilde{f},\mathcal{T})$, to denote, respectively, the set of all $(\gamma,\eta) \in \mathrm{Adm}^{\mathcal{C}^1}(\tilde{f})$ such that $\mathrm{Dom}(\eta)$ is an interval of the form $[T,0]$ with $T \le 0$, the set of those $(\gamma,\eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f})$ such that $\gamma(0) \in \mathcal{T}$, and the set of those $(\gamma,\eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f},\mathcal{T})$ such that $\gamma^- = x$. For a pair $(\gamma,\eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f},\mathcal{T})$, we use $J(\gamma,\eta)$ to denote the *cost* of (γ,η), given by

$$J(\gamma,\eta) = \tau(\gamma^+) + \int_{\mathrm{Dom}(\gamma)} L(\gamma(t),\eta(t))\,dt\,. \tag{2.5}$$

If $\Gamma \subseteq \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f},\mathcal{T})$ is an arbitrary set of $\mathcal{C}^1$-admissible pairs for $\tilde{f}$ ending at the target $\mathcal{T}$ at time 0, we define a function $V_\Gamma : \Omega \to I\!R \cup \{+\infty\}$, called the *cost function of* Γ, by letting

$$V_\Gamma(x) = \inf\{J(\gamma,\eta) : (\gamma,\eta) \in \Gamma,\ \gamma^- = x\} \tag{2.6}$$

for $x \in \Omega$. (In particular, $V_\Gamma(x) = +\infty$ iff there is no $(\gamma,\eta) \in \Gamma$ starting at x.) If $\Gamma = \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f},\mathcal{T})$, then the cost function V_Γ is the *value function* of our problem, and in that case we write $V_{\tilde{f},\mathcal{T}}$ rather than $V_{\mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f},\mathcal{T})}$. Naturally, $V_\Gamma \ge V_{\tilde{f},\mathcal{T}}$ pointwise for every subset Γ of $\mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f},\mathcal{T})$.

We can now formulate $\mathcal{P}(\Omega,U,f,L,\mathcal{T},\tau,x)$ precisely: $\mathcal{P}(\Omega,U,f,L,\mathcal{T},\tau,x)$ is the problem of finding a $(\gamma,\eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}_x(\tilde{f},\mathcal{T})$ such that $J(\gamma,\eta) = V_{\tilde{f},\mathcal{T}}(x)$. Any such pair (γ,η) is said to be a *solution* of $\mathcal{P}(\Omega,U,f,L,\mathcal{T},\tau,x)$. A pair $(\gamma,\eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f},\mathcal{T})$ is *optimal* if it is a solution of $\mathcal{P}(\Omega,U,f,L,\mathcal{T},\tau,\gamma^-)$. We use $\boldsymbol{\mathcal{P}}(\Omega,U,f,L,\mathcal{T},\tau)$—or, simply, $\boldsymbol{\mathcal{P}}$, in any context where the meaning of Ω, U, f, L, $\mathcal{T}$, τ is clear—to denote the family of problems $\Big\{\mathcal{P}(\Omega,U,f,L,\mathcal{T},\tau,x)\Big\}_{x\in\Omega}$.

Remark 2.1. We emphasize that in the optimal control problems studied here *the trajectories have to end at the target but could pass through it before.* For example, consider the dynamical law $\dot{x} = 1$ in $\mathbb{R}$, with Lagrangian $L(x) \equiv -1$ and target $\mathcal{T} = \{0, 1\}$. Then the value function V is given by $V(x) = x - 1$ for $x \leq 1$ and $V(x) = +\infty$ for $x > 1$. If $x < 1$, then the optimal trajectory γ_x from x to the target is given by $\gamma_x(t) = t + 1$ for $T_x \leq t \leq 0$, where $T_x = x - 1$. Notice that if $x < 0$ then $t = 0$ is not the first time when $\gamma_x(t)$ belongs to $\mathcal{T}$. ♢

Given $\lambda \in I\!R_n$ (the space of row n-vectors), $\lambda^0 \in I\!R$, $x \in \Omega$ and $u \in U$, we define:

$$H^c(\lambda, \lambda^0, x, u) = \langle \lambda, f(x, u) \rangle + \lambda^0 L(x, u)\,, \quad H(\lambda, \lambda^0, x) = \inf\{H^c(\lambda, \lambda^0, x, u) : u \in U\}\,. \tag{2.7}$$

The functions $H^c : I\!R_n \times I\!R \times \Omega \times U \to I\!R$ and $H : I\!R_n \times I\!R \times \Omega \to I\!R \cup \{-\infty\}$ are known, respectively, as the *control Hamiltonian* and the *minimized Hamiltonian* of $\tilde{f}$.

Definition 2.2. We say that the pair $(\gamma, \eta) \in \mathrm{Adm}^{\mathcal{C}^{1,0}}(\tilde{f}, \mathcal{T})$ is *extremal* if there exist an absolutely continuous map $\lambda : \mathrm{Dom}(\gamma) \to I\!R_n$ and a constant $\lambda^0 \geq 0$, that satisfy:

- (EX1) $(\lambda^0, \lambda(t)) \neq (0, 0)$ *for some (and hence every) t;*
- (EX2) $\dot{\lambda} = -\frac{\partial H^c}{\partial x}(\lambda(t), \lambda^0, \gamma(t), \eta(t))$ *for a.e.* $t \in \mathrm{Dom}(\eta)$*;*
- (EX3) $H(\lambda(t), \lambda^0, \gamma(t)) = H^c(\lambda(t), \lambda^0, \gamma(t), \eta(t)) = 0$ *for a.e.* $t \in \mathrm{Dom}(\eta)$;
- (EX4) $\langle \lambda(0), v \rangle = \lambda^0 \langle \nabla\tau(\gamma(0)), v \rangle$ *for all* $v \in \tilde{T}_{\gamma(0)}\mathcal{T}$. ♢

Properties (EX1), (EX2), (EX3), (EX4) are, respectively, the *nontriviality condition*, the *adjoint equation*, the *Hamiltonian minimization condition*, and the *transversality condition*. A pair (λ, λ^0) such that (EX2) holds is called an *adjoint vector for* (γ, η). An adjoint vector for which (EX3) holds is a *minimizing adjoint vector for* (γ, η).

The *maximum principle* says that, under some special conditions on the target set $\mathcal{T}$, *every optimal pair* (γ, η) *is extremal.* (This is true, for example, if $\mathcal{T}$ is a submanifold of Ω of class $\mathcal{C}^1$, or a closed convex subset of Ω, or a set which is locally equivalent to a convex set via a $\mathcal{C}^1$ diffeomorphism. More generally, it is true if the cone $\tilde{T}_{\gamma(0)}\mathcal{T}$ is convex—in which case it is of course a subspace, since $v \in \tilde{T}_{\gamma(0)}\mathcal{T} \Rightarrow -v \in \tilde{T}_{\gamma(0)}\mathcal{T}$—and is an approximating cone to $\mathcal{T}$ at $\gamma(0)$. Recall that a closed convex cone C is an *approximating cone* to a set S at a point $s \in S$ if there exist a neighborhood W of 0, and a continuous map $F : W \cap C \to S$, such that $F(v) = s + v + o(||v||)$ as $v \to 0$ via values in C.)

Definition 2.3. A *presynthesis* for $\boldsymbol{\mathcal{P}}$ is a subset Γ of $\mathrm{Adm}^{\mathcal{C}^{1,0}}(\tilde{f}, \mathcal{T})$ such that,

(PS) *Whenever* $(\gamma_1, \eta_1) \in \Gamma$, $(\gamma_2, \eta_2) \in \Gamma$, *and* $\gamma_1^- = \gamma_2^-$, *it follows that* $(\gamma_1, \eta_1) = (\gamma_2, \eta_2)$.

The set $\mathrm{Dom}(\Gamma) \stackrel{\mathrm{def}}{=} \{\gamma^- : (\gamma, \eta) \in \Gamma\}$ is called the *domain* of Γ. We say that Γ is a *presynthesis on* S if Γ is a presynthesis and $S = \mathrm{Dom}(\Gamma)$. ♢

Clearly, giving a presynthesis on S amounts to choosing, for each $x \in S$, a $(\gamma, \eta) \in \mathrm{Adm}^{\mathcal{C}^{1,0}}(\tilde{f}, \mathcal{T})$ such that $\gamma^- = x$.

If Γ is a presynthesis such that $\mathrm{Dom}(\Gamma)$ consists of all points that can be steered in finite time to a point of $\mathcal{T}$ by means of an admissible pair, then we say that Γ is *total*.

Given a presynthesis Γ and a point $x \in \mathrm{Dom}(\Gamma)$, we will always use $(\gamma_x^\Gamma, \eta_x^\Gamma)$—or, simply, (γ_x, η_x) when the meaning of Γ is clear from the context—to denote the unique pair $(\gamma, \eta) \in \Gamma$ such that $\gamma^- = x$, and will write $T_x = \min \mathrm{Dom}(\gamma_x)$, so $T_x \leq 0$ and $\mathrm{Dom}(\gamma_x) = \mathrm{Dom}(\eta_x) = [T_x, 0]$.

Definition 2.4. A presynthesis is *memoryless* if whenever $x \in S$ and $t \in \mathrm{Dom}(\eta_x)$ it follows that $y = \gamma_x(t)$ belongs to S and η_y is the restriction of η_x to the interval $[t, 0]$. A *synthesis* is a memoryless presynthesis. ♢

Definition 2.5. If each pair of a presynthesis Γ is optimal (resp. extremal) then we say that Γ is *optimal* (resp. *extremal*). $\diamondsuit$

In particular, a presynthesis Γ is a total optimal presynthesis if and only if $V_\Gamma \equiv V_{\tilde{f},\mathcal{T}}$.

As explained in the introduction, we wish to propose a definition of "regular synthesis" that will give rise to a sufficiency theorem. The theorem will say that a total extremal regular synthesis that satisfies the appropriate boundary conditions is optimal. We will actually define the more general notion of "regular presynthesis," and prove the sufficiency theorem in this broader setting, so the desired optimality result will be true in particular when the presynthesis is a true synthesis.

The definition will say that a "regular presynthesis" is a presynthesis that satisfies suitable regularity assumptions. So we begin by describing these regularity assumptions in detail.

Definition 2.6. We call a function $V : \Omega \to I\!R \cup \{-\infty, +\infty\}$ *weakly upper semicontinuous* (w.u.s.c) at a point $x \in \Omega$ if $\liminf_{y\to x} \limsup_{z\to y} V(z) \le V(x)$. $\diamondsuit$

Equivalently, V is w.u.s.c. at x if there exist sequences $\{x_j\}$, $\{\varepsilon_j\}$ and $\{\delta_j\}$ such that $x_j \in \Omega$, $\varepsilon_j > 0$, and $\delta_j > 0$ for all j, and $x_j \to x$ and $\delta_j \to 0$ as $j \to \infty$, having the property that $V(y) \le V(x) + \delta_j$ whenever $\|y - x_j\| < \varepsilon_j$.

Definition 2.7. Suppose we are given a Lipschitz-continuous vector field X on Ω and a function $V : \Omega \to I\!R \cup \{-\infty, +\infty\}$. We say that V has the *NDJ* ("no downward jumps") *property along X* if

(NDJ) *if $[a, b] \ni t \to \gamma(t) \in \Omega$ is an integral curve of X, then $\liminf_{h\downarrow 0} V(\gamma(t - h)) \le V(\gamma(t))$ for every $t \in]a, b]$.* $\diamondsuit$

Definition 2.8. We say that V satisfies the *weak continuity conditions* for the control system (2.1) if V is lower semicontinuous and weakly upper semicontinuous, and has the NDJ property along the vector field $x \to f(x, u)$ for every $u \in U$. $\diamondsuit$

We now define the "weak differentiability conditions" for Γ. In this definition, we let $\rho_{\tilde{f},\Gamma,\bar{x},T}(v)$ be the integrable $I\!R^{n+1}$-valued function on $[T, 0]$ given by

$$\rho_{\tilde{f},\Gamma,\bar{x},T}(v)(t) = \tilde{f}_{\eta_{\bar{x}+v}}(\gamma_{\bar{x}}(t), t) - \tilde{f}_{\eta_{\bar{x}}}(\gamma_{\bar{x}}(t), t)\,,$$

and use the symbol $\rho_{\tilde{f},\Gamma,\bar{x},T}$ to denote the correspondence $v \to \rho_{\tilde{f},\Gamma,\bar{x},T}(v)$, regarded as a map into the space $\mathrm{Bor}([T, 0], I\!R^{n+1})$ of $I\!R^{n+1}$-valued Borel measures on $[T, 0]$. (Recall that $\mathrm{Bor}([T, 0], I\!R^{n+1})$ is the dual of the space $C^0([T, 0], I\!R^{n+1})$ of continuous $I\!R^{n+1}$-valued functions on $[T, 0]$.)

Definition 2.9. A presynthesis $\Gamma = \{(\gamma_x, \eta_x) : x \in S\}$ is said to be *(f, L)–differentiable* at a point $\bar{x} \in \Omega$ if there exist: (a) a neighborhood N of $\bar{x}$ in Ω such that $N \subseteq \mathrm{Dom}(\Gamma)$, (b) an interval $[T, 0]$ such that $\mathrm{Dom}(\eta_x) \subseteq [T, 0]$ for all $x \in N$, and (c) an $\bar{\varepsilon} > 0$, for which

(DC1) there are integrable functions $\psi_\varepsilon : [T, 0] \to I\!R$, for $\varepsilon \in]0, \bar{\varepsilon}]$, with the property that $\lim_{\varepsilon\to 0} \int_T^0 \psi_\varepsilon(t)\, dt = 0$, for which the inequality

$$\left\| \tilde{f}_{\eta_x}(y, t) - \tilde{f}_{\eta_x}(\gamma_x(t), t) - D\tilde{f}_{\eta_{\bar{x}}}(\gamma_{\bar{x}}(t), t)\cdot(y - \gamma_{\bar{x}}(t)) \right\| \le \psi_\varepsilon(t).(||y - \gamma_{\bar{x}}(t)|| + ||x - \bar{x}||) \quad (2.8)$$

holds for every $y \in \Omega$, $x \in N$, $t \in [T, 0]$ such that $\|y - \gamma_{\bar{x}}(t)\| \le \varepsilon$ and $||x - \bar{x}|| \le \varepsilon$;

(DC2) the map $\rho_{\tilde{f},\Gamma,\bar{x},T}$ is weak-differentiable at $v = 0$, in the sense that, for every continuous function $\alpha : [T, 0] \to I\!R$, the map $I\!R^n \ni v \to \int_T^0 \alpha(t).\rho_{\tilde{f},\Gamma,\bar{x},T}(v)(t)\, dt \in I\!R^{n+1}$ is differentiable at $v = 0$.* $\diamondsuit$

Definition 2.10. A subset A of Ω is *thin* if there exist A_1, A_2 such that $A = A_1 \cup A_2$, A_1 is a finite or countable union of connected $\mathcal{C}^1$ submanifolds of positive codimension, and $\mathcal{H}^{n-1}(A_2) = 0$, where $\mathcal{H}^k$ denotes k-dimensional Hausdorff measure. $\diamondsuit$

If S is a subset of Ω, we use $\overset{\circ}{S}$ to denote the interior of S in Ω.

Definition 2.11. If n, Ω, U, f, L, $\mathcal{T}$, τ are such that Assumptions (A1) – (A6) hold, Γ is a presynthesis for $\boldsymbol{\mathcal{P}}(\Omega, U, f, L, \mathcal{T}, \tau)$, and $S = \mathrm{Dom}(\Gamma)$, we say that Γ has the *interior approximation property* if

(IAP) *If $(\gamma, \eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f})$ is such that $\gamma(t) \in S$ for all $t \in \mathrm{Dom}(\gamma)$, then there exists a sequence $\{(\gamma_j, \eta_j)\}_{j=1}^{\infty}$ in $\mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f})$ such that*

- *(IAP.1) $\gamma_j(t) \in \overset{\circ}{S}$ for every $t \in \mathrm{Dom}(\gamma_j)$;*
- *(IAP.2) $V_\Gamma(\gamma_j^+) \to V_\Gamma(\gamma^+)$;*
- *(IAP.3) $\gamma_j^- \to \gamma^-$;*
- *(IAP.4) $\int_{\mathrm{Dom}(\gamma_j)} L(\gamma_j(t), \eta_j(t))\, dt \to \int_{\mathrm{Dom}(\gamma)} L(\gamma(t), \eta(t))\, dt$.* $\diamondsuit$

We are now ready to present our definition or "regular presynthesis":

Definition 2.12. Let n, Ω, U, f, L, $\mathcal{T}$, τ be such that Assumptions (A1) – (A6) hold. Let Γ be a presynthesis for $\boldsymbol{\mathcal{P}}(\Omega, U, f, L, \mathcal{T}, \tau)$. We say that Γ is *regular* if

(2.12.a) *the associated cost function V_Γ satisfies the weak continuity conditions*;

(2.12.b) *there exists a thin subset A of Ω such that Γ is (f, L)–differentiable at all points of* $\mathrm{Dom}(\Gamma)\backslash A$. $\diamondsuit$

With this definition, our main sufficiency theorem (a weaker version of which was already stated in [27], with only an outline of the proof) says:

Theorem 2.13. *Let n, Ω, U, f, L, $\mathcal{T}$, τ be such that Assumptions (A1) – (A6) hold. Let Γ be an extremal regular presynthesis for $\boldsymbol{\mathcal{P}}(\Omega, U, f, L, \mathcal{T}, \tau)$ and let $S = \mathrm{Dom}(\Gamma)$. Then:*

(2.13.a) The "dynamic programming inequality"

$$V_\Gamma(\gamma^-) \leq \int_{\mathrm{Dom}(\gamma)} L(\gamma(t), \eta(t))\, dt + V_\Gamma(\gamma^+) \tag{2.9}$$

holds for every $(\gamma, \eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f})$ such that γ is contained in $\overset{\circ}{S}$.

(2.13.b) If V_Γ satisfies the boundary condition

$$V_\Gamma(x) \leq \tau(x) \quad \text{for} \quad \text{every} \quad x \in \mathcal{T}\,, \tag{2.10}$$

then $V_\Gamma(\gamma^-) \leq J(\gamma, \eta)$ for all $(\gamma, \eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f}, \mathcal{T})$ such that γ is contained in $\overset{\circ}{S}$.

(2.13.c) If Γ has the interior approximation property, then (2.9) holds for all $(\gamma, \eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f})$ such that γ is contained in S.

(2.13.d) If Γ has the interior approximation property and satisfies (2.10) then $V_\Gamma(\gamma^-) \leq J(\gamma, \eta)$ for all $(\gamma, \eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f}, \mathcal{T})$ such that γ is contained in S.

(2.13.e) If Γ is total and satisfies (2.10) then Γ is optimal. $\diamondsuit$

PROOF. We first establish the relation between the differential of the cost function V_Γ and the adjoint fields of covectors along the extremal trajectories of the presynthesis. This result is important in its own right, so we state it as a separate theorem:

Theorem 2.14. *Let n, Ω, U, f, L, $\mathcal{T}$, τ be such that Assumptions (A1) – (A6) hold, and let $\Gamma = \{(\gamma_x, \eta_x) : x \in S\}$ be a presynthesis. Let $\bar{x} \in S$ be such that Γ is (f, L)–differentiable at $\bar{x}$. Then*

(2.14.a) the function V_Γ is differentiable at $\bar{x}$;

(2.14.b) if the pair $(\gamma_{\bar{x}}, \eta_{\bar{x}})$ is extremal, then every minimizing adjoint vector (λ, λ^0) for $(\gamma_{\bar{x}}, \eta_{\bar{x}})$ satisfies the identity $\lambda^0(\nabla V_\Gamma)(\bar{x}) = \lambda(T_{\bar{x}})$. In particular, $\lambda^0 \neq 0$, and (λ, λ^0) is unique up to multiplication by a positive constant.

PROOF OF THEOREM 2.14. We apply Theorem A.11 in Appendix A, taking P to be $\mathbb{R}^n$ and letting $f_x = f_{\eta_x}$, $L_x = L_{\eta_x}$, $\tilde{f}_x = \tilde{f}_{\eta_x}$. Then, if $y \in \Omega$ and $x \in N$, we have $f_x(y,t) = f(y, \eta_x(t))$, $L_x(y,t) = L(y, \eta_x(t))$, $f_x(y,t) = f(y, \eta_x(t))$ if $t \in [T_x, 0]$, and $f_x(y,t) = 0$, $L_x(y,t) = 0$, $\tilde{f}_x(y,t) = 0$ if $T \notin [T_x, 0]$.

We let $\hat{V}_\Gamma(x) = \int_{T_x}^0 L(\gamma_x(t), \eta_x(t))\, dt$, so $\hat{V}_\Gamma$ is the "Lagrangian part" of our cost functional. We extend the curve γ_x to the interval $[T, 0]$ by letting $\gamma_x(t) = x$ for $T \leq t < T_x$, so γ_x still is an integral curve of f_{η_x} on $[T, 0]$, and $\hat{V}_\Gamma(x) = \int_T^0 L_x(\gamma_x(t), t)\, dt$. Also, we extend the adjoint vector λ to $[T, 0]$ by letting $\lambda(t) = \lambda(T_x)$ for $T \leq t < T_x$. We then define $H_x^c(p, p^0, y, t) = \langle p, f_x(y,t)\rangle + p^0 L_x(y,t)$. Then the adjoint equation $\dot{\lambda}(t) = -\nabla_y H_{\bar{x}}^c(\lambda(t), \lambda^0, \gamma_{\bar{x}}(t), t)$ holds on $[T, 0]$. Moreover, it is clear that

$$H_x^c(\lambda(t), \lambda^0, \gamma_{\bar{x}}(t), t) \geq H_{\bar{x}}^c(\lambda(t), \lambda^0, \gamma_{\bar{x}}(t), t) = 0 \ \text{ for a. e. } t \in [T, 0]\,. \tag{2.11}$$

The map $p \to \bar{x}_p$ of Theorem A.11 is the identity map, which is certainly differentiable at $\bar{x}$. So Theorem A.11 tells us that the "endpoint map" $x \to \mathcal{E}(x) \stackrel{\text{def}}{=} \gamma_x(0)$ and the function $\hat{V}_\Gamma$ are differentiable at $\bar{x}$, and the identity

$$\begin{aligned} &\lambda(0).D\mathcal{E}(\bar{x}).v + \lambda^0 \nabla \hat{V}_\Gamma(\bar{x}).v - \lambda(T_{\bar{x}}) \cdot v \\ &\qquad = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \int_T^0 \Big(H_{\bar{x}+\varepsilon v}^c(\lambda(t), \lambda^0, \gamma_{\bar{x}}(t), t) - H_{\bar{x}}^c(\lambda(t), \lambda^0, \gamma_{\bar{x}}(t), t) \Big)\, dt \end{aligned} \tag{2.12}$$

holds for every vector v if (λ, λ^0) is an adjoint vector for $(\gamma_{\bar{x}}, \eta_{\bar{x}})$. Then (2.11) tells us that the right-hand side of inequality (2.12) is nonnegative. So

$$\lambda(0).D\mathcal{E}(\bar{x}).v + \lambda^0 \nabla \hat{V}_\Gamma(\bar{x}).v - \lambda(T_{\bar{x}}) \cdot v \geq 0 \qquad \text{for all } v \in \mathbb{R}^n\,. \tag{2.13}$$

Given v, the curve $r \to \gamma_{\bar{x}+rv}(0)$ is defined for r in a neighborhood of 0 in $\mathbb{R}$, and is differentiable at 0 with derivative $D\mathcal{E}(\bar{x}).v$. On the other hand, this curve is contained in $\mathcal{T}$, so the vector $w = D\mathcal{E}(\bar{x}).v$ belongs to $T_{\gamma_{\bar{x}}(0)}\mathcal{T}$. It then follows from the transversality condition (EX4) that $\langle \lambda(0), w\rangle = \lambda^0 \langle \nabla \tau(\gamma(0)), w\rangle$. Therefore

$$\lambda^0 \Big(\nabla \tau(\gamma(0)).D\mathcal{E}(\bar{x}).v + \nabla \hat{V}_\Gamma(\bar{x}).v \Big) - \lambda(T_{\bar{x}}) \cdot v \geq 0 \qquad \text{for all } v \in \mathbb{R}^n\,. \tag{2.14}$$

Since the left-hand side of (2.14) is linear in v, it has to vanish for all v. So

$$\lambda^0 \nabla V_\Gamma(\bar{x}).v = \lambda^0 \Big(\nabla \tau(\gamma(0)).D\mathcal{E}(\bar{x}).v + \nabla \hat{V}_\Gamma(\bar{x}).v \Big) = \lambda(T_{\bar{x}}) \cdot v \qquad \text{for all } v \in \mathbb{R}^n\,. \tag{2.15}$$

Therefore $\lambda^0 \nabla V_\Gamma(\bar{x}) = \lambda(T_{\bar{x}})$.

In particular, λ^0 cannot vanish, because $\lambda^0 = 0$ would imply $\lambda(T_{\bar{x}}) = 0$, contradicting (EX1). So we may, after multiplication by a positive constant, assume that $\lambda^0 = 1$, and then $\nabla V_\Gamma(\bar{x}) = \lambda(T_{\bar{x}})$. This proves, in particular, the uniqueness of $(\lambda(\cdot), \lambda^0)$ up to multiplication by a positive constant, since the Cauchy problem for the adjoint equation has unique solutions. ♢

CONTINUATION OF THE PROOF OF THEOREM 2.13. We first prove (2.14.a). For any given $x \in S \setminus A$, let λ_x be the unique λ such that $(\lambda, 1)$ is a minimizing adjoint vector. (The existence and uniqueness of λ_x follows from Theorem 2.14.) The minimization condition (EX4) of the Pontryagin Maximum Principle then implies that, for any given $u \in U$, the inequality

$$\langle \lambda_x(t), f(\gamma_x(t), u)\rangle + L(\gamma_x(t), u) \geq 0$$

holds for almost all $t \in [T_x, 0]$. Since $t \to \langle \lambda(t), f(\gamma_x(t), u)\rangle + L(\gamma_x(t), u)$ is continuous on $[T_x, 0]$, the inequality is valid at time T_x. In view of Theorem 2.14,

$$\langle \nabla V_\Gamma(x), f(x, u)\rangle + L(x, u) \geq 0\,. \tag{2.16}$$

Now consider an arbitrary $(\gamma, \eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f})$, with domain $[T, 0]$, such that γ is contained in $\mathring{S}$. We want to prove that (2.9) holds. It follows from Lemma 4.1 of [25] that there exists a sequence of admissible pairs (γ_j, η_j), with domain $[T, 0]$, such that $\gamma_j(0) = \gamma(0)$, η_j is piecewise constant, $\gamma_j \to \gamma$ uniformly on $[T, 0]$, and $\int_T^0 L(\gamma_j(t), \eta_j(t))\,dt \to \int_T^0 L(\gamma(t), \eta(t))\,dt$.

Since γ is contained in $\mathring{S}$, we can clearly assume that the γ_j are also contained in $\mathring{S}$. Since V_Γ is lower semicontinuous, it follows that, if $V_\Gamma(\gamma_j(T)) + \int_T^0 L(\gamma_j(t), \eta_j(t))\,dt \leq V_\Gamma(\gamma(0))$ for each j, then $V_\Gamma(\gamma(T)) + \int_T^0 L(\gamma(t), \eta(t))\,dt \leq V_\Gamma(\gamma(0))$. Hence it suffices to assume that the control η is piecewise constant. Moreover, if (2.9) holds for two pairs $(\gamma_i, \eta_i) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f})$ such that $\mathrm{Dom}(\gamma_1) = [a, b]$, $\mathrm{Dom}(\gamma_2) = [b, c]$, and $\gamma_1(b) = \gamma_2(b)$ for some a, b, c, then it also holds for the concatenated pair (γ, η) defined in an obvious way. It therefore suffices to assume that η is in fact constant, so we make this assumption from now on.

Let $u \in U$ be such that $\eta(t) = u$ for $t \in [T, 0]$. For each $y \in \Omega$, let $t \to \zeta^y(t)$ be the trajectory of η that satisfies the terminal condition $\zeta^y(0) = y$. Let W be an open subset of Ω such that $\gamma(0) \in W$, $\zeta^y(T)$ is defined for all $y \in W$, and ζ^y is contained in $\mathring{S}$ whenever $y \in W$. We will show that

(#) *for almost every $y \in W$ the set $B^y = \{t : \zeta^y(t) \in A\}$ is finite or countable.*

Since A is thin, we can write $A = A_1 \cup A_2$, where $A_1 = \bigcup_j M_j$ and $\{M_j\}_{j \in J}$ is a finite or countable family of connected submanifolds of Ω of codimension $d_j > 0$, while $\mathcal{H}^{n-1}(A_2) = 0$. After replacing each M_j by a finite or countable family of open submanifolds of M_j, we may assume that the M_j are embedded.

Define $\tilde{W} = W \times]T, 0[$, and let Φ be the map $\tilde{W} \ni (y, t) \to \zeta^y(t) \in \Omega$. Then Φ is a $\mathcal{C}^1$ submersion, because for each fixed t the partial map $y \to \Phi(y, t)$ is a $\mathcal{C}^1$ diffeomorphism. Therefore each set $\tilde{M}_j = \Phi^{-1}(M_j)$ is an embedded submanifold of $\tilde{W}$ of codimension d_j, and $\mathcal{H}^n(\Phi^{-1}(A_2)) = 0$.

Let π be the projection $\tilde{W} \ni (y, t) \to y \in W$. Let S_j be the set of points $s \in \tilde{M}_j$ such that the restriction $\pi \lceil \tilde{M}_j$ of π to $\tilde{M}_j$ is not regular at s. (Recall that, if S_1, S_2 are $\mathcal{C}^1$ manifolds, then a $\mathcal{C}^1$ map $F : S_1 \to S_2$ is *regular* at $s \in S_1$ if the differential $d\pi(s)$ maps the tangent space $T_s S_1$ surjectively onto $T_{F(s)} S_2$.) Then $\mathcal{H}^n(\pi(\Phi^{-1}(A_2))) = 0$, and $\mathcal{H}^n(\pi(S_j)) = 0$ for each j, by Sard's Theorem. So the "bad" set $\mathcal{B} = \pi(\Phi^{-1}(A_2)) \cup \left(\bigcup_j \pi(S_j)\right)$ has Lebesgue measure zero.

Now let $y \in W \backslash \mathcal{B}$. Then $\zeta^y(t) \notin A_2$ if $T < t < 0$. (Otherwise, if $\zeta^y(t) \in A_2$, $T < t < 0$, it would follow that $(y, t) \in \Phi^{-1}(A_2)$, so $y \in \pi(\Phi^{-1}(A_2))$, and then $y \in \mathcal{B}$.) Also, we claim that the set $\{t : \zeta^y(t) \in A_1\}$ is at most countable. To see this, it suffices to show that, for each j, the set $E_j = \{t : T < t < 0\,, \zeta^y(t) \in M_j\}$ is finite or countable. Fix j, and suppose $t \in E_j$. The manifold $\tilde{M}_j$ has codimension d_j in $\tilde{W}$. Since $d_j > 0$, the dimension ν_j of $\tilde{M}_j$ is $\leq n$. Moreover, $d\pi(y, t)$ maps

$T_{y,t}\tilde{M}_j$ surjectively onto $I\!R^n$ (the tangent space to Ω at y), because $y \notin \mathcal{B}$. So $\nu_j = n$, and $d\pi(y,t)$ is injective on $T_{y,t}\tilde{M}_j$. Since $d\pi(y,t)\left(\frac{\partial}{\partial t}\right)$ is obviously equal to 0, it follows that $\frac{\partial}{\partial t} \notin T_{y,t}\tilde{M}_j$. Since $\tilde{M}_j$ is embedded, $(y,t') \notin \tilde{M}_j$ if $0 < |t'-t| \leq \varepsilon$, provided that ε is sufficiently small. Therefore t is an isolated point of E_j. So E_j is a discrete subset of the open interval $]T,0[$. Therefore E_j is finite or countable, as desired.

We have therefore proved that, if $y \in W\backslash\mathcal{B}$, then $\zeta^y(t) \notin A_2$ for all $t \in [T,0]$, and the set $\{t : \zeta^y(t) \in A_1\}$ is at most countable. Since $A = A_1 \cup A_2$, and $\mathcal{B}$ has measure zero, (#) follows.

In view of the weak upper semicontinuity of V_Γ, there exist x_j, $\varepsilon_j > 0$, $\delta_j > 0$ such that $x_j \to \gamma(0)$ and $\delta_j \to 0$ as $j \to \infty$, and $V_\Gamma(y) \leq V_\Gamma(\gamma(0)) + \delta_j$ whenever $\|y - x_j\| \leq \varepsilon_j$. Choose the ε_j so that $\|y - x_j\| \leq \varepsilon_j$ implies $y \in W$. Then, by (#), we can choose for each j a y_j such that $\|y_j - x_j\| \leq \varepsilon_j$ and the bad set B^{y_j} is at most finite or countable. Since $\zeta^{y_j}(t) \in \overset{\circ}{S}$ for every $t \in [T,0]$, (2.16) implies that the function $[T,0] \ni t \to \varphi_j(t) = V_\Gamma(\zeta^{y_j}(t)) + \int_T^t L(\zeta^{y_j}(s),\eta(s))ds \in I\!R$ is differentiable at every t, except for a finite or countable number of values, with a nonnegative derivative. Therefore φ_j satisfies assumption (B.1.a) of Lemma B.1 (Appendix B). Moreover, the lower semicontinuity of V_Γ ensures assumption (B.1.b) of Lemma B.1 for φ_j, and the (NDJ) condition ensures assumption (B.1.c). Hence we can apply Lemma B.1 to φ_j, and conclude that $\varphi_j(T) \leq \varphi_j(0)$. So

$$V_\Gamma(\zeta^{y_j}(T)) \leq \int_T^0 L(\zeta^{y_j}(t),\eta(t))dt + V_\Gamma(y_j) \leq \int_T^0 L(\zeta^{y_j}(t),\eta(t))dt + V_\Gamma(\gamma(0)) + \delta_j.$$

If we let $j \to \infty$ we find, using the lower semicontinuity of V_Γ, that (2.9) holds for (γ,η), as desired. So (2.9) holds when η is constant and, as explained earlier, this completes the proof of (2.14.a).

Statements (2.14.b), (2.14.c), and (2.14.d) are immediate consequences of (2.14.a).

To prove (2.14.e), we assume that Γ is total and prove that it satisfies the interior approximation property. So now S is the set of all points of Ω that can be steered to a point of $\mathcal{T}$ in finite time. Let $(\gamma,\eta) \in \mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f})$ be a trajectory-control pair such that γ is entirely contained in S. Then $\gamma(0) \in S$, so $\gamma(0) \in \mathrm{Dom}(\Gamma)$ and then $V_\Gamma(\gamma(0)) < +\infty$. The weak upper semicontinuity of V_Γ implies that there exist x_j, $\varepsilon_j > 0$, $\delta_j > 0$ such that $x_j \to \gamma(0)$ and $\delta_j \to 0$ as $j \to \infty$, and $V_\Gamma(y) \leq V_\Gamma(\gamma(0)) + \delta_j$ whenever $\|y - x_j\| \leq \varepsilon_j$. It then follows that $V_\Gamma(x_j) \to V_\Gamma(\gamma(0))$, since V_Γ is lower semicontinuous. Let $\mathrm{Dom}(\gamma) = [T,0]$. Let γ_j be the maximally defined trajectory for the control η that satisfies $\gamma_j(0) = x_j$. Then γ_j is defined on $[T,0]$ if j is large enough, and conditions (IAP.2,3,4) hold with $\eta_j = \eta$. We may then assume that γ_j is defined on $[T,0]$ for all j. To prove that (IAP.1) holds as well, we observe that $y \in S$ whenever $\|y - x_j\| \leq \varepsilon_j$, because in that case $V_\Gamma(y) < +\infty$, since $V_\Gamma(y) \leq V_\Gamma(\gamma(0)) + \delta_j$, and $V_\Gamma(\gamma(0)) < +\infty$ because $\gamma(0) \in S$. Let ζ^y be the trajectory for the control η that satisfies $\zeta^y(0) = y$. Then for each j there exists a $\tilde{\varepsilon}_j$ such that $0 < \tilde{\varepsilon}_j \leq \varepsilon_j$, with the property that the curve ζ^y is defined on $[T,0]$ for all y such that $\|y - x_j\| \leq \tilde{\varepsilon}_j$. Moreover, it is clear that y is reachable from $\zeta^y(t)$ for every $t \in [T,0]$, so $\zeta^y([T,0]) \subseteq S$ if $\|y - x_j\| \leq \tilde{\varepsilon}_j$, because $y \in S$. Now, if $T \leq t \leq 0$, the map $y \to \zeta^y(t)$ is a $\mathcal{C}^1$ diffeomorphism from the neighborhood $\{x : \|x - x_j\| < \tilde{\varepsilon}_j\}$ of x_j onto a neighborhood $N_j(t)$ of $\gamma_j(t)$. Clearly, $N_j(t) \subseteq S$, so $\gamma_j(t) \in \overset{\circ}{S}$. This completes the proof of (2.14.e). $\diamondsuit$

Remark 2.15. It is not hard to prove that *the function V_Γ is weakly upper semicontinuous everywhere if and only if it is weakly upper semicontinuous at all the points of the target $\mathcal{T}$.* $\diamondsuit$

§3. The Fuller Phenomenon

In this section we show that the famous control system considered by Fuller in [12]—and extensively studied by other authors, e.g. Zelikin-Borisov [28]—satisfies the hypotheses of Theorem 2.13.

Consider the system:

$$\dot{x}_1 = x_2\,, \qquad \dot{x}_2 = u\,, \qquad |u| \leq 1\,, \tag{3.1}$$

and the family $\hat{\boldsymbol{\mathcal{P}}} = \{\hat{\mathcal{P}}_x\}_{x\in\mathbb{R}^2}$ of minimization problems:

$$\hat{\mathcal{P}}_x : \qquad \text{minimize} \int_0^{+\infty} x_1^2(t)\,dt \ \text{ subject to } \ (x_1(T), x_2(T)) = x\,. \tag{3.2}$$

We proceed to describe the optimal synthesis for this problem, determined by Fuller, referring the reader to Zelikin-Borisov [28], Ch. 2, for further details. First define the "switching locus" $\zeta = \{(x_1, x_2) : |x_1| = Cx_2^2\,,\ x_1x_2 \leq 0\}$, where C is a constant whose precise definition will be given later. Write $\zeta = \zeta^+ \cup \zeta^- \cup \{(0,0)\}$, where $\zeta^\pm = \{(x_1, x_2) \in \zeta : \pm x_1 > 0\}$. Then the piecewise smooth curve ζ divides the plane into two regions A^- and A^+ consisting, respectively, of the points that lie above and below ζ. We define a discontinuous feedback control $k : \mathbb{R}^2 \to [-1, 1]$ by letting $k(x_1, x_2) = -1$ if $(x_1, x_2) \in A^- \cup \zeta^-$, $k(x_1, x_2) = 1$ if $(x_1, x_2) \in A^+ \cup \zeta^+$, and $k(0, 0) = 0$.

Then for every $x \in \mathbb{R}^2$ there exist a unique solution $\hat{\gamma}_x : [0, \infty[\to \mathbb{R}^2$ of the Cauchy problem $\dot{x}_1 = x_2$, $\dot{x}_2 = k(x_1, x_2)$, $(x_1(0), x_2(0)) = x$. The corresponding open-loop control $\hat{\eta}_x$ is related to $\hat{\gamma}_x$ by $\hat{\eta}_x(t) = +1$ when $\hat{\gamma}_x(t) \in A^+$, $\hat{\eta}_x(t) = -1$ when $\hat{\gamma}_x(t) \in A^-$, and $\hat{\eta}_x(t) = 0$ when $\hat{\gamma}_x(t) = 0$.

The switchings occur when $\hat{\gamma}(t) \in \zeta$. The curve $\hat{\gamma}_x$ reaches the origin in a finite time $\hat{T}_x$, and then stays there. Therefore $\hat{\eta}_x(t) = 0$ when $\hat{\gamma}_x(t) = 0$, which happens iff $t \geq \hat{T}_x$ (see Fig. 1.)

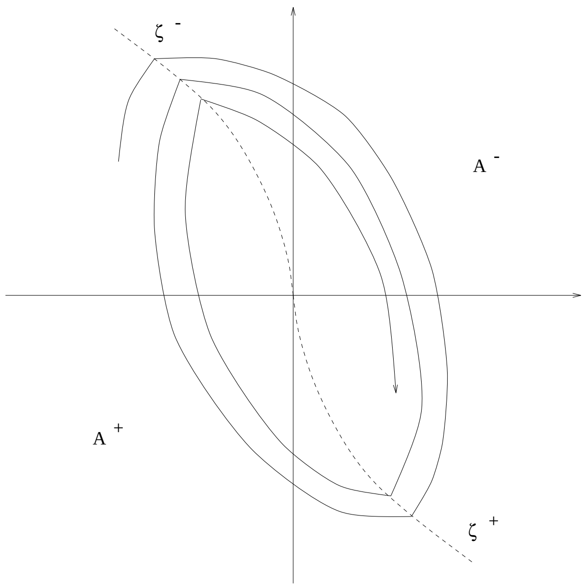

Figure 1

Therefore the admissible pairs $(\hat{\gamma}_x, \hat{\eta}_x)$ actually solve the family $\boldsymbol{\mathcal{P}} = \{\mathcal{P}_x\}_{x\in\mathbb{R}^2}$ of minimization problems

$$\mathcal{P}_x : \quad \text{minimize} \int_T^0 x_1^2(t)dt \text{ subject to } (x_1(T), x_2(T)) = x \text{ and } (x_1(0), x_2(0)) = (0,0)\,, \tag{3.3}$$

i.e., the problems of reaching the origin in finite time, subject to the dynamical constraints (3.1), and minimizing the cost given by (3.3).

So far, C was an arbitrary positive constant. It turns out that, if C is chosen to be equal to C^*, where C^* is the unique positive root of the polynomial $x^4 + x^2/12 - 1/18$, then the pairs $(\hat{\gamma}_x, \hat{\eta}_x)$ are extremal. (This is proved in [28]. In page 26, it is shown that C must be chosen so that $(1-2C)^{1/2}(1+2C)^{-1/2} = \mu$, where μ is such that, if $\sigma = \mu + \mu^{-1}$, then $\sigma^2 - 3\sigma - 6 = 0$. These equations easily imply $C^4 + C^2/12 - 1/18{=}0$.)

We now choose $C = C^*$, so in particular $0.4 < C < 0.5$. Then $\{\hat{\gamma}_x\}_{x\in I\!R^2}$ is a family of extremals reaching the target, and one would like to prove that the $\hat{\gamma}_x$ are optimal. The sufficiency theorems of [2] and [6] do not apply, since the controls $\hat{\eta}_x$ have an infinite number of switchings. For this reason, the optimality of the synthesis obtained by Fuller is usually established by other means. For example, one can prove optimality using the *dilation symmetry properties* of the problem. Precisely, there exists a one parameter family $\Delta = \{\Delta_\rho\}_{\rho>0}$ of maps from $I\!R^2$ to $I\!R^2$ such that, if (γ,η) is an admissible pair having cost c, and we let $\gamma_\rho(t) = \Delta_\rho(\gamma(\frac{t}{\rho}))$ and $\eta_\rho(t) = \eta(\frac{t}{\rho})$, then (γ_ρ, η_ρ) is an admissible pair having cost $\rho^5 c$. It then follows that the Δ_ρ map optimal trajectories to optimal trajectories.

The proof of optimality using the dilation symmetry is given, for example, in [28], and will not be repeated here. It will turn out, however, that the dilations Δ_ρ will play a role in our arguments, so we recall their definition and one simple identity that will be useful later:

(1) By definition, $\Delta_\rho(x_1,x_2) = (\rho^2 x_1, \rho x_2)$.
(2) It follows from (1), in particular, that $\Delta_\rho(x_1,x_2) - x = (\rho-1)(2x_1,x_2) + (\rho-1)^2(x_1,0)$ for all ρ, x_1, x_2.

In order to apply Theorem 2.13, we first make a time translation. We define $T_x = -\hat{T}_x$, and then let $\gamma_x(t) = \hat{\gamma}_x(t+\hat{T}_x)$, $\eta_x(t) = \hat{\eta}_x(t+\hat{T}_x)$, for $T_x \le t \le 0$. Then $\Gamma = \{(\gamma_x,\eta_x)\}_{x\in I\!R^2}$ is a memoryless family of admissible extremal pairs such that γ_x is defined on $[T_x, 0]$, $\gamma_x(T_x) = x$, and $\gamma_x(0) = 0$.

We want to prove that Γ is an optimal synthesis. For this purpose, we will verify the hypotheses of Theorem 2.13, and in particular the (f,L)–differentiability of Γ at every point $x \in I\!R^2\backslash\zeta$. Since it is very easy to check Property (DC1), we will only verify (DC2).

As in [28], page 26, we let $\mu = \sqrt{(1-2C)(1+2C)^{-1}}$. Then $0 < \mu < 1$.

Fix a point $\bar{x} = (\bar{x}_1, \bar{x}_2)$ belonging to $I\!R^2\backslash\{(0,0)\}$. Then there exist $s_0(\bar{x}), s_1(\bar{x})$ such that $T_{\bar{x}} = s_0(\bar{x}) < s_1(\bar{x}) \le \mu s_0(\bar{x}) < 0$, and $\eta_{\bar{x}} = \pm\sum_{k=0}^{+\infty}(-1)^k \chi_{I_k(\bar{x})}$, where (a) $I_0(\bar{x}) = [s_0(\bar{x}), s_1(\bar{x})]$, (b) $I_k(\bar{x}) = [\mu^{k-1}s_1(\bar{x}), \mu^k s_1(\bar{x})]$ for $k = 1,2,\ldots$, (c) χ_S is the indicator function of a set S, and (d) the $\pm$ sign depends on $\bar{x}$ as follows: it is $+$ if $\bar{x} \in A^+ \cup \zeta^+$, and $-$ if $\bar{x} \in A^- \cup \zeta^-$.

Let T be such that $T < 0$ and the interval $[T,0]$ contains $\mathrm{Dom}(\eta_x)$—i.e. $T \le T_x$—for x near $\bar{x}$. To prove (DC2), we must pick a point $\bar{x} \in I\!R^2 \setminus \zeta$ and a continuous function $\alpha : [T,0] \to I\!R$, and show that the map

$$I\!R^2 \ni x \longrightarrow J^\alpha(x) \stackrel{\mathrm{def}}{=} \int_T^0 \alpha(t).\tilde{f}_{\eta_x}(\gamma_{\bar{x}}(t),t)\,dt \in I\!R^3 , \tag{3.4}$$

is differentiable at $x = \bar{x}$. We will do this by showing that, for two linearly independent directions $v_1(x)$, $v_2(x)$, depending smoothly on x, the directional derivatives $D_{v_1(x)}J^\alpha$ and $D_{v_2(x)}J^\alpha$ exist and are continuous with respect to x. We choose $v_1(x)$ to be the direction of the curve $\rho \to \Delta_\rho(x)$ at $\rho = 1$, and take $v_2(x) = f(x, \eta_x(T_x))$. It is then easy to see that $v_1(x)$ and $v_2(x)$ are linearly independent for all $x \in A^+$. We will just determine the directional derivatives at $\bar{x}$, and the result will make it obvious that they depend continuously on $\bar{x}$, so differentiability will follow.

Let us assume, for simplicity, that $\bar{x} \in A^+$. (The case when $\bar{x} \in A^-$ is similar.) We first differentiate J^α in the direction of $v_1(\bar{x})$. If $\rho > 0$, we have (letting $\bar{s}_k = s_k(\bar{x})$, $I_k = [\bar{s}_k, \bar{s}_{k+1}]$, and $I_k^\rho = [\rho\bar{s}_k, \rho\bar{s}_{k+1}]$ for $k = 0, 1, \ldots$, and writing $\bar{x}_\rho = \Delta_\rho(\bar{x})$):

$$T_{\bar{x}_\rho} = \rho T_{\bar{x}}\,, \quad \eta_{\bar{x}_\rho} = \sum_{k=0}^{+\infty}(-1)^k \chi_{I_k^\rho}\,, \quad \text{and} \quad \tilde{f}_{\eta_{\bar{x}_\rho}}(\gamma_{\bar{x}}(t), t) - \tilde{f}_{\eta_{\bar{x}}}(\gamma_{\bar{x}}(t), t) = \begin{pmatrix} 0 \\ \eta_{\bar{x}_\rho}(t) - \eta_{\bar{x}}(t) \\ 0 \end{pmatrix} + R_\rho(t),$$

where $R_\rho(t)$ is a term arising because of the possibility that $T_{\bar{x}} \neq T_{\bar{x}_\rho}$.

Then a somewhat tedious but straightforward calculation shows that

$$\lim_{\rho\to 1} \tfrac{1}{\rho-1} \int_T^0 (\eta_{\bar{x}_\rho}(t) - \eta_{\bar{x}}(t))\alpha(t)\, dt = -s_0\alpha(s_0) - 2s_1 \textstyle\sum_{k=1}^{+\infty}(-1)^k \mu^{k-1}\alpha(\mu^{k-1}s_1)\,, \tag{3.6}$$

and

$$\lim_{\rho\to 1}(\rho - 1)^{-1} \int_T^0 \alpha(t).R_\rho(t)\, dt = -T_{\bar{x}}\alpha(T_{\bar{x}}).\begin{pmatrix} \bar{x}_2 \\ 0 \\ \bar{x}_1^2 \end{pmatrix}. \tag{3.7}$$

So the directional derivative of J^α at $x = \bar{x}$ in the direction of $v_1(\bar{x})$ exists and is given by the sums of the right-hand sides of (3.5) and (3.6), which depend continuously on $\bar{x}$, since s_0, s_1 and $T_{\bar{x}}$ (which is none other than $s_0(\bar{x})$) depend continuously on $\bar{x}$ as long as $\bar{x} \in A^+$.

A much simpler argument verifies the existence of the derivative of J^α in the direction of $v_2(\bar{x})$—i.e., of $f(\bar{x}, 1)$—whose value turns out to be the vector with components $\alpha(T_{\bar{x}})\bar{x}_2$, $\alpha(T_{\bar{x}})$, and $\alpha(T_{\bar{x}})\bar{x}_1^2$. Once again, this expression depends continuously on $\bar{x}$, so the proof of (DC2) is complete.

The other assumptions of Theorem 2.13 are easily checked. So the theorem applies, and the optimality of the synthesis described above follows. ♢

§4. Comparison with Other Definitions of Synthesis

The purpose of this section is to compare our concept of regular presynthesis with other definitions of synthesis that either have been proposed by other authors or are sufficiently natural to be worth considering as possible ways to define this notion. For simplicity, we will only consider the case of a single-point target set $\mathcal{T}$, and take this point to be the origin. (Both Boltyanskii and Brunovský, whose definitions will be discussed in detail, do the same, so this will facilitate the comparison.) Moreover, we will only consider nonnegative Lagrangians, so a trajectory containing a loop can never be optimal, and an optimal trajectory necessarily terminates when it hits the origin.

Perhaps the simplest conceivable notion of an "optimal synthesis" is that of an "optimal feedback control." In principle, one could define a "feedback control law" on a set S to be a map $v : S \to U$. If v is such that the vector field $S \ni x \to f(x, v(x)) \in I\!R^n$ is Lipschitz-continuous, and we assume for simplicity that S is open, then v gives rise to unique maximally defined trajectories $\hat{\gamma}_x$ starting at each $x \in S$, and corresponding open-loop controls $\hat{\eta}_x$, given by $\hat{\eta}_x(t) = v(\hat{\gamma}_x(t))$. If we make the additional assumption that every trajectory $\hat{\gamma}_x$ obtained in this way reaches the target at a time $\hat{T}_x$, define $T_x = -\hat{T}_x$, and let $\gamma_x(t) = \hat{\gamma}(t + \hat{T}_x)$ and $\eta_x(t) = \hat{\eta}(t + \hat{T}_x)$ for $T_x \le t \le 0$, then we will have constructed a synthesis $\Gamma = \{(\gamma_x, \eta_x)\}_{x\in S}$—in the sense of Definition 2.4—with domain S.

The main drawback of such a definition is that, as is well known, for most reasonable optimal control problems there does not exist an optimal feedback that renders the map $x \to f(x, v(x))$ Lipschitz-continuous, or even continuous. So it is absolutely essential to allow "discontinuous feedback laws," and when this is done one immediately runs into the problems of

(a) the lack of a truly satisfactory notion of solution

and

(b) the lack of good theorems guaranteeing existence of solutions.

Difficulty (b) can be handled in at least two ways, namely, by

(A) incorporating into the definition requirements that imply existence and uniqueness of solutions of the closed-loop equation (CLE) $\dot{x} = f(x, v(x))$;

(B) incorporating into the definition additional requirements that imply existence—but not necessarily uniqueness—of solutions of the CLE, and adding to the specification of v a definite prescription for choosing a solution when uniqueness fails, so that the "feedback law" is no longer the closed-loop control v alone, but the *pair* (v, Γ), where $\Gamma = \{(\gamma_x, \eta_x)\}$ is a family that selects one solution of the CLE for each initial condition x.

Moreover, whether we choose (A) or (B), one has to be precise about the concept of "solution." Here we will consider three such concepts, namely,

(1) classical solutions (i. e., absolutely continuous curves $t \to x(t)$ having the property that the equality $\dot{x}(t) = f(x(t), v(x(t)))$ holds for almost every t),

(2) Filippov solutions (cf. [9]),

(3) "CLSS solutions," that is, the "feedback solutions" defined by Clarke, Ledyaev, Subbotin and Sontag in [8].

The examples of §5 will show that the concepts of CLSS solution and Filippov solution are not adequate. Indeed, Example 5.4 shows an optimal synthesis whose trajectories are not Filippov solutions of the optimal closed-loop equation. In this example, the optimal trajectories are CLSS solutions, but the optimal closed-loop equation has many other CLSS solutions that are not optimal (cf. Remark 5.5). In Example 5.6, we exhibit an optimal control problem whose optimal trajectories are not CLSS solutions of the optimal closed-loop equation.

This leaves the notion of classical solution as the only viable candidate for the concept of solution to be used in the definition of optimal synthesis. Example 5.3 shows that it may happen that the set of classical solutions of the optimal closed-loop equation contains, in addition to all the optimal trajectories, some other arcs that are not optimal.

It then follows that (A) is not an adequate way of handling difficulty (b), and we are left with (B). This was indeed the strategy followed by Boltyanskii in [2], Brunovský in [6], and Sussmann in [16]. In all these cases, an "optimal synthesis" is defined by first specifying a—not necessarily continuous—feedback control law $x \to v(x)$. The feedback v is supposed to be "piecewise smooth" in a technical sense that, among other things, guarantees the existence of trajectories for every initial condition. Uniqueness is not assumed, but the specification of v is supplemented with a prescription for selecting a trajectory when non-uniqueness occurs. So, a "synthesis" in the sense of [2], [6], and [16] is more than just an optimal feedback: it is really a pair (v, Γ) of the kind discussed in (B).

We now review the three concepts of regular synthesis proposed in [2], [6], and [16], starting with Brunovský's definition as stated in [6], and then explaining how to modify this idea to obtain the alternative formulations suggested by Boltyanskii in [2] and Sussmann in [16]. Since the notations used in [2], [6], and [16] are different, we will give a unified account of the three definitions using a single set of symbols, and indicating which notations of [2] and [6] they correspond to. Furthermore, our accounts of [2] and [6] will be slightly modified versions of the text of the published papers, correcting what we believe are some minor imprecisions or typographical errors, and introducing some additional notations of our own for extra clarity.

Both Boltyanskii and Brunovský work with a system defined on an open subset Ω of a finite-dimensional real space (called X in [2], equal to $\mathbb{R}^n$ in [6]). Both assume that the dynamical behavior of the controlled system is given by a law $\dot{x} = f(x, u)$, where U is a subset of $\mathbb{R}^m$, and the maps f and L are of class C^1, in the sense that they can be extended to maps of class C^1 on an open subset of $\Omega \times I\!R^m$ that contains $\Omega \times clos_{I\!R^m}(U)$. Both assume that the synthesis is defined on an open subset S of Ω (called V in [2], G in [6]).

Brunovský's definition uses the concept of a stratification, so we review this notion first.

Definition 4.1. Given a $k \in \{1, 2, \ldots\} \cup \{+\infty, \omega\}$, and a manifold M of class $\mathcal{C}^k$, a $\mathcal{C}^k$ *stratification in* M of a subset S of M is a partition $\mathcal{P}$ of S into nonempty connected embedded submanifolds of M of class $\mathcal{C}^k$, such that $\mathcal{P}$ is locally finite in M (i.e., every compact subset of M intersects finitely many members of $\mathcal{P}$) and the following "frontier axiom" holds:

(FA) *if* $P_1, P_2 \in \mathcal{P}$, $P_1 \neq P_2$ *(so that* $P_1 \cap P_2 = \emptyset$*), and* $P_1 \cap clos_M(P_2) \neq \emptyset$*, then* $P_1 \subseteq clos_M(P_2)$ *and* $dim(P_1) < dim(P_2)$.

A $\mathcal{C}^k$*-stratified subset of* M is a pair $(S, \mathcal{P})$ having the property that $S \subseteq M$ and $\mathcal{P}$ is a $\mathcal{C}^k$ stratification of S in M. $\diamondsuit$

We are now ready to present Brunovský's definition, inserting some comments of our own—labeled "BP&HS"—in square brackets.

Definition 4.2. Let $S \subseteq \Omega$ be an open subset such that the origin belongs to S. A *Brunovský regular synthesis on* S *for the control problem (2.1), (2.2), with target the origin*, is a 6–tuple $\Xi = (\mathcal{P}, \mathcal{P}_1, \mathcal{P}_2, \Pi, \Sigma, v)$ (called $(\mathcal{S}, \mathcal{S}_1, \mathcal{S}_2, \Pi, \Sigma, v)$ in [6]) such that

(Br.1) $\mathcal{P}$ *is the union of* $\{\{0\}\}$ *and a locally finite (in* S*) partition* $\mathcal{P}\backslash\{\{0\}\}$ *of* $S\backslash\{0\}$ *into nonempty connected embedded* $\mathcal{C}^1$ *submanifolds of* S *(called "cells"),*

(Br.2) $\mathcal{P}\backslash\{\{0\}\}$ *is the disjoint union of* $\mathcal{P}_1$ *(the set of "type I cells") and* $\mathcal{P}_2$ *(the set of "type II cells"),*

(Br.3) $v : S \to U$ *(the "closed loop control"),* $\Pi : \mathcal{P}_1 \to \mathcal{P}$*, and* $\Sigma : \mathcal{P}_2 \to \mathcal{P}_1$ *are maps such that the following properties are satisfied:*

(Br.3.A) *If* $\mathcal{P}' = \{P \in \mathcal{P} : \dim\, P < n\,,\, P \neq \{0\}\}$*, and* $S' = \bigcup\{P \in \mathcal{P}'\}$*, then* $(S', \mathcal{P}')$ *is a stratified subset of* S *of dimension* $< n$. [BP&HS: This turns out to be equivalent to the much simpler statement that $\mathcal{P}$ *is a* $\mathcal{C}^1$ *stratification of* S.]

(Br.3.B) *The function* v *is of class* $\mathcal{C}^1$ *on each cell.*

(Br.3.C) *If* $P_1 \in \mathcal{P}_1$ *then* $f(x, v(x)) \in T_x P_1$ *(the tangent space to* P_1 *at* x*) for every* $x \in P_1$*. In addition, for each* $x \in P_1$*, if we let* ξ_x *be the maximally defined solution of the initial value problem*

$$\dot{\xi} = f(\xi, v(\xi))\,,\ \ \xi(0) = x\,,\ \ \xi \in P_1\,, \tag{4.1}$$

and define $t_x = \sup\, \mathrm{Dom}(\xi_x)$*, then the limit* $\xi_x(t_x-) \stackrel{\mathrm{def}}{=} \lim_{t \uparrow t_x} \xi_x(t)$ *exists (in* Ω*) and belongs to* $\Pi(P_1)$. [BP&HS: The limit $\xi_x(t_x-)$ cannot belong to P_1. It then follows that $P_1 \neq \Pi(P_1) \subseteq clos(P_1)$ *and, moreover,* $\dim(\Pi(P_1)) < \dim(P_1)$ *for all* $P_1 \in \mathcal{P}_1$.]

(Br.3.D) *If* $P_2 \in \mathcal{P}_2$*, then the control* v *is continuous on* $P_2 \cup \Sigma(P_2)$*, and for each* $x \in P_2$ *there exists a unique curve* $\xi_x : [0, t_x[\, \to \Omega$ *such that the restriction* $\xi_x \lceil\,]0, t_x[$ *is a maximally defined integral curve of the vector field* $f(\cdot, v(\cdot))$ *on* $\Sigma(P_2)$*, and* $\xi_x(0) = x$. [BP&HS: This implies that $\Sigma(P_2) \neq P_2$, $P_2 \subseteq clos(\Sigma(P_2))$, and $\dim\Big(\Sigma(P_2)\Big) > \dim(P_2)$ for every $P_2 \in \mathcal{P}_2$.]

(Br.3.E) *On every cell* P*,* $x \to t_x$ *is a continuously differentiable function, and* $(t, x) \to \xi_x(t)$*,* $(t, x) \to u_x(t) \stackrel{\mathrm{def}}{=} v(\xi_x(t))$ *are continuously differentiable maps on the set*

$$E(P) \stackrel{\mathrm{def}}{=} \{(t, x) : x \in P\,,\, t \in [0, t_x]\} \tag{4.2}$$

in the sense in the sense that they can be prolonged to maps of class $\mathcal{C}^1$ on some open subset of $\mathbb{R} \times P$ containing $E(P)$.

(Br.3.F) *For every $x \in S\backslash\{0\}$, if we let $\tilde{\xi}_x$ denote the curve obtained in an obvious way by piecing together the trajectories on every single cell, and write $\tilde{\eta}_x(t) = v(\tilde{\xi}_x(t))$, then the admissible pair $(\tilde{\xi}_x, \tilde{\eta}_x)$ ends at the origin after passing through a finite number of cells, and is extremal.*

(Br.3.G) *The cost function $V^{\Xi} \stackrel{\text{def}}{=} V_{\Gamma(\Xi)}$ corresponding to the synthesis $\Gamma(\Xi) = \{(\tilde{\xi}_x, \tilde{\eta}_x)\}_{x \in S}$ is continuous.* $\diamondsuit$

Boltyanskii's definition, as given in [2], includes an extra ingredient, namely, a subset N of S where the synthesis is allowed to have a more singular behavior. The definition is formulated in terms of a sequence $\mathbf{P} = (P^0, \ldots, P^n)$ of subsets of S such that

$$\{0\} = P^0 \subseteq P^1 \subseteq \ldots \subseteq P^n = S\,, \tag{4.3}$$

and the object that we called $\mathcal{P}$ here (that is, the set of "cells") is the union

$$\mathcal{P} = \bigcup_{i=0}^{n} \mathcal{P}^i\,, \tag{4.4}$$

where

(Bo.*) *$\mathcal{P}^i$, for $i > 0$, is the set of all connected components of $P^i \setminus (P^{i-1} \cup N)$, and $\mathcal{P}^0 = \{\{0\}\}$,*

so now $\mathcal{P}$ is a partition of $(S \setminus N) \cup \{0\}$ rather than of S. Boltyanskii does not explicitly require that this partition be locally finite in S. He only asks that the sets P^i and N be "piecewise smooth" (that is, locally finite unions of "curvilinear polyhedra," i. e. of sets that are $\mathcal{C}^1$-diffeomorphic to closed bounded polyhedra in finite-dimensional Euclidean spaces), but this does not imply that $\mathcal{P}$ is locally finite. (In fact, in §5, Example 5.2, we exhibit a Boltyanskii regular synthesis whose set of cells is not locally finite.)

In addition, Boltyanskii imposes some extra requirements, which imply some of the Brunovský conditions. In order to facilitate the comparison, we will include in our definition of a Boltyanskii synthesis all the Brunovský conditions that necessarily follow from the Boltyanskii assumptions, even though this will introduce a number of redundancies.

Definition 4.3. A *Boltyanskii regular synthesis* on the open subset S of the open set $\Omega \subseteq \mathbb{R}^n$, for the control problem (2.1), (2.2), with target the origin, is an 8-tuple $\Xi = (\mathbf{P}, \mathcal{P}_1, \mathcal{P}_2, \Pi, \Sigma, v, N, \Xi^N)$ such that

(Bo.1) *$\mathbf{P} = (P^0, \ldots, P^n)$ is a sequence of piecewise smooth subsets of S such that (4.3) holds.*

(Bo.2) *N is a piecewise smooth subset of S such that $\dim(N) < n$.*

(Bo.3) *If $\mathcal{P}$—the set of "cells"—is defined by (4.4), with the $\mathcal{P}^i$ defined by (Bo.*), then all the cells are embedded submanifolds of S, of class $\mathcal{C}^1$, such that $\dim(P) = i$ whenever $P \in \mathcal{P}^i$.* [BP&HS: The cells are obviously connected, by definition, and form a partition of $(S \setminus N) \cup \{0\}$. But $\mathcal{P}$ need not be locally finite and *a fortiori* $\mathcal{P}$ need not be a stratification.]

(Bo.4) *$dim(\Pi(P)) = \dim(P) - 1$ whenever $P \in \mathcal{P}_1$.*

(Bo.5) *The only zero-dimensional cell is $\{0\}$.*

(Bo.6) *Ξ^N is a family of pairs in $\mathrm{Adm}^{\mathcal{C}^1,0}(\tilde{f}, \{0\})$ such that (a) if $(\xi, \eta) \in \Xi^N$ then $\xi^- \in N$, (b) for every $x \in N$ there is a $(\xi, \eta) \in \Xi^N$ such that $\xi^- = x$, and (c) if (ξ_1, η_1) and (ξ_2, η_2) belong to Ξ^N, and $\xi_1^- = \xi_2^-$, then $J(\xi_1, \eta_1) = J(\xi_2, \eta_2)$.*

(Bo.7) Conditions (Br.2) and (Br.3) of Brunovský's definition hold, with the following modifications: (a) v is defined on $S \setminus N$ rather than on S, (b) the stratification requirement (Br3.A) is dropped, and (c) $V^{\Xi}(x)$ is defined to be equal to $J(\tilde{\xi}_x, \tilde{\eta}_x)$ if $x \in S \setminus N$, and to $J(\xi, \eta)$ if $x \in N$ and (ξ, η) is any member of Ξ^N such that $\xi^- = x$.

(Bo.8) Whenever the trajectory $\tilde{\xi}_x$ enters a new cell, it does so "at a nonzero angle". (That is, more precisely: whenever $P \in \mathcal{P}_1$ and $x \in P$, the tangent vector to $\tilde{\xi}_x$ at time t_x is not tangent to $\Pi(P)$ at $\tilde{\xi}_x(t_x)$.) ♢

It is then clear that the two concepts of a regular synthesis defined by Boltyanskii and Brunovský are not comparable. A Brunovský synthesis can fail to be a Boltyanskii synthesis because

a. it violates the "nonzero angle" condition (Bo.8), as in the example given by Brunovský in [6],

or

b. it violates the dimension condition (Bo.4), as in our Example 5.4 below,

or

c. it violates condition (Bo.5), as in Example 5.1 below.

On the other hand, a Boltyanskii synthesis can fail to be a Brunovský synthesis because, for example, the set of cells could fail to be locally finite, as in our Example 5.2. Also, the Brunovský definition, as stated in [6], does not allow for an extra "singular set" N.

The uniqueness requirement of $(Br.3.D)$ is rather strong. In [16] we proposed an even less restrictive formulation, as we now explain. For a $P \in \mathcal{P}_2$, we not only specify a cell $\Sigma(P) \in \mathcal{P}_1$ as in (Br.3.D), but also give a continuous "exiting map" $\mathcal{E}_P$, defined on the set $\{(x,t) : x \in P, 0 \le t < \varepsilon(x)\}$ and with values in $P \cup \Sigma(P)$. Here $\varepsilon : P \to I\!R$ is a continuous strictly positive function, and the map $\mathcal{E}_P$ is required to be such that, for every $x \in P$, (a) $\mathcal{E}_P(x,0) = x$, and (b) the map $]\,0, \varepsilon(x)\,[\, \ni t \to \mathcal{E}_P(x,t)$ takes values in $\Sigma(P)$ and is an integral curve of $f(\cdot, v(\cdot))$. So in [16] a *synthesis* is not just a 6-tuple $\Gamma = (\mathcal{P}, \mathcal{P}_1, \mathcal{P}_2, \Pi, \Sigma, v)$ as in the Brunovský definition, but a 7-tuple $\Gamma = (\mathcal{P}, \mathcal{P}_1, \mathcal{P}_2, \Pi, \Sigma, v, \mathcal{E})$, where the extra ingredient is the family $\mathcal{E} = \{\mathcal{E}_P\}_{P \in \mathcal{P}_2}$ of exiting maps.

Remark 4.4. The map $\mathcal{E}_P$ can be thought of as way of selecting in a continuous way, for each $x \in P$, an initial piece $]\,0, \varepsilon(x)\,[\, \ni t \to \mathcal{E}_P(x,t) \in \Sigma(P)$ of an integral curve of the feedback vector field $\Sigma(P) \ni y \to f(y, v(y)) \in T_y\Sigma(P)$. Naturally, if (Br.3.D) was satisfied we would just choose this curve to be ξ_x. ♢

It is not hard to produce a concept of regular synthesis that contains as special cases the three definitions of [2], [6], and [16] but does not differ too much from any of them. This can be done by removing from the definitions of [2], [6], and [16] conditions that not really needed, such as the local finiteness requirements and Boltyanskii's conditions (Bo.4), (Bo.5) and (Bo.8). One can also eliminate the requirement that the control space U be a subset of some Euclidean space $I\!R^m$, and take U to be an arbitrary set with no extra structure, provided only that all the conditions involving differentiability of the feedback control $x \to v(x)$ are replaced by a differentiability requirement for the map $x \to \tilde{f}(x, v(x))$. One possible result is the following concept, that we will call "BB-regular synthesis," using the letters BB to stand for "Boltyanskii and Brunovský":

Definition 4.5. A *BB-regular synthesis* on the open subset S of Ω, for the control problem (2.1), (2.2), with target $\{0\}$, is a 9-tuple $\Xi = (\mathcal{P}, \mathcal{P}_1, \mathcal{P}_2, \Pi, \Sigma, \mathcal{E}, v, N, \Xi^N)$ such that

(BB.1) $0 \in N \subseteq S$, *and* $\mathcal{P}$ *is a finite or countable partition of* $S \setminus N$ *into nonempty connected embedded submanifolds of* S *of class* $\mathcal{C}^1$*(called "cells");*

(BB.2) $\mathcal{P}$ *is the disjoint union of* $\mathcal{P}_1$ *(the set of "type I cells") and* $\mathcal{P}_2$ *(the set of "type II cells"),*

(BB.3) $v : \bigcup\{P : P \in \mathcal{P}_1\} \to U$ *(the "closed loop control"),* $\Pi : \mathcal{P}_1 \to \mathcal{P} \cup \{\{0\}\}$, *and* $\Sigma : \mathcal{P}_2 \to \mathcal{P}_1$, *are maps, and* $\mathcal{E}$ *is a family* $\{\mathcal{E}_P\}_{P \in \mathcal{P}_2}$, *such that the following properties are satisfied:*

(BB.3.A) *If* $P \in \mathcal{P}_1$, *then the map* $P \ni x \to \tilde{f}(x, v(x))$ *is of class* $\mathcal{C}^1$, *and such that* $f(x, v(x))$ *belongs to* $T_x P$ *(the tangent space to* P *at* x*) for every* $x \in P$.

(BB.3.B) *If* $P \in \mathcal{P}_2$, *then* $\Sigma(P) \in \mathcal{P}_1$, *and* $\mathcal{E}_P$ *is a map that assigns to each* $x \in P$ *a maximally defined integral curve* $\mathcal{E}_{P,x}$ *of the vector field* $\Sigma(P) \ni y \to f(y, v(y)) \in T_y\Sigma(P)$, *in such a way that* $0 = \inf \mathrm{Dom}(\mathcal{E}_{P,x})$ *and* $\lim_{t \downarrow 0} \mathcal{E}_{P,x}(t) = x$.

(BB.3.C) *For* $x \in P \in \mathcal{P}_2$, *define a map* $\xi_x : \{0\} \cup \mathrm{Dom}(\mathcal{E}_{P,x}) \to \Omega$ *by letting* $\xi_x(0) = x$ *and* $\xi_x(t) = \mathcal{E}_{P,x}(t)$ *for* $t \in \mathrm{Dom}(\mathcal{E}_{P,x})$. *For* $x \in P_1 \in \mathcal{P}_1$, *let* ξ_x *be the maximally defined solution of the initial value problem (4.1). Define* $t_x = \sup \mathrm{Dom}(\xi_x)$ *for* $x \in P \in \mathcal{P}$. *Then for every cell* $P \in \mathcal{P}$ *the function* $P \ni x \to t_x$ *is of class* $\mathcal{C}^1$, *and the maps* $\hat{E}(P) \ni (t, x) \to \xi_x(t) \in \Omega$ *and* $\hat{E}(P) \ni (t, x) \to \tilde{f}(\xi_x(t), v(\xi_x(t))) \in \mathbb{R}^{n+1}$ *have continuously differentiable extensions to a neighborhood of* $E(P)$ *in* $\mathbb{R} \times P$, *where* $E(P)$ *is the set defined by (4.2), and* $\hat{E}(P) = \{(t, x) \in E(P) : t < t_x\}$.

(BB.3.D) *If* $x \in P \in \mathcal{P}_1$, *then the limit* $\xi_x(t_x-) = \lim_{t \uparrow t_x} \xi_x(t)$*—which exists because of (BB.3.C)—belongs to* $\Pi(P)$.

(BB.4) *Boltyanskii's condition (Bo.6) holds.*

(BB.5) *Brunovský's condition (Br.3.F) holds for* $x \in S \setminus N$.

(BB.6) *Brunovský's condition (Br.3.G) holds, with* V^{Ξ} *defined as in (Bo.7.c).* ♢

It is then easy to show that if Ξ is a regular synthesis in the sense of Boltyanskii [2], Brunovský [6], or Sussmann [16], then Ξ is a BB-regular synthesis in the sense of Definition 4.5. Moreover, a BB-regular synthesis gives rise in a natural way to a regular presynthesis $\Gamma(\Xi)$ in the sense of our Definition 2.12, which is a synthesis if $N = \{0\}$.

Our Definition 2.12—which was already proposed in [27]—is even more general, because:

a. We do not require that V_Γ be continuous, and only assume the weak continuity conditions.

b. We do not require that the domain of the synthesis be an open set. This is quite important because one wants the theory to apply to systems that are not locally controllable, and in those cases one usually wants to take S to be the set $\hat{S}$ of all points of Ω that can be steered to the target, and in general $\hat{S}$ is not open.

c. We dot require the existence of a partition into "cells." From our point of view, such a partition is only needed to guarantee the weak differentiability condition (2.12.b).

d. Even when a good partition into "cells" exists, satisfying the most restrictive conditions of both definitions (that is, Brunovský's stratification condition (Br.3.A) and Boltyanskii's conditions (Bo.4), (Bo.5), and (Bo.8)), it may still happen that the crucial "finite number of steps" condition (Br.3.F) is violated, as in Fuller's example, but our requirements for a regular presynthesis are still met.

§5. Examples

We now present several examples of optimal regular syntheses to illustrate the differences between the various definitions.

One example showing how our notion of regular synthesis is more general than those of Boltyanskii and Brunovský's was already discussed in §3, where we showed that our theory applies to Fuller's problem while the other ones clearly do not, since they do not allow the trajectories of the synthesis to have infinitely many switchings. Another important example, relevant for the comparison of the Boltyanskii and Brunovský definitions, is the one given by Brunovský in [6], showing how the "nonzero angle" condition (Bo.8) can be violated.

We now show by means of an example how Boltyanskii's condition (Bo.5) can be violated. In principle, it is quite easy to produce examples of syntheses that satisfy all of Boltyanskii's conditions except for (Bo.5). It turns out, however, that in most such examples one can construct a new synthesis where (Bo.5) holds, for example by incorporating the zero-dimensional cells in the singular set N, or by drawing a one-dimensional arc A_q through every q such that $\{q\}$ is a cell and making the A_q's cells of the new synthesis. To produce an example where no synthesis can possibly satisfy (Bo.5) more work is needed, but it can be done, as we now show.

Example 5.1. Let

$$S_1 = \{(x,0) : x \le 0\}\,, \qquad S_2 = \{(x,x) : 0 \le x \le 1\}\,,$$
$$S_3 = \{(x,2x-1) : x \ge 1\}\,. \qquad S_4 = \{(x,0) : x \ge 0\}\,,$$

Let $\Lambda_1 = S_1 \cup S_2 \cup S_3$, $\Lambda_2 = S_1 \cup S_4$. Let A be the set of points (x,y) such that $x \ge 0$ and either $0 \le y \le x \le 1$ or $0 \le y \le 2x-1$, and let B be the set of those (x,y) such that either $y \le 0$, or $x \le 0$, or $0 \le x \le 1$ and $x \le y$, or $x \ge 1$ and $y \ge 2x-1$. Let $\varphi, \psi : \mathbb{R}^2 \to \mathbb{R}$ be smooth nonnegative functions such that $\varphi(x,y) = 0$ if and only if $(x,y) \in A$, and $\psi(x,y) = 0$ if and only if $(x,y) \in B$.

We consider the minimum time optimal control problem in $\mathbb{R}^2$ with target $\{(0,0)\}$, and dynamics given by

$$\dot{x} = \frac{u_1}{1+\varphi(x,y)}\,, \qquad \dot{y} = \frac{u_2}{1+\psi(x,y)}\,,$$

where the control constraint is $|u_1| + |u_2| \le 1$.

If $q = (\bar{x},\bar{y})$ is any point in $\mathbb{R}^2$, and $[a,b] \ni t \to \xi(t) = (x(t),y(t)) \in \mathbb{R}^2$ is any trajectory from q to the target, corresponding to a control $[a,b] \ni t \to \eta(t) = (u_1(t),u_2(t))$, then

$$b - a \ge \int_a^b \Big(|u_1(t)| + |u_2(t)|\Big)\,dt \ge \int_a^b \Big(|\dot{x}(t)| + |\dot{y}(t)|\Big)\,dt \ge |\bar{x}| + |\bar{y}|\,.$$

Moreover, a trajectory ξ_q from q to the origin whose cost is exactly $|\bar{x}| + |\bar{y}|$ can be constructed as follows:

1. If q belongs to Λ_1, then the point $\xi_q(t)$ moves along Λ_1 towards the origin, with a control vector $\eta_q(t) = (u_1(t),u_2(t))$ such that $|u_1(t)| + |u_2(t)| = 1$ (so that $\eta_q(t) = (1,0)$ if $q \in S_1$ and $\eta_q(t) = (-\frac{1}{2},-\frac{1}{2})$ if $q \in S_2$; if $q \in S_8$, then $\eta_q(t) = (-\frac{1}{3},-\frac{2}{3})$ until $\xi_q(t) = (1,1)$, and $\eta_q(t) = (-\frac{1}{2},-\frac{1}{2})$ from then on).
2. If $q \in A$, then $\xi_q(t)$ moves horizontally to the left (with control $(u_1,u_2) = (-1,0)$) until it reaches Λ_1, and then follows the trajectory described in item 1.
3. If $q \in B$ and $\bar{y} > 0$, then $\xi_q(t)$ moves vertically down (with control $(u_1,u_2) = (0,-1)$) until it reaches Λ_1, and then follows the trajectory described in item 1.
4. If $q \in B$ and $\bar{y} < 0$, then $\xi_q(t)$ moves vertically up (with control $(u_1,u_2) = (0,1)$) until it reaches Λ_2, and then moves horizontally towards the origin with control $(1,0)$ or $(-1,0)$.

The family $\Xi = \{\xi_q\}_{q\in\mathbb{R}^2}$ is an optimal synthesis. If we define P^0 to be the two-point set consisting of $(0,0)$ and $(1,1)$, and let $P^1 = \cup_{i=1}^4 S_i$, $P^2 = \mathbb{R}^2$, and $N = \emptyset$, then all the conditions of a Boltyanskii regular synthesis are satisfied (with 17 cells), except for (Bo.4). Moreover, it is easy to show that *this problem does not admit a regular synthesis that satisfies all the Boltyanskii conditions, including (Bo.4).* $\diamondsuit$

In our second example, we present a Boltyanskii optimal synthesis which is not a Brunovský synthesis, because the set of cells is not locally finite.

Example 5.2. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^\infty$ such that (a) $\psi(x) = 0$ whenever $x \le 0$, (b) $0 \le \psi(x) < x$ when $x > 0$, and (c) the set of zeros of ψ is the union of $\{0\}$ and the points of a decreasing infinite sequence $\{x_k\}_{k=1}^\infty$ having an accumulation point at 0. (For example, we could take

$$\psi(x) = \rho(x).\sin^2\frac{1}{x}\,, \qquad \text{where} \qquad \rho(x) = \begin{cases} e^{-\frac{1}{x}} & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \le 0 \end{cases},$$

and let $x_k = \frac{1}{\pi k}$.) Let

$$\begin{aligned} A &= \{(x,y)\in\mathbb{R}^2 : x \ge 0 \text{ and } -\psi(x) \le y \le \psi(x)\}\,, \\ B &= \{(x,y)\in\mathbb{R}^2 : x \le |y|\}\,, \\ E &= \Big\{(x,y)\in\mathbb{R}^2 : x \ge 0 \text{ and } \Big(\psi(x) \le y \le x \text{ or } -x \le y \le -\psi(x)\Big)\Big\}\,. \end{aligned}$$

Let L be the y axis. Then A, B, E, and L are closed subsets of $\mathbb{R}^2$. Let $\sigma : \mathbb{R}^2 \to \mathbb{R}$ be a function of class $\mathcal{C}^\infty$ such that $\sigma \equiv 0$ on $A\cup B$ and $\sigma > 0$ on $\mathbb{R}^2 \setminus (A\cup B)$. Let $\tau : \mathbb{R}^2 \to \mathbb{R}$ be a function of class $\mathcal{C}^\infty$ such that $\tau \equiv 0$ on $E\cup L$ and $\tau > 0$ on $\mathbb{R}^2\setminus(E\cup L)$.

Consider the optimal control problem in $\mathbb{R}^2$ with target $\{(0,0)\}$, in which the dynamics is given by

$$\dot{x} = u_1\,, \qquad \dot{y} = u_2\,,$$

the control constraints are $u_1 \in \{-1,0,1\}$ and $|u_1| + |u_2| = 1$, and the cost functional to be minimized is

$$J = \int_a^b \Big(u_1(t)^2\Big(1+\sigma(x(t),y(t))\Big) + u_2(t)^2\tau(x(t),y(t))\Big)\,dt\,.$$

If $q = (\bar{x},\bar{y}) \in \mathbb{R}^2$, and $[a,b] \ni t \to \xi(t) = (x(t),y(t))$ is any trajectory going from q to the origin, then $J(\xi) \ge \int_a^b u_1(t)^2\,dt = \int_a^b |u_1(t)|\,dt \ge |\int_a^b \dot{x}(t)\,dt| = |\bar{x}|$. On the other hand, there exists a trajectory ξ_q for which $J(\xi_q) = |\bar{x}|$. To see this, observe that purely vertical motion—i. e. $|u_2| = 1$ and $u_1 = 0$—costs nothing as long as it takes place in $E\cup L$, while horizontal motion—i. e. $|u_1| = 1$ and $u_2 = 0$—has a cost per unit time equal to 1 on $A\cup B$. If $q\in B$, then we can move horizontally towards the y axis—using $u_1 = -\mathrm{sgn}(\bar{x})$—and then move vertically along the y axis and end up at the origin, thereby obtaining a trajectory whose cost is exactly $|\bar{x}|$. If $q \in E$, then we can move vertically up or down—using $u_2 = \mathrm{sgn}(\bar{y})$—until we hit the boundary of B. Once we are in B, we follow the optimal trajectory already described for points of B, and end up producing a trajectory from q to the origin whose cost is $|\bar{x}|$. If $q \in A$ and $\bar{y} = 0$, then the constant control $u_1 = -1$, $u_2 = 0$ steers q to the origin with cost $\bar{x}$. If $q\in A$ and $\bar{y} \ne 0$, then we can use the constant control $u_1 = -1$, $u_2 = 0$ until we reach a point $(\hat{x},\hat{y})$ such that $|\hat{y}| = \psi(\hat{x})$. The cost of this is clearly $\bar{x} - \hat{x}$, and $(\hat{x},\hat{y})$ belongs to E, so we know how to go from $(\hat{x},\hat{y})$ to the origin with cost $\hat{x}$. So we can go from q to the origin with cost $\bar{x}$.

Therefore the value function V for this problem is given by $V(x,y) = |x|$. Moreover, we have given an explicit description of a family $\Xi = \{\xi_q\}_{q\in\mathbb{R}^2}$ which is a total optimal synthesis. It is not hard to show that *for every q, the arc ξ_q is the unique optimal trajectory from q to the origin.*

We now show that Ξ *is the family of trajectories of a Boltyanskii regular synthesis whose set of cells is not locally finite.* To see this, we define N to be the set $[0,\infty[\times\{0\}$, so N is clearly piecewise smooth. We take $P^0 = \{(0,0)\}$, and let P^1 be the union of the two coordinate axes, the two half-lines $L_1 = \{(x,x) : x \geq 0\}$ and $L_2 = \{(x,-x) : x \geq 0\}$, and the graphs of ψ and $-\psi$. Finally, we let $P^2 = \mathbb{R}^2$. It is then clear that P^0, P^1, and P^2 are piecewise smooth. The connected components of $P^1 \setminus (P^0 \cup N)$ are the open half-lines $\Lambda_1 = \{(x,x) : x > 0\}$, $\Lambda_2 = \{(0,y) : y > 0\}$, $\Lambda_3 = \{(x,0) : x < 0\}$, $\Lambda_4 = \{(0,y) : y < 0\}$, and $\Lambda_5 = \{(x,-x) : x > 0\}$, and the arcs $A_k^+ = \{(x,\psi(x)) : x_{k+1} < x < x_k\}$, $A_k^- = \{(x,-\psi(x)) : x_{k+1} < x < x_k\}$. The connected components of $P^2 \setminus (P^1 \cup N)$ are the sets $U_1 = \{(x,y) : x > 0\,, \psi(x) < y < x\}$, $U_2 = \{(x,y) : 0 < x < y\}$, $U_3 = \{(x,y) : x < 0 < y\}$, $U_4 = \{(x,y) : x < 0\,, y < 0\}$, $U_5 = \{(x,y) : x > 0\,, y < -x\}$, and $U_6 = \{(x,y) : x > 0\,, -x < y < -\psi(x)\}$, and the regions $W_k^+ = \{(x,y) : x_{k+1} < x < x_k\,, 0 < y < \psi(x)\}$, $W_k^- = \{(x,y) : x_{k+1} < x < x_k\,, 0 > y > -\psi(x)\}$. The arcs A_k^+ and A_k^- and the half-lines Λ_1, Λ_5 are declared to be type II cells, and we choose $\Sigma(A_k^+) = U_1$, $\Sigma(A_k^-) = U_6$, $\Sigma(\Lambda_1) = U_2$, $\Sigma(\Lambda_5) = U_5$. The remaining cells (that is, Λ_2, Λ_3, Λ_4, the U_i, the W_k^+ and the W_k^-) are type I cells. The feedback v is defined by letting $v \equiv (-1,0)$ on $N \setminus \{(0,0)\}$, the W_k^+, the W_k^-, U_2 and U_5, $v \equiv (1,0)$ on U_3 and U_4, $v \equiv (0,1)$ on U_1 and Λ_4, $v \equiv (0,-1)$ on U_5 and Λ_2. It is then clear that all the conditions of the Boltyanskii definition are satisfied, but the set of cells is not locally finite. $\diamondsuit$

Example 5.3. We now present an example of a synthesis that satisfies all the Brunovský conditions except for the continuity requirement (Br.3.G). This example will, in addition, exhibit the phenomenon of nonuniqueness of trajectories for the closed-loop equation arising from the optimal feedback, thereby providing a concrete illustration of the reasons for including as an extra ingredient a selection of trajectories, i. e., for following strategy (B) of §4.

We construct and study in detail the time-optimal synthesis for the planar system:

$$\dot{q} = F(q) + u\,G(q)\,, \qquad |u| \leq 1\,, \tag{5.1}$$

where

$$q \stackrel{\text{def}}{=} \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2\,, \qquad F(q) \stackrel{\text{def}}{=} \begin{pmatrix} 1 - \frac{y}{2} \\ \frac{x+1}{2} \end{pmatrix}\,, \qquad G(q) \stackrel{\text{def}}{=} \begin{pmatrix} -\frac{y}{2} \\ \frac{x+1}{2} \end{pmatrix}\,, \tag{5.2}$$

and the target is the origin of $\mathbb{R}^2$.

The trajectories of (5.1) corresponding to the constant control $u \equiv -1$ are straight horizontal lines going from left to right, while those corresponding to $u \equiv +1$ are circles centered at the point (-1,1), running counterclockwise. The optimal synthesis, to be determined below, is shown in Fig.2.

An arc $\gamma : [a,b] \to \mathbb{R}^2$ is *simple* if γ is an injective map. Clearly, all optimal trajectories are simple.

Since $G(q) = 0$ only for $q = (-1,0)$, it is clear that a trajectory $\gamma : [a,b] \to \mathbb{R}^2$ uniquely determines the corresponding control $u(\cdot)$, unless the set $\{t : \gamma(t) = (-1,0)\}$ is of positive measure. In particular, for simple trajectories γ—the only ones that will be of interest to us—the control is uniquely determined by γ.

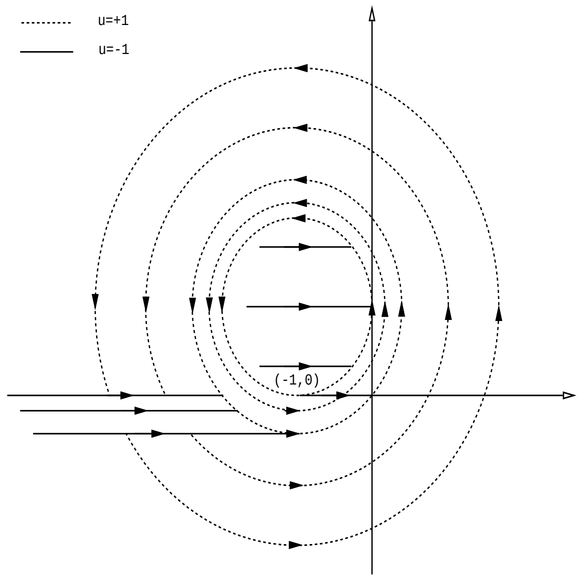

Figure 2

As in previous sections, we use $I\!R_k$ to denote the space of real row k-vectors. If Ω is an open subset of $I\!R^k$ and $X, Y : \Omega \to I\!R^k$ are vector fields of class C^1, then $[X,Y]$ will denote the Lie bracket of X and Y, so $[X,Y] = (DY).X - (DX).Y$.

If X, Y are $I\!R^2$-valued functions on a subset of $I\!R^2$, then (X,Y) will denote the 2×2 matrix with columns X, Y. In particular, following Sussmann [22], we define two real-valued functions Δ_A, Δ_B on $I\!R^2$, by letting

$$\Delta_A = det\,(F,G) \qquad \Delta_B = det\,(G,[F,G]) \tag{5.3}$$

where "*det*" stands for determinant. From the results of [22], we know that every extremal trajectory of (5.1) is a finite concatenation of "bang" and "singular" arcs. An arc is a *bang arc* if it is a B_- arc—i. e. a trajectory for the control $u = -1$—or a B_+ arc—i. e. a trajectory for $u = 1$. Singular arcs lie in the set of zeros of the function Δ_B. In our case,

$$\Delta_A = \frac{x+1}{2} \quad \text{and} \quad \Delta_B = -\frac{y}{4} \tag{5.4}$$

so the only singular arcs are those entirely contained in the x axis, and these correspond to the control -1, so they actually are bang. So *all the extremals of our problem are bang-bang.*

If $\Delta_A(q) \neq 0$ then the vectors $F(q)$ and $G(q)$ are linearly independent, and we can define the numbers $f(q)$ and $g(q)$ as the coefficients of the linear combination:

$$[F,G](q) = f(q)F(q) + g(q)G(q). \tag{5.5}$$

Let Ω_A, Ω_B be, respectively, the sets $\{q : \Delta_A(q) \neq 0\}$, $\{q : \Delta_B(q) \neq 0\}$, and define $\Omega_{AB} = \Omega_A \cap \Omega_B$. It was proved in [22] that, if γ is an extremal trajectory and $\gamma(t) \in \Omega_{AB}$ for every t, then γ is bang-bang with at most one switching, and this switching can only be from the control value -1 to the value $+1$ if $f > 0$, and from $+1$ to -1 if $f < 0$. In our case, it follows from (5.4) that

$$\Omega_A = \{(x,y) : x \neq -1\}\,, \qquad \Omega_B = \{(x,y) : y \neq 0\}\,, \tag{5.6}$$

so Ω_{AB} is the union of four open quadrants $\Omega_{AB,i}$, $i = 1,2,3,4$, defined by

$$\begin{array}{ll} \Omega_{AB,1} = \{(x,y) : x > -1\,,\, y > 0\}\,, & \Omega_{AB,2} = \{(x,y) : x < -1\,,\, y > 0\}\,, \\ \Omega_{AB,3} = \{(x,y) : (x,-y) \in \Omega_{AB,2}\}\,, & \Omega_{AB,4} = \{(x,y) : (x,-y) \in \Omega_{AB,1}\}\,. \end{array} \tag{5.7}$$

Moreover,

$$[F,G] = \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix} = \frac{y}{2(x+1)}F + \frac{2-y}{2(x+1)}G\,, \tag{5.8}$$

so $f > 0$ on $\Omega_{AB,1} \cup \Omega_{AB,3}$ and $f < 0$ on $\Omega_{AB,2} \cup \Omega_{AB,4}$. So the control switchings are from -1 to 1 on $\Omega_{AB,1} \cup \Omega_{AB,3}$ and from 1 to -1 on $\Omega_{AB,2} \cup \Omega_{AB,4}$.

To determine the optimal synthesis, we will first construct a "sufficient family" of trajectories for our problem, i. e. a collection $\mathcal{F}$ of simple trajectories such that every optimal trajectory ending at the target is in $\mathcal{F}$. We will then find, for every initial point q, a $\gamma_q \in \mathcal{F}$ such that if $\delta \in \mathcal{F}$ goes from q to 0 and $\delta \neq \gamma_q$ then δ is not optimal. Since the existence of optimal trajectories from any initial point follows by elementary arguments, we will be able to conclude that γ_q is optimal.

Suppose that $\gamma : [a,b] \to I\!R^2$ is time-optimal, $\gamma(b) = (0,0)$, $a < b$, and γ arises from an open-loop control $t \to u(t)$. Then γ is simple, and is a concatenation of finitely many maximal bang arcs. Since γ is an extremal, we can fix a nontrivial minimizing adjoint vector λ along γ and a constant $\lambda_0 \geq 0$ for which (EX4) holds. Define the *switching function* $\varphi(t) = \lambda(t) \cdot G(\gamma(t))$. The Hamiltonian minimization condition implies that

$$\varphi(t) \neq 0 \implies u(t) = -sgn(\varphi(t))\,. \tag{5.9}$$

Let $\delta : [c,d] \to I\!R^2$ be a maximal B_- piece of γ, so $c < d$ and δ is of the form:

$$\delta(t) = (x_0 + t, y_0)\,. \tag{5.10}$$

For a.e. $t \in [c,d]$, we have $\dot{\lambda} = -\lambda \cdot D(F - G) = 0$. So $\lambda = (\lambda_1, \lambda_2)$ is constant, $(\lambda_1, \lambda_2) \neq (0,0)$, and:

$$\varphi(t) = -\lambda_1 \frac{y_0}{2} + \lambda_2 \frac{x_0 + t + 1}{2} \tag{5.11}$$

is a linear function of t. Since a linear function cannot have more than one zero unless it vanishes identically, we conclude that $a < c < d < b$ can only be true if $\lambda_2 = y_0 = 0$. If $d = b$, then of course $y_0 = 0$, since $\gamma(b) = (0,0)$. So one of the following two possibilities occurs:

(I) δ is contained in the x axis,
(II) $a = c$.

This shows that all the maximal B_- pieces of γ are entirely contained in the x axis, with the only possible exception that the domain $[a,b]$ of γ may contain an initial segment $[a,d]$ such that $\gamma([a,d])$ is contained in a horizontal line for which $y \neq 0$.

Let us now analyze the adjoint equation for a maximal B_+ piece $\delta : [c,d] \to I\!R^2$ of γ, such that $c < d$. Clearly, δ is of the form:

$$x(t) = -1 + r\cos(t+\theta)\,, \qquad y(t) = 1 + r\sin(t+\theta)\,, \qquad c \leq t \leq d\,, \tag{5.12}$$

where r, θ are constants, $r > 0$, and $\theta \in [0, 2\pi]$. (The possibility that $r = 0$ is obviously excluded, because γ is time-optimal and $c < d$.) The components λ_1, λ_2 of λ satisfy $\dot{\lambda}_1 = -\lambda_2$ and $\dot{\lambda}_2 = \lambda_1$, so there exist constants $A > 0$, $\theta_1 \in [0, 2\pi]$, such that

$$\lambda_1(t) = A\cos(t + \theta_1)\,, \qquad \lambda_2(t) = A\sin(t+\theta_1)\,, \qquad c \leq t \leq d\,. \tag{5.13}$$

Therefore, up to multiplication by a positive constant,

$$\varphi(t) = -\cos(t+\theta_1) + r\sin(\theta_1 - \theta)\,. \tag{5.14}$$

Now suppose that one of the two endpoints c, d of the domain of δ is a switching time and the corresponding point in $\mathbb{R}^2$ lies in the x axis. That is, we assume that $h \in \{c, d\}$ is such that $\varphi(h) = 0$ and $y(h) = 0$. Then

$$r\sin(h+\theta) = -1\,, \tag{5.15}$$

so

$$\begin{aligned}\varphi(t) &= -\cos(t+\theta_1) + r\Big(\sin(h+\theta_1)\cos(h+\theta) - \cos(h+\theta_1)\sin(h+\theta)\Big)\\ &= -\cos(t+\theta_1) + \cos(h+\theta_1) + r\sin(h+\theta_1)\cos(h+\theta)\,.\end{aligned} \tag{5.16}$$

In particular,

$$0 = \varphi(h) = r\sin(h+\theta_1)\cos(h+\theta)\,, \tag{5.17}$$

so $\varphi(t) = -\cos(t+\theta_1) + \cos(h+\theta)$. If $r \neq 1$ then (5.15) implies $0 \geq \sin(h+\theta) \neq -1$, so $\cos(h+\theta) \neq 0$. But then (5.17) implies $\sin(h+\theta_1) = 0$, so $\cos(h+\theta_1) = \pm 1$. The possibility that $\cos(h+\theta_1) = 1$ is excluded, for in that case φ would be > 0 throughout $]c, d[$, contradicting the minimization condition. So $\varphi(t) = -1 - \cos(t+\theta_1)$. This shows that $\varphi(t) < 0$ for $t \in [c, d] - \{h\}$, unless $d - c = 2\pi$. But if $d - c = 2\pi$ then γ would not be simple. So $d - c < 2\pi$, and we have shown that the other endpoint h' of $[c, d]$ cannot be a switching point, and therefore $h' \in \{a, b\}$.

It then follows that one of the following possibilities must occur:

(i) $d = b$;
(ii) $a = c < d < b$ and $\gamma(d)$ lies in the x axis;
(iii) $a < c < d < b$ and $\gamma(d) = (-1, 0)$.

Indeed, if $d < b$ but $\gamma(d)$ is not in the x axis, then we could let δ' be the maximal B_- piece of γ that starts at time d, and our analysis of the B_- pieces of γ, applied to δ', would show that δ' cannot in fact occur, since δ' is not an initial segment of γ and is not contained in the x axis. So, if $d < b$ then $\gamma(d)$ lies in the x axis, and d is a switching time of γ. Our previous argument shows that $c = a$, unless $r = 1$. But if $r = 1$, and $\gamma(d)$ is in the x axis, then $\gamma(d) = (-1, 0)$. So $d < b$ implies that either (ii) or (iii) must hold.

We now combine the results about the two types of pieces and provide a complete description of the optimal trajectories ending at the origin. Let us use L, L_0, C to denote, respectively, the trajectory types "B_- and not contained in the x axis," "B_- and contained in the x axis," and "B_+." (The symbols L and C stand for "line" and "circle.") Let us use $T_1T_2\ldots T_m$ to denote the trajectory type "T_1 followed by T_2 ... followed by T_m".

With this notation, we now show that *a nontrivial optimal trajectory γ ending at the origin is of one of the following seven types*: L_0, C, L_0C, LC, CL_0, LCL_0, CL_0C. First, concatenations of six or more types are ruled out as follows: any such concatenation would have to contain at least three different C pieces; at most one of them can satisfy (i) and at most one can satisfy (ii); if both of these possibilities occurred, then we would have a concatenation of the form $C * C * C$, i. e. five pieces rather than six; so there must be two different pieces satisfying (iii), and then γ would go at least twice through the point $(-1, 0)$, and would not be simple. Next, all concatenations containing an L_0CL_0 sequence are excluded, because if γ is any such concatenation then the C piece that lies between the two L_0's would have to satisfy (iii), so its last switching would have to happen at $(-1, 0)$; then the C piece would have to be an arc of the circle with center $(-1, 1)$ and radius 1, which intersects the x axis only at $(-1, 0)$; then the other switching would also happen at $(-1, 0)$, and γ would not be simple. A five-piece concatenation must be CL_0CL_0C, LCL_0CL_0 or

$L_0CL_0CL_0$, because every B_- piece must satisfy (I) or (II), so it cannot be an L unless it is the first piece. So all such concatenations contain an L_0CL_0 piece and are therefore excluded. A four-piece concatenation must be CL_0CL_0 or L_0CL_0C or LCL_0C. The types CL_0CL_0 and L_0CL_0C are ruled out because they contain an L_0CL_0 triple. The type LCL_0C is excluded because the first C would have to satisfy (iii), so the switching from C to L_0 would have to happen at $(-1,0)$; then the L_0 piece following the switching has to be a segment going from $(-1,0)$ to a point $(\alpha,0)$ and α must satisfy $-1<\alpha<0$, for if α was ≥ 0 then the L_0 piece would already reach the origin before switching to the last C; and if $-1<\alpha<0$ then the C arc starting at $(\alpha,0)$ does not go through the origin. So all four-piece concatenations are excluded. Of the three-piece possibilities, CLC, LCL and L_0CL are excluded because an L piece can only occur at the beginning of the sequence, and we also know that L_0CL_0 is excluded. So only CL_0C and LCL_0 are left. Of the two and one-piece concatenations, CL and L are obviously excluded, and we are left with L_0, C, CL_0, L_0C and LC.

The LCL_0 trajectories are further restricted by the condition that the switching from C to L_0 must happen at the point $(-1,0)$. Let us call those LCL_0 trajectories that satisfy this extra condition "good."

Let $\mathcal{F}$ be the set of all simple trajectories that end at the origin and are either of one of the six types L_0, C, L_0C, LC, CL_0, CL_0C, or of type LCL_0 and good. We have shown that every optimal trajectory belongs to $\mathcal{F}$. It will turn out that the optimal synthesis involves all seven types listed above, but not all trajectories of one of these types are optimal, so we need a finer analysis.

To begin with, observe that $\dot{x}=1-y$ along a B_+ arc, so $\dot{x}>1$ along a B_+ arc as long as $y<0$. Therefore, if $q_1=(x_1,y_1)$ and $q_2=(x_2,y_2)$ satisfy $y_1=y_2\leq 0$, $x_1<x_2$, and $x_1+x_2=-2$, so that q_1 and q_2 can be joined both by a C arc and an L (or L_0) arc, then the C arc is contained in the lower half-plane, and is faster than the L or L_0 arc. In particular, if we let $L_0(\beta)$ denote, for $\beta>0$, the L_0 arc going from $(-\beta,0)$ to $(0,0)$, we see that $L_0(\beta)$ *is not optimal if* $1<\beta\leq 2$, because the piece of L_0 going from $(-\beta,0)$ to $(-2+\beta,0)$ can be replaced by a faster type C arc. Then the principle of optimality implies that $L_0(\beta)$ *is not optimal if* $1<\beta$, because if $\beta>2$ then $L_0(\beta)$ contains $L_0(2)$, which is not optimal.

Next we show that $L_0(1)$ *is optimal.* It suffices to notice that there is no trajectory from $(-1,0)$ to $(0,0)$ of any of the types C, LC, LCL_0, and exactly one trajectory of each of the types L_0, L_0C, CL_0, and CL_0C, but the last three are not simple, so $L_0(1)$ is left as the only possible candidate. It then follows from the principle of optimality that $L_0(\beta)$ *is optimal for* $0\leq\beta\leq 1$.

We now study the optimal trajectories that are of one of the types C, L_0C, LC, and CL_0C. Let $\gamma:[a,b]\to I\!R^2$ be such a trajectory, and take $b=0$, so $\gamma(0)=(0,0)$.

If γ is a C arc, then $a>-2\pi$, for otherwise γ would contain a full loop. For $0<a<2\pi$, we let $C(a)$ be the C arc $\gamma:[-a,0]\to I\!R^2$ such that $\gamma(0)=(0,0)$. It will be shown below that the arcs $C(a)$ are all optimal.

Next suppose that γ is LC. Since the C piece satisfies (5.12), we must have $r=\sqrt{2}$ and $\theta=\frac{7\pi}{4}$. Let c be the time when the switching from L to C occurs, so $-2\pi<c<0$, for if $c\leq-2\pi$ then γ would not be simple. On $[c,0]$, the switching function φ is given, up to multiplication by a positive constant, by:

$$\varphi(t)=\sqrt{2}\sin\Big(\theta_1-\frac{7\pi}{4}\Big)-\cos(t+\theta_1)=\sin\theta_1+\cos\theta_1-\cos(t+\theta_1).$$

So

$$\varphi(t)=(1+\sin t)\sin\theta_1+(1-\cos t).\cos\theta_1$$

Suppose that $\sin\theta_1 = 0$. Then $\cos\theta_1 = \pm 1$, so $\cos\theta_1 = -1$, because $\varphi(t) < 0$ when t is negative and close to 0, since $u(t) = +1$ for such t. So $\varphi(t) = \cos t - 1$, and then $\varphi(c) \neq 0$, since $-2\pi < c < 0$. Since $\varphi(c) =$, we conclude that $\sin\theta_1 \neq 0$. We can then define

$$\psi(t) = \frac{1+\sin t}{1-\cos t}, \qquad \nu = \frac{\cos\theta_1}{\sin\theta_1}.$$

Then ψ is well defined and smooth on $[c, 0[$, and $\psi(t)$ goes to $+\infty$ as $t \uparrow 0$. Moreover, φ can be expressed as $\varphi(t) = \sin\theta_1.(1-\cos t).(\nu + \psi(t))$ for $t \in [c, 0[$. Therefore $\sin\theta_1 < 0$, and $\psi(c) = -\nu$, since $1 - \cos c \neq 0$, because $-2\pi < c < 0$. Clearly, $\psi(t) \geq 0$ for all $t \in]-2\pi, 0[$, and ψ attains the value 0 on $]-2\pi, 0[$ at $t = -\frac{\pi}{2}$ and nowhere else. A simple calculation shows that ψ is strictly increasing on $]-\frac{\pi}{2}, 0[$. If $\nu > 0$, then $\psi + \nu$ never vanishes on $]-2\pi, 0[$, so $\varphi(c) \neq 0$. So $\nu \leq 0$. If $\nu = 0$, then $\psi(c) = -\nu$ implies $c = -\frac{\pi}{2}$. If $\nu < 0$, then ψ must take the value $-\nu$ at some point $t(\nu)$ such that $-\frac{\pi}{2} < t(\nu) < 0$, because $\psi(\frac{\pi}{2}) = 0$ and $\psi(0-) = +\infty$. Since ψ is strictly increasing on $]-\frac{\pi}{2}, 0[$, we see that $t(\nu)$ is unique, and φ changes sign at $t(\nu)$. If $c < t(\nu)$, we would contradict the fact that φ has constant sign on $[c, 0]$. On the other hand, it is impossible that $c > t(\nu)$, because then the equation $\psi(t) = -\nu$ would have two solutions on $]-\frac{\pi}{2}, 0[$. So $c = t(\nu)$, and we have proved that $-\frac{\pi}{2} \leq c < 0$. If $\frac{\pi}{4} < c$, then $\gamma(c) \in \Omega_{AB,4}$, and the switching is not permitted. So the only possibility left to us is $-\frac{\pi}{2} \leq c \leq -\frac{\pi}{4}$. If $c = -\frac{\pi}{2}$, then we really have an L_0C arc rather than an LC arc. So the only LC arcs that could be optimal are those for which $-\frac{\pi}{2} < c \leq -\frac{\pi}{4}$.

For $0 < \sigma < 2\pi$, $s > 0$, let $LC(s,\sigma)$ denote the trajectory defined on the interval $[-s-\sigma, 0]$ that ends at the origin and corresponds to a $u = -1$ control on $[-s-\sigma, -\sigma]$ followed by a $u - 1$ control on $[\sigma, 0]$. We have shown that *$LC(s,\sigma)$ is not optimal unless $\frac{\pi}{4} \leq \sigma < \frac{\pi}{2}$*. We now show that *if $\frac{\pi}{4} \leq \sigma < \frac{\pi}{2}$ then $LC(s,\sigma)$ is optimal.* To see this, let $q = (x,y)$ be the starting point of $LC(s,\sigma)$, for $s > 0$, $\frac{\pi}{4} \leq \sigma < \frac{\pi}{2}$. Then $1 - \sqrt{2} < y < 0$. It is clear that $LC(s,\sigma)$ is the only LC trajectory from q to $(0,0)$. Also, it is easy to see that there are no trajectories from q to $(0,0)$ of the types L_0, C, L_0C, or good LCL_0. There is exactly one CL_0 trajectory, but it is easy to see that the time $\hat\tau$ along this trajectory is larger than the time τ along $LC(s,\sigma)$. (Actually, $\tau = s + \sigma < |x| < \hat\tau - \frac{3\pi}{2}$.) Finally, there is exactly one CL_0C trajectory γ, which can also be ruled out. (Let $\bar x = -1 - \sqrt{2-(y-1)^2}$, so $\bar q = (\bar x, y)$ is the point where $LC(s,\sigma)$ switches from L to C. Then both $LC(s,\sigma)$ and γ go through $\bar q$ and coincide from that point on. So it suffices to compare the parts of $LC(s,\sigma)$ and γ up to $\bar q$. The time along $LC(s,\sigma)$ up to $\bar q$ is $s = \bar x - x$, and that along γ is $> \sqrt{(1+x)^2 + (1-y)^2} - \sqrt{2} + \frac{3\pi}{2}$. So we have to show that $\alpha > 0$, where $\alpha = \sqrt{(1+x)^2+(1-y)^2} - \sqrt{2} + \frac{3\pi}{2} - (\bar x - x) = \sqrt{(1+x)^2+(1-y)^2} - \sqrt{2} + \frac{3\pi}{2} + x + 1 + \sqrt{2-(y-1)^2}$. Then, using the fact that $\sqrt{a} + \sqrt{b} \geq \sqrt{a+b}$ when $a \geq 0$ and $b \geq 0$, we get the inequalities $\alpha \geq \sqrt{(1+x)^2+2} - \sqrt{2} + \frac{3\pi}{2} + x + 1 > \sqrt{(1+x)^2} + x + 1 = |1+x| + 1 + x \geq 0$, as desired.)

For $s > 0$, we let $L_0C(s)$ denote the trajectory that consists of a B_- arc during time s followed by a B_+ arc during time $\frac{\pi}{2}$, and ending at the origin. Then it is easily shown that *$L_0C(s)$ is optimal for every $s > 0$.* (Indeed, if $q = (x, 0)$ is the starting point of $L_0C(s)$, for $s > 0$, then $L_0C(s)$ is the only L_0C trajectory from q to $(0,0)$, and there are no simple trajectories from q to $(0,0)$ of the types C, LC, CL_0, LCL_0 or CL_0C. There is an L_0 arc, namely, $L_0(|x|)$, but we know that this arc is not optimal, because $|x| > 1$.)

Now define $R_1 = \{(0,0)\}$, $R_2 =]-1, 0[\times \{0\}$, $R_3 = \{(-1,0)\}$. For $q = (0,0)$, let γ_q be the trajectory γ with domain $\{0\}$ such that $\gamma(0) = (0,0)$. For $q = (x,0) \in R_2 \cup R_3$, let $\gamma_q = L_0(|x|)$. Then the γ_q, for $q \in R_1 \cup R_3 \cup R_3$, are optimal.

For $r > 0$, let $D(r)$ be the open disc with center $(-1,1)$ and radius r, and let $\partial D(r)$ denote the boundary of $D(r)$. Let $R_4 = D(1)$, $R_5 = \partial D(1) \backslash \{(-1,0)\}$. Then every point $q \in R_4$ can be joined to $(0,0)$ by a unique good LCL_0 trajectory. We use γ_q to denote this trajectory. Then *the arcs γ_q, for $q \in R_4$, are optimal.* (To see this, observe that if $q \in R_4$ then the only trajectories other than

γ_q that are of one of our seven types and go from q to the origin are either LCL_0 but not good, or LC with a switching from L to C at a time c that violates the requirement that $-\frac{\pi}{2} < c$. So all the alternative candidates are excluded, and γ_q must be the optimal trajectory.)

A similar argument shows that *if $q \in R_5$ and γ_q is the unique CL_0 trajectory going from q to $(0,0)$ then γ_q is optimal.*

Next, let $R_6 = D(\sqrt{2})\backslash(R_2 \cup R_3 \cup R_4 \cup R_5)$. If $q \in R_6$, then the obvious candidate for optimality is the arc γ_q of type CL_0 obtained by following the C arc that starts at q until R_2 is reached, and then continuing along R_2 towards the origin. We now prove that *the arcs γ_q, $q \in R_6$, are optimal.* To see this, observe that if $q = (x,y) \in R_6$ then there are no simple arcs from q to $(0,0)$ of the types C, L_0C, CL_0C. If $y \neq 0$ then there are no L_0 arcs either. If $y = 0$ then there is an L_0 arc, but it is $L_0(\beta)$ for $\beta = -|x| > 1$, so it is not optimal. If $y \neq 0$ then there is an LC arc, but (i) if $y < 0$ then this arc switches from L to C in $\Omega_{AB,4}$, so it is not optimal; (ii) if $y > 0$, then the arc is $LC(s,\sigma)$ for some $\sigma \in]\,\pi, 2\pi\,[$, so it is not optimal either. If $y \leq 0$ then γ_q is the only CL_0 arc, and if $y > 0$ then there are two such arcs, but the other one is not optimal, because it contains $L_0(\beta)$ for some $\beta > 1$. So we have ruled out all the alternative candidates to γ_q other than good LCL_0 arcs. We now exclude this possibility too, by doing it first for the case when $y \neq 2$. If $y \leq 0$ or $y > 2$ or $x > -1$, then there is no good LCL_0 arc. If $0 < y < 2$ and $x < -1$ then there is a good LCL_0 arc, but it switches from L to C at a point in $\Omega_{AB,2}$, so this arc is not optimal. So now all the alternative possibilities have been ruled out if $q = (x,y) \in R_6$ is such that $y \neq 2$ or $y = 2$ and $x > -1$. So γ_q is optimal for all such q's. Suppose now that $y = 2$ and $x < -1$. Then there is a good LCL_0 arc γ, which switches from L to C at the point $p = (-1,2)$. This arc can be ruled out in a number of ways, of which the following qualitative argument appears to be the shortest. Let γ be defined on $[T,0]$. Let α be a small positive number—actually, any α such that $0 < \alpha < \pi$ will do—and let $\tilde{\gamma} : [T-\alpha, 0] \to I\!R^2$ be the trajectory such that $\tilde{\gamma}(t) = \gamma(t)$ for $t \in [T,0]$ and $\tilde{\gamma}$ is a B_+ arc on $[T-\alpha, T]$. Let $(\tilde{x},\tilde{y}) = \tilde{q} = \tilde{\gamma}(T-\alpha)$. Then $\tilde{q}$ also belongs to R_6, but it's not true that $\tilde{y} = 2$ and $\tilde{x} < -1$. So we already know that $\gamma_{\tilde{q}}$ is the unique optimal arc from $\tilde{q}$ to the origin. Since γ_q is a cofinal piece of $\gamma_{\tilde{q}}$, the principle of optimality implies that γ_q is optimal as well. If γ was optimal, then γ_q and γ would go from q to 0 in the same time, so $\gamma_q(T) = \gamma_{\tilde{q}}(T) = q$. Let $[\tilde{T},0]$ be the domain of $\gamma_{\tilde{q}}$, and define $\delta : [\tilde{T},0] \to I\!R^2$ by letting δ agree with $\gamma_{\tilde{q}}$ on $[\tilde{T},T]$ and with γ on $[T,0]$. Then δ is an optimal arc from $\tilde{q}$ to 0, and δ is $CLCL_0$. This, however, is impossible, because we know that $CLCL_0$ is excluded. So we have now completed the proof that *for each $q \in R_6$, γ_q is the unique optimal arc from q to the origin.*

Next, we define five sets R_7, R_8, R_9, R_{10}, R_{11}, by letting $R_7 = \partial D(\sqrt{2}) \cap \Big(I\!R \times]\,0,+\infty\,[\Big)$, $R_8 = \Big\{(-2,0)\Big\}$, $R_9 = \Big\{(x,y) \in \partial D(\sqrt{2}) : -2 < x < -1\Big\}$, $R_{10} = \Big\{(-1, 1-\sqrt{2})\Big\}$ and, finally, $R_{11} = \Big\{(x,y) \in \partial D(\sqrt{2}) : -1 < x < 0\Big\}$. Then the sets R_1, R_7, R_8, R_9, R_{10} and R_{11} are pairwise disjoint and $D(\sqrt{2}) = R_1 \cup R_7 \cup R_8 \cup R_9 \cup R_{10} \cup R_{11}$. For $q \in R_7 \cup R_8 \cup R_9 \cup R_{10} \cup R_{11}$ there is a unique trajectory γ_q of type C going from q to $(0,0)$. The optimality of γ_q for $q \in R_7 \cup R_8 \cup R_9 \cup R_{10} \cup R_{11}$ is proved by considerations similar to those used for R_6.

We let $R_{12} =]-\infty, -2\,[\, \times \{0\}$, $R_{13} = \{(x,y) : 1-\sqrt{2} < y < 0\,,\, x < -1\,,\, (x+1)^2 + (1-y)^2 > 2\}$, $R_{14} =]-\infty, -1\,[\, \times \{1-\sqrt{2}\}$. Then $R_{13} \cup R_{14}$ is exactly the set of starting points q of the trajectories $LC(s,\sigma)$, for $s > 0$, $\frac{\pi}{4} \leq \sigma < \frac{\pi}{2}$, and we know that for each $q \in R_{13} \cup R_{14}$ the $LC(s,\sigma)$ trajectory is unique and optimal. We define γ_q to be this trajectory. A similar fact is true for R_{12}, with $L_0C(s)$ rather than $LC(s,\sigma)$ in the role of γ_q.

Finally, we let $R_{15} = I\!R^2 \backslash (R_1 \cup \ldots \cup R_{14})$. For $q \in R_{15}$, we let γ_q be the unique simple CL_0C trajectory from q to $(0,0)$. We show that γ_q *is optimal for all $q \in R_{15}$.* Let $q \in R_{15}$, and let $\gamma \in \mathcal{F}$ be an extremal trajectory starting at q and verifying $\gamma(0) = 0$. It is then obvious that γ is not of type L_0, C or L_0C. The case LC is impossible, because in that case γ would have a switching at a time c such that $c < -\pi/2$. If γ is of type CL_0 then it contains an $L_0(\beta)$ piece, with $\beta > 1$, and we know that any such piece is nonoptimal, so γ is not optimal. Finally, if γ is of type LCL_0 and good then $\gamma(t) = p = (x_p, y_p) \in R_5$ for some t. It is clear that $y_p > 0$. If $x_p > -1$, then γ is not simple, so this possibility is excluded. If $x_p \leq -1$, then we can pick a positive ε such that $\hat{q} = \gamma(t-\varepsilon)$ is in R_6. Then the restriction $\hat{\gamma}$ of γ to the interval $[t-\varepsilon, 0]$ is a good LCL_0 trajectory from $\hat{q}$ to 0. Since $\hat{q} \in I\!R_6$, we know that $\hat{\gamma}$ is not optimal. So γ is not optimal either, by the principle of optimality.

We have now completely analyzed the minimum time problem with target set $\{(0,0)\}$ for every initial condition q, and proved in all possible cases that γ_q is the unique solution whose terminal time is 0. So the synthesis $\Gamma = \{\gamma_q\}_{q \in I\!R^2}$ shown in Figure 2 is optimal, and is the unique optimal presynthesis. The partition $\mathcal{P} = (R_1, \ldots, R_{15})$ is a C^1—and, actually, C^ω—stratification of $I\!R^2$.

Let us define the feedback control $(x,y) \to v(x,y)$ by letting $v = 0$ on R_1, $v = -1$ on the set $R_2 \cup R_3 \cup R_4 \cup R_{12} \cup R_{13} \cup R_{14}$, and $v = +1$ on $R_5 \cup R_6 \cup R_7 \cup R_8 \cup R_9 \cup R_{10} \cup R_{11} \cup R_{15}$. We let $\mathcal{P}_1 = \{R_2, R_4, R_5, R_6, R_7, R_9, R_{11}, R_{12}, R_{13}, R_{14}, R_{15}\}$, $\mathcal{P}_2 = \{R_3, R_8, R_{10}\}$. We define $\Pi(R_2) = R_1$, $\Pi(R_4) = R_5$, $\Pi(R_5) = R_3$, $\Pi(R_6) = R_2$, $\Pi(R_7) = R_8$, $\Pi(R_9) = R_{10}$, $\Pi(R_{11}) = R_1$, $\Pi(R_{12}) = R_8$, $\Pi(R_{13}) = R_9$, $\Pi(R_{14}) = R_{10}$, $\Pi(R_{15}) = R_{12}$, $\Sigma(R_3) = R_2$, $\Sigma(R_8) = R_9$, $\Sigma(R_{10}) = R_{11}$.

Then $(\mathcal{P}, \mathcal{P}_1, \mathcal{P}_2, \Pi, \Sigma, v)$ is a regular synthesis that satisfies all the conditions of Brunovský's definition, except for the continuity requirement (Br.3.G) on the cost function V_Γ. On the other hand, it is easy to verify that V_Γ *satisfies the weak continuity conditions.* So $(\mathcal{P}, \mathcal{P}_1, \mathcal{P}_2, \Pi, \Sigma, v)$ gives rise to a "regular synthesis" in the sense of our definition.

Moreover, it is clear that *the* $+1$ *trajectory starting from the point* $(-1,0)$ *is an admissible trajectory for the optimal feedback control but is not optimal.* So this example illustrates a very important positive aspect of the Boltyanskii-Brunovský's definition, namely, the role of the "instantaneous exit map" Σ. To understand how Σ matters, let us suppose we did not specify Σ, and allowed all the trajectories of the discontinuous feedback v. Then we could, for example, start from the Type II cell $R_3 = \{(-1,0)\}$ and follow the trajectory corresponding to the control $+1$ up to time 2π, returning to the starting point, after which we could go to the origin using the control -1. The curve γ_1 defined in this way is also an admissible trajectory of v, but it is clear that γ_1 is clearly not time optimal. Even worse, we can define a trajectory γ_2 of v, with domain $[0, +\infty[$, corresponding to the constant control $+1$ and starting from $(-1,0)$. This trajectory neither reaches the origin in finite time nor approaches it as time goes to infinity. Notice that the trajectories γ_1 and γ_2 satisfy the equation $\dot{\gamma}(t) = f(\gamma(t), v(\gamma(t)))$ for every t in their domain, so they satisfy the equation corresponding to the discontinuous feedback v. However, if we regard it as part of the specification of our synthesis that Σ has to be given as well, then the difficulty disappears. $\diamondsuit$

Example 5.4. We now exhibit an optimal synthesis in Brunovský's sense that does not satisfy Condition (Bo.4) of Boltyanskii's definition. We consider the minimum time problem for the system:

$$\dot{x} = u\,, \qquad \dot{y} = v\,, \qquad \dot{z} = \frac{w}{\sqrt{1+x^2+y^2}}\,,$$

with target $\{(0,0,0)\}$ and control constraint $(u,v,w) \in U$, where U is the set of those $(u,v,w) \in \mathbb{R}^3$ such that $(u^2+v^2)^{1/2} + |w| \leq 1$, i. e. the solid of revolution obtained by rotating the square $\{(u,w) : |u| + |w| \leq 1\}$ in the u,w-plane about the w axis. Define the sets:

$$R_0 = \{(0,0,0)\}\,, \qquad R_3 = \{(x,y,z) : x^2+y^2 > 0,\ z = 0\}\,,$$
$$R_1 = \{(x,y,z) : x = y = 0,\ z > 0\}\,, \qquad R_4 = \{(x,y,z) : x^2+y^2 > 0,\ z > 0\}\,,$$
$$R_2 = \{(x,y,z) : x = y = 0,\ z < 0\}\,, \qquad R_5 = \{(x,y,z) : x^2+y^2 > 0,\ z < 0\}\,.$$

We then define a feedback control law $(u, v, w) : \mathbb{R}^3 \to U$ by specifying its restriction $R_i \ni (x, y, z) \to (u_i(x, y, z), v_i(x, y, z), w_i(x, y, z)) \in I\!R^3$ for every R_i as follows:

$$\begin{aligned} &u_0 = v_0 = w_0 = 0\,, &&u_3 = u_4 = u_5 = -\frac{x}{\sqrt{x^2+y^2}}\,, \\ &u_1 = v_1 = 0,\ w_1 = -1\,, &&v_3 = v_4 = v_5 = -\frac{y}{\sqrt{x^2+y^2}}\,, \\ &u_2 = v_2 = 0,\ w_2 = 1\,, &&w_3 = w_4 = w_5 = 0\,. \end{aligned} \tag{5.18}$$

If we follow the trajectories of this feedback starting from a point $(x, y, z) \in R_3 \cup R_4 \cup R_5$, we move horizontally towards the z axis with speed 1, and then, if we are not yet at the origin, we move to the origin along the z axis with speed 1. This obviously gives rise to a synthesis Γ, and it is possible to prove that Γ is optimal by applying Theorem 2.13. However, in this case it is easy to verify directly that the associated cost function

$$V_\Gamma = |z| + \sqrt{x^2+y^2}$$

is the value function of our problem. Indeed, since V_Γ is the cost function arising from the synthesis Γ, all we need is to show that the cost $b - a$ of any trajectory $[a, b] \ni t \to \xi(t) = (x(t), y(t), z(t))$ from a point $\bar{q} = (\bar{x}, \bar{y}, \bar{z})$ to the origin is bounded below by $V_\Gamma(\bar{q})$.

Assume first that $x(t)^2 + y(t)^2 > 0$ for all $t \in]\,a, b\,[$, and let $t \to (u(t), v(t), w(t))$ be the corresponding control. Write $\rho(t) = \sqrt{u(t)^2 + v(t)^2}$. Then an elementary calculation shows that

$$\begin{aligned} < \frac{d}{dt}\Big(V_\Gamma(\xi(t))\Big) &= \frac{u(t).x(t) + v(t).y(t)}{\sqrt{x(t)^2+y(t)^2}} + \operatorname{sgn}(z(t)).\frac{w(t)}{\sqrt{1+x(t)^2+y(t)^2}} \\ &\geq -\rho(t) - |w(t)| \\ &\geq -1\,. \end{aligned}$$

Therefore, since the function $t \to V_\Gamma(\xi(t))$ is Lipschitz, and hence absolutely continuous, we can conclude that

$$V_\Gamma(\bar{q}) = -\int_a^b \frac{d}{dt}\Big(V_\Gamma(\xi(t))\Big)\,dt \leq b - a\,.$$

Now, if $\xi : [a, b] \to \mathbb{R}^3$ is an arbitrary trajectory from $\bar{q}$ to the origin, then ξ can be approximated by a sequence of trajectories $\xi_j : [a_j, b_j] \to \mathbb{R}^3$ from $\bar{q}$ to $(0,0,0)$ such that $\xi_j(t)$ is not in the z axis for $a_j < t < b_j$. Then $b_j - a_j \geq V_\Gamma(\bar{q})$ for all j, so $b - a \geq V_\Gamma(\bar{q})$. Therefore Γ is an optimal synthesis.

Notice that in this case $\Pi(R_4) = R_1$, $\Pi(R_5) = R_2$, and $\Pi(R_3) = R_0$. So the optimal synthesis Γ does not satisfy the conditions of Boltyanskii's definition, because (Bo.4) is violated, since the dimension of the exit manifold from R_i is equal to $\dim(R_i) - 2$ for $i = 3, 4, 5$.

Remark 5.5. In Example 5.4 the optimal pair (γ_q, η_q) starting at q is unique for every q, and is a classical solution of the closed-loop equation corresponding to the optimal feedback law defined by 5.18.

On the other hand, it is easy to see that *the optimal trajectories are not Filippov solutions of the optimal closed-loop equation.* Finally, we point out that, *although the optimal trajectories are CLSS solutions, it not true that all CLSS solutions of the optimal closed-loop equation are optimal trajectories.* For example, if the initial condition is a point q which is not in the z axis and also not in the plane $z = 0$, then most CLSS solutions $t \to \gamma(t)$ will never switch at a time t such that $\gamma(t)$ lies in the z axis, and will therefore be entirely contained in a plane $z =$ constant, without ever reaching the target. ♢

Example 5.6. We now exhibit an example of an optimal synthesis for which it is not true that all the optimal trajectories are CLSS solutions of the optimal closed-loop equation.

Consider the optimal control problem in $\mathbb{R}$, with control space $U = \mathbb{R}$, dynamics given by $\dot{x} = u$, Lagrangian $L = (u^4 - x^2)^2$, and target $\{1\}$. The value function V is given by $V(x) \equiv 0$. The feedback control law v defined by

$$v(x) = \begin{cases} \sqrt{|x|} & \text{if} \quad x < 1\,, \\ -\sqrt{x} & \text{if} \quad x > 1\,, \\ 0 & \text{if} \quad x = 1 \end{cases}$$

leads to trajectories $\gamma_x : [T_x, 0] \to \mathbb{R}$, corresponding to open-loop controls $\eta_x : [T_x, 0] \to \mathbb{R}$, given by the formulas

$$\begin{aligned} T_x = \;& 2.\mathrm{sgn}(1-x).\Big(\mathrm{sgn}(x).\sqrt{|x|} - 1\Big)\,, \\ \gamma_x(t) = \;& \begin{cases} \frac{(2+t)^2}{4} & \text{if} \quad x \le 1 \quad \text{and} \quad \max(T_x, -2) \le t \le 0\,, \\ -\frac{(2+t)^2}{4} & \text{if} \quad x \le 1 \quad \text{and} \quad T_x \le t \le -2\,, \\ \frac{(2-t)^2}{4} & \text{if} \quad x \ge 1 \quad \text{and} \quad T_x \le t \le 0\,, \end{cases} \\ \eta_x(t) = \;& \begin{cases} \frac{1}{2}|2+t| & \text{if} \quad x \le 1 \quad \text{and} \quad T_x \le t \le 0\,, \\ \frac{1}{2}(2-t) & \text{if} \quad x \ge 1 \quad \text{and} \quad T_x \le t \le 0\,. \end{cases} \end{aligned}$$

It is easy to see that the family $\Gamma = \{(\gamma_x, \eta_x)\}_{x \in \mathbb{R}}$ is a regular synthesis in the sense of our definition. Moreover, although the optimal trajectories are not unique, it is not hard to show that Γ *is the only synthesis for our problem.*

If we let $f(x, u) = u$, then the arcs γ_x are classical solutions of the closed-loop equation $\dot{x} = f(x, v(x))$. On the other hand, *the trajectory γ_0 corresponding to the initial condition $x = 0$ is not a CLSS solution of the closed-loop equation $\dot{x} = f(x, v(x))$* because, on an interval $[a, b]$, the only CLSS solution γ of this equation with initial condition $\gamma(a) = 0$ is $\gamma(t) \equiv 0$. ♢

Appendix A. Differentiation of trajectories with respect to a parameter

The purpose of this appendix is to present a general theorem on differentiation with respect to a parameter p, taking values in a normed space P, of a family of trajectories $x_p : [a, b] \to \mathbb{R}^n$ of time-varying vector fields f_p. We will state the theorem under minimal hypotheses, much weaker than what is actually needed for the main results of this paper, because this result has other applications, such as the theory of envelopes (cf. Sussmann [26]), where the more general statement is useful.

If $\dot{x}_p(t) = f_p(x_p(t), t)$, then formal differentiation with respect to p at $p = p_0$ in the direction of a vector v yields the variational equation

$$\dot{y}_{p_0,v}(t) = \frac{\partial f_{p_0}}{\partial x}\Big(x^{p_0}(t), t\Big).y_{p_0,v}(t) + \lim_{\varepsilon \downarrow 0} \frac{w_{p_0+\varepsilon v}(t) - w_{p_0}(t)}{\varepsilon}\,, \tag{A.1}$$

where $y_{p_0,v}(t)$ is the directional derivative of $x_p(t)$ at p_0 in the v-direction, and $w_p(t) = f_p(x_{p_0}(t), t)$.

The technical problem that will concern us here is to make (A.1) rigorous even when the limit in the right-hand side only exists in a weak sense. It can then happen that the variational equation is a linear equation $\dot{y}(t) = A(t)y(t) + w(t)$ where the "input" w is the formal derivative of a function W which is not necessarily absolutely continuous or even continuous. So our first task will be to study the solutions of linear systems with "generalized inputs" w.

Let $-\infty < a < b < \infty$, and let n be a positive integer. If $A \in L^1([a,b]; I\!R^{n\times n})$, and $\bar{Y} \in I\!R^n$, then it is well known that the initial value problem

$$\dot{Y}(t) = A(t)\cdot Y(t) + w(t)\,, \qquad Y(a) = \bar{Y} \tag{A.2}$$

has a unique solution for every initial condition and every map $w \in L^1([a,b], I\!R^n)$. Since (A.2) can also be written formally as

$$dY(t) = A(t)\cdot Y(t)\,dt + w(t)\,dt\,, \qquad Y(a) = \bar{Y}\,, \tag{A.3}$$

we will use $Y_{A,w\,dt,\bar{Y}}$ to denote the unique solution of (A.2).

As explained before, we want to solve (A.2) for "inputs" w more general than integrable functions. The appropriate class of inputs turns out to be that of *formal derivatives of bounded measurable functions.* Precisely, we let $BM([a,b], I\!R^n)$ denote the space of all bounded Lebesgue measurable functions $W : [a,b] \to I\!R^n$, and use $BM_0([a,b], I\!R^n)$ to denote the set of those W in $BM([a,b], I\!R^n)$ such that $W(a) = 0$. (In $BM([a,b], I\!R^n)$ we do *not* identify two functions that are equal almost everywhere.)

We endow $BM([a,b], I\!R^n)$ with the norm $||W||_{BM} \stackrel{\text{def}}{=} \sup\{||W(t)|| : t \in [a,b]\}$. Then $BM([a,b], I\!R^n)$ is a Banach space, and $BM_0([a,b], I\!R^n)$ is a closed subspace of $BM([a,b], I\!R^n)$. The space $C^0([a,b], I\!R^n)$ of continuous maps $W : [a,b] \to I\!R^n$ is of course a closed subspace of $BM([a,b], I\!R^n)$, but $L^\infty([a,b], I\!R^n)$ is *not* a subspace of $BM([a,b], I\!R^n)$, because in $L^\infty([a,b], I\!R^n)$ we identify functions that are equal a.e., but in $BM([a,b], I\!R^n)$ we do not.

If $W \in BM_0([a,b], I\!R^n)$ is absolutely continuous, and $w = \dot{W}$, so that, formally, $dW = w\,dt$, then we can rewrite (A.2) formally as

$$dY(t) = A(t)\cdot Y(t)\,dt + dW(t) \qquad Y(a) = \bar{Y}\,. \tag{A.4}$$

A solution of (A.4) is then a continuous map $t \to Y(t)$ such that

$$Y(t) = \bar{Y} + \int_a^t A(s)\cdot Y(s)\,ds + W(t) \qquad \text{for all} \quad t \in [a,b]\,. \tag{A.5}$$

Now (A.5) makes sense for arbitrary $Y \in BM([a,b], I\!R^n)$, $W \in BM_0([a,b], I\!R^n)$, even if W is not absolutely continuous. So we turn it into a definition: for a pair $(W, \bar{Y}) \in BM_0([a,b], I\!R^n) \times I\!R^n$, a *solution* of (A.4) is a map $Y \in BM([a,b], R^n)$ such that (A.5) holds. It then follows that

(EU) *for every $W \in BM_0([a,b], I\!R^n)$ and every $\bar{Y} \in I\!R^n$ there exists a unique solution Y of (A.4).*

Indeed, uniqueness is trivial, since the difference $Y = Y^1 - Y^2$ of two solutions is a solution of the homogeneous problem (i.e. of (A.2) with $w \equiv 0$ and $\bar{Y} = 0$), so $Y^1 - Y^2 \equiv 0$. To prove existence, we observe that (A.5) is equivalent to

$$Y(t) - W(t) = \bar{Y} + \int_a^t A(s)\cdot(Y(s) - W(s))\,ds + \int_a^t A(s)\cdot W(s)\,ds \qquad \text{for all} \quad t \in [a,b]\,, \tag{A.6}$$

i. e. to $dZ = AZ\,dt + AW\,dt$, $Z(a) = \bar{Y}$, with $Z = Y - W$. Since $AW \in L^1$, because $A \in L^1$ and W is bounded and measurable, the equation $\dot{Z}(t) = A(t)\cdot Z(t) + A(t)W(t)$ has a unique solution Z^* for which $Z^*(a) = \bar{Y}$. If we let $Y(t) = Z^*(t) + W(t)$, and recall that $W(a) = 0$, we see that Y is the desired solution of (A.4).

We will use $Y_{A,dW,\bar{Y}}$ to denote the solution of (A.4), whose existence and uniqueness has just been established. In addition to proving (EU), the preceding argument also gives an explicit formula for $Y_{A,dW,\bar{Y}}$, namely:

$$Y_{A,dW,\bar{Y}} = W + Y_{A,AW\,dt,\bar{Y}}\,. \tag{A.7}$$

Clearly, the function $Y_{A,AW\,dt,\bar{Y}}$ is absolutely continuous, so the regularity properties of $Y_{A,dW,\bar{Y}}$—modulo absolutely continuous functions—are exactly the same as those of W. (For example: Y is continuous, right-continuous, or left-continuous at a point $t \in [a,b]$ if and only if W is; Y is of bounded variation iff W is; Y is absolutely continuous iff W is.)

Gronwall's inequality, applied to (A.6), tells us that

$$||Y_{A,dW,\bar{Y}}(t) - W(t)|| \le e^{||A||_{L^1}}\Big(||\bar{Y}|| + ||A||_{L^1}||W||_{BM}\Big)\,. \tag{A.8}$$

Therefore

$$||Y_{A,dW,\bar{Y}}||_{BM} \le ||W||_{BM} + e^{||A||_{L^1}}\Big(||\bar{Y}|| + ||A||_{L^1}||W||_{BM}\Big)\,. \tag{A.9}$$

On the other hand, (A.5) also implies that

$$||W||_{BM} \le ||\bar{Y}|| + ||Y_{A,dW,\bar{Y}}||_{BM}\Big(1 + ||A||_{L^1}\Big)\,. \tag{A.10}$$

It is not hard to prove the following technical fact:

Theorem A.1. *Let $\{W_j\}$, $\{\bar{Y}_j\}$, $\{A_j\}$ be sequences in $BM_0([a,b],\mathbb{R}^n)$, $\mathbb{R}^n$, $L^1([a,b],\mathbb{R}^{n\times n})$, respectively. Assume that $\{A_j\}$ is bounded in $L^1([a,b],\mathbb{R}^{n\times n})$. Let $Y_j = Y_{A_j,dW_j,\bar{Y}_j}$. Then*

(A.1.1) *$\{Y_j\}$ is bounded in $BM([a,b],\mathbb{R}^n)$ if and only if $\{W_j\}$ is bounded in $BM([a,b],\mathbb{R}^n)$ and $\{\bar{Y}_j\}$ is bounded in $\mathbb{R}^n$.*

(A.1.2) *If $\{A_j\}$ converges in $L^1([a,b],\mathbb{R}^n)$ to a limit A, $\bar{Y} \in \mathbb{R}^n$, $\bar{Y}_j \to \bar{Y}$, $\{W_j\}$ is bounded in $BM([a,b],\mathbb{R}^n)$, $W \in BM_0([a,b],\mathbb{R}^n)$, and $Y = Y_{A,dW,\bar{Y}}$, then $Y_j \to Y$ almost everywhere iff $W_j \to W$ almost everywhere. In that case, (a) $Y_j - W_j$ converges to $Y - W$ uniformly on $[a,b]$ and, in particular, (b) for every $t \in [a,b]$, $Y_j(t) \to Y(t)$ iff $W_j(t) \to W(t)$.* $\diamondsuit$

A very important subspace of $BM([a,b],\mathbb{R}^n)$ is $BV([a,b],\mathbb{R}^n)$, the space of all functions $W : [a,b] \to \mathbb{R}^n$ such that (a) W is of bounded variation, (b) W is right-continuous at every point of $]a,b]$, and (c) $W(a) = 0$. There is a canonical correspondence between $BV([a,b],\mathbb{R}^n)$ and the dual space $C^0([a,b],\mathbb{R}_n)^*$ of the Banach space $C^0([a,b],\mathbb{R}_n)$ of continuous functions $[a,b] \to \mathbb{R}_n$. (We write R_n rather than R^n because we want to think of the members of $C^0([a,b],\mathbb{R}_n)$ as *row-vector-valued functions*, cf. below.)

The members of $C^0([a,b],\mathbb{R}_n)^*$ are the *$\mathbb{R}^n$-valued Borel measures on* $[a,b]$. When $n = 1$, they are the *finite signed Borel measures on* $[a,b]$. The identification map from $BV([a,b],\mathbb{R}^n)$ to $C^0([a,b],\mathbb{R}_n)^*$ is the one that assigns to a function $W \in BV([a,b],\mathbb{R}^n)$ the unique Borel measure μ_W such that $\mu_W([a,t]) = W(t)$ for $a < t \le b$.

From now on we identify a function $W \in BV([a,b],\mathbb{R}^n)$ with its corresponding measure μ_W, and write $\int \alpha \cdot dW$ for $\int \alpha \cdot d\mu_W$. Clearly, $\int \alpha \cdot dW$ is defined, more generally, for an arbitrary bounded Borel measurable $\mathbb{R}_n$-valued map α on $[a,b]$.

Naturally, the integral $\int \alpha\, dW$ of a scalar bounded Borel measurable function $\alpha : [a,b] \to \mathbb{R}$ is also well defined, and is a vector in $\mathbb{R}^n$.

On $BV([a,b],R^n)$ there are at least three important topologies: (a) the one induced by the norm $||\ldots||_{BM}$, (b) the strong topology of $BV([a,b],\mathbb{R}^n)$ as the dual of the Banach space $C^0([a,b],\mathbb{R}_n)$,

(c) the weak* topology, also corresponding to the duality with $C^0([a,b],\mathbb{R}_n)$. The second one is induced by the *total variation norm*:

$$||W||_{TV} = \sup\left\{ \left| \int \alpha \cdot dW \right| : \alpha \in C^0([a,b],\mathbb{R}_n)\,,\ ||\alpha||_{sup} \leq 1 \right\}. \tag{A.11}$$

It is clear that $||W||_{BM} \leq ||W||_{TV}$ for all W in $BV([a,b],\mathbb{R}^n)$. Therefore, *if a subset S of $BV([a,b],\mathbb{R}^n)$ is bounded in TV norm then S is also bounded in BM norm.*

We will need the following result, whose proof is carried out using the uniform boundedness theorem and standard compactness properties:

Lemma A.2. *If a sequence $\{W_j\}$ in $BV([a,b],\mathbb{R}^n)$ weak*-converges to a $W \in BV([a,b],\mathbb{R}^n)$, then (a) $\{W_j\}$ is bounded in TV norm, (b) $W_j(a) \to W(a)$ and $W_j(b) \to W(b)$, (c) every subsequence of $\{W_j\}$ has a subsequence that converges to W at all but at most countably many points of $[a,b]$, (d) $\int_a^b \beta(s)||W_j(t) - W(t)||\,dt \to 0$ for every function $\beta \in L^1([a,b],\mathbb{R})$.* ♢

As a corollary of Lemma A.2, we get the integration by parts formula:

Lemma A.3. *If $\lambda : [a,b] \to \mathbb{R}_n$ is absolutely continuous, and $W \in BV([a,b],\mathbb{R}^n)$, then*

$$\lambda(b) \cdot W(b) = \int \lambda \cdot dW + \int_a^b \dot{\lambda}(t) \cdot W(t)\,dt\,. \tag{A.12}$$

PROOF. The result is obviously true, because of the standard product rule for derivatives, if W is also absolutely continuous. In the general case, we use the fact that an arbitrary $W \in BV([a,b],\mathbb{R}^n)$ is the weak*-limit of a sequence $\{W_j\}$ of absolutely continuous functions. It then follows that (i) $\lambda(b) \cdot W_j(b) \to \lambda(b) \cdot W(b)$, because $W_j(b) \to W(b)$, (ii) $\int \lambda \cdot dW_j \to \int \lambda \cdot dW$, because $\{W_j\}$ weak*-converges to W, and λ is continuous, and (iii) $\int_a^b \dot{\lambda}(t) \cdot W_j(t)\,dt \to \int_a^b \dot{\lambda}(t) \cdot W(t)\,dt$, because Part (d) of Lemma A.2 implies that $\int_a^b ||\dot{\lambda}(t)||.||W_j(t) - W(t)||\,dt \to 0$, since $\dot{\lambda} \in L^1$. ♢

An important consequence of Lemma A.3 is the following

Lemma A.4. *Let $W \in BV([a,b],\mathbb{R}^n)$, $A \in L^1([a,b],R^{n\times n})$, $\bar{Y} \in \mathbb{R}^n$, and write $Y = Y_{A,dW,\bar{Y}}$. Suppose $\lambda : [a,b] \to \mathbb{R}_n$ is absolutely continuous and satisfies $\dot{\lambda}(t) = -\lambda(t)A(t)$ for almost all t. Then*

$$\lambda(b) \cdot Y(b) - \lambda(a) \cdot Y(a) = \int \lambda \cdot dW\,. \tag{A.13}$$

PROOF. Write $Y(t) = V(t) + W(t)$, where $V(t) = \bar{Y} + \int_a^t A(s).Y(s)\,ds$. Then

$$\lambda(b) \cdot Y(b) - \lambda(a) \cdot Y(a) = \lambda(b) \cdot V(b) - \lambda(a) \cdot V(a) + \lambda(b) \cdot W(b)\,. \tag{A.14}$$

The function $t \to \lambda(t) \cdot V(t)$ is absolutely continuous and its derivative is $\dot{\lambda}(t) \cdot V(t) + \lambda(t) \cdot \dot{V}(t)$, which equals $-\lambda(t)A(t)Y(t) + \lambda(t)A(t)Y(t) - \dot{\lambda}(t) \cdot W(t)$, i.e. $-\dot{\lambda}(t) \cdot W(t)$. So

$$\lambda(b) \cdot Y(b) - \lambda(a) \cdot Y(a) = \lambda(b) \cdot W(b) - \int_a^b \dot{\lambda}(t) \cdot W(t)\,dt\,. \tag{A.15}$$

Equation (A.13) follows from this and the integration by parts formula (A.12). ♢

It is clear that, if $A \in L^1([a,b],\mathbb{R}^{n\times n})$, $\bar{Y} \in \mathbb{R}^n$, and $W \in BM_0([a,b],\mathbb{R}^n)$, then $Y_{A,dW,\bar{Y}} - W$ is absolutely continuous, so $W \in BV([a,b],\mathbb{R}^n)$ iff $Y_{A,dW,\bar{Y}}$ is of bounded variation, which happens iff $Y_{A,dW,\bar{Y}} - \bar{Y} \in BV([a,b],\mathbb{R}^n)$.

The following theorem is proved by repeated use of Gronwall's inequality:

Theorem A.5. *Let $\{W_j\}$, $\{\bar{Y}_j\}$, $\{A_j\}$ be sequences in $BV([a,b],\mathbb{R}^n)$, $\mathbb{R}^n$, $L^1([a,b],\mathbb{R}^{n\times n})$, respectively. Assume that $\{A_j\}$ is bounded in $L^1([a,b],\mathbb{R}^{n\times n})$ and $\{\bar{Y}_j\}$ is bounded in $\mathbb{R}^n$. Let $Y_j = Y_{A_j,dW_j,\bar{Y}_j}$. Then*

(A.5.1) *$\{Y_j - \bar{Y}_j\}$ is bounded in $BV([a,b],\mathbb{R}^n)$ if and only if $\{W_j\}$ is bounded in $BV([a,b],\mathbb{R}^n)$.*

(A.5.2) *If $\{A_j\}$ converges in $L^1([a,b],\mathbb{R}^n)$ to a limit A, $\bar{Y}_j \to \bar{Y}$, $W \in BV([a,b],\mathbb{R}^n)$, and we write $Y = Y_{A,dW,\bar{Y}}$, then $\{W_j\}$ weak*-converges to W iff $\{Y_j - \bar{Y}_j\}$ weak*-converges to $Y - \bar{Y}$. In that case, $Y_j - W_j - \bar{Y}_j$ converges to $Y - W - \bar{Y}$ strongly in $BV([a,b],\mathbb{R}^n)$. In particular, (a) $W_j(b) \to W(b)$, (b) $Y_j(b) \to Y(b)$, and (c) for every $t \in [a,b]$, $Y_j(t) \to Y(t)$ iff $W_j(t) \to W(t)$.* $\diamondsuit$

Recall that, if P, Q are normed spaces, and $F : \Omega \to Q$ is a map defined on a neighborhood Ω of a point p of P, then F is said to be *Fréchet differentiable* at p if there exists a bounded linear map L—called the *Fréchet differential* of F at p—such that $F(p') = F(p) + L(p'-p) + o(||p'-p||)$ as $p' \to p$. If X is a Banach space and $Q = X^*$, then we call F *weak* Fréchet differentiable* at p if for every $x \in X$ the map $\Omega \ni p' \to \langle F(p'), x\rangle \in \mathbb{R}$ is Fréchet differentiable at p. (We use $\langle x^*, x\rangle$, for $x \in X$, $x^* \in X^*$, as an alternative notation for $x^*(x)$.)

Suppose P is a normed space, X is a Banach space, $p \in P$, $F : \Omega \to X^*$ is a map defined on a neighborhood Ω of p, and F is weak* Fréchet differentiable at p. Then for each $x \in X$ there is a bounded linear functional θ_x on P such that $\langle F(p+v), x\rangle = \langle F(p), x\rangle + \theta_x(v) + o(||v||)$ as $v \to 0$ in P. Clearly, θ_x is unique, and is given by the formula

$$\theta_x(v) = \lim_{h\to 0} h^{-1}\langle F(p+hv) - F(p), x\rangle . \tag{A.16}$$

For any given $v \in P$, use θ^v to denote the map $x \to \theta_x(v)$. Then θ^v is a limit of linear maps, so it is linear. If $\{h_j\}$ is any sequence going to 0, then the sequence of linear functionals ν_j on X given by $\nu_j(x) = h_j^{-1}\langle F(p+h_j v) - F(p), x\rangle$ converges pointwise to θ^v. Since each ν_j is a bounded linear functional on X, it follows for the uniform boundedness theorem that θ^v is bounded. Since $\theta^v(x) = \theta_x(v)$, which is linear with respect to v, we can conclude that $v \to \theta^v$ is a linear map from P to X^*. We use $D^wF(p)$ to denote this map, and $D^wF(p).v$ to denote its value for a particular v, so $D^wF(p).v = \theta^v$, and

$$\langle D^wF(p).v, x\rangle = \lim_{h\to 0} h^{-1}\langle F(p+hv) - F(p), x\rangle . \tag{A.17}$$

The following lemma follows by applying the uniform boundedness principle:

Lemma A.6. *If P is a normed space, X is a Banach space, $p \in P$, $F : \Omega \to X^*$ is a map defined on a neighborhood Ω of p, and F is weak* Fréchet differentiable at p. Then:*

(A.6.i) *there is a $C > 0$ such that*

$$||F(p+v) - F(p)|| \le C||v|| \qquad \textit{for all sufficiently small} \qquad v \in P\,, \tag{A.18}$$

(A.6.ii) *the linear map $D^wF(p) : P \to X^*$ defined by (A.17) is bounded.* $\diamondsuit$

We are now ready to study the differentiability properties with respect to a parameter p, belonging to a normed space P, of a family of trajectories $x_p : [a,b] \to I\!R^n$ of time-varying vector fields f_p. We will prove, under minimal technical hypotheses, that if the initial condition map $p \to x_p(a)$ is differentiable at $p = 0$, and the map $p \to f_p$ satisfies a "weak differentiability condition" at the trajectory x_0, then the map $p \to x_p$ is weak*- differentiable at $p = 0$, as a map into $BV([a,b], I\!R^n)$, and in particular the endpoint map $p \to x_p(b)$ is differentiable at $p = 0$.

A *curve in* $I\!R^n$ is a continuous map $\xi : I \to I\!R^n$, whose domain $I = \mathrm{Dom}(\xi)$ is a nonempty interval. A curve ξ is an *arc* if $\mathrm{Dom}(\xi)$ is compact. The *graph* $G(\xi)$ of a curve ξ is the set $G(\xi) = \{(\xi(t), t) : t \in \mathrm{Dom}(\xi)\}$. If $\xi : [a,b] \to I\!R^n$ is an arc in $I\!R^n$, and $\varepsilon > 0$, then the ε*-tube about* ξ is the set $\mathcal{T}(\xi,\varepsilon) = \{(x,t) : x \in I\!R^n, a \le t \le b, ||x - \xi(t)|| \le \varepsilon\}$.

A *time-varying vector field* on $I\!R^n$ is an $I\!R^n$-valued map f whose domain $\mathrm{Dom}(f)$ is a—possibly empty—subset S of the product $I\!R^n \times I\!R$. We use $TVVF(I\!R^n)$ to denote the set of all time-varying vector fields on $I\!R^n$. A *trajectory*, or *integral curve*, of an $f \in TVVF(I\!R^n)$ is a locally absolutely continuous curve ξ in $I\!R^n$ such that $G(\xi) \subseteq \mathrm{Dom}(f)$ and $\dot\xi(t) = f(\xi(t), t)$ for almost all $t \in \mathrm{Dom}(\xi)$. We use $\mathrm{Traj}(f)$ to denote the set of all trajectories of f, and $\mathrm{Traj}_c(f)$ to denote the set of all $\xi \in \mathrm{Traj}(f)$ such that $\mathrm{Dom}(\xi)$ is compact.

If $f \in TVVF(I\!R^n)$ and $\xi \in \mathrm{Traj}(f)$, we say that ξ is a *maximal* trajectory of f if it cannot be extended to a trajectory $\tilde\xi : \tilde I \to I\!R^n$ of f defined on an interval $\tilde I$ such that $I \subseteq \tilde I$ but $I \ne \tilde I$. We use $\mathrm{MTraj}(f)$ to denote the set of all maximal trajectories of f. Given $\bar x \in I\!R^n$, $\bar t \in I\!R$, we use $\mathrm{Traj}(f, \bar x, \bar t)$, $\mathrm{MTraj}(f, \bar x, \bar t)$ to denote, respectively, the set of all $\xi \in \mathrm{Traj}(f)$ such that $\xi(\bar t) = \bar x$, and the set $\mathrm{Traj}(f, \bar x, \bar t) \cap \mathrm{MTraj}(f)$. Zorn's lemma implies that every $\xi \in \mathrm{Traj}(f, \bar x, \bar t)$ can be extended to a $\tilde\xi \in \mathrm{MTraj}(f, \bar x, \bar t)$, so $\mathrm{Traj}(f, \bar x, \bar t) \ne \emptyset$ iff $\mathrm{MTraj}(f, \bar x, \bar t) \ne \emptyset$. Clearly, $\mathrm{Traj}(f, \bar x, \bar t) \ne \emptyset$ iff $(\bar x, \bar t) \in \mathrm{Dom}(f)$. (Indeed, if $(\bar x, \bar t) \in \mathrm{Dom}(f)$, then the map ξ with domain $\{\bar t\}$ such that $\xi(\bar t) = \bar x$ is in $\mathrm{Traj}(f)$.)

We say that an $f \in TVVF(I\!R^n)$ has the *forward existence property* at a point $(\bar x, \bar t)$ if for some $\varepsilon > 0$ there exists a $\xi \in \mathrm{Traj}(f, \bar x, \bar t)$ with domain $[\bar t, \bar t + \varepsilon]$. We say that f has the *forward limit property* on a subset S of $\mathrm{Dom}(f)$ if, whenever $\xi : [a, b[\to I\!R^n$ is a trajectory of f which is contained in a compact subset of S, it follows that $\lim_{t\uparrow b} \xi(t)$ exists.

Suppose we are given the following data:

(D.1) a normed space P,
(D.2) a family $\mathbf{f} = \{f_p\}_{p\in P}$ of time-varying vector fields on $I\!R^n$, depending on a parameter $p \in P$,
(D.3) a $p_0 \in P$,
(D.4) a trajectory $\xi : [a,b] \to I\!R^n$ of f_{p_0}.

We then define functions $w_p : [a,b] \to I\!R^n$ by

$$w_p(t) = f_p(\xi(t), t) - f_{p_0}(\xi(t), t) \quad \text{for} \quad t \in [a,b]\,, \tag{A.19}$$

so w_p is defined on $[a,b]$ whenever $G(\xi) \subseteq \mathrm{Dom}(f_p)$. If w_p is integrable, we let

$$W_p(t) = \int_a^t w_p(s)\,ds \qquad \text{for} \qquad t \in [a,b]\,. \tag{A.20}$$

We then let $\mathcal{W}$ be the map that sends p to $\mathcal{W}(p) = W_p$, with domain

$$\mathrm{Dom}(\mathcal{W}) = \{p \in P : G(\xi) \subseteq \mathrm{Dom}(f_p),\ w_p \in L^1([a,b], I\!R^n)\}\,. \tag{A.21}$$

We regard $\mathcal{W}$ as a map from $\mathrm{Dom}(\mathcal{W})$ to $BV([a,b], I\!R^n)$.

We then say that $\mathbf{f}$ is *weakly differentiable at* p_0 *along* ξ if there exists $\bar\varepsilon > 0$ such that

(WD1) $\mathcal{T}(\xi,\bar{\varepsilon}) \subseteq \mathrm{Dom}(f_p)$ *for every p such that* $\|p-p_0\| \le \bar{\varepsilon}$.

(WD2) *For every p such that $\|p-p_0\| \le \bar{\varepsilon}$, f_p has the forward limit property on $\mathcal{T}(\xi,\bar{\varepsilon})$ and the forward existence property at every (x,t) such that $a \le t < b$ and $\|x-\xi(t)\| < \bar{\varepsilon}$.*

(WD3) *There exist $A \in L^1([a,b],\mathbb{R}^{n\times n})$, and functions $\psi_\varepsilon \in L^1([a,b],\mathbb{R})$, for $\varepsilon \in]0,\bar{\varepsilon}]$, such that*

$$\|f_p(x,t)-f_p(\xi(t),t)-A(t).(x-\xi(t))\| \le \psi_\varepsilon(t)(\|x-\xi(t)\|+\|p-p_0\|) \qquad (A.22)$$

whenever $a \le t \le b$, $\|x-\xi(t)\| \le \varepsilon$, $\|p-p_0\| \le \varepsilon$, and

$$\lim_{\varepsilon \downarrow 0} \int_a^b \psi_\varepsilon(t)\,dt = 0\,. \qquad (A.23)$$

(WD4) *The function w_p defined by (A.19) is integrable for every p such that $\|p-p_0\| \le \bar{\varepsilon}$ and, if $\mathcal{W}$ is defined as above, then $\mathcal{W}$ is weak* Fréchet differentiable at p_0.*

The second part of condition (WD4) says that for each $\alpha \in C^0([a,b],\mathbb{R}_n)$ the map $p \to \int \alpha \cdot dW_p$ is Fréchet differentiable at p_0. Lemma A.6 then implies—since $W_{p_0}=0$—that there exists a bounded linear map $D^w\mathcal{W}(p_0): P \to BV([a,b],\mathbb{R}^n)$ having the property that

$$\int \alpha\cdot dW_p = \langle \mathcal{W}(p)-\mathcal{W}(p_0),\alpha\rangle = \int \alpha\cdot\Big(D^w\mathcal{W}(p_0).(p-p_0)\Big)+o(\|p-p_0\|) \qquad \text{as} \qquad p\to p_0 \quad (A.24)$$

for each $\alpha \in C^0([a,b],\mathbb{R}_n)$, and also that there is a constant C such that

$$\|W_p\|_{TV} \le C\|p-p_0\| \qquad \text{for all } p\in P \text{ such that } \|p-p_0\| \text{ is sufficiently small.} \qquad (A.25)$$

The following result is the main theorem on differentiation of trajectories with respect to a parameter. The proof involves a number of estimates using Gronwall's inequality.

Theorem A.7. *Let P, $\mathbf{f}$, p_0, ξ be data as in (D.1,2,3,4). Assume that $\mathbf{f}$ is weakly differentiable at p_0 along ξ. Let $\bar{\varepsilon}>0$ be such that (WD1) to (WD4) hold. Then*

(A.7.a) *For every $\varepsilon \in]0,\bar{\varepsilon}]$ there exists a $\delta \in]0,\varepsilon[$ such that, whenever $\|p-p_0\| \le \delta$, $\|\bar{x}-\xi(a)\| \le \delta$, and $\zeta \in \mathrm{MTraj}(f_p,\bar{x},a)$, it follows that $[a,b] \subseteq \mathrm{Dom}(\zeta)$ and $\|\zeta(t)-\xi(t)\| \le \varepsilon$ for all $t \in [a,b]$.*

(A.7.b) *If $\{\bar{x}_p\}_{\|p-p_0\|\le\bar{\varepsilon}}$ is a family of points of $\mathbb{R}^n$ such that the map $p \to \bar{x}_p$ is Fréchet differentiable at p_0 and $\bar{x}_{p_0}=\xi(a)$, then, if $\{\xi_p\}_{\|p-p_0\|\le\bar{\varepsilon}}$ is a family of curves such that $\xi_p \in \mathrm{MTraj}(f_p,\bar{x}_p,a)$, and x_p denotes the restriction of ξ_p to $[a,b]$—which is well defined for small enough $\|p-p_0\|$ by (A.7.a)—then the map $\mathcal{X}$ that sends each p to the function x_p is weak* Fréchet differentiable at p_0 as a map from P to $BV([a,b],\mathbb{R}^n)$. Moreover, $D^w\mathcal{X}(p_0)$ can be computed as follows. For $v \in P$,*

$$D^w\mathcal{X}(p_0).v = Y_{A,\,d(D^w\mathcal{W}(0).v),\,L.(p-p_0)}\,, \qquad (A.26)$$

where $L: P \to \mathbb{R}^n$ is the differential at p_0 of the map $p \to \bar{x}_p$. ♢

If $f \in TVVF(\mathbb{R}^n)$ and $\xi:[a,b]\to\mathbb{R}^n$ is a trajectory of f, we say that f *is differentiable along* ξ if the map $x \to f(x,t)$ is differentiable at $\xi(t)$ for almost every t. In that case, we write $D_\xi f(t) \stackrel{\mathrm{def}}{=} \frac{\partial f}{\partial x}(\xi(t),t)$, and refer to the map $[a,b] \ni t \to D_\xi f(t) \in \mathbb{R}^{n\times n}$ as the *differential of f along ξ*.

Lemma A.8. *If $\mathbf{f}$ is weakly differentiable at p_0 along a trajectory $\xi:[a,b]\to\mathbb{R}^n$ of f_{p_0}, then f_{p_0} is differentiable along ξ and the map A of (WD3) is such that $A(t)=D_\xi f_{p_0}(t)$ for almost all $t \in [a,b]$.*

PROOF. We begin by showing that the functions ψ_ε for which (WD3) holds can be chosen so that $\psi_\varepsilon(t)$ is a monotonically nonincreasing function of ε for each t. To see this, start with some family $\{\psi_\varepsilon\}_{\varepsilon\in]0,\bar\varepsilon]}$ of functions for which (WD3) holds, and define $\hat\psi_\varepsilon = \psi_{2^{-k}\bar\varepsilon}$ whenever $2^{-k-1}\bar\varepsilon < \varepsilon \le 2^{-k}\bar\varepsilon$ and k is a nonnegative integer. Then (WD3) also holds if the $\hat\psi_\varepsilon$ are substituted for the ψ_ε. Next define $\bar\psi_\varepsilon(t) = \min\{\hat\psi_{\varepsilon'}(t) : \varepsilon \le \varepsilon' \le \bar\varepsilon\}$. Then the $\bar\psi_\varepsilon$ are well defined integrable functions of t, because for each ε the set of functions $\{\hat\psi_{\varepsilon'} : \varepsilon \le \varepsilon' \le \bar\varepsilon\}$ is finite. It is clear that (WD3) also holds with the $\bar\psi_\varepsilon$ instead of the $\hat\psi_\varepsilon$, and in addition $\hat\psi_\varepsilon(t)$ is nonincreasing as a function of ε.

Now, if we pick the ψ_ε to be monotonically nonincreasing with respect to ε, it follows that the limit $\rho(t) = \lim_{\varepsilon\downarrow 0}\psi_\varepsilon(t)$ exists for every t. By the dominated convergence theorem, $\int_a^b \rho(t)dt = 0$. Since $\rho(t) \ge 0$ for all t, it follows that $\rho(t) = 0$ almost everywhere.

Let t be such that $\rho(t) = 0$. Let $\theta(x) = \psi_{||x-\xi(t)||}(t)$. Then $\lim_{x\to\xi(t)}\theta(x) = 0$. On the other hand, (WD3) implies that

$$||f_{p_0}(x,t) - f_{p_0}(\xi(t),t) - A(t).(x-\xi(t))|| \le \theta(x)||x-\xi(t)||\,. \tag{A.27}$$

Since $\theta(x) \to 0$ as $x \to \xi(t)$, (A.27) says precisely that $A(t)$ is the differential at $\xi(t)$ of the map $x \to f_{p_0}(x,t)$, completing our proof. $\diamondsuit$

If $f \in TVVF(\mathbb{R}^n)$ is differentiable along a trajectory $\xi : [a,b] \to \mathbb{R}^n$ of f, and the matrix-valued function $D_\xi f$ is integrable, then the ordinary differential equation

$$\dot\lambda(t) = -\lambda(t).D_\xi f(t)\,, \tag{A.28}$$

for an $\mathbb{R}_n$-valued function λ on $[a,b]$, is the *adjoint variational equation*—or, simply, the *adjoint equation*—for f along ξ. A solution of the adjoint equation (i.e. an absolutely continuous function $\lambda : [a,b] \to \mathbb{R}_n$ such that (A.28) holds for almost all t) is known as an *adjoint vector for f along ξ*.

Theorem A.9. *Let P, $\mathbf{f} = \{f_p\}_{p\in P}$, $p_0 \in P$, $\xi : [a,b] \to \mathbb{R}^n$ be as in Theorem A.7. Let $P \ni p \to \bar x_p \subset \mathbb{R}^n$ be a map which is Fréchet differentiable at p_0 and such that $\bar x_{p_0} = \xi(a)$, and let L_a be its differential at p_0. Let ξ_p be, for p close enough to p_0, maximal trajectories of f_p such that $\xi_p(a) = \bar x_p$, and let x_p be the restriction of ξ_p to $[a,b]$. Then the map $p \to x_p(b)$ is differentiable at p_0. Let L_b be its differential at p_0. Let $\lambda : [a,b] \to \mathbb{R}^n$ be an adjoint vector for f_{p_0} along ξ. Then*

$$\lambda(b)\cdot L_b(v) - \lambda(a)\cdot L_a(v) = \lim_{\varepsilon\downarrow 0}\varepsilon^{-1}\int_a^b \Big\langle \lambda(t), f_{p_0+\varepsilon v}(\xi(t),t) - f_{p_0}(\xi(t),t)\Big\rangle dt \tag{A.29}$$

for every $v \in \mathbb{R}^n$.

PROOF. Assume, without loss of generality, that $p_0 = 0$. According to Theorem A.7 and Lemma A.8, the map $p \to x_p$ is weak* Fréchet differentiable at 0—as a map from P to $BV([a,b],\mathbb{R}^n)$—and its differential at 0 is the map that sends $v \in P$ to $Y_{A,d(U.v),L_a(v)}$, where $U = D^w\mathcal{W}(0)$, and $\mathcal{W}$ is the map $p \to W_p$, where $W_p(t) = \int_a^t \Big(f_p(\xi(s),s) - f_0(\xi(s),s)\Big)dt$, and $A = D_\xi f$. Then $p \to x_p(b)$ is Fréchet differentiable at 0, and the differential L_b is given by $L_b(v) = Y_{A,d(U.v),L_a(v)}(b)$. Clearly, $L_a(v) = Y_{A,d(U.v),L_a(v)}(a)$. Since λ satisfies $\dot\lambda = -\lambda.A$, Lemma A.4 tells us that

$$\lambda(b)\cdot L_b(v) - \lambda(a)\cdot L_a(v) = \int_a^b \lambda.d(U.v)\,.$$

Since $U.v$ is the weak*-limit of $\varepsilon^{-1}W_{\varepsilon v}$ as $\varepsilon \downarrow 0$, our conclusion follows. $\diamondsuit$

We now turn to the differentiation of a Lagrangian cost functional.

If $f \in TVVF(I\!R^n)$, a *Lagrangian* for f in $I\!R^n$ is a real-valued function L such that

(L.1) $\mathrm{Dom}(f) \subseteq \mathrm{Dom}(L)$;
(L.2) the function $t \to L(\xi(t), t)$ is locally integrable for every $\xi \in \mathrm{Traj}(f)$.

If $f \in TVVF(I\!R^n)$ has components $f^1, \ldots, f^n$, and L is a Lagrangian for f, then we can consider the time-varying vector field $[L; f]$ in $I\!R^{n+1}$ whose components are $(f^0, f^1, \ldots, f^n)$, where $f^0 \equiv L$. A trajectory of $[L; f]$ can then be regarded a pair (ξ^0, ξ), where $\xi \in \mathrm{Traj}(f)$ and ξ^0 is a real-valued locally absolutely continuous function on $\mathrm{Dom}(\xi)$ such that $\dot{\xi}^0(t) = L(\xi(t), t)$ for a.e. t. Any such function ξ^0 will be called a *running L-cost along ξ*. Our definition of a Lagrangian implies that a running L-cost along ξ exists for all $\xi \in \mathrm{Traj}(f)$, and it is clear that any two running L-costs along ξ must differ by a constant. If ξ is such that $\mathrm{Dom}(\xi)$ is of the form $[a, b]$ or $[a, b[$ or $[a, \infty[$, then there is a canonical choice of running L-cost along ξ, obtained by letting $\xi^0(t) = \int_a^t L(\xi(s), s)\, ds$. In that case, we will write ξ^L to denote the curve $t \to (\xi^0(t), \xi(t))$, where ξ^0 is the canonical running L-cost.

Now suppose we are given data P, $\mathbf{f}$, p_0, ξ as in (D.1,2,3,4), as well as

(D.5) a family $\mathbf{L} = \{L_p\}_{p \in P}$ such that each L_p is a Lagrangian for f_p.

Given a trajectory $\xi : [a, b] \to I\!R^n$ of f_{p_0}, we say that the pair $(\mathbf{f}, \mathbf{L})$ is *weakly differentiable at p_0 along ξ* if the family $\{[L_p; f_p]\}_{p \in P}$ is weakly differentiable along $\xi^{L_{p_0}}$.

It is easy to see that $(\mathbf{f}, \mathbf{L})$ is weakly differentiable at p_0 along ξ iff there exists $\bar{\varepsilon} > 0$ such that (WD1,2,3,4) hold, and in addition

(WD5) *There exists $h \in L^1([a, b], I\!R_n)$ such that*

$$||L_p(x, t) - L_p(\xi(t), t) - h(t).(x - \xi(t))|| \le \psi_\varepsilon(t)(||x - \xi(t)|| + ||p - p_0||) \qquad \text{(A.30)}$$

whenever $a \le t \le b$, $||x - \xi(t)|| \le \varepsilon$ and $||p - p_0|| \le \varepsilon$.

(WD6) *If we let $w_p^0(t) = L_p(\xi(t), t) - L_{p_0}(\xi(t), t)$, $W_p^0(t) = \int_a^t w_p^0(s)\, ds$, then the map $\mathcal{W}^0$ that sends p to $\mathcal{W}^0(p) = W_p^0 \in BV([a, b], I\!R)$ is weak* Fréchet differentiable at p_0.*

If $f \in TVVF(I\!R^n)$, $\xi : [a, b] \to I\!R^n$ is a trajectory of f, and L is a Lagrangian for f, we say that *L is differentiable along ξ* if the map $x \to L(x, t)$ is differentiable at $\xi(t)$ for almost every t. In that case, we write $\nabla_\xi L(t) \stackrel{\mathrm{def}}{=} \nabla_x L(\xi(t), t)$, and refer to the map $[a, b] \ni t \to \nabla_\xi L(t) \in I\!R^{n \times n}$ as the *differential of L along ξ*.

Lemma A.8, applied to the curve $\xi^{L_{p_0}}$, yields

Lemma A.10. *If $(\mathbf{f}, \mathbf{L})$ is weakly differentiable at p_0 along a trajectory $\xi : [a, b] \to I\!R^n$ of f_{p_0}, then f_{p_0} and L_{p_0} are differentiable along ξ and the maps A, h of (WD3,5) are such that $A(t) = D_\xi f(t)$ and $h(t) = \nabla_\xi L_{p_0}(\xi(t), t)$ for almost all $t \in [a, b]$.* ♢

Under the hypothesis of Lemma A.10, we can apply Theorem A.9 to the family $\{[L_p; f_p]\}_{p \in P}$ and the curve $\xi^{L_{p_0}}$. Define

$$J_p(\zeta) = \int_a^b L_p(\zeta(t), t)\, dt\,, \qquad \text{(A.31)}$$

if $\zeta : [a, b] \to I\!R^n$ is a curve such that the integral exists. (We will only be using this when $\zeta \in \mathrm{Traj}(f_p)$, in which case $J_p(\zeta)$ is well defined.) An adjoint vector for $[L_{p_0}; f_{p_0}]$ along $\xi^{L_{p_0}}$ is then a pair (λ^0, λ) such that $\lambda^0 \in I\!R$, $\lambda : [a, b] \to I\!R_n$ is absolutely continuous, and

$$\dot{\lambda}(t) = -\lambda(t).A(t) - \lambda^0 h(t) \qquad \text{for} \qquad \text{a.e.} \qquad t \in [a, b]\,. \qquad \text{(A.32)}$$

Theorem A.9 then implies:

Theorem A.11. *Assume that we are given data P, $\mathbf{f}$, p_0, ξ, $\mathbf{L}$ as in (D.1,2,3,4,5). Assume that $(\mathbf{f}, \mathbf{L})$ is weakly differentiable at p_0 along ξ. Let $P \ni p \to \bar{x}_p \subset \mathbb{R}^n$ be a map which is Fréchet differentiable at p_0 and such that $\bar{x}_{p_0} = \xi(a)$, and let L_a be its differential at p_0. Let ξ_p be, for p close enough to p_0, maximal trajectories of f_p such that $\xi_p(a) = \bar{x}_p$, and let x_p be the restriction of ξ_p to $[a,b]$. Then the maps $p \to x_p(b)$ and $p \to J_p(x_p)$ are differentiable at p_0. Let L_b, L_b^0 be their differentials at p_0. Then, if (λ^0, λ) is an adjoint vector for $[L_{p_0}, f_{p_0}]$ along ξ, it follows that*

$$\lambda(b) \cdot L_b(v) + \lambda^0 L_b^0(v) - \lambda(a) \cdot L_a(v) = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \int_a^b \Big(H_{p_0+\varepsilon v}(\xi(t), \lambda(t), \lambda_0, t) - H_{p_0}(\xi(t), \lambda(t), \lambda_0, t) \Big)\, dt\,, \tag{A.33}$$

where

$$H_p(x, z, z^0, t) = \langle z, f_p(x,t) \rangle + z^0 L_p(x,t) \tag{A.34}$$

for $x \in \mathbb{R}^n$, $z \in \mathbb{R}_n$, $z^0 \in \mathbb{R}$, $t \in [a,b]$. ◊

Appendix B. A lemma

Lemma B.1. *Let f be a real-valued function on a compact interval $[a,b]$. Assume that there exists a finite or countable subset E of $[a,b]$ with the following properties:*

(B.*1.a*) $\liminf_{h \downarrow 0} \frac{f(x+h)-f(x)}{h} \geq 0$ *for all* $x \in [a,b[\setminus E$,
(B.*1.b*) $\liminf_{h \downarrow 0} f(x+h) \geq f(x)$ *for all* $x \in [a,b[$,
(B.*1.c*) $\liminf_{h \downarrow 0} f(x-h) \leq f(x)$ *for all* $x \in]\,a,b]$.

Then $f(b) \geq f(a)$.

PROOF. Let $\varepsilon > 0$ be arbitrary. We will prove that $f(b) \geq f(a) - \varepsilon$. Let $\varepsilon' = \frac{\varepsilon}{1+b-a}$ and $E = \{x_k : k \in \mathbb{N}\}$.

For each $x \in [a,b]$, let $I(x) = \{k : x_k < x\}$. Define $g : [a,b] \to \mathbb{R}$ by $g(x) = \sum_{k \in I(x)} 2^{-k}\varepsilon'$. Then g is left-continuous and monotonically nondecreasing.

Let S be the set of all $x \in [a,b]$ such that $f(y) \geq f(a) - (y-a)\varepsilon' - g(y)$ for all $y \in [a,x]$. It is clear that $a \in S$, and also that S is an interval. So S is of the form $[a,c[$ or $[a,c]$, with $c \in [a,b]$.

Suppose $S = [a,c\,[$. Then $c > a$. By (B.1.c), given $\delta > 0$ there is a sequence $\{z_j\}$ of points of $[a,b]$ such that $z_j < c$, $z_j \to c$, and $f(c) \geq f(z_j) - \delta$. Then the z_j belong to S, so the inequalities $f(z_j) \geq f(a) - (z_j - a)\varepsilon' - g(z_j) \geq f(a) - (c-a)\varepsilon' - g(c)$ hold. It then follows that $f(c) \geq f(a) - (c-a)\varepsilon' - g(c) - \delta$. Since δ is arbitrary, we have $f(c) \geq f(a) - (c-a)\varepsilon' - g(c)$, so $c \in S$, which is a contradiction.

Therefore $S = [a,c]$ for some $c \in [a,b]$. We now show that $c = b$. Assume that $c < b$, and consider separately the cases $c \notin E$, $c \in E$.

If $c \notin E$, then (B.1.a) implies that there exists a $c' > c$ such that $f(x) - f(c) \geq -(x-c)\varepsilon'$ for every $x \in [c,c']$. But then, since $f(c) \geq f(a) - (c-a)\varepsilon' - g(c)$, we can conclude that $f(x) \geq f(a) - (x-c)\varepsilon' - g(c) \geq f(a) - (x-a)\varepsilon' - g(x)$ for $x \in [c,c']$. Therefore $c' \in S$, contradicting the fact that $S = [a,c]$.

Now suppose $c \in E$. Then $c = x_k$ for some k, and $x_k \neq b$. By (B.1.b), there exists $h > 0$ such that $c + h \leq b$ and $f(y) - f(c) \geq -2^{-k}\varepsilon'$ for all $y \in [c, c+h]$. If $y \in]\,c, c+h]$, then $k \notin I(c)$, and $I(c) \cup \{k\} \subseteq I(y)$. Therefore $g(y) \geq g(c) + 2^{-k}\varepsilon'$. Since $f(c) \geq f(a) - (c-a)\varepsilon' - g(c)$, we have $f(y) \geq f(c) - 2^{-k}\varepsilon' \geq f(a) - (c-a)\varepsilon' - 2^{-k}\varepsilon' - g(c) \geq f(a) - (y-a)\varepsilon' - g(y)$. So $c+h \in S$, which is a contradiction.

So we have shown that $c = b$. Then $f(b) \geq f(a) - (b-a)\varepsilon' - g(b) \geq f(a) - (b-a)\varepsilon' - \varepsilon' = f(a) - \varepsilon$. Since ε is arbitrary, we have established that $f(b) \geq f(a)$, as desired. ◊

REFERENCES

1. Bardi, M., and I. Capuzzo-Dolcetta, "Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equation," Birkäuser Boston, Inc., Boston MA, 1997.
2. Boltyanskii, V. G., "Sufficient conditions for optimality and the justification of the dynamic programming principle," *SIAM J. Control Optim.* **4**, (1966), pp. 326-361.
3. Bressan, A., and B. Piccoli, "Structural stability for time-optimal planar syntheses," *Dynam. Contin. Discrete Impuls. Systems*2 **3** (1997), pp. 335–371.
4. Bressan, A., and B. Piccoli, "A generic classification of time optimal planar stabilizing feedbacks," *SIAM J. Control Optim.* **36** (1998), no. 1, pp. 12-32.
5. Brunovský, P., "Every normal linear system has a regular time–optimal synthesis," *Math. Slovaca* **28**, (1978), pp. 81-100.
6. Brunovský, P., "Existence of regular syntheses for general problems," *J. Diff. Equations* **38**, (1980), pp. 317-343.
7. Cesari, L., "Optimization – Theory and applications," Springer–Verlag, New York, 1983.
8. Clarke, F. H., Yu. S. Ledyaev, A. I. Subbotin and E. D. Sontag, "Asymptotic controllability implies feedback stabilization," *IEEE Trans. Autom. Control* **42** (1997), pp. 1394-1407.
9. Filippov, A. F., "Differential equations with discontinuous right-hand side," Reidel, 1989.
10. Fleming, W.H., and R. W. Rishel, "Deterministic and stochastic optimal control," Springer–Verlag, New York, 1975.
11. Fleming, W.H., and M. Soner, "Controlled Markov processes and viscosity solutions," Springer–Verlag, New York, 1993.
12. Fuller, A. T., "Relay control systems optimized for various performance criteria," Automatic and remote control, Proc. first world congress IFAC Moscow, vol. 1, Butterworths, 1961, pp. 510-519.
13. Marchal, C., "Chattering arcs and chattering controls," *J. Optim Theory Appl.* **11** (1973), pp. 441-468.
14. Piccoli, B., "Regular time-optimal syntheses for smooth planar systems," *Rend. Sem. Mat. Univ. Padova* **95**, (1996), pp. 59-79.
15. Piccoli, B., "Classification of generic singularities for the planar time optimal syntheses," *SIAM J. Control Optim.* **34** (1996), no. 6, pp. 1914-1946.
16. Sussmann, H. J., "Subanalytic sets and feedback control," *J. Diff. Equations* **31** (1979), no. 1, pp. 31-52.
17. Sussmann, H. J., "Piecewise analyticity of optimal cost functions and optimal feedback," in *Proc. Joint Automatic Control Conference*, Denver, Colo. (August 1979).
18. Sussmann, H. J., "Analytic stratifications and control theory," in *Proc. 1978 Int. Congress of Mathematicians*, Helsinki (1980), pp. 865-871.
19. Sussmann, H.. J., "Time-optimal control in the plane," in *Feedback Control of Linear and Nonlinear Systems*, D. Hinrichsen and A. Isidori Eds., Springer-Verlag, New York, 1982, pp. 244-260.
20. Sussmann, H. J., "Subanalytic sets and optimal control in the plane," *Proc. 21st I.E.E.E. Conference on Decision and Control*, Orlando, Fla. (Dec 1982), pp. 295-299.

21. Sussmann, H. J., "Lie brackets, real analyticity and geometric control theory," in *Differential Geometric Control Theory*, R. W. Brockett, R. S. Millman and H.J. Sussmann Eds., Birkhäuser Boston Inc. (1983), pp. 1-115.

22. Sussmann, H. J., "The structure of time-optimal trajectories for single-input systems in the plane : the C^∞ nonsingular case, *SIAM J. Control Optim.,* **25**, no. 2 (1987), pp. 433-465.

23. Sussmann, H. J., "The structure of time-optimal trajectories for single-input systems in the plane: the general real-analytic case," *SIAM J. Control Optim.* **25**, no. 4 (1987), pp. 868-904.

24. Sussmann, H. J., "Regular synthesis for time-optimal control of single-input real-analytic systems in the plane," *SIAM J. Control Optim.* **25**, no. 5 (1987), pp. 1145-1162.

25. Sussmann, H. J., "Recent developments in the regularity theory of optimal trajectories", in Linear and nonlinear mathematical control theory, *Rend. Sem. Mat. Univ. e Pol. Torino*, Fascicolo speciale, 1987, pp. 149-182.

26. Sussmann, H. J., "Envelopes, conjugate points and optimal bang-bang extremals," in *Algebraic and Geometric Methods in Nonlinear Control Theory*, M. Fliess and M. Hazewinkel Eds., Math. Appl., 29, D. Reidel Publishing Co., Dordrecht-Boston, Mass., 1986, pp. 325-346.

27. Sussmann, H. J., "Synthesis, presynthesis, sufficient conditions for optimality and subanalytic sets", in Nonlinear controllability and optimal control, H. J. Sussmann ed., Marcel Dekker, New York, 1990, pp. 1-19.

28. Zelikin, M. I., and V. F. Borisov, *Theory of Chattering Control, with applications to astronautics, robotics, economics and engineering*, Birkhäuser, Boston, 1994.