Nonsmooth Analysis in Systems and Control Theory

Francis Clarke
Institut universitaire de France et Université de Lyon

[January 2008. To appear in the
Encyclopedia of Complexity and System Science, Springer.]

Article Outline

Glossary

Generalized gradients and subgradients These terms refer to various set-valued replacements for the usual derivative which are used in developing differential calculus for functions which are not differentiable in the classical sense. The subject itself is known as *nonsmooth analysis*. One of the best-known theories of this type is that of *generalized gradients*. Another basic construct is the *subgradient*, of which there are several variants. The approach also features *generalized tangent and normal vectors* which apply to sets which are not classical manifolds. A short summary of the essential definitions is given in §III.

Pontryagin Maximum Principle The main theorem on necessary conditions in optimal control was developed in the 1950s by the Russian mathematician L. Pontryagin and his associates. The Maximum Principle unifies and extends to the control setting the classical necessary conditions of Euler and

Weierstrass from the calculus of variations, as well as the transversality conditions. There have been numerous extensions since then, as the need to consider new types of problems continues to arise, and as discussed in §IV.

VERIFICATION FUNCTIONS In attempting to prove that a certain control is indeed the solution to a given optimal control problem, one important approach hinges upon exhibiting a function having certain properties implying the optimality of the given control. Such a function is termed a verification function (see §V). The approach becomes widely applicable if one allows nonsmooth verification functions.

DYNAMIC PROGRAMMING A well-known technique in dynamic problems of optimization is to solve (in a discrete context) a backwards recursion for a certain value function related to the problem. This technique, which was developed notably by Bellman, can be applied in particular to optimal control problems. In the continuous setting, the recursion corresponds to the *Hamilton-Jacobi equation.* This partial differential equation does not generally admit smooth classical solutions. The theory of *viscosity solutions* uses subgradients to define generalized solutions, and obtains their existence and uniqueness (see §VI).

LYAPUNOV FUNCTION In the classical theory of ordinary differential equations, global asymptotic stability is most often verified by exhibiting a *Lyapunov function*, a function which decreases along trajectories. In that setting, the existence of a smooth Lyapunov function is both necessary and sufficient for stability. The Lyapunov function concept can be extended to control systems, but in that case it turns out that nonsmooth functions are essential. These generalized *control Lyapunov functions* play an important role in designing optimal or stabilizing feedback (see §§VII,VIII).

I. Definition of the Subject and Its Importance

The term *nonsmooth analysis* refers to the body of theory which develops differential calculus for functions which are not differentiable in the usual sense, and for sets which are not classical smooth manifolds. There are several different (but related) approaches to doing this. Among the better-known constructs of the theory are the following: generalized gradients and Jacobians, proximal subgradients, subdifferentials, generalized directional (or Dini) derivates, together with various associated tangent and normal cones. Nonsmooth analysis is a subject in itself, within the larger mathematical field of differential (variational) analysis or functional analysis, but it has also played an increasingly important role in several areas of application, notably in optimization, calculus of variations, differential equations, mechanics, and control theory. Among those who have participated in its development (in addition to the author) are J. Borwein, A. D. Ioffe, B.

Mordukhovich, R. T. Rockafellar, and R. B. Vinter, but many more have contributed as well.

In the case of control theory, the need for nonsmooth analysis first came to light in connection with finding proofs of necessary conditions for optimal control, notably in connection with the Pontryagin Maximum Principle. This necessity holds even for problems which are expressed entirely in terms of smooth data. Subsequently, it became clear that problems with intrinsically nonsmooth data arise naturally in a variety of optimal control settings. Generally, nonsmooth analysis enters the picture as soon as we consider problems which are truly nonlinear or nonlinearizable, whether for deriving or expressing necessary conditions, in applying sufficient conditions, or in studying the sensitivity of the problem.

The need to consider nonsmoothness in the case of stabilizing (as opposed to optimal) control has come to light more recently. It appears in particular that in the analysis of truly nonlinear control systems, the consideration of nonsmooth Lyapunov functions and discontinuous feedbacks becomes unavoidable.

II. Introduction

The basic object in the control theory of ordinary differential equations is the system

$$\dot{x}(t) = f(x(t), u(t)) \text{ a.e.}, \quad 0 \le t \le T, \tag{1}$$

where the (measurable) *control function* $u(\cdot)$ is chosen subject to the constraint

$$u(t) \in U \text{ a.e.} \tag{2}$$

(In this article, U is a given set in a Euclidean space.) The ensuing *state* $x(\cdot)$ (a function with values in $\mathbb{R}^n$) is subject to certain conditions, including most often an initial one of the form $x(0) = x_0$, and perhaps other constraints, either throughout the interval (pointwise) or at the terminal time. A control function $u(\cdot)$ of this type is referred to as an *open loop control.* This indirect control of $x(\cdot)$ via the choice of $u(\cdot)$ is to be exercised for a purpose, of which there are two principal sorts:

- *positional*: $x(t)$ is to remain in a given set in $\mathbb{R}^n$, or approach that set;
- *optimal*: $x(\cdot)$, together with $u(\cdot)$, is to minimize a given functional.

The second of these criteria follows directly in the tradition of the calculus of variations, and gives rise to the subject of *optimal control*, in which the dominant issues are those of optimization: necessary conditions for optimality, sufficient conditions, regularity of the optimal control, sensitivity. We shall discuss the role of nonsmooth analysis in optimal control in Sections III and IV; this was the setting of many of the earliest applications.

In contrast, a prototypical control problem of purely positional sort would be the following:

Find a control $u(\cdot)$ such that $x(\cdot)$ goes to 0.

This rather vaguely worded goal is often more precisely expressed as that of finding a *stabilizing feedback control* $k(x)$; that is, a function k with values in U (this is referred to as a closed-loop control) such that, for any initial condition α, all solutions of the differential equation

$$\dot{x}(t) = f\big(x(t), k\big(x(t)\big)\big), \quad x(0) = \alpha \tag{3}$$

converge to 0 in a suitable sense.

The most common approach to designing the required stabilizing feedback uses the technique that is central to most of applied mathematics: *linearization.* In this case, one examines the linearized system

$$\dot{x}(t) = Ax(t) + Bu(t)$$

where

$$A := f_x(0,0),\ B := f_u(0,0).$$

If the linearized system satisfies certain controllability properties, then classical linear systems theory provides well-known and powerful tools for designing (linear) feedbacks that stabilize the linearized system. Under further mild hypotheses, this yields a feedback that stabilizes the original nonlinear system (1) *locally*; that is, for initial values $x(0)$ sufficiently near 0.

This approach has been feasible in a large number of cases, and in fact it underlies the very successful role that control theory has played in a great variety of applications. Still, linearization does require that a certain number of conditions be met:

- The function f must be smooth (differentiable) so that the linear system can be constructed;
- The linear system must be a 'nondegenerate' approximation of the nonlinear one (that is, it must be controllable);
- The control set U must contain a neighborhood of 0, so that near 0 the choice of controls is unconstrained, and all state values near 0 must be acceptable (no state constraints);
- Both x and u must remain small so that the linear approximation remains relevant (in dealing with errors or perturbations, the feedback is operative only when they are sufficiently small).

It is not hard to envisage situations in which the last two conditions fail, and indeed such challenging problems are beginning to arise increasingly often. The first condition fails for simple problems involving electrical circuits in which a diode is present, for example (see [14]). A famous (smooth) mechanical system for which the second condition fails is the *nonholonomic integrator*, a term which refers to the following system, which is linear (separately) in the state and in the control variables:

$$\left\{ \begin{array}{lcc} \dot{x}_1 & = & u_1 \\ \dot{x}_2 & = & u_2 \\ \dot{x}_3 & = & x_1 u_2 - x_2 u_1 \end{array} \right. \qquad \|(u_1, u_2)\| \le 1.$$

(Thus $n = 3$ here, and U is the closed unit ball in $\mathbb{R}^2$.) Here the linearized system is degenerate, since its third component is $\dot{x}_3 = 0$. As discussed in Section VIII, there is in fact no continuous feeback law $u = k(x)$ which will stabilize this system (even locally about the origin), but certain discontinuous stabilizing feedbacks do exist.

This illustrates the moral that when linearization is not applicable to a given nonlinear system (for whatever reason), nonsmoothness generally arises. (This has been observed in other contexts in recent decades: catastrophes, chaos, fractals.) Consider for example the issue of whether a (control) Lyapunov function exists. This refers to a pair (V, W) of positive definite functions satisfying notably the following *Infinitesimal Decrease* condition:

$$\inf_{u \in U} \langle \nabla V(x), f(x,u) \rangle \;\le\; -W(x) \quad x \neq 0.$$

The existence of such (smooth) functions implies that the underlying control system is *globally asymptotically controllable* (GAC), which is a necessary condition for the existence of a stabilizing feedback (and also sufficient, as recently proved [21]). In fact, exhibiting a Lyapunov function is the principal technique for proving that a given system is GAC; the function V in question then goes on to play a role in designing the stabilizing feedback.

It turns out, however, that even smooth systems that are GAC need not admit a smooth Lyapunov function. (The nonholonomic integrator is an example of this phenomenon.) But if one extends in a suitable way the concept of Lyapunov function to nonsmooth functions, then the existence of a Lyapunov function becomes a necessary and sufficient condition for a given system to be GAC. One such extension involves replacing the gradient that appears in the Infinitesimal Decrease condition above by elements of the *proximal subdifferential* (see below). How to use such extended Lyapunov functions to design a stabilizing feedback is a nontrivial topic that has only recently been successfully addressed, and one that we discuss in Section VIII.

In the next section we define a few basic constructs of nonsmooth analysis; knowledge of these is sufficient for interpreting the statements of the results discussed in this article.

III. Elements of Nonsmooth Analysis

This section summarizes some basic notions in nonsmooth analysis, in fact a minimum so that the statements of the results to come can be understood. This minimum corresponds to three types of set-valued generalized derivative (generalized gradients, proximal subgradients, limiting subgradients), together with a notion of normal vector applicable to any closed (not necessarily smooth or convex) set. (See [24] for a thorough treatment and detailed references.)

Generalized gradients. For *smooth* real-valued functions f on $\mathbb{R}^n$ we have a well-known formula linking the usual directional derivative to the gradient:

$$f'(x;v) := \lim_{t\downarrow 0} \frac{f(x+tv)-f(x)}{t} = \langle \nabla f(x), v \rangle .$$

We can extend this pattern to *Lipschitz* functions: f is said to be Lipschitz on a set S if there is a constant K such that $|f(y) - f(z)| \leq K\|y - z\|$ whenever y and z belong to S. A function f that is Lipschitz in a neighborhood of a point x is not necessarily differentiable at x, but we can define a *generalized directional derivative* as follows:

$$f^\circ(x;v) := \limsup_{y\to x,\ t\downarrow 0} \frac{f(y+tv)-f(y)}{t}.$$

Having done so, we proceed to define the *generalized gradient*:

$$\partial_C f(x) := \{\zeta \in \mathbb{R}^n : f^\circ(x;v) \geq \langle \zeta, v\rangle \quad \forall\ v \in X\}.$$

It turns out that $\partial_C f(x)$ is a compact convex nonempty set that has a calculus reminiscent of the usual differential calculus; for example, we have $\partial(-f)(x) = -\partial f(x)$. We also have $\partial(f+g)(x) \subset \partial f(x) + \partial g(x)$; note that (as is often the case) this is an inclusion, not an equation. There is also a useful analogue of the Mean Value Theorem, and other familiar results. In addition, though, there are new formulas having no smooth counterpart, such as one for the generalized gradient of the pointwise maximum of locally Lipschitz functions ($\partial\{\max_{1\leq i\leq n} f_i(x)\} \subset \ldots$).

A very useful fact for actually calculating $\partial_C f(x)$ is the following Gradient Formula:

$$\partial_C f(x) = \mathrm{co}\left\{\lim_{i\to\infty} \nabla f(x_i) : x_i \to x, x_i \notin \Gamma\right\},$$

where Γ is *any* set of measure zero containing the local points of nondifferentiability of f. Thus the generalized gradient is 'blind to sets of measure zero'.

Generalized gradients and their calculus were defined by Clarke [12] in 1973. The theory can be developed on any Banach space; the infinite-dimensional context is essential in certain control applications, but for our present purposes it suffices to limit attention to functions defined on $\mathbb{R}^n$. There is in addition a corresponding theory of tangent and normal vectors to arbitrary closed sets; we give some elements of this below.

Proximal subgradients. We now present a different approach to developing nonsmooth calculus, one that uses the notion of *proximal subgradient.* Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a given function (note that the value $+\infty$ is admitted here), and let x be a point where $f(x)$ is finite. A vector ζ in $\mathbb{R}^n$ is said to be a proximal subgradient of f at x provided that there exist a neighborhood Ω of x and a number $\sigma \geq 0$ such that

$$f(y) \geq f(x) + \langle \zeta, y - x \rangle - \sigma \|y - x\|^2 \ \forall\, y \in \Omega.$$

Thus the existence of a proximal subgradient ζ at x corresponds to the possibility of approximating f from below (thus in a *one-sided* manner) by a function whose graph is a parabola. The point $\big(x, f(x)\big)$ is a contact point between the graph of f and the parabola, and ζ is the slope of the parabola at that point. Compare this with the usual derivative, in which the graph of f is approximated (in a two-sided way) by an affine function.

The set of proximal subgradients at x (which may be empty, and which is not necessarily closed, open, or bounded but which is convex) is denoted $\partial_P f(x)$, and is referred to as the *proximal subdifferential.* If f is differentiable at x, then we have $\partial_P f(x) \subset \{f'(x)\}$; equality holds if f is of class C^2 at x.

As a guide to understanding, the reader may wish to carry out the following exercise (in dimension $n = 1$): the proximal subdifferential at 0 of the function $f_1(x) := -|x|$ is empty, while that of $f_2(x) := |x|$ is the interval $[-1, 1]$.

The *proximal density theorem* asserts that $\partial_P f(x)$ is nonempty for all x in a dense subset of

$$\operatorname{dom} f := \{x : f(x) < \infty\}.$$

Although it can be empty at many points, the proximal subgradient admits a very complete calculus for the class of lower semicontinuous functions: all the usual calculus rules that the reader knows (and more) have their counterpart in terms of $\partial_P f$. Let us quote for example Ioffe's *fuzzy sum rule*: if $\zeta \in \partial_P(f + g)(x)$, then for any $\epsilon > 0$ there exist x' and x'' within ϵ of x, together with points $\zeta' \in \partial_P f(x')$ and $\zeta'' \in \partial_P g(x'')$ such that

$$\zeta \in \zeta' + \zeta'' + \epsilon B.$$

The limiting subdifferential. A sequential closure operation applied to $\partial_P f$ gives rise to the *limiting subdifferential*, useful for stating results:

$$\partial_L f(x) := \{\lim \zeta_i : \zeta_i \in \partial_P f(x_i), x_i \to x, f(x_i) \to f(x)\}.$$

If f is Lipschitz near x, then $\partial_L f(x)$ is nonempty, and (for any lower semicontinuous g finite at x) we have

$$\partial_L(f+g)(x) \subset \partial_L f(x) + \partial_L g(x).$$

When the function f is Lipschitz near x, then both the approaches given above (generalized gradients, proximal subdifferential) apply, and the corresponding constructs are related as follows:

$$\partial_C f(x) = \operatorname{co} \partial_L f(x).$$

Normal vectors. Given a nonempty closed subset S of $\mathbb{R}^n$ and a point x in S, we say that $\zeta \in X$ is a *proximal normal* (vector) to S at x if there exists $\sigma \geq 0$ such that

$$\left\langle \zeta, x' - x \right\rangle \leq \sigma \left\| x' - x \right\|^2 \quad \forall\, x' \in S.$$

(This is the *proximal normal inequality.*) The set (convex cone) of such ζ, which always contains 0, is denoted $N_S^P(x)$ and referred to as the proximal normal cone. We apply to $N_S^P(x)$ a sequential closure operation in order to obtain the *limiting normal cone*:

$$N_S^L(x) := \left\{ \lim \zeta_i : \zeta_i \in N_S^P(x_i), x_i \to x, x_i \in S \right\}.$$

These geometric notions are consistent with the analytical ones, as illustrated by the formulas

$$\partial_P I_S(x) = N_S^P(x),\ \partial_L I_S(x) = N_S^L(x),$$

where I_S denotes the *indicator* of the set S: the function which equals 0 on S and $+\infty$ elsewhere. They are also consistent with the more traditional ones: When S is either a convex set, a smooth manifold, or a manifold with boundary, then both $N_S^P(x)$ and $N_S^L(x)$ coincide with the usual normal vectors (a cone, space, or half-space respectively).

Viscosity subdifferentials. We remark that the *viscosity subdifferential* of f at x (commonly employed in pde's, but not used in this article) corresponds to the set of ζ for which we have

$$f(y) \geq f(x) + \langle \zeta, y - x \rangle + \theta(|y - x|),$$

where θ is a function such that $\lim_{t\downarrow 0} \theta(t)/t = 0$. This gives a potentially larger set than $\partial_P f(x)$; however, the viscosity subdifferential satisfies the same (fuzzy) calculus rules as does $\partial_P f$. In addition, the sequential closure operation described above gives rise to the *same* limiting subdifferential $\partial_L f(x)$ (see [24]). In most cases, therefore, it is equivalent to work with viscosity or proximal subgradients.

IV. Necessary Conditions in Optimal Control

The central theorem on necessary conditions for optimal control is the Pontryagin Maximum Principle. (The literature on necessary conditions in optimal control is now very extensive; we cite some standard references in the bibliography.) Even in the somewhat special (by current standards) smooth context in which it was first proved [47], an element of generalized differentiability was required. With the subsequent increase in both the generality of the model and weakening of the hypotheses (all driven by real applications), the need for nonsmooth analysis is all the greater, even for the very statement of the result.

We give here just one example, a broadly applicable *hybrid maximum principle* taken from [11], in order to highlight the essential role played by nonsmooth analysis as well as the resulting versatility of the results obtained.

The problem and basic hypotheses. We consider the minimization of the functional

$$\ell(x(a), x(b)) + \int_a^b F(t, x(t), u(t))\, dt$$

subject to the boundary conditions $(x(a), x(b)) \in S$ and the standard control dynamics

$$\dot{x}(t) = f(t, x(t), u(t)) \;\text{ a.e.}$$

The minimization takes place with respect to (absolutely continuous) state arcs x and measurable control functions $u : [a, b] \to \mathbb{R}^m$. Note that no explicit constraints are placed upon $u(t)$; if such constraints exist, they are accounted for by assigning to the extended-valued integrand F the value $+\infty$ whenever the constraints are violated. It is part of our intention here to demonstrate the utility of indirectly taking account of constraints in this way.

As basic hypotheses, we assume that F is measurable and lower semicontinuous in (x, u); ℓ is taken to be locally Lipschitz and S closed. In the theorem below, we suppose that f is measurable in t and continuously differentiable in (x, u) (but see the remarks for a weakening of the smoothness).

The growth conditions. The remaining hypothesis (the main one) is that f and F satisfy the following: for every bounded subset X of $\mathbb{R}^n$, there exist a constant c and a summable function d such that, for almost every t, for every $(x,u) \in \operatorname{dom} F(t,\cdot,\cdot)$ with $x \in X$, we have

$$\|D_x f(t,x,u)\| \leq c\,\{|f(t,x,u)| + F(t,x,u)\} + d(t),$$

and for all (ζ,ψ) in $\partial_P F(t,x,u)$ (if any) we have

$$\frac{|\zeta|\,(1+\|D_u f(t,x,u)\|)}{1+|\psi|} \leq c\,\{|f(t,x,u)| + F(t,x,u)\} + d(t).$$

In the following, $(*)$ denotes evaluation at $(x_*(t), u_*(t))$.

Theorem 1 Let the control function u_* give rise to an arc x_* which is a local minimum for the problem above. Then there exist an arc p on $[a,b]$ such that the following *transversality condition* holds

$$(p(a), -p(b)) \in \partial_L \ell(x_*(a), x_*(b)) + N_S^L(x_*(a), x_*(b)),$$

and such that p satisfies the *adjoint inclusion*: $\dot{p}(t)$ belongs almost everywhere to the set

$$\text{co}\left\{\omega : (\omega + D_x f(t,*)^T p(t), D_u f(t,*)^T p(t)) \in \partial_L F(t,*)\right\},$$

as well as the *maximum condition*: for almost every t, for every u in $\operatorname{dom} F(t, x_*(t), \cdot)$, one has

$$\langle p(t), f(t, x_*(t), u)\rangle - F(t, x_*(t), u) \leq \langle p(t), f(t,*)\rangle - F(t,*). \qquad \blacklozenge$$

We proceed to make a few remarks on this theorem, beginning with the fact that it can fail in the absence of the growth condition, even for smooth problems of standard type [30]. The statement of the theorem is not complete, since in general the necessary conditions may hold only in *abnormal* form; we do not discuss this technical point here for reasons of economy. The phrase 'local minimum' (which we have also not defined) can be interpreted very generally, see [11]. The 'maximum condition' above is of course responsible for the name 'maximum principle', while the exotic-looking adjoint inclusion reduces to more familiar conditions in a variety of special cases (see below). The transversality condition illustrates well the utility of framing the conclusion in nonsmooth analysis terms, since the given form encapsulates simultaneously a wide variety of conclusions obtainable in special cases. To give but one example, consider an optimal control problem in which $x(a)$ is prescribed, $x(b)$ does not appear in the cost, but is subject to an inequality constraint $g(x(b)) \leq 0$ for a certain smooth scalar function g. Then the transversality condition of the theorem (interpreted via standard facts in nonsmooth analysis) asserts that $p(b)$ is of the form $\lambda \nabla g(x_*(b))$ for some nonpositive number λ (a Lagrange multiplier).

To illustrate the versatility of the theorem, we look at some special cases.

The standard problem. The first case we examine is that in which for each t,

$$F(t,x,u) = I_{U(t)}(u),$$

the indicator function which takes the value 0 when $u \in U(t)$ and $+\infty$ otherwise. This simply corresponds to imposing the condition $u(t) \in U(t)$ on the admissible controls u; this is the *Mayer form* of the problem (no integral term in the cost, obtained by reformulation if necessary). Note that in this case the second growth condition is trivially satisfied (since $\zeta = 0$). The first growth condition is active only on $U(t)$, and certainly holds if f is smooth in (t,x,u) and $U(t)$ is uniformly bounded. As regards the conclusions, the hybrid adjoint inclusion immediately implies the standard adjoint equation

$$-\dot{p}(t) = D_x f(t, x_*(t), u_*(t))^T p(t),$$

and we recover the classical Maximum Principle of Pontryagin. (An extension is obtained when $U(t)$ not bounded, see [10].) When f is not assumed to be differentiable, but merely locally Lipschitz with respect to x, there is a variant of the theorem in which the adjoint inclusion is expressed as follows:

$$-\dot{p}(t) \in \partial_C \left\langle p(t), f(t, \cdot, u_*(t)) \right\rangle (x_*(t)),$$

where the generalized gradient ∂_C (see §III) is taken with respect to the x variable (note the connection to the standard adjoint equation above). This is an early form of the nonsmooth maximum principle [13].

The calculus of variations. When we take $f(t,x,u) = u$, the problem reduces to the *problem of Bolza* in the calculus of variations. The first growth condition is trivially satisfied, and the second coincides with the generalized Tonelli-Morrey growth condition introduced in [11]; in this way we recover the state-of-the-art necessary conditions for the generalized problem of Bolza, which include as a special case the multiplier rule for problems with pointwise and/or isoperimetric constraints.

Differential inclusions. When we specialize further by taking $F(t, \cdot)$ to be the indicator of the graph of a multifunction $M(t, \cdot)$, we obtain the principal necessary conditions for the *differential inclusion problem.* These in turn lead to necessary conditions for *generalized control systems* [11].

Mixed constraints. Consider again the optimal control problem in Mayer form (no integral term in the cost), but now in the presence of mixed state/control pointwise constraints of the form $(x(t), u(t)) \in \Omega$ a.e. for a given closed set Ω. Obtaining general necessary conditions for such problems is a well-known challenge in the subject; see [33, 34, 46]. We treat this case by taking $F(t, \cdot) = I_\Omega(\cdot)$. Then the second growth condition reduces to the following geometric assumption: for every $(x, u) \in \Omega$, for every $(\zeta, \psi) \in N^P_\Omega(x, u)$, one has

$$\frac{|\zeta|\,(1 + \|D_u f(t, x, u)\|)}{|\psi|} \le c\,|f(t, x, u)| + d(t).$$

By taking a suitable representation for Ω in terms of functional equalities and/or inequalities, sufficient conditions in terms of rank can be adduced which imply this property, leading to explicit multiplier rules (see [10] for details). With an appropriate 'transversal intersection' condition, we can also treat the case in which *both* the constraints $(x(t), u(t)) \in \Omega$ and $u(t) \in U(t)$ are present.

Sensitivity. The multiplier functions p that appear in necessary conditions such as the ones above play a central role in analyzing the sensitivity of optimal control problems that depend on parameters. To take but one example, consider the presence of a perturbation $\alpha(\cdot)$ in the dynamics of the problem:

$$\dot{x}(t) = f(t, x(t), u(t)) + \alpha(t) \text{ a.e.}$$

Clearly the minimum in the problem depends upon α; we denote it $V(\alpha)$. The function V is an example of what is referred to as a *value function.* Knowledge of the derivative of V would be highly relevant in studying the sensitivity of the problem to perturbations (errors) in the dynamics. Generally, however, value functions are not differentiable, so instead one uses nonsmooth analysis; the multipliers p give rise to estimates for the generalized gradient of V. We refer to [15, 16] for examples and references.

V. Verification Functions

We consider now a special case of the problem considered in the preceding section: to minimize the integral cost functional

$$J(x, u) := \int_a^b F(t, x(t), u(t))\, dt$$

subject to the prescribed boundary conditions

$$x(a) = A,\ x(b) = B$$

and the standard control dynamics

$$\dot{x}(t) = f(t, x(t), u(t)) \text{ a.e., } u(t) \in U(t) \text{ a.e.}$$

Suppose now that a candidate (x_*, u_*) has been identified as a possible solution to this problem (perhaps through a partial analysis of the necessary conditions, or otherwise). An elementary yet powerful way (traceable to Legendre) to prove that (x_*, u_*) actually does solve the problem is to exhibit a function $\phi(t, x)$ such that:

$$F(t, x, u) \geq \phi_t(t, x) + \langle \phi_x(t, x), f(t, x, u) \rangle \ \forall t \in [a, b],\ (x, u) \in \mathbb{R}^n \times U(t),$$

with equality along $(t, x_*(t), u_*(t))$.

The mere existence of such a function verifies that (x_*, u_*) is optimal, as we now show. For any admissible state/control pair (x, u), we have

$$\begin{aligned} F(t, x(t), u(t)) &\geq \phi_t(t, x(t)) + \langle \phi_x(t, x(t)), \dot{x}(t) \rangle \\ &= d/dt\ \{\phi(t, x(t))\} \\ \Rightarrow J(x, u) := \int_a^b F(t, x(t), u(t))\, dt &\geq \phi(t, x(t))\big|_{t=a}^{t=b} \\ &= \phi(b, B) - \phi(a, A). \end{aligned}$$

But this lower bound on J holds with equality when $(x, u) = (x_*, u_*)$, which proves that (x_*, u_*) is optimal.

In this argument, we have implicitly supposed that the *verification function* ϕ is smooth. It is a fact that if we limit ourselves to smooth verification functions, then there may not exist such a ϕ (even when the problem itself has smooth data). However, if we admit nonsmooth (locally Lipschitz) verification functions, then (under mild hypotheses on the data) the existence of a verification function ϕ becomes necessary and sufficient for (x_*, u_*) to be optimal.

An appropriate way to extend the smooth inequality above uses the generalized gradient $\partial_C \phi$ of ϕ as follows:

$$F(t, x, u) \geq \theta + \langle \zeta, f(t, x, u) \rangle \ \forall\, (\theta, \zeta) \in \partial_C \phi(t, x),$$

together with the requirement

$$J(x_*, u_*) = \phi(b, B) - \phi(a, A).$$

It is a feature of this approach to proving optimality that it extends readily to more general problems, for example those involving unilateral state constraints, boundary costs, or isoperimetric conditions. It can also be interpreted in terms of duality theory, as shown notably by Vinter. We refer to [16, 26] for details.

The basic inequality that lies at the heart of this approach is of Hamilton–Jacobi type, and the fact that we are led to consider nonsmooth verification functions is related to the phenomenon that Hamilton–Jacobi equations may not admit smooth solutions. This is one of the main themes of the next section.

VI. Dynamic Programming and Viscosity Solutions

The Minimal Time problem. By a *trajectory* of the standard control system

$$\dot{x}(t) = f(x(t), u(t)) \text{ a.e., } u(t) \in U \text{ a.e.}$$

we mean a state function $x(\cdot)$ corresponding to some choice of admissible control function $u(\cdot)$. The *minimal time problem* refers to finding a trajectory that reaches the origin as quickly as possible from a given point. Thus we seek the least $T \geq 0$ admitting a control function $u(\cdot)$ on $[0, T]$ having the property that the resulting trajectory x begins at the prescribed point and satisfies $x(T) = 0$. We proceed now to describe the well-known *dynamic programming* approach to solving the problem.

We begin by introducing the *minimal time function* $T(\cdot)$, defined on $\mathbb{R}^n$ as follows: $T(\alpha)$ is the least time $T \geq 0$ such that some trajectory $x(\cdot)$ satisfies

$$x(0) = \alpha,\ x(T) = 0.$$

An issue of *controllability* arises here: Is it always possible to steer α to 0 in finite time? When such is not the case, then in accord with the usual convention we set $T(\alpha) = +\infty$.

The *principle of optimality* is the dual observation that if $x(\cdot)$ is any trajectory, then we have, for $s < t$,

$$T\big(x(t)\big) - T\big(x(s)\big) \geq s - t$$

(that is, the function $t \mapsto T\big(x(t)\big) + t$ is increasing), while if x is optimal, then equality holds (that is, the same function is constant).

Let us explain this in other terms (for $s = 0$): if $x(\cdot)$ is an optimal trajectory joining α to 0, then

$$T\big(x(t)\big) = T(\alpha) - t \quad \text{for} \quad 0 \leq t \leq T(\alpha),$$

since an optimal trajectory from the point $x(t)$ is furnished by the truncation of $x(\cdot)$ to the interval $\big[t, T(\alpha)\big]$. If $x(\cdot)$ is any trajectory, then the inequality

$$T\big(x(t)\big) \geq T(\alpha) - t$$

is a reflection of the fact that in going to the point $x(t)$ from α (in time t), we may have acted optimally (in which case equality holds) or not (then inequality holds).

Since $t \mapsto T\big(x(t)\big) + t$ is increasing, we expect to have

$$\big\langle \nabla T\big(x(t)\big), \dot{x}(t)\big\rangle + 1 \geq 0,$$

with equality when $x(\cdot)$ is an optimal trajectory. The possible values of $\dot{x}(t)$ for a trajectory being precisely the elements of the set $f\big(x(t), U\big)$, we arrive at

$$\min_{u\in U}\big\langle \nabla T(x), f(x(t), u)\big\rangle + 1 = 0. \tag{4}$$

We define the (lower) *Hamiltonian function* h as follows:

$$h(x, p) := \min_{u\in U} \langle p, f(x, u)\rangle.$$

In terms of h, the partial differential equation obtained above reads

$$h\big(x, \nabla T(x)\big) + 1 = 0, \tag{5}$$

a special case of the *Hamilton–Jacobi equation*.

Here is the first step in the dynamic programming heuristic: use the Hamilton–Jacobi equation (5), together with the boundary condition $T(0) = 0$, to find $T(\cdot)$. How will this help us find the optimal trajectory?

To answer this question, we recall that an optimal trajectory is such that equality holds in (4). This suggests the following procedure: for each x, let $k(x)$ be a point in U satisfying

$$\min_{u\in U} \big\langle \nabla T(x), f(x, u)\big\rangle = \big\langle \nabla T(x), f(x, k(x))\big\rangle = -1. \tag{6}$$

Then, if we construct $x(\cdot)$ via the initial-value problem

$$\dot{x}(t) = f\big(x(t), k\big(x(t)\big)\big), \quad x(0) = \alpha, \tag{7}$$

we will have a trajectory that is optimal (from α)!

Here is why: Let $x(\cdot)$ satisfy (7); then $x(\cdot)$ is a trajectory, and

$$\begin{aligned} \frac{d}{dt}T\big(x(t)\big) &= \big\langle \nabla T\big(x(t)\big), \dot{x}(t)\big\rangle \\ &= \big\langle \nabla T\big(x(t)\big), f\big(x(t), k\big(x(t)\big)\big)\big\rangle = -1. \end{aligned}$$

Integrating, we find

$$T\big(x(t)\big) = T(\alpha) - t,$$

which implies that at $t = T(\alpha)$, we have $T(x(t)) = 0$, whence $x(t) = 0$. Therefore $x(\cdot)$ is an optimal trajectory.

Let us stress the important point that $k(\cdot)$ generates the optimal trajectory from *any* initial value α (via (7)), and so constitutes what can be considered the ultimate solution for this problem: an *optimal feedback synthesis.* There can be no more satisfying answer to the problem: If you find yourself at x, just choose the control value $k(x)$ to approach the origin as fast as possible. This goes well beyond finding a single open-loop optimal control.

Unfortunately, there are serious obstacles to following the route that we have just outlined, beginning with the fact that T is nondifferentiable, as simple examples show, even when it is finite everywhere (which it generally fails to be).

We will therefore have to examine anew the argument that led to the Hamilton–Jacobi equation (5), which, in any case, will have to be recast in some way to accommodate nonsmooth solutions. Having done so, will the generalized Hamilton–Jacobi equation admit T as the unique solution?

The next step (after characterizing T) offers fresh difficulties of its own. Even if T were smooth, there would be in general no *continuous* function $k(\cdot)$ satisfying (6) for each x. The meaning and existence of a trajectory $x(\cdot)$ generated by $k(\cdot)$ via the differential equation (7), in which the right-hand side is discontinuous in the state variable, is therefore problematic in itself.

The intrinsic difficulties of this approach to the minimal-time problem have made it a historical focal point of activity in differential equations and control, and it is only recently that fully satisfying answers to all the questions raised above have been found. We begin with generalized solutions of the Hamilton-Jacobi equation.

Subdifferentials and viscosity solutions. We shall say that ϕ is a *proximal solution* of the Hamilton–Jacobi equation (5) provided that

$$h(x, \partial_P \phi(x)) = -1 \quad \forall\, x \in \mathbb{R}^n, \tag{8}$$

a 'multivalued equation' which means that for all x, for all $\zeta \in \partial_P \phi(x)$ (if any), we have $h(x, \zeta) = -1$. (Recall that the proximal subdifferential $\partial_P \phi$ was defined in §III.)

Note that the equation holds automatically at a point x for which $\partial_P \phi(x)$ is empty; such points play an important role, in fact, as we now illustrate. Consider the case in which $f(x, U)$ is equal to the unit ball for all x, in dimension $n = 1$. Then $h(x, p) \equiv -|p|$ (and the equation is of *eikonal* type). Let us examine the functions $\phi_1(x) := -|x|$ and $\phi_2(x) := |x|$; they both satisfy $h(x, \nabla \phi(x)) = -1$ at all points $x \neq 0$, since (for each of these functions) the proximal subdifferential at points different from 0 reduces to the singleton consisting of the derivative. However, we have (see the exercise in §III)

$$\partial_P \phi_1(0) = \emptyset,\ \partial_P \phi_2(0) = [-1, 1],$$

and it follows that ϕ_1 is (but ϕ_2 is not) a proximal solution of the Hamilton-Jacobi equation (8).

A lesson to be drawn from this example is that in defining generalized solutions we need to look closely at the differential behavior at specific and individual points; we cannot argue in an 'almost everywhere' fashion, or by 'smearing' via integration (as is done for linear partial differential equations via distributional derivatives).

Proximal solutions are just one of the ways to define generalized solutions of certain partial differential equations, a topic of considerable interest and activity, and one which seems to have begun with the Hamilton-Jacobi equation in every case. The first 'subdifferential type' of definition was given by the author in the 1970s, using generalized gradients and for locally Lipschitz solutions. While no uniqueness theorem holds for that solution concept, it was shown that the value function of the associated optimal control problem is a solution (hence existence holds), and is indeed a special solution: it is the maximal one. In 1980 A.I. Subbotin defined his 'minimax solutions', which are couched in terms of Dini derivates rather than subdifferentials, and which introduced the important feature of being 'two-sided'. This work featured existence and uniqueness in the class of Lipschitz functions, the solution being characterized as the value of a differential game. Subsequently, M. Crandall and P.-L. Lions incorporated both subdifferentials and two-sidedness in their *viscosity solutions*, a theory which they developed for merely continuous functions. In the current context, and under mild hypotheses on the data, it can be shown that minimax, viscosity, and proximal solutions all coincide [19, 24].

Recall that our goal (within the dynamic programming approach) is to characterize the minimal time function. This is now attained, as shown by the following (we omit the hypotheses; see [63], and also the extensive discussion in Bardi and Capuzzo-Dolcetta [4]):

Theorem 2 There exists a unique lower semicontinuous function $\phi : \mathbb{R}^n \to (-\infty, +\infty]$ bounded below on $\mathbb{R}^n$ and satisfying the following:

[HJ equation] $\quad h(x, \partial_P \phi(x)) = -1 \quad \forall\, x \neq 0;$

[Boundary condition] $\quad \phi(0) = 0$ and $\quad h(0, \partial_P \phi(0)) \geq -1.$

That unique function is $T(\cdot)$. ♦

The proof of this theorem is based upon proximal characterizations of certain monotonicity properties of trajectories related to the inequality forms of the Hamilton–Jacobi equation (see Section 4.7 of [24]). The fact that monotonicity is closely related to the solution of the minimal time problem is already evident in the following elementary assertion: a trajectory x joining α to 0 is optimal iff the rate of change of the function $t \mapsto T(x(t))$ is -1 a.e.

The characterization of T given by the theorem can be applied to verify the validity of a general conjecture regarding the time-optimal trajectories from arbitrary initial values. This works as follows: we use the conjecture to calculate the supposedly optimal time $T(\alpha)$ from any initial value α, and then we see whether or not the function T constructed in this way satisfies the conditions of the theorem. If so, then (by uniqueness) T is indeed the minimal time function and the conjecture is verified (the same reasoning as in the preceding section is in fact involved here). Otherwise, the conjecture is necessarily incorrect (but the way in which it fails can provide information on how it needs to be modified; this is another story).

Another way in which the reasoning above is exploited is to discretize the underlying problem as well as the Hamilton–Jacobi equation for T. This gives rise to the backwards recursion numerical method developed and popularized by Bellman under the name of *dynamic programming*; of course, the approach applies to problems other than minimal time.

With respect to the goal of finding an optimal feedback synthesis for our problem, we have reached the following point in our quest: given that T satisfies the proximal Hamilton–Jacobi equation $h(x, \partial_P T(x)) = -1$, which can be written in the form

$$\min_{u \in U} \big\langle \zeta, f(x(t), u) \big\rangle = -1 \quad \forall \zeta \in \partial_P T(x), \quad \forall x \neq 0, \tag{9}$$

how does one proceed to construct a feedback $k(x)$ having the property that any trajectory x generated by it via (7) is such that $t \mapsto T(x(t))$ decreases at a unit rate? There will also arise the issue of defining the very sense of (7) when $k(\cdot)$ is discontinuous, as it must be in general.

This is a rather complex question to answer, and it turns out to be a special case of the issue of designing stabilizing feedback (where, instead of T, we employ a *Lyapunov function*). We shall address this in §VIII, and return there to the minimal time synthesis, which will be revealed to be a special case of the procedure. We need first to examine the concept of Lyapunov function.

VII. Lyapunov Functions

In this section we consider the standard control system

$$\dot{x}(t) = f(x(t), u(t)) \text{ a.e.,} \ \ u(t) \in U \text{ a.e.,}$$

under a mild local Lipschitz hypothesis: for every bounded set S in $\mathbb{R}^n$ there exists a constant $K = K(S)$ such that

$$|f(x,u) - f(x',u)| \leq K|x - x'| \ \ \forall x, x' \in S, u \in U.$$

We also suppose that $f(0, U)$ is bounded.

A point α is *asymptotically guidable* to the origin if there is a trajectory x satisfying $x(0) = \alpha$ and $\lim_{t\to\infty} x(t) = 0$. When every point has this property, and when additionally the origin has the familiar local stability property known as *Lyapunov stability*, it is said in the literature to be GAC: (open loop) *globally asymptotically controllable* (to 0). A well-known *sufficient* condition for this property is the existence of a smooth (C^1, say) pair of functions

$$V : \mathbb{R}^n \to \mathbb{R}, \quad W : \mathbb{R}^n \backslash \{0\} \to \mathbb{R}$$

satisfying the following conditions:

1. **Positive Definiteness:**

$$V(x) > 0 \text{ and } W(x) > 0 \quad \forall x \neq 0, \text{ and } V(0) \geq 0.$$

2. **Properness:** The sublevel sets $\{x : V(x) \leq c\}$ are bounded $\forall c$.

3. **Weak Infinitesimal Decrease:**

$$\inf_{u \in U} \langle \nabla V(x), f(x,u) \rangle \ \leq \ -W(x) \quad x \neq 0.$$

The last condition asserts that V decreases in *some* available direction. We refer to V as a (weak) *Lyapunov function*; it is also referred to in the literature as a *control Lyapunov function*.

It is a fact, as demonstrated by simple examples (see [17] or [57]), that the existence of a smooth function V with the above properties fails to be a necessary condition for global asymptotic controllability; that is, the familiar converse Lyapunov theorems of Massera, Barbashin and Krasovskii, and Kurzweil (in the setting of a differential equation with no control) do not extend to this weak controllability setting, at least not in smooth terms. This may be a rather general phenomenon, in view of the following result [22], which holds under the additional hypothesis that the sets $f(x, U)$ are closed and convex:

Theorem 3 If the system admits a C^1 weak Lyapunov function, then it has the following surjectivity property: for every $\epsilon > 0$, there exists $\delta > 0$ such that $f(B(0, \epsilon), U) \supset B(0, \delta)$. ⧫

It is not difficult to check that the nonholonomic integrator system (see §II) is GAC. Since it fails to have the surjectivity property described in the theorem, it cannot admit a smooth Lyapunov function.

It is natural therefore to seek to weaken the smoothness requirement on V so as to obtain a necessary (and still sufficient) condition for a system to be GAC. This necessitates the use of some construct of nonsmooth analysis to replace the gradient of V that appears in the infinitesimal decrease condition. In this connection we use the proximal subgradient (§III) $\partial_P V(x)$, which requires only that the (extended-valued) function V be lower semicontinuous. In proximal terms, the Weak Infinitesimal Decrease condition becomes

$$\sup_{\zeta \in \partial_P V(x)} \inf_{u \in U} \langle \zeta, f(x,u) \rangle \leq -W(x) \quad x \neq 0.$$

Note that this last condition is trivially satisfied when x is such that $\partial_P V(x)$ is empty, in particular when $V(x) = +\infty$. (The supremum over the empty set is $-\infty$.) A *general Lyapunov pair* (V, W) refers to extended-valued lower semicontinuous functions $V : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $W : \mathbb{R}^n \backslash \{0\} \to \mathbb{R} \cup \{+\infty\}$ satisfying the positive definiteness and properness conditions above, together with proximal weak infinitesimal decrease.

The following is proved in [24].

Theorem 4 Let (V, W) be a general Lyapunov pair for the system. Then any $\alpha \in \operatorname{dom} V$ is asymptotically guidable to 0. ♦

It follows from the theorem that the existence of a lower semicontinuous Lyapunov pair (V, W) with V everywhere finite-valued implies the global asymptotic guidability to 0 of the system. When V is not continuous, this does not imply Lyapunov stability at the origin, however, so it cannot characterize global asymptotic controllability. An early and seminal result due to Sontag [56] considers *continuous* functions V, with the infinitesimal decrease condition expressed in terms of Dini derivates. Here is a version of that result in proximal subdifferential terms:

Theorem 5 The system is GAC if and only if there exists a *continuous* Lyapunov pair (V, W). ♦

There is an advantage to being able to replace continuity in such a result by a stronger regularity property (particularly in connection with using V to design a stabilizing feedback, as we shall see). Of course, we cannot assert smoothness, as pointed out above. In [20] it was shown that certain locally Lipschitz *value functions* give rise to *practical* Lyapunov functions, that is, assuring stable controllability to arbitrary neighborhoods of 0. Building upon this, L. Rifford [49] was able to combine a countable family of such functions in order to construct a global *locally Lipschitz* Lyapunov function. This answered a long-standing open question in the subject. Rifford [51] also went on to show the existence of a *semiconcave* Lyapunov function, a stronger property (familiar in pde's) whose relevance to feedback construction will be seen in the next section.

The role of Lyapunov functions in characterizing various types of convergence to 0 (including guidability in finite time) is discussed in detail in [18].

VIII. Stabilizing Feedback

We now address the issue of finding a *stabilizing feedback control* $k(x)$; that is, a function k with values in U such that all solutions of the differential equation

$$\dot{x} = g(x),\ x(0) = \alpha, \quad \text{where} \quad g(x) := f(x, k(x)) \tag{10}$$

converge to 0 (for all values of α) in a suitable sense. When this is possible, the system is termed *stabilizable*. Here, the origin is supposed to be an equilibrium of the system; to be precise, we take $0 \in U$ and $f(0,0) = 0$.

A necessary condition for the system to be stabilizable is that it be GAC. A central question in the subject has been whether this is sufficient as well. An early observation of Sontag and Sussmann [58] showed that the answer is negative if one requires the feedback k to be continuous, which provides the easiest (classical) context in which to interpret the differential equation that appears in (10). Later, Brockett showed that the nonholonomic integrator fails to admit a continuous stabilizing feedback.

One is therefore led to consider the use of discontinuous feedback, together with the attendant need to define an appropriate solution concept for a differential equation in which the dynamics fail to be continuous in the state. The best-known solution concept in this regard is that of Filippov; it turns out, however, that the nonholonomic integrator fails to admit a (discontinuous) feedback which stabilizes it in the Filippov sense [32, 55]. Clarke, Ledyaev, Sontag and Subbotin [21] gave a positive answer when the (discontinuous) feedbacks are implemented in the *closed-loop system sampling* sense (also referred to as *sample-and-hold*). We proceed now to describe the sample-and-hold implementation of a feedback.

Let $\pi = \{t_i\}_{i\geq 0}$ be a partition of $[0, \infty)$, by which we mean a countable, strictly increasing sequence t_i with $t_0 = 0$ such that $t_i \to +\infty$ as $i \to \infty$. The *diameter* of π, denoted diam$\,(\pi)$, is defined as $sup_{i\geq 0}(t_{i+1} - t_i)$. Given an initial condition x_0, the π-trajectory $x(\cdot)$ corresponding to π and an arbitrary feedback law $k : \mathbb{R}^n \to U$ is defined in a step-by-step fashion as follows. Between t_0 and t_1, x is a classical solution of the differential equation

$$\dot{x}(t) = f(x(t), k(x_0)), \quad x(0) = x_0, \quad t_0 \leq t \leq t_1.$$

(In the present context we have existence and uniqueness of x, and blow-up cannot occur.) We then set $x_1 := x(t_1)$ and restart the system at $t = t_1$ with control value $k(x_1)$:

$$\dot{x}(t) = f(x(t), k(x_1)), \quad x(t_1) = x_1, \quad t_1 \leq t \leq t_2,$$

and so on in this fashion. The trajectory x that results from this procedure is an actual state trajectory corresponding to a piecewise constant open-loop control; thus it is a physically meaningful one. When results are couched in terms of π-trajectories, the issue of defining a solution concept for discontinuous differential equations is effectively sidestepped. Making the diameter of the partition smaller corresponds to increasing the sampling rate in the implementation.

We remark that the use of possibly discontinuous feedback has arisen in other contexts. In linear time-optimal control, one can find discontinuous feedback syntheses as far back as the classical book of Pontryagin et alii [47]; in these cases the feedback is invariably piecewise constant relative to certain partitions of state space, and solutions either follow the switching surfaces or cross them transversally, so the issue of defining the solution in other than a classical sense does not arise. Somewhat related to this is the approach that defines a multivalued feedback law [6]. In stochastic control, discontinuous feedbacks are the norm, with the solution understood in terms of stochastic differential equations. In a similar vein, in the control of certain linear partial differential equations, discontinuous feedbacks can be interpreted in a distributional sense. These cases are all unrelated to the one under discussion. We remark too that the use of discontinuous pursuit strategies in differential games [41] is well-known, together with examples to show that, in general, it is not possible to achieve the result of a discontinuous optimal strategy to within any tolerance by means of a continuous stategy (thus there can be a positive unbridgeable gap between the performance of continuous and discontinuous feedbacks).

It is natural to say that a feedback $k(x)$ (continuous or not) *stabilizes* the system in the sample-and-hold sense provided that for every initial value x_0, for all $\epsilon > 0$, there exists $\delta > 0$ and $T > 0$ such that whenever the diameter of the partition π is less than δ, then the corresponding π-trajectory x beginning at x_0 satisfies

$$\|x(t)\| \leq \epsilon \;\; \forall\, t \geq T.$$

The following theorem is proven in [21].

Theorem 6 The system is open loop globally asymptotically controllable if and only if there exists a (possibly discontinuous) feedback $k : \mathbb{R}^n \to U$ which stabilizes it in the sample-and-hold sense. ♦

The proof of the theorem used the method of *proximal aiming*, which can be viewed as a geometric version of the Lyapunov technique. We now discuss how to define stabilizing feedbacks if one has in hand a sufficiently regular Lyapunov function.

The smooth case. We begin with the case in which a C^1 smooth Lyapunov function exists, where a very natural approach can be used. For $x \neq 0$, we simply define $k(x)$ to be any element $u \in U$ satisfying

$$\langle \nabla V(x), f(x,u) \rangle \leq -W(x)/2.$$

Note that at least one such u does exist, in light of the Infinitesimal Decrease condition. It is then elementary to prove [18] that the pointwise feedback k described above stabilizes the system in the sample-and-hold sense.

We remark that Rifford [50] has shown that the existence of a smooth Lyapunov pair is equivalent to the existence of a locally Lipschitz one satisfying Infinitesimal Decrease in the sense of generalized gradients (that is, with $\partial_P V$ replaced by $\partial_C V$), and that this in turn is equivalent to the existence of a stabilizing feedback in the Filippov (as well as sample-and-hold) sense.

Semiconcavity. We have seen that a smooth Lyapunov function generates a stabilizing feedback in a simple and natural way. But since a smooth Lyapunov function does not necessarily exist, we still require a way to handle the general case. It turns out that the smooth and the general case can be treated in a unified fashion through the notion of *semiconcavity*, which is a certain regularity property (not implying smoothness). Rifford has proven that any GAC system admits a semiconcave Lyapunov function; we shall see that this property permits a natural extension of the pointwise definition of a stabilizing feedback that was used in the smooth case.

A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is said to be (globally) semiconcave provided that for every ball $B(0,r)$ there exists $\gamma = \gamma(r) \geq 0$ such that the function $x \mapsto \phi(x) - \gamma|x|^2$ is (finite and) concave on $B(0,r)$. (Hence ϕ is locally the sum of a concave function and a quadratic one.) Observe that any function of class C^2 is semiconcave; also, any semiconcave function is locally Lipschitz, since both concave functions and smooth functions have that property. (There is a local definition of semiconcavity that we omit for present purposes.)

Semiconcavity is an important regularity property in partial differential equations (see for example [9]). The fact that the semiconcavity of a Lyapunov function V turns out to be useful in stabilization is a new observation, and may be counterintuitive: V often has an interpretation in terms of energy, and it may seem more appropriate to seek a *convex* Lyapunov function V. We proceed now to explain why semiconcavity is a highly desirable property, and why a convex V would be of less interest (unless it were smooth, but then it would be semiconcave too).

Recall the ideal case discussed above, in which (for a smooth V) we select a function $k(x)$ such that

$$\langle \nabla V(x), f(x, k(x)) \rangle \le -W(x)/2.$$

How might this appealing idea be adapted to the case in which V is nonsmooth? We cannot use the proximal subdifferential $\partial_P V(x)$ directly, since it may be empty for some values of x. We are led to consider the limiting subdifferential $\partial_L V(x)$ (see §III), which is nonempty when V is locally Lipschitz. By passing to the limit, the Weak Infinitesimal Decrease Condition for proximal subgradients implies the following:

$$\inf_{u \in U} \langle f(x,u), \zeta \rangle \le -W(x) \;\; \forall \zeta \in \partial_L V(x), \forall x \ne 0.$$

Accordingly, let us consider the following idea: For each $x \ne 0$, choose some element $\zeta \in \partial_L V(x)$, then choose $k(x) \in U$ such that

$$\langle f(x, k(x)), \zeta \rangle \le -W(x)/2.$$

Does this lead to a stabilizing feedback, when (of course) the discontinuous differential equation is interpreted in the sample-and-hold sense? When V is smooth, the answer is 'yes', as we have seen. But when V is merely locally Lipschitz, a certain 'dithering' phenomenon may arise to prevent k from being stabilizing. However, if V is semiconcave (locally on $\mathbb{R}^n \backslash \{0\}$), this does not occur, and stabilization is guaranteed. This explains the interest in finding a semiconcave Lyapunov function.

When V is a locally Lipschitz Lyapunov function with no additional regularity (neither smooth nor semiconcave), then it can still be used for defining stabilizing feedback, but less directly. It is possible to *regularize* V: to approximate it by a semiconcave function through the process of *inf-convolution* (see [24]). This leads to *practical semiglobal* stabilizing feedbacks; that is, for any $0 < r < R$, we derive a feedback which stabilizes all initial values in $B(0,R)$ to $B(0,r)$ [20].

Optimal feedback. The strategy described above for defining stabilizing feedbacks via Lyapunov functions can also be applied to construct (nearly) optimal feedbacks as well as stabilizing ones. The key is to use an appropriate value function instead of an arbitrary Lyapunov function. We obtain in this way a unification of optimal control and feedback control, at least at the mathematical level, and as regards feedback design.

To illustrate, consider again the Minimal Time problem, at the point at which we had left it at the end of §VII (thus, unresolved as regards the time-optimal feedback synthesis). As pointed out there, T satisfies the

Hamilton–Jacobi equation: this yields Infinitesimal Decrease in subgradient terms. Consequently, if the Minimal Time function T happens to be finite everywhere as well as smooth or semiconcave, then we can use the same direct definition as above to design a feedback which (in the limiting sample-and-hold sense) produces trajectories along which T decreases at rate 1; that is, which are time-optimal. This of course yields the fastest possible stabilization (and in finite time). In general, however, T may lack such regularity, or (when the controllability to the origin is only asymptotic) not even be finite everywhere. Then it is necessary to apply an approximation (regularization) procedure in order to obtain a variant of T, and use that instead. When T is finite, we can obtain in this way an approximate time-optimal synthesis (to any given tolerance).

The whole approach described here can be carried out for a variety of optimal control contexts [45, 27], and also for finding optimal strategies in *differential games* [25]. It also carries over to problems in which *unilateral state constraints* are imposed: $x(t) \in X$, where X is a given closed set [27, 28, 29]. The issue of robustness, not discussed here, is particularly important in the presence of discontinuity; see [42, 57, 18].

IX. Future Directions

A lesson of the past appears to be that nonsmooth analysis is likely to be required whenever linearization is not adequate or is inapplicable. It seems likely therefore to accompany the subject of control theory as it sets out to conquer new nonlinear horizons, in ways that cannot be fully anticipated. Let us nonetheless identify a few directions for future work.

The extensions of most of the results cited above to problems on manifolds, or of tracking, or with partial information (as in adaptive control) remain to be carried out to a great extent. There are a number of currently evolving contexts not discussed above in which nonsmooth analysis is highly likely to play a role, notably *hybrid control*; an example here is provided by *multiprocesses* [31, 8]. Distributed control (of pde's) is another area which requires development. There is also considerable work to be done on numerical implementation; in this connection see [36, 37].

References

[1] Z. Artstein. Stabilization with relaxed controls. *Nonlinear Analysis TMA*, 7:1163–1173, 1983.

[2] A. Astolfi. Discontinuous control of nonholonomic systems. *System Control Lett.*, 27:37–45, 1996.

[3] A. Astolfi. Discontinuous control of the Brockett integrator. *European J. Control*, 1998.

[4] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Birkhäuser, Boston, 1997.

[5] M. Bardi and V. Staicu. The Bellman equation for time-optimal control of noncontrollable nonlinear systems. *Acta Appl. Math.*, 31:201–223, 1993.

[6] L. D. Berkovitz. Optimal feedback controls. *SIAM J. Control Optim.*, 27:991–1006, 1989.

[7] J. M. Borwein and Q. J. Zhu. A survey of subdifferential calculus with applications. *Nonlinear Anal.*, 38:687–773, 1999.

[8] P. E. Caines, F. H. Clarke, X. Liu, and R. B. Vinter. A maximum principle for hybrid optimal control problems with pathwise state constraints. In *Proc. Conf. Decision and Control*. IEEE, 2006.

[9] P. Cannarsa and C. Sinestrari. *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*. Birkhäuser, Boston, 2004.

[10] F. Clarke. The maximum principle in optimal control. *J. Cybernetics and Control*, 34:709–722, 2005.

[11] F. Clarke. *Necessary Conditions in Dynamic Optimization*. Memoirs of the Amer. Math. Soc., No. 816, Vol. 173. 2005.

[12] F. H. Clarke. *Necessary Conditions for Nonsmooth Problems in Optimal Control and the Calculus of Variations*. Doctoral thesis, University of Washington, 1973. (Thesis director: R. T. Rockafellar).

[13] F. H. Clarke. The maximum principle under minimal hypotheses. *SIAM J. Control Optim.*, 14:1078–1091, 1976.

[14] F. H. Clarke. *Optimization and Nonsmooth Analysis*. Wiley-Interscience, New York, 1983. Republished as vol. 5 of *Classics in Applied Mathematics*, SIAM, 1990.

[15] F. H. Clarke. Perturbed optimal control problems. *IEEE Trans. Aut. Control*, 31:535–542, 1986.

[16] F. H. Clarke. *Methods of Dynamic and Nonsmooth Optimization*. S.I.A.M., Philadelphia, 1989. Regional Conference Series in Applied Mathematics vol. 57.

[17] F. H. Clarke. Nonsmooth analysis in control theory: a survey. *European J. Control*, 7:63–78, 2001.

[18] F. H. Clarke. Lyapunov functions and feedback in nonlinear control. In M.S. de Queiroz, M. Malisoff, and P. Wolenski, editors, *Optimal Control, Stabilization and Nonsmooth Analysis*, volume 301 of *Lecture Notes in Control and Information Sciences*, pages 267–282, New York, 2004. Springer-Verlag.

[19] F. H. Clarke and Yu. S. Ledyaev. Mean value inequalities in Hilbert space. *Trans. Amer. Math. Soc.*, 344:307–324, 1994.

[20] F. H. Clarke, Yu. S. Ledyaev, L. Rifford, and R. J. Stern. Feedback stabilization and Lyapunov functions. *SIAM J. Control Optim.*, 39:25–48, 2000.

[21] F. H. Clarke, Yu. S. Ledyaev, E. D. Sontag, and A. I. Subbotin. Asymptotic controllability implies feedback stabilization. *IEEE Trans. Aut. Control*, 42:1394–1407, 1997.

[22] F. H. Clarke, Yu. S. Ledyaev, and R. J. Stern. Asymptotic stability and smooth Lyapunov functions. *J. Differential Equations*, 149:69–114, 1998.

[23] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski. Qualitative properties of trajectories of control systems: a survey. *J. Dyn. Control Sys.*, 1:1–48, 1995.

[24] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory.* Graduate Texts in Mathematics, vol. 178. Springer-Verlag, New York, 1998.

[25] F. H. Clarke, Yu. S. Ledyaev, and A. I. Subbotin. The synthesis of universal pursuit strategies in differential games. *SIAM J. Control Optim.*, 35:552–561, 1997.

[26] F. H. Clarke and C. Nour. Nonconvex duality in optimal control. *SIAM J. Control Optim.*, 43:2036–2048, 2005.

[27] F. H. Clarke, L. Rifford, and R. J. Stern. Feedback in state constrained optimal control. *ESAIM Control Optim. Calc. Var.*, 7:97–133, 2002.

[28] F. H. Clarke and R. J. Stern. Hamilton-Jacobi characterization of the state-constrained value. *Nonlinear Anal.*, 61:725–734, 2005.

[29] F. H. Clarke and R. J. Stern. Lyapunov and feedback characterizations of state constrained controllability and stabilization. *Systems and Control Letters*, 54:747–752, 2005.

[30] F. H. Clarke and R. B. Vinter. On the conditions under which the Euler equation or the maximum principle hold. *Applied Math. and Optimization*, 12:73–79, 1984.

[31] F. H. Clarke and R. B. Vinter. Applications of optimal multiprocesses. *SIAM J. Control Optim.*, 27:1048–1071, 1989.

[32] J.-M. Coron and L. Rosier. A relation between continuous time-varying and discontinuous feedback stabilization. *J. Math. Syst., Estimation, Control*, 4:67–84, 1994.

[33] M. R. de Pinho. Mixed constrained control problems. *J. Math. Anal. Appl.*, 278:293–307, 2003.

[34] A. V. Dmitruk. Maximum principle for a general optimal control problem with state and regular mixed constraints. *Comp. Math. and Modeling*, 4:364–377, 1993.

[35] M. M. A. Ferreira. On the regularity of optimal controls for a class of problems with state constraints. *Internat. J. Systems Sci.*, 37:495–502, 2006.

[36] F. A. C. C. Fontes. A general framework to design stabilizing nonlinear model predictive controllers. *System Control Lett*, 42:127–143, 2001.

[37] F. A. C. C. Fontes and L. Magni. Min-max model predictive control of nonlinear systems using discontinuous feedbacks. New directions on nonlinear control. *IEEE Trans. Automt. Control*, 48:1750–1755, 2003.

[38] B. Hamzi and L. Praly. Ignored input dynamics and a new characterization of control Lyapunov functions. *Automatica J. IFAC*, 37:831–841, 2001.

[39] A. D. Ioffe and V. Tikhomirov. *Theory of Extremal Problems*. Nauka, Moscow, 1974. English translation, North-Holland (1979).

[40] C. M. Kellett and A. R. Teel. On the robustness of $\mathcal{KL}$-stability for difference inclusions: smooth discrete-time Lyapunov functions. *SIAM J. Control Optim.*, 44:777–800, 2005.

[41] N. N. Krasovskii and A. I. Subbotin. *Game-Theoretical Control Problems*. Springer-Verlag, 1988.

[42] Yu. S. Ledyaev and E. D. Sontag. A Lyapunov characterization of robust stabilization. *Nonlinear Analysis*, 37:813–840, 1999.

[43] A. A. Milyutin and N. P. Osmolovskii. *Calculus of Variations and Optimal Control*. American Math. Soc., Providence, 1998.

[44] L. W. Neustadt. *Optimization*. Princeton University Press, Princeton, 1976.

[45] S. Nobakhtian and R. J. Stern. Universal near-optimal feedbacks. *J. Optim. Theory Appl.*, 107:89–122, 2000.

[46] Z. Páles and V. Zeidan. Optimal control problems with set-valued control and state constraints. *SIAM J. Optim.*, 14:334–358, 2003.

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

[48] C. Prieur and E. Trélat. Robust optimal stabilization of the Brockett integrator via a hybrid feedback. *Math. Control Signals Systems*, 17:201–216, 2005.

[49] L. Rifford. Existence of Lipschitz and semiconcave control-Lyapunov functions. *SIAM J. Control Optim*, 39:1043–1064, 2000.

[50] L. Rifford. On the existence of nonsmooth control-Lyapunov functions in the sense of generalized gradients. *ESAIM Control Optim. Calc. Var.*, 6:539–611, 2001.

[51] L. Rifford. Semiconcave control-Lyapunov functions and stabilizing feedbacks. *SIAM J. Control Optim*, 41:659–681, 2002.

[52] L. Rifford. Singularities of viscosity solutions and the stabilization problem in the plane. *Indiana Univ. Math. J.*, 52:1373–1396, 2003.

[53] R. T. Rockafellar and R. Wets. *Variational Analysis*. Springer-Verlag, New York, 1998.

[54] H. Rodríguez, A. Astolfi, and R. Ortega. On the construction of static stabilizers and static output trackers for dynamically linearizable systems, related results and applications. *Internat. J. Control*, 79:1523–1537, 2006.

[55] E. P. Ryan. On Brockett's condition for smooth stabilizability and its necessity in a context of nonsmooth feedback. *SIAM J. Control Optim.*, 32:1597 –1604, 1994.

[56] E. D. Sontag. A Lyapunov-like characterization of asymptotic controllability. *SIAM J. Control Optim.*, 21:462–471, 1983.

[57] E. D. Sontag. Stability and stabilization: discontinuities and the effect of disturbances. In F. H. Clarke and R. J. Stern, editors, *Nonlinear Analysis, Differential Equations and Control (NATO ASI, Montreal 1998)*, pages 551–598. Kluwer Acad. Publ., Dordrecht, 1999.

[58] E. D. Sontag and H. J. Sussmann. Remarks on continuous feedback. In *Proc. IEEE Conf. Decision and Control, Albuquerque*, pages 916–921, Piscataway, December 1980. IEEE Publications.

[59] A. I. Subbotin. *Generalized Solutions of First-Order PDEs.* Birkhäuser, Boston, 1995.

[60] E. Trélat. Singular trajectories and subanalyticity in optimal control and Hamilton-Jacobi theory. *Rend. Semin. Mat. Univ. Politec. Torino*, 64:97–109, 2006.

[61] R. B. Vinter. *Optimal Control.* Birkhäuser, Boston, 2000.

[62] J. Warga. *Optimal Control of Differential and Functional Equations.* Academic Press, New York, 1972.

[63] P. R. Wolenski and Y. Zhuang. Proximal analysis and the minimal time function. *SIAM J. Control Optim.*, 36:1048–1072, 1998.