Nonsmooth analysis in control theory: a survey

Francis Clarke *

Abstract

In the classical calculus of variations, the question of regularity (smoothness or otherwise of certain functions) plays a dominant role. This same issue, although it emerges in different guises, has turned out to be crucial in nonlinear control theory, in contexts as various as necessary conditions for optimal control, the existence of Lyapunov functions, and the construction of stabilizing feedbacks. In this report we give an overview of the subject, and of some recent developments.

Contents

1 Introduction

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

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

*Institut Girard Desargues, Université Claude Bernard Lyon I, et professeur à l'Institut universitaire de France. Paper presented at the European Control Conference, Porto, September 2001. E-mail address: Clarke@desargues.univ-lyon1.fr

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

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

and where 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 or at the terminal time. 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 optimization, and gives rise to the subject of *optimal control*. Let us proceed to make explicit such an optimal control problem. We are given two functions: $\ell : \mathbb{R}^n \to \mathbb{R}$ and $L : \mathbb{R}^n \times U \to \mathbb{R}$, a subset C of $\mathbb{R}^n$ and a point $x_0 \in \mathbb{R}^n$, and we consider the problem

$$\text{minimize} \quad \ell(x(T)) + \int_0^T L(x(t), u(t))\, dt \qquad \text{[OCP]}$$

subject to $x(0) = x_0, x(T) \in C$. The system constraints (1) and (2) are also imposed, of course, and the minimization is relative to the choice of controls $u(\cdot)$. The problem is said to be one of fixed-time or free-time depending on whether the horizon T is prescribed or not.

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. [PCP]

It is clear that [OCP] is much more precisely stated than [PCP]; for example, it may often have a unique solution, which is not typical for [PCP]. This may be one of the reasons that mathematicians have generally paid more attention to optimal control, along with the fact that it naturally generalizes the Lagrange problem in the calculus of variations. On the other hand, [PCP], related as it is to the issue of stabilization, is of greater interest to control engineers, who are content with 'reasonable' solutions. While the two types of problems have always interacted to some extent, the underlying philosophies have generally been quite distinct. One important facet of this distinction has been the use of open-loop controls in optimal control, as opposed to closed-loop controls in systems applications.

It is one of our main purposes here to give an account of an approach to control theory which leads to a more unified point of view. One of the essential ingredients of this approach, perhaps surprisingly, is nonsmooth analysis. Another is the use of possibly discontinuous feedbacks.

The main topics we will discuss are listed above in the table of contents. Of course, even a modestly complete account of the above topics would require several volumes, some of which may not be ready to be written. So our report will in truth be an outline, with its inevitable arbitrariness regarding what to leave in and what to leave out. We provide references to the literature in which can be found all the details as well as further bibliographic pointers.

2 The dynamic programming method

We consider optimal control first. As is the case in optimization generally, certain problems arise in which the underlying data itself is nonsmooth, for example the system function f in (1), or the integral cost functional L of [OCP]. Minimax criteria give one example of this, and others are given in [16]. In this section, however, we wish to convey to the reader how considerations of nondifferentiability arise from the very way in which we might hope to solve the problem, even if the data is smooth. To unburden the discussion, we shall simplify [OCP] to a famous special case, namely the *minimal time* problem.

It consists of finding the least $T \geq 0$ admitting a control function $u(\cdot)$ on $[0, T]$ having the property that the resulting state x satisfies $x(T) = 0$. This corresponds to the free-time case of [OCP] in which $C = \{0\},\ \ell \equiv 0,\ L \equiv 1$.

By a *trajectory* of the system we mean a state function $x(\cdot)$ corresponding to some choice of admissible control function $u(\cdot)$. In terms of trajectories, then, the problem is to find one which is optimal from x_0; that is, one which reaches the origin as quickly as possible. We proceed 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 $s < t$,

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

Equivalently, the function

$$t \mapsto T\big(x(t)\big) + t$$

is increasing. Furthermore, if x is optimal, then the same function is constant.

Let us explain this in other terms: 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

$$\min_{u \in U} \big\langle \nabla T(x), f(x(t), u) \big\rangle + 1 = 0. \tag{3}$$

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

$$h\big(x, \nabla T(x)\big) + 1 = 0, \tag{4}$$

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

Here is the first step in the dynamic programming method: use the Hamilton–Jacobi equation (4), 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 (3). This suggests the following procedure: for each x, let $k(x)$ be a point in U satisfying

$$\begin{aligned} \min_{u \in U} \big\langle \nabla T(x), f(x, u) \big\rangle &= \\ \big\langle \nabla T(x), f(x, k(x)) \big\rangle &= -1. \end{aligned} \tag{5}$$

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

$$\dot{x}(t) = f\big(x(t), k\big(x(t)\big)\big), \quad x(0) = \alpha, \tag{6}$$

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

Here is why: Let $x(\cdot)$ satisfy (6); 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 must have $x = 0$. 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 (6)), 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 (4), 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 (5) for each x. The meaning and existence of a trajectory $x(\cdot)$ generated by $k(\cdot)$ via the differential equation (6), 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 an 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.

3 Subdifferentials and viscosity solutions

We require a differential construct that applies to functions that are not differentiable in the usual sense. Experience now indicates that several such notions are useful and necessary in general. One very basic and useful tool of this type is the *proximal subgradient*. Let $\phi : \mathbb{R}^n \to (-\infty, \infty]$ be a given function, and x a point where $\phi(x)$ is finite. A vector ζ in $\mathbb{R}^n$ is said to be a proximal subgradient of ϕ at x provided that there exist a neighborhood Ω of x and a number $\sigma \geq 0$ such that

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

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 $\partial_P \phi(x)$, and is referred to as the *proximal subdifferential*. If ϕ is differentiable at x, then we have $\partial_P \phi(x) \subset \{\phi'(x)\}$; equality holds if ϕ is of class C^2 at x. The *proximal density theorem* asserts that $\partial_P \phi(x)$ is nonempty for all x in a dense subset of

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

The existence of a proximal subgradient ζ at x corresponds to the possibility of approximating ϕ from below (thus in a *one-sided* manner) by a function whose graph is a parabola. The point $\big(x, \phi(x)\big)$ is a contact point between the graph of ϕ 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 ϕ is approximated by an affine function. As a guide to understanding, we ask the reader to do the following exercise (in dimension $n = 1$): the proximal subdifferential at 0 of the function $\phi_1(x) := -|x|$ is empty, while that of $\phi_2(x) := |x|$ is the interval $[-1, 1]$.

Surprisingly enough, the proximal subgradient admits a very complete calculus: all the usual calculus rules that the reader knows (and more) have their counterpart in terms of $\partial_P \phi$, and it is enough to assume merely that ϕ is lower semicontinuous. In a thorough treatment, there are some complexities to be dealt with, of the sort that mathematicians delight in , but for present purposes, we only need the object itself to help us define a generalized solution concept of the Hamilton-Jacobi equation.

We shall say that ϕ is a *proximal solution* of the Hamilton–Jacobi equation (4) provided that

$$h(x, \partial_P \phi(x)) = -1 \ \ \forall x \in \mathbb{R}^n, \tag{7}$$

a 'multivalued equation' which means that for all x, for all $\zeta \in \partial_P \phi(x)$ (if any), we have $h(x, \zeta) = -1$.

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. Consider, for example, the case in which $f(x, U)$ is equal to the unit ball for all x, in dimension $n = 1$. Then $h(x, p) \equiv -|p|$. The functions ϕ_1 and ϕ_2 defined above both satisfy $h(x, \nabla \phi(x)) = -1$ at all points $x \neq 0$, since for both ϕ_1 and ϕ_2, at all points different from 0, the proximal subdifferential reduces to the singleton consisting of the derivative. However, we have (see the exercise above)

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

and it follows that ϕ_1 is (but ϕ_2 is not) a proximal solution of the Hamilton-Jacobi equation (7).

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 1970's, with the 'generalized gradient' (see Section 7 below) and for locally Lipschitz solutions. While no uniqueness theorem holds for that solution concept, it was shown that the value function of a related 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 [51] 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.

Let us now explain the relationship between viscosity and proximal solutions. As introduced by Crandall and Lions [30], a viscosity solution ϕ of $F(x, \phi, \phi') = 0$ is a continuous function having the property that, for all x, whenever g is a smooth function such that $\phi - g$ admits a local minimum at x, then we have $F(x, \phi(x), g'(x)) \geq 0$, and whenever g is a smooth function such that $\phi - g$ admits a local maximum at x, then $F(x, \phi(x), g'(x)) \leq 0$.

The set of values $g'(x)$, where g is a smooth function such that $\phi - g$ admits a local minimum at x, is known as the *Dini subdifferential* (or *viscosity subdifferential*) of ϕ at x, denoted $\partial_D \phi(x)$. The corresponding object with 'maximum' instead is the Dini superdifferential $\partial^D \phi(x)$. The reason for this terminolgy is that (in the subdifferential case) one has $\zeta \in \partial_D \phi(x)$ iff

$$D\phi(x; v) \geq \langle \zeta, v \rangle \quad \forall\, v \in \mathbb{R}^n,$$

where the Dini subderivate is defined by

$$D\phi(x; v) := \liminf_{\substack{t \downarrow 0 \\ v' \to v}} \frac{\phi(x + tv') - \phi(x)}{t} \tag{8}$$

(see [24] for details).

In terms of subdifferentials, a viscosity solution of $F(x, \phi, \phi') = 0$ is a continuous function satisfying

$$\left\{ \begin{array}{lcl} F(x, \phi(x), \partial_D \phi(x)) & \geq & 0 \\ F(x, \phi(x), \partial^D \phi(x)) & \leq & 0 \end{array} \right. \quad \forall\, x \in \mathbb{R}^n.$$

Observe now that any element $\zeta \in \partial_P \phi(x)$ is such that the function

$$y \to \phi(y) - [\langle \zeta, y \rangle - \sigma \|y - x\|^2]$$

admits a local minimum at x; further, the derivative at x of the function in brackets equals ζ. Hence a viscosity solution necessarily satisfies $F(x, \phi(x), \zeta) \geq 0$. Since this holds for each $\zeta \in \partial_P \phi(x)$), we conclude $F(x, \phi(x), \partial_P \phi(x)) \geq 0$. A similar argument shows that a viscosity solution ϕ must satisfy $F(x, \phi(x), \partial^P \phi(x)) \leq 0$, where the *proximal superdifferential* $\partial^P \phi(x)$ is given by $-\partial_P(-\phi)(x)$. (Alternatively, it can be defined directly as $\partial_P \phi(x)$ was, by reversing the defining inequality and taking $\sigma \leq 0$.)

Of course, the proximal subgradients and supergradients correspond to the special case in which the functions g appearing in the definition of viscosity solution are linear/quadratic, but the sufficiency of the two proximal inequalities for ϕ to be a viscosity solution can be proven (under mild hypotheses) by a theorem of Subbotin (see [52] and [24]) which asserts, roughly speaking, that 'any Dini subgradient can be approximated by a proximal subgradient'.

To summarize to this point, ϕ is a viscosity solution of $F(x, \phi, \phi') = 0$ iff it satisfies

$$\left\{ \begin{array}{lcl} F(x, \phi(x), \partial_P \phi(x)) & \geq & 0 \\ F(x, \phi(x), \partial^P \phi(x)) & \leq & 0 \end{array} \right. \quad \forall\, x \in \mathbb{R}^n.$$

Now suppose that F is convex in its third argument. Then it can be shown (under mild hypotheses) that the two inequalities

$$F(x, \phi(x), \partial^P \phi(x)) \leq 0, F(x, \phi(x), \partial_P \phi(x)) \leq 0$$

are equivalent (this is the case because every element of $\partial_P \phi(x)$ is *almost* a convex combination of elements of $\partial^P \phi(x)$, and vice versa (see [22] for the proximal inversion formula that is being alluded to). In this case, then, it follows that a viscosity solution of $F(x, \phi, \phi') = 0$ is a function satisfying the 'single' (or unilateral) condition

$$F(x, \phi(x), \partial_P \phi(x)) = 0 \quad \forall\, x \in \mathbb{R}^n.$$

Since the lower Hamiltonian h is concave in its last argument, we see that the proximal solutions defined via (7) above correspond to viscosity solutions of the partial differential equation $-h(x, \phi'(x)) = 0$ (but they are not necessarily continuous, just lower semicontinuous). This type of 'unilateral' characterization was first derived by E.N. Barron and R. Jensen [7]. It should be stressed that in the absence of some convexity hypothesis, viscosity solutions do not generally admit a unilateral proximal characterization like the above (see [6]).

Recall that our goal (within the dynamic programming approach) is to characterize the minimal time function. Using either the proximal or the viscosity concept, this is now attained, as shown by the following, whose mild hypotheses on the data we omit (see [55], and also the extensive discussion in Bardi and Capuzzo-Dolcetta [5]).

Theorem 3.1 *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] $h(x, \partial_P \phi(x)) = -1 \quad \forall\, x \neq 0;$

[Boundary condition]

$$\phi(0) = 0 \textit{ and } \quad h(0, \partial_P \phi(0)) \geq -1.$$

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..

We have reached, then, 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

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

how does one proceed to construct a feedback $k(x)$ having the property that any trajectory x generated by it via (6) (for which we still have to define the sense) is such that $t \mapsto T(x(t))$ decreases at a unit rate ?

We shall leave the optimal control problem [OCP] temporarily at this point in order to discuss [PCP] and feedback stabilization, for we shall see that in analyzing that apparently quite different issue, we shall arrive at precisely the same point as above.

4 Stabilizing feedback

We continue to study the control system (1)(2), but now from the point of view of [PCP], which is concerned with position rather than optimality. It is convenient to suppose that 0 is an equilibrium point of the system; in fact, let us posit $f(0, 0) = 0$, and $0 \in U$. In this context, an important property of the system is that it be Globally Asymptotically Controllable [GAC] (to the origin). This means that from any initial condition x_0 there is an open-loop control $u(t)$ with corresponding trajectory x beginning at x_0 such that $x(t) \to 0$ as $t \to +\infty$. (The definition also includes a local stability property at 0 that we ignore in this discussion.)

In contrast to the optimal control problem, the intended applications of [PCP] force us to consider feedback from the start. To put it briefly, an open-loop control $u(t)$, even if it has the property of steering x_0 to 0, is of little interest since it has no robustness with respect to the starting point: it does not steer the system to 0 except from the one initial condition (which will not be known *a priori* in any case). A feedback mechanism is itself the actual goal.

The natural question to pose, then, is the following: if the system is GAC, is there a *stabilizing feedback* $k(x)$; that is, a function with values in U such that all solutions of the differential equation

$$\dot{x}(t) = f\big(x(t), k\big(x(t)\big)\big), \quad x(0) = x_0 \tag{10}$$

converge asymptotically to 0 (for all values of x_0)? In some sense, we would wish to 'piece together' the various open-loop controls that each steer certain points to 0 into one coherent feedback law.

The classical linear case of our system is that in which $f(x,u)$ has the form $Ax + Bu$ for certain matrices A and B, and in which $U = \mathbb{R}^m$. Linear systems theory has a resoundingly positive answer to the question in that case: there is even a linear feedback $k(x) := Kx$ having the desired effect.

The situation for nonlinear systems is much more complex. As is always the case in nonlinear settings, one approach that suggests itself is to approximate the nonlinear system through a linear one obtained by *linearization*. Thus we set

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

and attempt to find a linear (locally) stabilizing feedback for the system $\dot{x} = f(x,u)$ by constructing one for the approximating system $\dot{x} = Ax + Bu$.

This approach has been quite 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;
- Both x and u must remain small so that the linear approximation remains relevant (the feedback is operative only for small perturbations from the equilibrium state).

The real nonlinear nature of the stabilizing feedback problem becomes a factor when these conditions are not met, as in the following simple example taken from [49].

Example

We consider a state (x,y) in two dimensions and a one–dimensional control u in which the underlying system is

$$\left\{ \begin{array}{rcl} \dot{x} & = & u(x^2 - y^2) \\ \dot{y} & = & 2uxy \end{array} \right. \qquad u \in [-1,1].$$

Let us understand the behavior of the trajectories of the system. Given an initial condition $(x_0, y_0) \in \mathbb{R}^2$ with $y_0 \neq 0$, consider the unique circle of the form

$$\{(x',y') : {x'}^2 + (y'-c)^2 = c^2\}$$

which is centered on the y-axis and contains both the origin and (x_0, y_0). At any point (x,y) of that circle, the vector $(x^2 - y^2, 2xy)$ is tangent to the circle at (x,y). It follows that the circle is an invariant set, so that any trajectory of the given control system that starts at (x_0,y_0) must remain on the circle. For $y_0 = 0$ we obtain a different invariant set: the x-axis.

The choice of $u \equiv +1$ (or more generally, $u > 0$) gives rise to counter–clockwise movement on the circles above the x-axis and clockwise movement on those situated below the x-axis; along the x-axis itself we get movement toward the right. For the case $u < 0$, all these motions are simply reversed. We observe that this system is GAC.

Let us now suppose that a continuous stabilizing feedback k exists for it. Then $k(x,y)$ must be different from 0 for all $(x,y) \neq 0$, for otherwise the system admits an equilibrium different from 0. We deduce from this that k must be strictly negative along the positive x-axis, and strictly positive along the negative x-axis. But if we now consider the values of k along any circle centered at the origin, it follows from the Intermediate Value Theorem that k vanishes at some point on that circle, a contradiction which proves that no continuous stabilizing feedback exists. (The continuity of k at the origin is not an issue in the analysis.) Observe that the argument also shows that no continuous feedback could 'approximately stabilize' the system (to some ball around the origin, say).

We remark that a simple *discontinuous* stabilizing feedback law is available here: use $k = 1$ to the left of the y-axis and $k = -1$ to the right (and either value on the axis itself). In applying this rule we immediately leave the set upon which k is discontinuous (the y-axis), never to return, so that no difficulty arises in applying k: the control is always constant ($+1$ or -1, depending on the initial condition).

The nonholonomic integrator

Another, and rather famous, example in nonlinear systems theory is that known as the *nonholonomic integrator*, a term which refers to the following system, which is linear (separately) in the state and in the control variables:

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

(Thus U is the closed unit ball in $\mathbb{R}^2$.) The fact that this system is globally asymptotically controllable can be proven by an elementary ad hoc argument. It was proven by Brockett in 1983 that, nonetheless, there is no continuous feeback law $u = k(x)$ which will stabilize the system (even locally about the origin). In fact, he obtained the following necessary condition for the existence of such a stabilizer: if $\dot{x} = f(x, u)$ admits a continuous stabilizing feedback, then for every neighborhood Ω of 0, the set $f(\Omega, U)$ is also a neighborhood of 0. In the case of the nonholonomic integrator, we invite the reader to verify that $f(\Omega, U)$ can contain no point of the form $(0, 0, \mu)$ for any $\mu \neq 0$, so that Brockett's condition rules out the existence of a continuous stabilizing feedback (even one that could be discontinuous at 0). We remark that the Example analyzed earlier does satisfy Brockett's condition, although it fails to admit a continuous stabilizer.

Discontinuous feedbacks

The reader may have observed that in each of the last three instances in which we used the word 'feedback', it was modified by the word 'continuous'. What about using *discontinuous* stabilizing feedback laws, since examples as seemingly benign as the nonholonomic integrator do not admit continuous ones?

In considering this possibility, there arises what may seem to be a merely technical point, but one which in fact turns out to be critical: if $k(x)$ is a discontinuous function, what is meant by a solution of the differential equation $\dot{x} = f(x, k(x))$? In the continuous case, of course, we simply mean that $x(\cdot)$ is a smooth function whose derivative $\dot{x}(t)$ coincides for each t with the value $f\big(x(t), k(x(t))\big)$. In the discontinuous case, we might reasonably try the following definition: a solution $x(\cdot)$ is an absolutely continuous function whose derivative $\dot{x}(t)$ coincides *almost everywhere* with the value $f\big(x(t), k(x(t))\big)$. This *Caratheodory solution* concept, however, turns out to be entirely unsatisfactory from several points of view, notably that of existence. Because of this, there have been various proposals for what a solution should mean in a discontinuous setting.

The best known of these is due to Filippov [31]: a *Filippov solution* $x(\cdot)$ of $\dot{x} = g(x)$ (where g is possibly discontinuous) means an absolutely continuous function whose derivative $\dot{x}(t)$ satisfies

$$\dot{x}(t) \in \mathrm{co} \left\{ \bigcap_{\delta > 0,\, m(\Omega) = 0} g\Big(B(x(t), \delta) \backslash \Omega\Big) \right\} \text{ a.e..}$$

Here, co stands for 'convex hull', $m(\Omega)$ is the Lebesgue measure of Ω, and the notation $B(x, r)$ stands for the open ball of radius r centered at x. In words, we require that, almost everywhere, $\dot{x}(t)$ belong to the convex hull of all values of g that are (at $x(t)$) 'essential and persistent locally in measure'. As an example, consider for $n = 1$ the function $g(x)$ equal to -1 for $x > 0$ and $+1$ for $x \leq 0$ (with initial condition $x(0) = 0$. Then there is no Caratheodory solution, but a unique Filippov solution ($x \equiv 0$).

The Filippov solution concept has a number of desirable properties from the mathematical point of view, although from the practical point of view it is not clear that such system trajectories are necessarily physically meaningful. In any case, armed with this notion, we are now led to the natural question: if a system is GAC, does it necessarily admit a stabilizing feedback, possibly discontinuous, if one interprets the ensuing trajectories in the Filippov sense?

The answer is 'no'. Indeed, as shown by Ryan [46] and by Coron and Rosier [29], Brockett's condition continues to hold for the Filippov solution concept, so that the nonholonomic integrator (for example) cannot be stabilized by a discontinuous feedback in the Filippov sense.

A recent result of a positive nature appears in [21], where it was shown that any globally asymptotically controllable system is stabilizable by a (possibly discontinuous) feedback if the trajectory $x(\cdot)$ associated to the feedback is defined not in the Filippov sense, but rather in a natural way that involves discretizing the control law. We refer to this solution concept as *closed-loop system sampling*; it has been used in other contexts, notably in differential games by Krasovskii and Subbotin [37]. We proceed now to describe it.

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 $\operatorname{diam}(\pi)$, 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

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

(Of course in general we do not have uniqueness of the solution, nor is there necessarily even one solution, although nonexistence can be ruled out when blow-up of the solution in finite time cannot occur, as is the case in the stabilization problem.) 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 *sample and hold* 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 [42]; 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 [8, 10]. 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 [37] 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).

Returning now to π-trajectories, it is natural to say that a feedback $k(x)$ (continuous or not) stabilizes the system (1)(2) 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

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

The following theorem is proven in [21].

Theorem 4.1 *Let the system (1)(2) be GAC. Then there exists a (possibly discontinuous) feedback k : $\mathbb{R}^n \to U$ which stabilizes it in the sense of closed-loop system sampling.*

The proof of the theorem actually yields precise estimates regarding how small the step size diam (π) must be for a prescribed stabilization tolerance to ensue, and of the resulting stabilization time, in terms of the given data. These estimates are uniform on bounded sets of initial conditions, and are relative to a given *Lyapunov function* for the system. This important concept is the next topic for discussion.

5 Lyapunov functions

A smooth function $V : \mathbb{R}^n \to \mathbb{R}$ is said to be a *Lyapunov function* for the system (1)(2) if it satisfies the following three properties:

V is positive definite:

$$V(x) \geq 0, V(x) = 0 \text{ iff } x = 0.$$

V is proper:

$$V(x) \to \infty \text{ as } \|x\| \to \infty.$$

Infinitesimal decrease: For some continuous function W that is positive on $\mathbb{R}^n \backslash 0$ we have

$$\min_{u \in U} \langle \nabla V(x), f(x,u) \rangle \leq -W(x), x \neq 0. \quad (11)$$

It is well-known that the existence of such a 'control-Lyapunov function' (V,W) implies global asymptotic controllability to the origin: for every $\alpha \in \mathbb{R}^n$, there is a control $u(t)$ such that the solution $x(\cdot)$ of (1) with initial condition $x(0) = \alpha$ satisfies $x(t) \to 0$ as $t \to \infty$. (In addition, convergence to zero takes place in a certain uniform and stable manner that we will not dwell upon here.) In fact, the most common way to verify that the system is GAC is to exhibit such a function.

Recall now our desire to produce a state feedback $k(\cdot) : \mathbb{R}^n \to U$ which stabilizes the system; i.e., such that the system $\dot{x} = f(x, k(x))$ is globally asymptotically stable. A natural question is whether one can define such a feedback law through the use of a given Lyapunov function V.

The *ideal case*, a well-known heuristic useful for motivational purposes, is the one in which we can find a continuous function $k(x)$ that selects a value of $u \in U$ attaining (or almost) the minimum in (11):

$$\langle \nabla V(x), f(x, k(x)) \rangle \leq -W(x) \quad \forall x \neq 0. \quad (12)$$

Then any (classical) solution of $\dot{x} = f(x, k(x))$ is such that

$$\frac{d}{dt} V(x(t)) = \langle \nabla V(x(t)), \dot{x}(t) \rangle \leq -W(x(t)) < 0,$$

a monotonicity conclusion that, together with the growth property of V, assures that $x(t) \to 0$ as $t \to \infty$.

There are two fundamental difficulties with this ideal picture, and again both concern regularity issues. The first is whether a smooth Lyapunov function actually exists in the first place (when the system is GAC), and the second is whether, even assuming a smooth V does exist, the continuous selection $k(\cdot)$ exists. If we have recourse to a discontinuous selection $k(\cdot)$, then, as above, the issue arises of how to interpret the discontinuous differential equation $\dot{x} = f(x, k(x))$.

Let us first discuss the existence or otherwise of a smooth Lyapunov function V.

An early and important result of Artstein [2] implies in particular that the nonholonomic integrator fails to admit a smooth V (see [23] for related results).

Let us show directly for the Example of Section 4 the nonexistence of a smooth Lyapunov function V. If such a one exists, then consider the maximization of V over a given circle centered on the y-axis and passing through the origin. The maximum is attained at a point $(x, y) \neq 0$, and we have $\nabla V(x, y)$ normal to the circle at (x, y). But then $\nabla V(x, y)$ makes a zero inner product with any element of the velocity set $f(x, y, U)$, which implies that the Infinitesimal Decrease Condition must fail.

Although a smooth Lyapunov function may not exist for a GAC system, it has been shown by Sontag [47] that globally asymptotically controllable systems always admit a continuous Lyapunov function V satisfying, instead of (11), the following nonsmooth version of the Infinitesimal Decrease Condition:

$$\inf_{v \in \operatorname{co} f(x,U)} DV(x;v) \leq -W(x) < 0, x \neq 0, \tag{13}$$

where the lower Dini derivate DV has been defined previously (see (8)).

The theory of nonsmooth analysis asserts the equivalence to (13) of another, and for our purposes more useful, form of the Infinitesimal Decrease Condition:

$$\inf_{u \in U} \langle f(x,u), \zeta \rangle \leq -W(x) < 0$$
$$\forall \zeta \in \partial_P V(x), \forall x \neq 0. \tag{14}$$

The equivalence of (13) and (14) is a consequence of Subbotin's Theorem, which links Dini derivates to proximal calculus (see [24]). In terms of the lower Hamiltonian h introduced in Section 2, (14) can be written

$$h(x, \partial_P V(x)) \leq -W(x),$$

which is to be compared to (4)

It is at this precise point that we can observe the confluence of the problems [OCP] and [PCP]. Recall that we suspended our analysis of the problem [OCP] (at the end of Section 2) before its conclusion. In that case, as in the present one, we have a function (T or V) satisfying a Hamilton-Jacobi condition (14) (with $W \equiv 1$ in the case of [OCP], see (9)), and the object in both cases is to find a feedback which induces trajectories x along which V decreases at rate W: in differential form, we seek

$$\frac{d}{dt} V(x(t)) = -W(x(t)).$$

(For the minimal time problem we want $d/dt\, T\big(x(t)\big) = -1$; recall that this is equivalent to the optimality of the trajectory.)

Let us now discuss how such a feedback is constructed in [20]. We shall work in the stabilization setting, and assume that we have a (nonsmooth) Lyapunov function V; optimal control synthesis is then just a special case.

The essential reason for which proximal calculus is well-suited to our approach is because of its relation to metric projection onto sets, upon which is based the 'proximal aiming' method that we employ. The crux is this: when $x(t_i) = x$ lies outside a level set $S = S(c) := \{V \leq c\}$ and admits closest point (or projection) s in S, then $x - s$ is a 'proximal normal' vector to S at s, and for some $\lambda > 0$ we have $\lambda(x - s) \in \partial_P V(s)$ (a fact from proximal calculus). Then (14) can be invoked at s to find a suitable value of the control u which moves the state toward S, in the sense that the Euclidian distance d_S decreases at a certain positive rate Δ:

$$d_S(x(t)) - d_S(x(t_i)) \leq -\Delta(t - t_i), \quad t_i \leq t \leq t_{i+1},$$

provided $x(t_i)$ is close enough to S to start with, and provided $\operatorname{diam}(\pi)$ is small enough. A sequence of such sets S and associated local feedbacks is 'glued together' to produce the global stabilizing feedback $k(\cdot)$ that is sought.

The details of this construction require the Lyapunov function V to be locally Lipschitz. In the case of a system which is globally asymptotically controllable to the origin, the theorem of Sontag cited above provides one which is only continuous. And if we dealing with the minimal time function (for example), that may not be locally Lipschitz either. There are different ways of dealing with this gap.

In [21] the proof uses Moreau-Yosida inf convolution to make a continuous Lyapunov function Lipschitz as an intermediate step. This methodology was also employed earlier in [25], in a differential game setting.

Another approach is taken in [20] , where it is shown that one may directly construct an explicit locally Lipschitz function which serves as a Lyapunov function for so-called *practical stabilization*. 'Practical stabilization' refers to the possibility of stabilizing the system via feeback to an arbitrarily small ball about the origin (in finite time), instead of asymptotically to zero. Thus the function V constructed in this way only has the Lyapunov properties outside a ball around the origin. It is locally Lipschitz, however, so that the proximal aiming construction sketched above can be carried out.

A conclusion that appears to be emerging from recent work is that the continuity of stabilizing feedback may not in fact be as important a consideration as was at first believed. If one is implementing a feedback in the sample-and-hold method used in system-sampling (which now appears to be the right thing to do), then its continuity or otherwise is irrelevant. An issue of greater importance seems to be the degree of regularity that one can obtain for a control Lyapunov function.

In this connection, a major step in the theory has been taken by L. Rifford in [43, 45]. He proved that a GAC system always admits a global Lyapunov function which is locally Lipschitz, which settles an open question in the subject. In fact, his results go beyond this in asserting the existence of a Lyapunov function which is *semiconcave*.

Semiconcavity

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 concave on $B(0, r)$. (Hence ϕ is 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 also a local definition of semiconcavity that we omit for present purposes.)

Semiconcavity is an important regularity property in differential equations. 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 (12) holds.

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 'many' x. We are led to consider the *limiting subdifferential* $\partial_L V(x)$, which, when V is continuous, is defined by applying a natural limiting operation to $\partial_P V$:

$$\partial_L V(x) := \left\{ \zeta = \lim_{i\to\infty} \zeta_i : \zeta_i \in \partial_P V(x_i), \lim_{i\to\infty} x_i = x \right\}. \tag{15}$$

It follows readily that, when V is locally Lipschitz, $\partial_L V(x)$ is nonempty for all x. And the Infinitesimal Decrease Condition (14) for proximal subgradients implies the following:

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

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

$$\langle f(x, k(x)), \zeta\rangle \leq -W(x).$$

Does this lead to a stabilizing feedback, when (of course) the discontinuous differential equation is interpreted in the closed-loop system sampling sense? When V is smooth, the answer is 'yes', but when V is merely locally Lipschitz, a certain 'dithering' phenomenon may arise to prevent k from being stabilizing. However, if V is semiconcave (on $\mathbb{R}^n\backslash\{0\}$), this does not occur, and stabilization is guaranteed. This accounts in part for the desirability of a semiconcave Lyapunov function, and the importance of knowing one always exists.

The properties of a semiconcave Lyapunov function also play an important role in another theorem of Rifford [43, 44] in which it is shown that for a certain class of systems (which includes the nonholonomic integrator), there is always a stabilizing feedback having the pleasant properties of the simple discontinuous one found in connection with the Example of Section 4. That is, the feedback in question is continuous on an open dense set, and when it is applied, the points of discontinuity are *repulsive*: the state lies in that set at most at the initial point. Thus the stabilization is 'essentially' achieved by continuous feedback.

The proof of this result uses the properties of semiconcave functions, together with both proximal and generalized gradient calculus (see below).

6 Robustness of discontinuous feedback

In applying a given stabilizing feedback law to the system, there will in general arise some inaccuracy. The nominal system $\dot{x} = f(x, k(x))$ may become

$$\dot{x} = f(x + e, k(x + p) + q) + r, \tag{16}$$

where e, p, q and r are unknown perturbations or errors. The effect of the external disturbance r is relatively benign: if r is small, then (in the absence of other errors) although the system may not be stabilized, the state is nonetheless driven to a ball around the origin whose radius is proportional to the size of r (thus, a small error has a small effect). The effect of e and q can be reduced to that of r if f is a continuous function. The truly troublesome term in (16) is p, which is a measurement error: when the state is x we measure it as $x + p$, and hence apply the 'wrong' value of k. Because k may not be continuous, even a small p could lead to a big change in the resulting control value; this explains in part the natural preference that a systems engineer would have for continuous feedbacks. In the general discontinuous case, we wish to know whether stabilization still occurs in the presence of measurement error. If so, we refer to this as a robustness property of the given feedback.

If a feedback k is defined according to the 'ideal case' presented above at the beginning of Section 5, where V is smooth, then some robustness is to be expected, as evidenced by the following observation due to Sontag [49]. We defined k so as to have

$$\langle \nabla V(x), f(x, k(x)) \rangle \leq -W(x).$$

With measurement error present, the decrease rate is actually given by

$$\langle \nabla V(x), f(x, k(x + p)) \rangle,$$

and this may seem to be potentially quite different if k is discontinuous, even when p is small. However, bear in mind that we also have

$$\langle \nabla V(x + p), f(x + p, k(x + p)) \rangle \leq -W(x + p).$$

Thus, if ∇V is continuous (that is, if V is smooth),and since W and f are continuous, we shall indeed have

$$\langle \nabla V(x), f(x, k(x + p)) \rangle \leq -W(x)/2$$

provided that p is small, and this gives adequate decrease to imply stabilization. This informal argument indicates that the existence of a smooth Lyapunov function implies robust stabilization, even if no continuous stabilizing feedback exists. But a smooth Lyapunov function may not exist; what then is the situation more generally?

Ledyaev and Sontag [38] have proved that there is a close relationship between the issues of 'how regular a Lyapunov function does the system admit' and 'how robust a stabilizing feedback does the system admit'. Consider for example a perturbed equation $\dot{x} = f(x, k(x + p))$, where, as above, p represents a measurement error in applying the feedback k. *Full robustness* of the feedback k is taken to mean that for any ϵ, there is a $\delta > 0$ such that whenever the perturbation $p(t)$ satisfies $\|p(t)\| \leq \delta$ for all t, then stabilization to the ϵ-ball takes place.

It is shown in [38] that the system admits a fully robust stabilizing feedback iff it admits a smooth (C^1 or C^∞) Lyapunov function. Thus the nonholonomic integrator, which *can* be stabilized by a discontinuous feedback (as we have seen), does *not* admit a fully robust stabilizing feedback. We remark that it is possible to recover robustness, however, through the use of a *dynamic* feedback . This involves the construction of an auxiliary state variable z and of a feedback law depending on both x and z; since z depends on (all) the past values of x (and so indirectly on the initial condition), this is also referred to as 'feedback with memory'. We remark that the issue of the robustness of discontinuous feedbacks with respect to measurement error seems to have been raised first by Hermes [35].

The above concept of full robustness is not closely related to the system sampling method that we employ. Let us instead introduce a type of *relative robustness* in which we relate the size of the measurement error to the maximum step size δ of the underlying partition. To begin, we ask whether for small but nonzero p, sufficiently small steps (or equivalently, sufficiently high sampling rate) will lead to stabilization.

A look at the simple discontinuous stabilizing feedback k found for the Example of Section 4 can lead to a troubling conclusion in this regard. Suppose that the state is currently at $(x, y) = (\epsilon_1, 1)$,where ϵ_1 is small but positive. The presence of a measurement error $p = (-2\epsilon_1, 0)$ may lead us to believe that the state lies to the left of the y-axis; accordingly we apply $u = 1$, in keeping with k. This has the effect of moving the state counterclockwise at the next partition point, let us say to $(-\epsilon_2, 1)$. Although the state is now to the left of the axis, a new small measurement error may lead us to believe that it lies to the right; in this case we would apply $k = -1$ on the next partition interval, thereby inducing clockwise movement and possibly returning to the initial point $(\epsilon_1, 1)$. We could keep 'dithering' back-and-forth in this way, so that an arbitrarily small (and insidiously clever) error has blocked stabilization.

The problem above can be viewed as being due to 'oversampling', in the sense that if we had allowed the state to evolve for a longer period before measuring it anew, the stabilizing effect would have had a chance to make itself felt adequately. (To be precise in our example, the problem is avoided if ϵ_2 is greater than the next measurement error can be.) This is, in effect, an argument in favor of *bigger* steps in the system sampling scheme.

This heuristic can be made precise and a positive result proven for the stabilizing feedbacks constructed either through proximal aiming or, more directly, through the limiting subdifferential of a semiconcave Lyapunov function, as explained above. For these feedbacks, it turns out that by suitably limiting the partition step size *from below*, the dithering can be excluded. At the same time, of course, the step size must be limited from above, in order for the feedback to be stabilizing. This insight leads in [20] to the stipulation of *reasonably uniform* partitions. This is taken to mean that the following holds for some $\delta > 0$:

$$\frac{\delta}{2} \leq t_{i+1} - t_i \leq \delta \quad \forall i \geq 0,$$

although it is possible to replace the factor 1/2 by any constant in $(0, 1]$.

Theorem 3 of [20] affirms that if the system (1)(2) is GAC, it admits a relatively robust stabilizing feedback k to any prescribed tolerance $r > 0$. Thus, for all initial conditions in a bounded set, for some positive T and δ_0, we will have $\|x(t)\| \leq r$ for all $t \geq T$, whenever x is a π-trajectory, if π is a reasonably uniform partition whose diameter δ satisfies $\delta < \delta_0$, and whenever measurement error p does not exceed a critical level $E(\delta)$ related to the sampling rate. We remark that the values of T and δ_0 can be estimated in terms of r and the data of the system; the 'maximum error' function $E(\delta)$ is linear in δ.

Among the many important variants of the stabilization issue that can be considered is the one in which it is required to constrain the state to lie in a given closed set. Progress on such state-constrained feedback is reported in [27]. In connection with robustness, the issue of dealing with measurement error that is big in relation to the step sizes required for (errorless) stabilization appears to be open, as do a number of substantial issues stemming from incomplete information (for example, when only a partial observation of the state x is available). We remark that there exists a general method for calculating stabilizing feedback in the absence of an explicit Lyapunov function. It is optimization-based, and referred to as the *receding horizon* (or *model predictive*) method; see for example the recent article of Fontes [32]

7 Necessary conditions for optimality

We now return to the optimal control problem [OCP] of Section 1, in order to discuss the *deductive method* for solving it (as opposed to, say, the dynamic programming method). This refers to the use of necessary conditions in an optimization problem in order to identify its solutions. In the case of [OCP], it is the celebrated Maximum Principle of Pontryagin that is best known; we proceed to recall its statement, in the original smooth setting. To ease the exposition, we shall consider only the case in which the horizon T is fixed, and $C = \mathbb{R}^n$ (thus, the terminal state is unconstrained).

It is convenient to introduce two functions, the (upper) Hamiltonian H and the pseudo-Hamiltonian $\overline{H}$:

$$\overline{H}(x, p, u) := \langle p, f(x, u)\rangle - L(x, u),$$

$$H(x, p) := \sup_{u \in U} \overline{H}(x, p, u).$$

Theorem 7.1 *Let the trajectory x and its control u solve [OCP]. Then there is a function p on $[0, T]$ which satisfies*

(a) $-\dot{p}(t) = \nabla_x\big(\overline{H}(\cdot, p(t), u(t))\big)(x(t))$ a.e.,

(b) $\overline{H}(x(t), p(t), u(t)) = H(x(t), p(t))$ a.e.,

(c) $H(x(t), p(t)) = c =$ *constant on* $[0, T]$,

(d) $-p(T) = \nabla \ell(x(T))$.

There is much to say concerning the meaning of this result, its classical origins in the calculus of variations, and its applications. For present purposes, we refer the reader to [19] or [54] for these matters, and we assume some familiarity with the theorem on the part of the reader. The issue we address is the following: suppose now that the functions f, L, and ℓ are nondifferentiable; how does one extend the Maximum Principle to that setting?

The problem is not merely one of stating a generalization, since the classical methodology for proving the theorem is precisely the linearization technique that is unavailable for nonsmooth data. Certain alternative approaches to the problem were developed by Clarke [12, 16], together with the calculus of *generalized gradients* [13], which has been widely applied in nonsmooth analysis and optimization

Generalized gradients

Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function, and define its generalized directional derivative as follows:

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

A vector ζ belongs to the generalized gradient of ϕ at x, denoted $\partial_C \phi(x)$, iff one has

$$\phi^\circ(x; v) \geq \langle \zeta, v\rangle \quad \forall\, v \in \mathbb{R}^n.$$

In case the reader has lost count, we mention that this is the *fourth* subdifferential to be introduced in this article. We believe that no others are required, for all current purposes. These objects together possess geometric counterparts and a complete interrelated calculus that we are unable to delve into here. Nonetheless, let us remark that the generalized gradient possesses a few characteristics that make its calculus especially useful in certain applications (we cited one earlier: Rifford's theorem on discontinuous but 'nice' stabilizing feedback). An example of such a property is the formula

$$\partial_C(-\phi)(x) = -\partial_C \phi(x),$$

which does not hold for the three other subdifferentials we have seen (and which shows that $\partial_C \phi$ is not really a subdifferential). Another is the following Gradient Formula:

$$\partial_C \phi(x) = \operatorname{co}\left\{\lim_{i \to \infty} \nabla \phi(x_i) : x_i \to x, x_i \notin \Omega\right\},$$

where Ω is *any* set of measure zero containing the local points of nondifferentiability of ϕ. Thus the generalized gradient is 'blind to sets of measure zero'. This formula resembles (15), the defining one for $\partial_L \phi$.

However, the latter fails to have that insensitivity property. It can be shown that [24]

$$\partial_C \phi(x) = \mathrm{co}\, \partial_L \phi(x).$$

In [12] and [14] it is proven that when the data of [OCP] are locally Lipschitz, the Maximum Principle holds with derivatives replaced by generalized gradients. Thus, for example, the 'adjoint equation' (assertion (a) of the Maximum Principle) becomes a *differential inclusion*

$$-\dot{p}(t) \in \partial_C \big(\overline{H}(\cdot, p(t), u(t))\big)(x(t)) \text{ a.e..}$$

Many other extensions and refinements of the Maximum Principle have been obtained since this initial work, to the point that the theory of generalized necessary conditions could be viewed as rather complete for the moment. For these and related matters, we refer the interested reader to [19] (which is rather expository in style) and to R. B. Vinter's book [54] (which has very thorough discussions of the recent literature).

Hamiltonian multipliers; sensitivity

We shall close our necessarily brief discussion of optimality conditions with the subject of *Hamiltonian multipliers*, which at its origins reflected a desire to express necessary conditions entirely in terms similar to those of the classical calculus of variations (and hence in terms of the true Hamiltonian H). This also involves viewing [OCP] as a 'generalized problem of Bolza'; we refer to [19] for a detailed discussion of the topics of this section. The following result was proven in parallel to the extended Maximum Principle [12, 15]:

Theorem 7.2 *Let the trajectory x and its control u solve [OCP]. Then there is a function p on $[0,T]$ which satisfies*

(i) $\big(-\dot{p}(t), \dot{x}(t)\big) \in \partial_C H(x(t), p(t))$ a.e.,

(ii) $H(x(t), p(t)) = c =$ *constant on* $[0,T]$,

(iii) $-p(T) = \nabla \ell(x(T))$.

The point here is that the single condition (i) encapsulates both (a) and (b) of the Maximum Principle (as well as the state equation (1)). These Hamiltonian necessary conditions also apply to situations in which the control set U depends on x, in contrast to the Maximum Principle. Being in true Hamiltonian form, they are more closely related to classical mechanics; this has made them useful in , for example, the study of periodic orbits of Hamiltonian sytems [19]. There have been a number of interesting developments in connection with the Lagrangian and Hamiltonian approach to problems of Bolza, (by Rockafellar, Loewen, Mordukhovic, Vinter, and others), and of course other issues arise (such as existence and regularity); we refer the reader to [54].

A function p satisfying the three conditions of the theorem is referred to as a *Hamiltonian multiplier* for the trajectory x. We denote by $M(x)$ the set of such multipliers associated with a given trajectory x. It turns out that a great deal of sensitivity information resides in these mutipliers, in the sense that they yield information on the rate of change of the value of the problem with respect to perturbation. Let us illustrate this in a simple case, one in which we consider a perturbation of the underlying system (1)(2) by a nonautonomous function α in $L^2\big([0,T], \mathbb{R}^n\big)$:

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

Let the value of the resulting problem [OCP] (which is otherwise unchanged) be denoted $V(\alpha)$. As we know, value functions like V will not be differentiable in general, even if the problem data is smooth. However, we can estimate the subdifferential of V, as in the following result taken from [17]. (We denote by Σ the set of trajectories x which solve the original *unperturbed* problem [OCP].)

Theorem 7.3 *The function V is Lipschitz (in the L^2 norm) in a neighborhood of* 0*, and one has*

$$\emptyset \neq \partial_L V(0) \subset -\bigcup_{x \in \Sigma} M(x).$$

The function V will be differentiable at 0 if $\partial_L V(0)$ reduces to a singleton. The theorem gives an evident sufficient condition for this: when the problem admits a unique solution x, and x in turn admits a unique Hamiltonian multiplier p. In that case, the theorem yields the following information on the asymptotic effect of small (in L^2) perturbations of the dynamics:

$$V(\alpha) \approx V(0) - \int_0^T p(t) \cdot \alpha(t)\, dt$$

Sensitivity formulas of this general type can be obtained for a variety of parametric perturbations (see [26]). The most classic scheme of all is the one that varies the initial condition, so that the corresponding value function is given by

$$V(\tau, \alpha) := \min\{\ell(x(T)) + \int_\tau^T L(x(t), u(t))\, dt : x(\tau) = \alpha\},$$

the minimum being taken over the trajectories x of the control system on $[\tau, T]$ and corresponding control u. In this case we find:

Theorem 7.4 *The function V is Lipschitz in a neighborhood of $(0, x_0)$, and one has*

$$\emptyset \neq \partial_L V(0, x_0) \subset - \bigcup_{x \in \Sigma} \{(c, p(0)) : p \in M(x)\}.$$

(It is understood above that c refers to the constant value of the Hamiltonian along the given (x, p).) This result not only leads to a sensitivity estimate, but actually subsumes the necessary conditions of Theorem 7.2 above, since (as a simple argument shows) it yields that for any solution x to [OCP], we have $M(x) \neq \emptyset$. Furthermore, it follows from the formula that for any $(\theta, \zeta) \in \partial_L V(0, x_0)$, we have

$$\theta + H(x_0, -\zeta) = 0.$$

Since $(0, x_0)$ can be replaced by any initial condition here, we see that V satisfies a Hamilton-Jacobi equation in the proximal sense. We recognize this as the initial point of the article: we have, quite appropriately, described a closed loop.

References

[1] F. Ancona and A. Bressan. Patchy vector fields and asymptotic stabilization. 2000. Preprint.

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

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

[4] A. Bacciotti and L. Rosier. Lyapunov and Lagrange stability: inverse theorems for discontinuous systems. *Math. Control Signals Systems*, 11:101–128, 1998.

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

[6] G. Barles. Discontinuous solutions of first order Hamilton-Jacobi equations: a guided visit. *Nonlinear Anal.*, 20:1123–1134, 1993.

[7] E. N. Barron and R. Jensen. Optimal control and semicontinuous viscosity solutions. *Proc. Amer.Math. Soc.*, 113:397–402, 1991.

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

[9] J. M. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization*, volume 3 of *CMS Books in Mathematics*. Springer-Verlag, 2000.

[10] P. Cannarsa and H. Frankowska. Some characterizations of optimal trajectories in control theory. *SIAM J. Control Optim.*, 29:1322–1347, 1991.

[11] P. Cannarsa and C. Sinestrari. Convexity properties of the minimum time function. *Calc. Var. Partial Diff. Equations*, 3:273–298, 1995.

[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. Generalized gradients and applications. *Trans. Amer. Math. Soc.*, 205:247–262, 1975.

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

[15] F. H. Clarke. Extremal arcs and extended Hamiltonian systems. *Trans. Amer. Math. Soc.*, 231:349–367, 1977. Russian translation: Differentsialnye Uravnenia 13 (1977) 427-442.

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

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

[18] F. H. Clarke. Hamiltonian analysis of the generalized problem of Bolza. *Trans. Amer. Math. Soc.*, 301:385–400, 1987.

[19] F. H. Clarke. Methods of dynamic and nonsmooth optimization. In *Regional conference Series in Applied Mathematics*, volume 57, Philadelphia, 1989. S.I.A.M.

[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. *I.E.E.E. Trans. Aut. Control*, 42:1394–1407, 1997.

[22] F. H. Clarke, Yu. S. Ledyaev, and R. J. Stern. Complements, approximations, smoothings and invariance properties. *J. Convex Analysis*, 4:189–219, 1997.

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

[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 P. D. Loewen. The value function in optimal control: Sensitivity, controllability and time-optimality. *SIAM J. Control Optim.*, 24:243–263, 1986.

[27] F. H. Clarke, L. Rifford, and R. J. Stern. Feedback in state constrained optimal control. *ESAIM:COCV*, 2001. To appear.

[28] J.-M. Coron, L. Praly, and A. Teel. Feedback stabilization of nonlinear systems: Sufficient conditions and Lyapunov and input-output techniques. In A. Isidori, editor, *Trends in Control*. Springer-Verlag, New York, 1995.

[29] 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.

[30] M. G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton-Jacobi equations. *Trans. Amer. Math. Soc.*, 277:1–42, 1983.

[31] A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides*. Kluwer Academic Publishers, 1988.

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

[33] R. Freeman and P. V. Kokotovic. *Robust Nonlinear Control Design. State-Space and Lyapunov Techniques*. Birkhäuser, 1996.

[34] D. Havelock. A generalization of the Hamilton-Jacobi equation. Master's thesis, University of British Columbia, 1977. (Thesis director: F.H. Clarke).

[35] H. Hermes. Discontinuous vector fields and feedback control. In J.K. Hale and J.P. LaSalle, editors, *Differential Equations and Dynamic Systems*. Academic Press, New York, 1967.

[36] C. M. Kellett and A. R. Teel. Uniform asymptotic controllability to a set implies locally Lipschitz control-Lyapunov function. In *Proc. Conf. Decision and Control (Sydney)*. IEEE Publications, 2000.

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

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

[39] Y. Lin, E. D. Sontag, and Y. Wang. A smooth converse Lyapunov theorem for robust stability. *SIAM J. Control Optim.*, 34:124–160, 1996.

[40] P. D. Loewen. *Optimal Control via Nonsmooth Analysis*, volume 2 of *CRM Proc. & Lecture Notes*. American Mathematical Society, Providence, R.I., 1993.

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

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

[43] L. Rifford. *Problèmes de stabilisation en théorie du contrôle*. PhD thesis, Université Claude Bernard Lyon I, 2000. (Directeur: Francis Clarke).

[44] L. Rifford. Semiconcave control-Lyapunov functions and stabilizing feedbacks. 2000. Preprint.

[45] L. Rifford. Existence of Lipschitz and semiconcave control-Lyapunov functions. *SIAM J. Control Optim*, 39:1043–1064, 2001.

[46] 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.

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

[48] E. D. Sontag. *Mathematical Control Theory*. Texts in Applied Mathematics, vol.6. Springer-Verlag, New York, 1990. (Second Edition, 1998).

[49] 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.

[50] E. D. Sontag and H. J. Sussmann. Nonsmooth control-Lyapunov functions. In *Proc. I.E.E.E. Conf. Decision and Control, New Orleans*, pages 61–81, Boston, 1990. Birkhäuser.

[51] A. I. Subbotin. A generalization of the basic equation of the theory of differential games. *Soviet Math. Dokl.*, 22, 1980. Original article in Russian: Doklady AN SSSR 254 (1980) 293-297.

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

[53] A. R. Teel and L. Praly. A smooth Lyapunov function from a class-$\mathcal{KL}$ estimate involving two positive semidefinite functions. *ESAIM: COCV*, 5:313–367, 2000.

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

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