On Lagrange Manifolds and Viscosity Solutions

Martin V. Day *
Department of Mathematics
Virginia Tech
Blacksburg, VA 24061-0123

January, 1997

Abstract

We consider the use of Lagrange manifolds to construct viscosity solutions of first order Hamiltonian-Jacobi equations. Recent work of several authors is indicated in which the essential underlying structure consists of a Lagrange manifold on which 1) the desired Hamiltonian function vanishes and 2) the canonical 1-form $p \cdot dx$ of classical mechanics has an integral $S(x, p)$. We explore the proposition that a viscosity solution $W(x)$ of the Hamiltonian-Jacobi equation is obtained by minimizing the function S over points in the Lagrange manifold that project to the state x. We prove that the function $W(x)$ produced by this construction is necessarily a viscosity supersolution, and if Lipschitz is also a subsolution. Elementary examples illustrate the construction, including situations in which the subsolution property fails. Connections with Riccati PDEs, L_2-gain in nonlinear systems, small-noise quasipotentials, and simple variational examples are all described.

1 Introduction

Hamilton-Jacobi equations of the form

$$D_t W(t, x) + H(t, x; D_x W(t, x)) = 0 \tag{1}$$

or

$$H(x, DW(x)) = 0 \tag{2}$$

*We are grateful for helpful conversations with C. Byrnes and Bill McEneaney.

arise naturally as descriptions of value functions associated with control or variational problems, or differential games. It is typical to define W as the value function of a control problem and then characterize it as a solution of (1) or (2). Since such functions $W(\cdot)$ are generally nonsmooth, the equations must be understood in a weak sense. The notion of viscosity solutions [11] has been developed for this purpose.

The classical method of characteristics, on the other hand, describes smooth solutions of these equations in terms of a family of trajectories of the Hamiltonian system

$$\dot{x}_t = H_p(x_t, p_t), \quad \dot{p}_t = -H_x(x_t, p_t) \tag{3}$$

associated with H. One is typically led to consider a particular family of these trajectories, making up what is known as a Lagrange manifold $\mathcal{M}$ in (x,p)-space. Recently several authors have exploited this Lagrange manifold structure in variational or control contexts for which viscosity sense solutions are generally called for.

Our purpose in this paper is to consider how (continuous) viscosity solutions result directly from the Lagrange manifold structure, *in the absence of any control or variational interpretation.* Control and/or variational problems certainly motivate most interest in viscosity solutions, and have been the context in which many aspects have been previously studied. It is no surprise that most of our conclusions below are familiar facts in the context of, say, Bolza problems in the calculus of variations. (See [11] for a nice summary of classical results for simple problems, or [7] and references for recent work.) Our purpose here is not to offer new results in that highly developed subject, but to explore the extent to which some of the familiar features of those problems follow from the Lagrange manifold structure alone, apart from any variational interpretation. The main artifact of the control-theoretic motivation will be the assumption that $H(x,p)$ is convex in p. (This excludes problems arising in differential games, however.)

1.1 Terminology and Classical Characteristics

Our analysis will take place in *phase space*, which consists of all $(x,p) \in \mathbb{R}^n \times \mathbb{R}^n$. The x component is called the *state* and p is the *costate* (sometimes *momentum*). We assume throughout that $H(x,p)$ is a C^2 function defined on phase space. In Section 3 we will add the assumption that H is convex in p for each x, but that is not needed initially. A solution (x_t, p_t) of (3) is called a *bicharacteristic.* The state component x_t of a bicharacteristic is called a *characteristic* or *extremal.*

A brief summary of the classical method of characteristics will introduce our point of view. Suppose $V(x)$ is a classical (C^2) solution to (2). Associated with $V(\cdot)$ is the following n-dimensional submanifold of phase space:

$$\mathcal{M} = \{(x,p) : \; p = DV(x)\}.$$

($DV(x)$ denotes the vector of partial derivatives of V at x. For $V(t,x)$ depending on a time variable as well, $D_t V$ denotes $\partial V/\partial t$ while $D_x V$ is the vector of partial derivatives with respect to x_i. H_p and H_x in (3) are the vectors of partial derivatives of H with respect to the x_i and p_i.) Equation (2) says that

$$H(x,p) = 0, \quad \text{all } (x,p) \in \mathcal{M}. \tag{4}$$

In addition, $\mathcal{M}$ has the property that

$$\int p \cdot dx \text{ is independent of path on } \mathcal{M}. \tag{5}$$

Indeed, if (x_t, p_t), $t \in [0,1]$ is a closed (piecewise smooth) path on $\mathcal{M}$, then $p_t = DV(x_t)$ and so $\int_0^1 p_t \cdot dx_t = V(x_1) - V(x_0) = 0$. This independence of path is what it means to say that $\mathcal{M}$ is a Lagrange manifold. (In general a Lagrange manifold need only have this property locally; see [18]. Here, since $\mathcal{M}$ is a graph over state space, the independence of path is global.)

Another important feature of $\mathcal{M}$ is its invariance with respect to the Hamiltonian system (3). To see this consider any $(x_0, p_0) \in \mathcal{M}$ and let x_t be the solution of $\dot{x}_t = H_p(x_t, DV(x_t))$ with the selected initial value x_0. Define $p_t = DV(x_t)$. Then differentiating (2) with respect to x produces an identity (the Riccati PDE below) which implies $\dot{p}_t = -H_x(x_t, p_t)$. This produces the solution of (3) through the prescribed $(x_0, p_0) \in \mathcal{M}$ in a way that makes $(x_t, p_t) \in \mathcal{M}$ manifest.

Conversely, given $\mathcal{M}$ with these properties, a solution of (2) is determined (up to a constant) by

$$V(x_T) - V(x_0) = \int_0^T p_t \cdot dx_t, \tag{6}$$

where (x_t, p_t) is any piecewise smooth curve on $\mathcal{M}$ joining x_0 to x_T (not necessarily a solution of (3)). The classical method of characteristics is essentially to find V by constructing $\mathcal{M}$. If some kind of boundary data is prescribed for (2), that data often determines a subset Σ of $\mathcal{M}$. By then including the solutions of (3) for all $(x_0, p_0) \in \Sigma$ one hopes to obtain a suitable manifold $\mathcal{M}$. The problem is that the resulting $\mathcal{M}$ may fail to be a graph $p = p(x)$ over state space; for a given x there may be more (or less) than one p with $(x,p) \in \mathcal{M}$ so that $V(x)$ is multiple-valued (or undefined). If, however, one can identify a piece $\mathcal{M}_0 \subset \mathcal{M}$ which is a graph over some domain $\Omega_0 \subseteq \mathbb{R}^n$, then at least the construction will produce a solution of (2) for $x \in \Omega_0$.

1.2 Lagrange Manifold Properties

The properties (4) and (5) of $\mathcal{M}$ cited above are not limited to the "ideal" situation in which $\mathcal{M}$ is a graph over state space. Rather they are very natural properties to expect of $\mathcal{M}$ in general, because they are associated with fundamental invariance properties of Hamiltonian systems. To be concrete, suppose that $\mathcal{M}$ is a smooth manifold, made up of a family of bicharacteristics (3), and that there is an open subset $\mathcal{M}_0 \subset \mathcal{M}$ in which (4) and (5) hold and through which every bicharacteristic on $\mathcal{M}$ passes. (This is common in the applications to be cited shortly, in which $\mathcal{M}$ is often taken to be the unstable manifold of a critical point of (3).) An elementary invariance property of (3) is that $H(x_t, p_t)$ is constant along every bicharacteristic. It follows that (4) extends from $\mathcal{M}_0$ to all of $\mathcal{M}$. A more profound invariance property is that (3) preserves the differential 2-form

$$d(p \cdot dx) = dp \wedge dx = \sum_{i=1}^{n} dp_i \wedge dx_i.$$

(See [1].) This means that if Γ_0 is a two-dimensional surface in phase space, bounded by a simple closed curve γ_0, and we let (3) transport Γ_0 through t time units to obtain a new two-dimensional surface Γ_t, bounded by the simple closed curve γ_t, then

$$\int_{\Gamma_0} dp \wedge dx = \int_{\Gamma_t} dp \wedge dx.$$

Stokes' formula says that for each t.

$$\int_{\Gamma_t} dp \wedge dx = \int_{\gamma_t} p \cdot dx.$$

Thus if the line integral $\int_{\gamma_t} p \cdot dx$ vanishes for $t = 0$ then it does for all t. Consider a given $(x, p) \in \mathcal{M}$. There exists a bicharacteristic with $(x_0, p_0) \in \mathcal{M}_0$ and $(x_t, p_t) = (x, p)$ for some t. (3) maps $\mathcal{M}_0$ to a neighborhood $\mathcal{M}_t \subseteq \mathcal{M}$ of (x, p). If γ_t is any closed path in $\mathcal{M}_t$ then $\int_{\gamma_t} p \cdot dx = \int_{\gamma_0} p \cdot dx = 0$, since γ_0 is in $\mathcal{M}_0$. The point is that the (local) independence of path of $\int p \cdot dx$ is inherited by $\mathcal{M}$ from that property in the initial section $\mathcal{M}_0$. Notice from Stokes' formula that the local path independence on $\mathcal{M}$ is equivalent to the property that $dp \wedge dx$ vanishes on $T\mathcal{M}$. This, and that $\mathcal{M}$ has dimension n, is the usual definition of a Lagrange manifold. See [1] and [18]. In general a Lagrange manifold only has the independence of path property in sufficiently small neighborhoods, not globally. (See Example 2.1.1 of the next section.) For our purposes it is essential that this independence of path in fact be global.

Suppose then that we do indeed have a manifold $\mathcal{M}$ satisfying (4) and (5), but not necessarily the graph of a function $p = p(x)$ over state space.

The global independence of path means that there is a smooth function $S(x,p)$ defined (up to a constant) on $\mathcal{M}$ by

$$S(x_T, p_T) = S(x_0, p_0) + \int p_t \cdot dx_t, \tag{7}$$

where the line integral is over any piecewise smooth curve $(x_t, p_t),\ t \in [0,T]$ on $\mathcal{M}$. This is abbreviated as

$$dS = p \cdot dx \quad \text{on } \mathcal{M}.$$

We propose the following simple formula as a natural candidate for a viscosity solution of (2):

$$\begin{aligned} W(x) &= \inf\{S(x,p) :\ p \text{ such that } (x,p) \in \mathcal{M}\} \\ &= \inf_{p:(x,p)\in\mathcal{M}} S(x,\cdot),\ \text{ as it is denoted below.} \end{aligned} \tag{8}$$

It is the veracity of this proposition that we explore in this paper. We emphasize that this point of view makes $\mathcal{M}$ the fundamental object underlying the proposed viscosity solution. We will assume we have $\mathcal{M}$ in hand (satisfying the technical hypotheses outlined in the next section).

We now describe briefly some previous studies that have exploited this Lagrange manifold point of view.

1.3 Riccati Equations

The recent paper of C. Byrnes [5] provides a nice discussion of the Lagrange manifold structure of the family of extremals determined by the maximum principle in basic control problems. The focus is on the Riccati PDE, viewed as a nonlinear generalization of the matrix equations of linear systems theory. The Riccati PDE turns out to be another expression of the invariance of $\mathcal{M}$ with respect to the Hamiltonian flow, but one which is valid only where $\mathcal{M}$ is a graph over state space. In our context, if $\mathcal{M}$ is given by $p = p(x)$, the Riccati PDE would be

$$H_x(x, p(x)) + \frac{\partial p(x)}{\partial x} H_p(x, p(x)) = 0,$$

which is the same as $\dot{p} = -H_x$ in the bicharacteristic equations. We note that the Lagrange property of $\mathcal{M}$ is equivalent to the symmetry of $\frac{\partial p(x)}{\partial x}$, and differentiating $H(x, p(x)) = 0$ with respect to x implies $H_x + (\frac{\partial p}{\partial x})^T H_p$. Thus the Riccati equation is a consequence of (4) and (5) above. In general, when $\mathcal{M}$ is not a graph, [5] calls $\mathcal{M}$ a "weak solution" of this equation. Generalizations of the Riccati equation are possible in such circumstances, by using some of the p_j instead of x_j as independent variables. See the discussion at the end of Section 2 on invariance with respect to the Hamiltonian flow.

One of the primary contributions of the Lagrange manifold point of view has been "geometric" existence proofs, based on taking $\mathcal{M}$ to be a stable (or unstable) invariant manifold associated with the Hamiltonian system (6), as mentioned briefly above. The use of invariant manifold theory to provide existence results, in particular the construction of solutions by taking $\mathcal{M}$ to be the stable/unstable manifold of the appropriate Hamiltonian system (in time independent settings), is described and traced to work of Brunovsky [4] and Lukes [17] in the 1960's. Numerous other references in the control theory literature are cited. Byrnes also notes that Burgers' equation is a particular instance of the Riccati PDEs he considers. This connection with scalar conservation laws goes back many years, as we will comment in Section 4.

1.4 L_2-gain of Input-Output Systems

Nonlinear systems theory offers a particular context in which the effort to construct viscosity solutions from Lagrange manifolds is natural in light of recent work. Consider a control system

$$\dot{x}_t = f(x_t) + g(x_t)u_t.$$

The control $u_t \in \mathbb{R}^m$ (locally L_2) is viewed as an input and a function $y_t = h(x_t)$ is considered as the output. Assume $f(\cdot)$, $g(\cdot)$ and $h(\cdot)$ are C^2 with appropriate dimensions, and that $x = 0$ is an equilibrium of the uncontrolled system: $f(0) = 0$, $h(0) = 0$. The goal is to establish a bound γ on the L_2-gain of the map $u_\cdot \mapsto y_\cdot$. A. J. van der Schaft [25] has formulated a version of this problem which is equivalent to the existence of a nonnegative function $V(x)$ satisfying $V(0) = 0$ and obeying the Dissipation Inequality:

$$V(x_T) - V(x_0) \leq \frac{1}{2}\int_0^T \gamma^2|u_t|^2 - |h(x_t)|^2\,dt, \tag{9}$$

for all controlled trajectories x_t. The appropriate Hamiltonian (see Section 5) is

$$H(x,p) = \frac{1}{2}\gamma^{-2}p\cdot g(x)g(x)^T p + p\cdot f(x) + \frac{1}{2}|h(x)|^2.$$

If V is smooth, (9) is equivalent to the differential inequality

$$H(x, DV(x)) \leq 0. \tag{10}$$

Van der Schaft shows how appropriate controllability and observability assumptions imply that the stable manifold of (6) for the equilibrium $(0,0)$

is the graph of a smooth function over state space in a neighborhood of $(0,0)$, providing (locally) the existence of a classical solution of (11) below. This provides a local verification of L_2-gain $\leq \gamma$, holding for paths limited to the domain Ω_0 in which the solution V so constructed is smooth.

The limitation to smooth functions V is too stringent, however. M. James [15] has shown that for a lower semicontinuous $V(\cdot)$, (9) is equivalent to the viscosity sense inequality

$$H(x, D^{-}V(x)) \leq 0.$$

Here $D^{-}V(x)$ is the usual set of subdifferentials of V at x. (See Fleming and Soner [11] for this and other background on viscosity notion solutions.) We use the notation "$H(x, D^{-}V(x)) \leq 0$" above to mean $H(x,p) \leq 0$ for all $p \in D^{-}V(x)$. In the usual terminology, this is equivalent to saying that V is a viscosity supersolution to the equation

$$-H(x, DV(x)) = 0. \tag{11}$$

Moreover results of Ball and Helton [2] and Soravia [23] imply that the minimal nonnegative function satisfying (9) (the available storage function – see Section 5) is in fact a viscosity solution of (11). The point is that (11) should be considered in the viscosity sense. Even though van der Schaft only considered smooth solutions, (8) provides a natural extension of his construction which offers the prospect of being a viscosity solution.

We note that (11) is not of the form (2) that we are considering here, because $-H(x,\cdot)$ is concave rather than convex. We will explain in Section 5 how our recipe (8) translates in this case to

$$V(x) = \sup_{p:(x,p)\in\mathcal{M}} S(x,\cdot)$$

as a construction of solutions to (11). Section 5 below is devoted to this particular application.

1.5 Quasipotentials in Small Noise Asymptotics

The Lagrange manifold structure has also been exploited in the study of the quasipotential functions which arise in the study of small Brownian perturbations of a dynamical system [13]. The equations there can be formulated as special cases of those for the the L_2-gain problem above. In fact take $g(\cdot) \equiv I$, $h(\cdot) \equiv 0$ and $\gamma = 1$. The control system is then

$$\dot{x}_t = f(x_t) + u_t, \tag{12}$$

and the Hamiltonian becomes

$$H(x,p) = \frac{1}{2}|p|^2 + p \cdot f(x).$$

Assume that 0 is an exponentially stable critical point for the uncontrolled system $\dot{x}_t = f(x_t)$. The original Wentzel-Freidlin quasipotential W [13] is defined by

$$W(x_T) = \inf_{x_0=0} \int_0^T \frac{1}{2}|u_t|^2 \, dt. \tag{13}$$

The infimum is over all $T > 0$ and controlled paths $x_t : [0, T] \to \mathbb{R}^d$ which join $x_0 = 0$ to the specified x_T and have $u_t \in L^2$. (In the language of dissipative systems, W so constructed is generally called the *required supply* function. In contrast to the available storage function, the required supply typically provides the maximal solution of the Dissipation Inequality (9).) Day and Darden [10] showed (with some growth assumptions on $f(x)$) that W is given by our recipe (8) using the unstable manifold $\mathcal{M}$ for the equilibrium at $(0, 0)$ of the Hamiltonian system. The terminology of Lagrange manifolds was not used in [10], but is none the less the structure that was exploited. The smoothness of W in a neighborhood of 0 was obtained as a consequence. A few years later Perthame [21] showed that W is a viscosity solution of (2). Thus for the Wentzel-Freidlin quasipotential specifically, we know that (8) does indeed produce a viscosity solution, by virtue of the combined results of [10] and [21]. More recently Day [9] reworked the same kind of analysis, but taking $\mathcal{M}$ to be the stable manifold associated with a periodic orbit of $\dot{x}_t = b(x_t)$, to produce a different quasipotential function (solving $H(x, -DW(x)) = 0$) and establishing smoothness in a neighborhood of the orbit.

Maier and Stein [19] also consider the same quasipotential function (13) in their studies of small noise phenomena. They too observe its relation to the unstable manifold and recognize that the possibility of a given x having multiple p with $(x, p) \in \mathcal{M}$ posed a problem for classical solutions. However the consideration of viscosity solutions to the Hamilton-Jacobi equation was not part of their discussion.

1.6 Overview

General hypotheses for our treatment and their implications are presented in Section 2. Some elementary examples are offered to clarify (8) above and certain technical issues. In Section 3 we show that (8) always produces a lower-semicontinuous supersolution of (2), under the hypotheses of Section 2. (8) can fail to produce a subsolution of (2), as the examples show. We establish some simple results on the regularity of W and prove that it is a subsolution under an additional Lipschitz continuity assumption. Section 4 presents two examples in which $W(\cdot)$ is discontinuous, clarifying the discrepancy between the hypotheses of Section 2 and features of a typical variational problem. Finally, in Section 5 we look more closely at the application to L_2-gain estimation.

2 Fundamentals

The purpose of this section is to define concisely the basic objects of our consideration, establish some of their fundamental properties, and identify those additional hypotheses under which we will consider

$$W(x) = \inf_{p:(x,p)\in\mathcal{M}} S(x,\cdot) \tag{14}$$

as a possible viscosity solution. We also present some elementary examples to illustrate (14) and some aspects of the hypotheses.

2.1 Basic Hypotheses

We are assuming that the Hamiltonian $H(x,p)$ is a smooth (C^2) real-valued function defined on phase space. In the sections to follow we will assume that H is convex in p for each x. This is natural in many situations, but unnatural in others. In particular nonconvex H are important in nonlinear H_∞ control. So at least in the present section we allow the possibility of a nonconvex H. However our discussions of the viscosity sub/supersolution properties below will depend on a convexity hypothesis.

We assume $\mathcal{M}$ is a smooth submanifold of phase space (without boundary) satisfying hypotheses (A1) — (A6) below. We will elaborate on these hypotheses in the paragraphs which follow.

(A1) $H = 0$ at all points of $\mathcal{M}$.

(A2) $\mathcal{M}$ is Lagrangian.

(A3) $p \cdot dx$ is globally independent of path on $\mathcal{M}$.

(A4) $\mathcal{M}$ is embedded in phase space.

(A5) $\mathcal{M}$ is locally bounded.

(A6) $\mathcal{M}$ covers an open region Ω of state space and has no boundary points over Ω

The usual definition of a Lagrangian manifold $\mathcal{M}$ is that it be a submanifold of phase space, of dimension n, such that the differential 2-form $dp \wedge dx$ vanishes on its tangent space $T\mathcal{M}$. See [18] or [1] for instance. As pointed out in the introduction, this is equivalent to local independence of path of $p \cdot dx$ on $\mathcal{M}$. An alternate characterization of the Lagrangian property can be given which will be useful in our discussion. A general feature of a Lagrange manifold is that, at any point, some complimentary selection of state and costate variables will provide a coordinate chart. (See

[18] Proposition 4.6.) That is, we can find $I \subseteq \{1, \ldots, n\}$ so that the x_i, $i \in I$ and p_j, $j \in I^c$ provide coordinates for $\mathcal{M}$ in some neighborhood of the prescribed point. Moreover ([18] Proposition 4.21) in such a neighborhood the Lagrange property is equivalent to the existence of a (smooth) generating function $G(x_i, p_j)$, $i \in I$, $j \in I^c$ in terms of which $\mathcal{M}$ is described by

$$x_j = -\frac{\partial G}{\partial p_j},\ j \in I^c \qquad p_i = \frac{\partial G}{\partial x_i},\ i \in I. \tag{15}$$

(Some authors, such as Arnold [1], call $F = -G$ the generating function, and the above relations are negated.) Of particular significance for us is the function defined in this neighborhood of $\mathcal{M}$ by

$$S(x,p) = G + \sum_{j \in I^c} x_j p_j.$$

On $\mathcal{M}$ we have, using (15),

$$\begin{aligned} dS &= \sum_{i \in I} \frac{\partial G}{\partial x_i} dx_i + \sum_{j \in I^c} \frac{\partial G}{\partial p_j} dp_j + \sum_{j \in I^c} (x_j dp_j + p_j dx_j) \\ &= \sum_{j=1}^{d} p_j dx_j = p \cdot dx. \end{aligned}$$

In particular $\int p \cdot dx$ is independent of paths remaining inside the region where (15) holds. The description of $\mathcal{M}$ in terms of generating functions will be a convenient way to describe simple examples. Note that for $n = 1$ any smooth curve in phase space ($\mathbb{R}^2$) is a Lagrange manifold $\mathcal{M}$. Where $\mathcal{M}$ is of the form $p = f(x)$ a generating function is given by $G(x) = \int f\, dx$; where $\mathcal{M}$ is the graph of $x = \phi(p)$, $G(p) = -\int \phi\, dp$ is a generating function.

2.1.1 Example

Consider $H(x,p) = x^2 + p^2 - 1$. The unit circle $\mathcal{M} = \{(x,p) :\ x^2 + p^2 = 1\}$ is a Lagrange manifold (since $n = 1$) on which H vanishes. Thus (A1) and (A2) are satisfied.

This example fails to satisfy (A3), however. That $p \cdot dx$ be globally independent of path on $\mathcal{M}$ is equivalent to the existence of a single function S defined on all of $\mathcal{M}$ with $dS = p \cdot dx$. This is not possible in Example 2.1.1, since integrating $p \cdot dx$ once around the circumference does not vanish. As indicated in the introduction, the desired $\mathcal{M}$ is frequently obtained as the stable (or unstable) invariant manifold of a critical point of the Hamiltonian system. Several authors, [5], [9] and [24] for instance, have noted that the stable (or unstable) invariant manifold of a hyperbolic critical point of a Hamiltonian system always has the Lagrangian property. A sufficient condition for (A3) is that $\mathcal{M}$ be simply connected. (See [18], pg.9.)

The term "embedded" in (A4) means that the intrinsic topology of $\mathcal{M}$, i.e. that induced by its coordinate charts, is the same as the relative topology it inherits as a subset of phase space. There is some genuine content to this assumption, as the next example illustrates.

2.1.2 Example

$H(x,p) = x^3 - x^2 + \frac{1}{2}p^2$, and take $\mathcal{M}$ to be the unstable manifold of the critical point at the origin for the associated Hamiltonian system: $p = x\sqrt{2(1-x)}$ for $x \leq 1$ and $p = \pm x\sqrt{2(1-x)}$ for $0 < x \leq 1$. $\mathcal{M}$ is Lagrangian because it is one-dimensional, and as the unstable invariant manifold. However in its intrinsic topology the points $p = -x\sqrt{2(1-x)}$ as $x \downarrow 0$ do not converge to the origin, while they do in the subspace topology from $\mathbb{R}^2$. The embedded hypothesis is thus violated. In general if the level set $H^{-1}\{0\}$ contains a homoclinic loop of the associated Hamiltonian system, then the full stable (unstable) manifold will not satisfy the embedded assumption. However a reduced version of $\mathcal{M}$, obtained say by removing a small piece of $p = -x\sqrt{2(1-x)}$, $x > 0$ near the origin, will satisfy (A4).

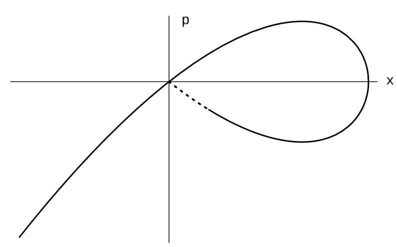

Figure 1: Example 2.1.2

It makes sense to consider (14) only at states x which are "covered" by $\mathcal{M}$, i.e. for which there exists some p with $(x,p) \in \mathcal{M}$. In (A6) we limit our consideration of (14) to an open region Ω of state space all of whose points are covered by $\mathcal{M}$ in this sense. By "boundary points" of $\mathcal{M}$ in (A6) we mean points (x,p) in phase space which occur as limits $(x,p) = \lim(x_n, p_n)$ with $(x_n, p_n) \in \mathcal{M}$ but for which $(x,p) \notin \mathcal{M}$. (In this sense $\partial\mathcal{M} = \bar{\mathcal{M}} \setminus \mathcal{M}$, where the closure $\bar{\mathcal{M}}$ is with respect to the phase space topology.) In (A6) we assume that no such $(x,p) \in \partial\mathcal{M}$ has $x \in \Omega$. The significance of (A4) and (A6) together is this: if $x_n \to x \in \Omega$, $(x_n, p_n) \in \mathcal{M}$ for each n and $p_n \to p$, then $(x,p) \in \mathcal{M}$ and $(x_n, p_n) \to (x,p)$ in the topology of $\mathcal{M}$. I.e. in the portion of $\mathcal{M}$ over Ω convergence of (x_n, p_n) in phase space is equivalent to convergence in $\mathcal{M}$.

Finally we come to the assumption (A5) that $\mathcal{M}$ is "locally bounded". By this we mean that for each $x_0 \in \Omega$ there exists a $\delta > 0$ and $K < \infty$ so that $|p| \leq K$ for all $(x,p) \in \mathcal{M}$ with $x \in B_\delta(x_0)$. Under (A1) a sufficient condition would be that $H(x,p) \to +\infty$ as $|p| \to \infty$ uniformly for x in compacts. This is satisfied in the two preceding examples for instance. Example 4.1.2 illustrates some of the pathological features possible when local boundedness fails. An important implication of local boundedness is that the infimum in (14) is achieved for every $x \in \Omega$: S is continuous on $\mathcal{M} \cap (\{x\} \times \mathbb{R}^d)$, which is compact by (A5) and the agreement of phase space and $\mathcal{M}$ topologies on $\mathcal{M}$. The following lemma collects several consequences of (A1) — (A6) for future use.

Lemma 1 *For every $x \in \Omega$ there exists $(x,p_*) \in \mathcal{M}$ such that*

$$S(x,p_*) \leq S(x,p) \quad \textit{all other } (x,p) \in \mathcal{M}.$$

If $x_n \to x \in \Omega$ and $(x_n,p_n) \in \mathcal{M}$, then the p_n form a bounded sequence and $(x,p) \in \mathcal{M}$ for all limit points p of p_n.

2.1.3 Example

The construction (14) is easily illustrated by examples with $n = 1$ and $\mathcal{M}$ described by

$$x = \phi(p).$$

If ϕ is not monotonic the projection $(x,p) \in \mathcal{M} \mapsto x$ is many-to-one so that the $\inf_p$ comes into play. Let $\Phi(p) = \int \phi(p)\,dp$. S is given on $\mathcal{M}$ by

$$S(x,p) = xp - \Phi(p), \quad \text{for } (x,p) \in \mathcal{M}.$$

To be specific, consider

$$\Phi(p) = \frac{1}{4}(2p^2 - p^4), \quad \phi(p) = p - p^3.$$

Figure 2 shows, first $\mathcal{M} = \{x = \phi(p)\}$ in the x,p plane, then S plotted over the points of $\mathcal{M}$, and then S plotted as a multiple-valued function of x (i.e. p suppressed). The graph of the resulting $W(x)$ is easy to pick out as the curve with a "peak" on the vertical axis.

2.1.4 Example

Contrast Example 2.1.3 with what happens if Φ is negated:

$$\Phi(p) = -\frac{1}{4}(2p^2 - p^4), \quad \phi(p) = -p + p^3.$$

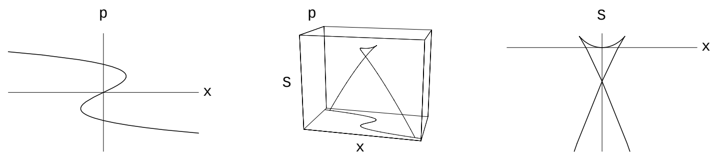

Figure 2: Example 2.1.3

Views of this example are shown in Figure 3. Here the resulting $W(x)$ is clearly discontinuous. This is typical of cases in which the graph of $x = \phi(p)$ loops back above itself as we move from left to right. We will refer to this general configuration an an "overloop", and that of Example 2.1.3 as an "underloop".

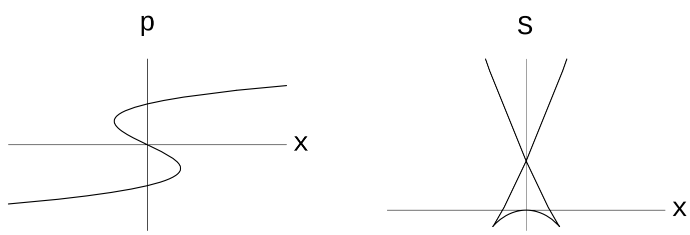

Figure 3: Example 2.1.4

We note that the Hamiltonian H plays no role in the construction (14). Rather, H is involved in the identification of the appropriate $\mathcal{M}$ for a particular equation (2) and boundary conditions. In Examples 2.1.3 and 2.1.4 we could simply take

$$H(x, p) = \phi(p) - x,$$

which obviously vanishes on $\mathcal{M}$. This H is not convex in p, however. In one (space) dimension clearly no Hamiltonian which is strictly convex in p can vanish on a curve $\mathcal{M}$ having these over/underloop configurations since the convex function $p \mapsto H(x, p)$ would vanish at 3 distinct p, for x under the loop. However we can easily embed these examples in a higher dimensional context (with convex Hamiltonian) so that the figures above occur as one dimensional cross-sections. Indeed consider $\mathcal{M}$ determined by the generating function

$$\begin{aligned} G(x_2, p_1) &= -(\frac{1}{2}x_2 p_1^2 + \Phi(p_1)) : \\ x_1 &= -\frac{\partial G}{\partial p_1} = x_2 p_1 + \phi(p_1), \end{aligned}$$

$$\begin{aligned} p_2 &= \frac{\partial G}{\partial x_2} = -\frac{1}{2}p_1^2 \\ S &= x_1 p_1 - \frac{1}{2}x_2 p_1^2 - \Phi(p_1). \end{aligned}$$

The Hamiltonian $H(x,p) = \frac{1}{2}p_1^2 + p_2$ is convex and clearly vanishes on $\mathcal{M}$. We note that this Hamiltonian is the same as that associated with the examples of Section 4 below.

2.2 Invariance with Respect to the Hamiltonian Flow

The property that $\mathcal{M}$ is invariant with respect to the Hamiltonian system

$$\dot{x} = H_p(x,p), \quad \dot{p} = -H_x(x,p) \tag{16}$$

was demonstrated in the introduction, assuming $\mathcal{M}$ is described by $p = DV(x)$ where $H(x, DV(x)) = 0$. Using generating functions (15) the argument we indicated earlier extends to show the invariance is a general consequence of (A1) and (A2). Consider a neighborhood in $\mathcal{M}$ in which $x_i, i \in I$ and $p_j, j \in I^c$ provide coordinates, some $I \subseteq \{1, 2, \ldots, n\}$, and let $G(x_i, p_j), i \in I, j \in I^c$ be a generating function, so that in this neighborhood $\mathcal{M}$ is described by (15). In particular on $\mathcal{M}$ we can view H, H_x, H_p as functions of the coordinates x_i, p_j alone ($i \in I, j \in I^c$). Through an initial point on $\mathcal{M}$ we can compute a trajectory by solving the system of n equations for the coordinate variables,

$$\dot{x}_i = H_{p_i}, \quad \dot{p}_j = -H_{x_j}, \tag{17}$$

and then use (15) to determine the values of the remaining variables x_j, p_i of the trajectory on $\mathcal{M}$. We need to check that the resulting trajectory on $\mathcal{M}$ is a bicharacteristic. Equations (17) are half of the bicharacteristic equations (16); we need to check the equations for the dependent variables x_j and p_i. Writing out $H = 0$ using (15) we have

$$0 = H(x_i, -\frac{\partial G}{\partial p_j}, \frac{\partial G}{\partial x_i}, p_j), \tag{18}$$

holding identically as a function of the coordinate variables x_i, p_j. Differentiating (18) with respect to an $x_i, i \in I$ yields

$$0 = H_{x_i} + \sum_{j \in I^c} -H_{x_j} \frac{\partial^2 G}{\partial p_j \partial x_i} + \sum_{i' \in I} H_{p_{i'}} \frac{\partial^2 G}{\partial x_{i'} \partial x_i},$$

which in light of (15) and (17) is equivalent to

$$\dot{p}_i = -H_{x_i}. \tag{19}$$

Similarly, differentiating (18) with respect to a $p_j, j \in I^c$ produces

$$0 = \sum_{j' \in I^c} -H_{p_{j'}} \frac{\partial^2 G}{\partial p_{j'} \partial p_j} + \sum_{i \in I} H_{p_i} \frac{\partial^2 G}{\partial x_i \partial p_j} + H_{p_j},$$

which is equivalent to

$$\dot{x}_j = H_{p_j}. \tag{20}$$

(17), (19), and (20) show that the trajectory we have constructed on $\mathcal{M}$ through a prescribed initial point is in fact the bicharacteristic through the initial point. This establishes the invariance of $\mathcal{M}$ with respect to (16), as claimed.

3 Continuity and Subsolution Properties

We now assume that $H(x, \cdot)$ is convex and begin to consider

$$W(x) = \inf_{p:(x,p) \in \mathcal{M}} S(x, \cdot), \quad x \in \Omega \tag{21}$$

as a possible viscosity solution of $H(x, DW(x)) = 0$. First we show, under the general hypotheses presented in Section 2, that W is lower semi-continuous and is a viscosity supersolution, namely that

$$H(x, p) \geq 0$$

for all p in the set of subdifferentials $D^-W(x)$. The proof of the supersolution property below was communicated by Bill McEneaney, although in a slightly different form. A very similar argument was given for Theorem 3.1 of [15]. Notice that the convexity of $H(x, \cdot)$ is used significantly.

Theorem 1 *$W(\cdot)$ is a lower semi-continuous viscosity supersolution of*

$$H(x, DW(x)) = 0 \quad \textit{in } \Omega.$$

Proof. The lower semi-continuity follows easily by considering sequences $x_n \to x$ and p_n with $(x_n, p_n) \in \mathcal{M}$ and $W(x_n) = S(x_n, p_n)$. By passing to a subsequence Lemma 1 implies $\liminf W(x_n) \geq \inf_{p:(x,p)\in\mathcal{M}} S(x, p) = W(x)$.

Consider $x_0 \in \Omega$ and suppose $\phi(\cdot)$ is C^1 such that $W \geq \phi$ with equality at x_0. Since $D^-W(x_0)$ consists of the set of $D\phi(x_0)$ for all such $\phi(\cdot)$, we need to show $H(x_0, D\phi(x_0)) \geq 0$. There exists p_0 with $(x_0, p_0) \in \mathcal{M}$ and $W(x_0) = S(x_0, p_0)$. For an arbitrary $(x, p) \in \mathcal{M}$ we have

$$S(x, p) \geq W(x) \geq \phi(x), \quad \text{with equality at } (x_0, p_0).$$

Consider the bicharacteristic (x_t, p_t) through (x_0, p_0). For $t < 0$

$$\int_t^0 D\phi(x_s) \cdot \dot{x}_s \, ds = \phi(x_0) - \phi(x_t) \geq S(x_0, p_0) - S(x_t, p_t) = \int_t^0 p_s \cdot \dot{x}_s \, ds \tag{22}$$

Since $H(x, p)$ is convex in p, $\dot{x}_s = H_p(x_s, p_s)$ and $H(x_s, p_s) = 0$ we know that

$$H(x_s, q) \geq (q - p_s) \cdot \dot{x}_s + H(x_s, p_s) = (q - p_s) \cdot \dot{x}_s \quad \text{for all } q.$$

Applying this with $q = D\phi(x_s)$ and using (22) yields, for all $t < 0$,

$$\int_t^0 H(x_s, D\phi(x_s)) \, ds \geq \int_t^0 (D\phi(x_s) - p_s) \cdot \dot{x}_s \, ds \geq 0.$$

The continuity of $H(x_s, D\phi(x_s))$ now implies that

$$H(x_0, D\phi(x_0)) \geq 0,$$

completing the proof.

We observe that the above proof only used (local) backwards invariance of Ω with respect to the bicharacteristics of $\mathcal{M}$: $(x_t, p_t) \in \mathcal{M}$ and $x_0 \in \Omega$ implies $x_t \in \Omega$ for all $t \in (-\epsilon, 0)$, some $\epsilon > 0$. Suppose the closure $\bar{\Omega}$ is likewise (locally) backward invariant and that we can extend the definition (21) to $x \in \bar{\Omega}$. Then the argument in the proof applies for $x_0 \in \partial\Omega$ with any C^1 function ϕ such that $W(x) \geq \phi(x)$ for all $x \in \bar{\Omega}$ with equality at x_0. We would conclude that W satisfies Soner's boundary condition on $\partial\Omega$ for problems with state constrained to $\bar{\Omega}$. (See [11] Section II.12.)

The subsolution property of (21) is more involved. Indeed it is false in general under only the hypotheses of Section 2. The shortcoming is that $W(x)$ so defined can easily be discontinuous, as we have seen in Example 2.1.4. The notion of viscosity solution does extend to discontinuous functions. (See [11] and references.) The subsolution property for discontinuous W requires that $H(x, D^+W^*(x)) \leq 0$ where W^* is the upper semicontinuous envelope of W. At a discontinuity $D^+W^*(x)$ is typically unbounded, as our examples illustrate. This makes $H(x, D^+W^*(x)) \leq 0$ impossible for Hamiltonians with $H(x, p) \to +\infty$ as $|p| \to \infty$, such as nondegenerate quadratics.

Our proof of the subsolution property requires the additional assumption that W is (locally) Lipschitz continuous in Ω. Before coming to the proof itself we will explore some continuity properties of W which do follow from our assumptions in Section 2. To this end we distinguish several types of points in Ω. First are the *regular points*, $x \in \Omega$ for which there is a unique $(x, p) \in \mathcal{M}$ giving the minimum of S:

$$W(x) = S(x, p) < S(x, p') \quad \text{for all } (x, p') \in \mathcal{M}, \ p \neq p'.$$

(This terminology is standard; see [11] pg. 48.) The set of all regular points will be denoted U. The points in $\Omega \setminus U$ might reasonably be called *multiple points*, since there are multiple $(x, p) \in \mathcal{M}$ achieving the value $W(x)$.

Let $\pi : \mathcal{M} \to \mathbb{R}^n$ be the projection to state space: $\pi(x, p) = x$. A point $x \in \Omega$ is called a *caustic point* if, there is some $(x, p) \in \mathcal{M}$ at which the state projection π is singular (i.e. has vanishing Jacobian with respect to any set of coordinates for a neighborhood of (x, p) in $\mathcal{M}$). C will denote the set of caustic points. We define the set of *essential caustics* C_* to consist of those x such that π is singular at every $(x, p) \in \mathcal{M}$ which is minimizing, $W(x) = S(x, p)$. Of course $C_* \subseteq C$.

We note that if the projection π is nonsingular at $(x_0, p_0) \in \mathcal{M}$, then the state variables, $x_i, i = 1, \ldots, n$ provide a coordinate chart in some neighborhood of (x_0, p_0). Following our discussion in Section 2, there exists a generating function $G(x)$ defined in some ball $B_\delta(x_0)$ about x_0 so that so that the set of (x, p) with $x \in B_\delta(x_0)$ with $p = DG(x)$ is a neighborhood of (x_0, p_0) in $\mathcal{M}$, and $S(x, p) = G(x)$ in this neighborhood. This fact will be used several times below.

The results in the following theorem and corollary are familiar in the context of variational and control problems; see [11] (Theorem I.10.4 in particular) and [7] for instance. Our definition of caustic point is related to the notion of conjugate point. However the actual definition of conjugate point makes reference to the subset of state space on which initial or boundary data is prescribed. Our hypotheses do not identify any such distinguished subset of state space or of $\mathcal{M}$, so we cannot quite define conjugate points in the general context of Section 2. Thus the following result is slightly different than the usual results on the region of strong regularity. Our main point however is not that these features are new, but that they follow from the general hypotheses in section 2 apart from any variational interpretation.

Theorem 2

a) $U \setminus C_*$ is open, and $W(\cdot)$ is smooth in $U \setminus C_*$.

b) $W(\cdot)$ is continuous at every point of $\Omega \setminus C_*$.

c) $W(\cdot)$ is locally Lipschitz in the interior of $\Omega \setminus C_*$.

Corollary 1 *All discontinuities of W occur at essential caustics.*

Proof. Consider $x_0 \in U \setminus C_*$. There exists a (unique) $(x_0, p_0) \in \mathcal{M}$ which minimizes $S(x_0, \cdot)$ over $(x_0, p) \in \mathcal{M}$ and at which π is nonsingular. By the observation just above there exists a smooth $G(\cdot)$ defined in a neighborhood of x_0 so that $p = DG(x)$ and $S(x, p) = G(x)$ in a neighborhood of (x_0, p_0) in $\mathcal{M}$. Since (x_0, p_0) is minimizing, $W(x_0) = G(x_0)$. We claim there exists $\delta > 0$ so that $p = DG(x)$ is the unique minimizer of $S(x, \cdot)$ in $\mathcal{M}$ for each $x \in B_\delta(x_0)$. If not, there would exist a sequence of $(x_n, p_n) \in \mathcal{M}$ with $x_n \to x_0$ such that p_n minimizes $S(x_n, \cdot)$ in $\mathcal{M}$ but $p_n \neq DG(x_n)$. By Lemma 1 the p_n are bounded, and any limit point $p_* = \lim p_{n'}$ has $(x_0, p_*) \in \mathcal{M}$. Since $W(x_n) = S(x_n, p_n) \leq G(x_n)$, we can use the lower semi-continuity of $W(\cdot)$ to deduce that

$$G(x_0) = W(x_0) \leq \liminf W(x_{n'}) = \lim S(x_{n'}, p_{n'}) \leq \lim G(x_{n'}) = G(x_0).$$

Therefore,

$$W(x_0) = G(x_0) = \lim S(x_{n'}, p_{n'}) = S(x_0, p_*).$$

In other words (x_0, p_*) is minimizing at x_0, which means $p_* = p_0$, since by hypothesis the minimizer at x_0 is unique. It follows then that $p_n \to p_0$. Therefore $(x_n, p_n) \to (x_0, p_0)$ in $\mathcal{M}$ and so, for all sufficiently large n, (x_n, p_n) is in the neighborhood of (x_0, p_0) in which $p = DG(x)$. Hence $p_n = DG(x_n)$ for all sufficiently large n, contrary to our construction. This proves the existence of $\delta > 0$ so that $p = DG(x)$ is the unique minimizer of $S(x, \cdot)$ in $\mathcal{M}$ for each $x \in B_\delta(x_0)$, showing both that x_0 has a neighborhood $B_\delta(x_0)$ contained in $U \setminus C_*$, and that $W(x) = S(x, p) = G(x)$ is smooth in $B_\delta(x_0)$, proving a).

Next, consider $x_0 \in \Omega \setminus C_*$. There exists $(x_0, p_0) \in \mathcal{M}$ minimizing $S(x_0, \cdot)$ at which π is nonsingular. Let $G(x)$ be defined in a neighborhood of x_0, as described prior to the theorem statement above. Then since $G(x) = S(x, p)$ in a neighborhood of (x_0, p_0) we have

$$W(x) \leq G(x), \quad \text{with equality at } x_0.$$

Hence

$$\limsup_{x \to x_0} W(x) \leq G(x_0) = W(x_0),$$

showing that W is upper-semicontinuous at x_0. Since lower-semicontinuity holds in general, this proves b).

Now suppose x_0 is an interior point of $\Omega \setminus C_*$. The local boundedness assumption implies that for some $\delta, K > 0$ we have

$$|p| \leq K \quad \text{for all } (x, p) \in \mathcal{M} \text{ with } x \in B_\delta(x_0).$$

For simplicity denote $B = B_\delta(x_0)$. We note that B is convex, and we may assume $B \subseteq \Omega \setminus C_*$. We will show that

$$|W(x_1) - W(x_2)| \leq K|x_1 - x_2|, \quad \text{for all } x_1, x_2 \in B. \tag{23}$$

Consider any $x_1 \in B$. Since $x_1 \notin C_*$ there exists $(x_1, p_1) \in \mathcal{M}$ minimizing $S(x_1, \cdot)$ and a smooth $G(\cdot)$ defined on some $B_{\delta_1}(x_1) \subseteq B$ as previously. For any $x_2 \in B_{\delta_1}(x_1)$ we have $|DG(x_2)| \leq K$ and so

$$\begin{aligned} W(x_2) &\leq G(x_2) \\ &\leq G(x_1) + K|x_2 - x_1| = W(x_1) + K|x_2 - x_1|. \end{aligned} \tag{24}$$

Pick a unit vector u and consider the maximal interval $0 \leq t \leq T$ of t such that $x_2 = x_1 + tu \in B$ and (24) holds. It must be that $x_1 + Tu \in \partial B$, else we could take $x_2 = x_1 + Tu$ and repeat the above argument, finding $\delta_2 > 0$ so that

$$W(x_3) \leq W(x_2) + K|x_3 - x_2| \quad \text{for all } x_3 \in B_{\delta_2}(x_2).$$

We could then combine this with (24) to find that (24) also holds for $x_3 = x_1 + (T + t)u$, $0 \leq t < \delta_2$, contradicting the maximality of T. In this way we find that (24) holds for x_2 along any ray from x_1 up to ∂B. Since B is convex, this means (24) holds for all $x_1, x_2 \in B$. Interchanging x_1 and x_2 we conclude (23), as claimed.

Lemma 2 *The set C of all caustics has measure 0.*

Proof This is a direct application of Sard's Theorem. (See [20].)

We observe that the set of caustics is closed, relative to Ω, because it is the image under π of the set of points where the Jacobian of π vanishes. Indeed $\mathcal{C} = \{(x,p) \in \mathcal{M} : \ \pi \text{ is singular at } (x,p)\}$ is a closed subset of $\mathcal{M}$. Any $x_0 \in C$ has a closed neighborhood $\bar{B}_\delta(x_0) \subseteq \Omega$. Since $\mathcal{M}$ is locally bounded, $\mathcal{C} \cap (\bar{B}_\delta(x_0) \times \mathbb{R}^d)$ is compact. Hence its image $C \cap \bar{B}_\delta(x_0)$ under π is closed. Since $x_0 \in C$ was arbitrary, it follows that C is closed, relative to Ω. As a result of this and part c) of the theorem above, the points at which $W(\cdot)$ fails to be Lipschitz is a subset of the caustics and thus has measure 0. However, even if we knew W was continuous in Ω, this would not be enough to deduce that W is Lipschitz in Ω. (Indeed the Cantor function is locally Lipschitz in the compliment of the Cantor set, which has measure 0, but is not Lipschitz overall.) The following stronger hypothesis on C_* implies Lipschitz continuity and will be adequate for simple examples, although it is too strong to be useful in the general case.

Lemma 3 *Suppose $n \geq 2$ and C_* has no accumulation points. Then W is (locally) Lipschitz continuous in Ω.*

Note that for $n = 1$ this result is false, as the overloop of Example 2.1.4 illustrates.

Proof. The hypotheses imply that C_* is closed, so that $\Omega \setminus C_*$ is open and Theorem 2 c) assures us that W is Lipschitz there. We need only consider a neighborhood $B_\delta(x_0)$ containing a single point x_0 of C_*. Repeating the reasoning of Theorem 2 b), it follows that for some $K > 0$,

$$|W(x_1) - W(x_2)| \leq K|x_1 - x_2|, \quad x_1, x_2 \in B_\delta(x_0) \tag{25}$$

provided x_0 is not on the line segment joining x_1 and x_2. If x_0 is on this line segment, introduce x_3 close to x_0 but off the line segment. Pass to the limit as $x_3 \to x_0$ in

$$|W(x_1) - W(x_2)| \le K|x_1 - x_3| + K|x_3 - x_2|$$

to see that (25) holds for any x_1, x_2 in the punctured ball $B_\delta(x_0) \setminus \{x_0\}$. Hence

$$w_0 = \lim_{x \to x_0} W(x)$$

exists. Once we show $w_0 = W(x_0)$ it will follow that (25) holds on the full open ball $B_\delta(x_0)$, and the proof of the lemma will be complete. We know

$$W(x_0) \le w_0 \tag{26}$$

by lower semi-continuity.

Let $(x_0, p_0) \in \mathcal{M}$ such that $S(x_0, p_0) = W(x_0)$. Consider the set $P_{x_0} = \{p : (x_0, p) \in \mathcal{M}\}$. The local boundedness hypothesis tells us that $P_{x_0} \subseteq \mathbb{R}^n$ is bounded. Let p_* be a boundary point of P_{x_0} whose distance from p_0 is as small as possible. We must have $(x_0, p_*) \in \mathcal{M}$ because otherwise it would be a boundary point of $\mathcal{M}$ over Ω, contrary to hypothesis (A6). On the other hand, no neighborhood of (x_0, p_*) in $\mathcal{M}$ can be contained in $\{x_0\} \times \mathbb{R}^d$ or else p_* would be interior to P_{x_0}. Hence there exits a sequence of points in $\mathcal{M}$, $(x_n, p_n) \to (x_0, p_*)$ with $x_n \neq x_0$. Since $W(x_n) \le S(x_n, p_n)$ we can argue that

$$w_0 = \lim W(x_n) \le \lim S(x_n, p_n) = S(x_0, p_*).$$

Because no boundary points of P_{x_0} are closer to p_0 than p_*, the line segment between (x_0, p_0) and (x_0, p_*) is contained in $\mathcal{M}$ and along it $dS = p \cdot dx = 0$, so that

$$w_0 \le S(x_0, p_*) = S(x_0, p_0) = W(x_0).$$

Together with (26) this completes the proof.

The main result of this section is the following.

Theorem 3 *If $H(x,p)$ is convex in p for each x and $W(x)$ is (locally) Lipschitz continuous in Ω, then W is a viscosity solution of $H(x, DW(x)) = 0$ in Ω.*

Proof. We need to show that

$$H(x,p) \le 0$$

for all p in the set $D^+W(x)$ of superdifferentials of W at x. Frankowska [12] showed that for (locally) Lipschitz functions W,

$$D^+W(x) \subseteq \partial W(x),$$

where $\partial W(x)$ is the generalized gradient in the sense of Clarke. Clarke [8] has shown (Theorem 2.5.1) that for any set G of Lebesgue measure 0

$$\partial W(x) = \text{co}\{\lim DW(x_i) : \; x_i \to x, \; x_i \notin G, DW(x_i) \text{ converges}\}. \tag{27}$$

(The "co" indicates the convex hull.) We take

$$G = \bar{C}^* \cup \{x \in \Omega : \; W \text{ is not differentiable at } x\}.$$

G does have measure 0 by the Lipschitz hypothesis, Lemma 2 and the observation following it which implies $\bar{C}_* \subseteq C$. We claim that for any $x_0 \in \Omega \setminus G$, $(x_0, p_0) \in \mathcal{M}$ where $p_0 = DW(x_0)$ and $S(x_0, p_0) = W(x_0)$. Indeed $x_0 \notin C_*$ means there exists $(x_0, p_0) \in \mathcal{M}$ at which π is nonsingular and $S(x_0, p_0) = W(x_0)$. There exists a smooth $G(x)$ defined in a neighborhood of x_0 with $p_0 = DG(x_0)$ and $(x, DG(x))$ defines a neighborhood of (x_0, p_0) in $\mathcal{M}$, with $G(x) = S(x, DG(x))$. Therefore in this neighborhood we have

$$W(x) \leq G(x)$$

with equality at x_0. Since both sides are differentiable at x_0 it follows that

$$DW(x_0) = DG(x_0) = p_0.$$

We already know $S(x_0, p_0) = W(x_0)$ by the choice of p_0. This justifies our claim.

Consider then any such sequence $x_i \to x$, $x_i \in G$ as in (27). We have $(x_i, p_i) \in \mathcal{M}$ where $p_i = DW(x_i)$ converges to some p. By Lemma 1, $(x, p) \in \mathcal{M}$. We conclude that

$$D^+W(x) \subseteq \text{co}\{p : \; (x, p) \in \mathcal{M}\}.$$

(We comment that this fact is in Cannarsa and Soner [6], at least for semi-convex (or concave) functions; see Def.4.3, Prop.4.9. The same is true of D^-W. This provides at a fundamental level the observation of Ball and Helton [2] about the equivalence of $H(x, D^{\pm}W(x)) \leq 0$, which results from time-reversal in the argument of M. James.) Since $H(x, p) = 0$ on $\mathcal{M}$ and $H(x, \cdot)$ is convex, it follows that

$$H(x, p) \leq 0$$

for all $p \in D^+W(x)$, which completes our proof.

4 Examples with Discontinuity

Our intent has been to develop viscosity solution properties directly from

$$W(x) = \inf_{p:(x,p)\in\mathcal{M}} S(x, p) \tag{28}$$

under the basic hypotheses of Section 2, divorced from any variational interpretation. The obvious shortcoming of those hypotheses is that they do not imply the continuity that is needed in Theorem 3. In this section we provide two examples which offer some insight into how discontinuities can arise in (28), and an important feature of variational problems that our basic hypotheses fail to capture.

Consider time-dependent equations of the following form in one space dimension ($x \in \mathbb{R}^1$):

$$D_t W + H(D_x W) = 0, \quad W(0,x) = \Phi(x) \tag{29}$$

Assume that $H(\cdot)$ is a proper convex function in the sense of convex analysis, with conjugate function

$$L(v) = \sup_p \{p \cdot v - H(p)\}, \tag{30}$$

and moreover that both $L(\cdot)$ and $H(\cdot)$ are finite-valued and sufficiently smooth. Assume that $\Phi(\cdot)$ is smooth with $\Phi'(x)$ bounded. We note that time-dependent equations such as this may be put in the form (2) by augmenting the state and costate variables

$$y = (t,x), \quad q = (\sigma, p)$$

and defining an augmented Hamiltonian,

$$H^+(y,q) = \sigma + H(p). \tag{31}$$

The Cauchy problem (29) has been thoroughly studied, since it describes the value function for the following basic variational problem:

$$W(t,x) = \inf_{x_t = x} \left\{ \Phi(x_0) + \int_0^t L(\dot{x}_s)\, ds \right\} \tag{32}$$

(The infimum is over $x_s : [0,t] \to \mathbb{R}^1$ which are absolutely continuous with the prescribed terminal position $x_t = x$.) An efficient discussion can be found in Fleming and Soner [11], Sections I.8, I.9. In particular, for $t > 0$ (32) defines a Lipschitz continuous function and the infimum is achieved by one of the bicharacteristic curves that will make up $\mathcal{M}$ below, so that for $t > 0$ (32) defines the same $W(t,x)$ that we would construct following (28). The variational representation implies Lipschitz continuity. E. Hopf [14] studied (29) long ago. Among other things, he showed that the formula

$$W(t,x) = \inf_z \sup_p \{\Phi(z) + p \cdot (x - z) - tH(p)\} \tag{33}$$

provides a generalized solution to (29) which is Lipschitz. Later, after the theory of viscosity solutions was in place, Bardi and Evans [3] revisited

Hopf's formula, showing that it provides the unique continuous viscosity solution for $t > 0$ to the Cauchy problem. One can check that Hopf's formula also reduces to (28) above in the context of the augmented variables (31).

The manifold $\mathcal{M}$ that we would construct in solving this problem is exactly what results from the method of characteristics. Start with the initial manifold consisting of $t = 0$, $p = \Phi'(x)$ and $\sigma = -H(p)$. Take these as initial conditions for the bicharacteristic system associated with H^+, which reduces to

$$\begin{aligned} \dot{p}_t &= 0 \\ \dot{x}_t &= H'(p_t) \\ \sigma_t &= -H(p_t). \end{aligned}$$

$\mathcal{M}$ is the union of points on this family of bicharacteristic curves. Note that $\dot{x}_t$ and p_t are constant. For $v = \dot{x}_t$ the supremum in (30) is achieved by $p = p_t$ (since $\dot{x}_t = H'(p_t)$). Therefore, the function S associated with $\mathcal{M}$ as in (7) obeys the following along a bicharacteristic:

$$dS = q_t \cdot dy_t = p_t \cdot dx_t - H(p_t)\,dt = [p_t H_p(p_t) - H(p_t)]dt = L(\dot{x}_t)dt. \quad (34)$$

We can therefore express the function S on $\mathcal{M}$ by

$$S(t, x_t, \sigma_t, p_t) = \Phi(x_0) + \int_0^t L(\dot{x}_s)\,ds$$

along the bicharacteristic (x_t, p_t). This makes plain the correspondence of (28) with (32) for $t > 0$.

The variational interpretation of $W(t,x)$ is limited to $t > 0$, while the construction of $\mathcal{M}$, $S(\cdot)$ and $W(\cdot)$ by (28) extend equally well to $t < 0$. The hypotheses of Section 2 remain satisfied. We have lost the variational interpretation for $t < 0$ and, as our first example will show, we can lose continuity of $W(\cdot)$ as well.

4.1 Examples

We now offer two examples, both using $H(p) = \frac{1}{2}p^2$. The corresponding Lagrangian is $L(v) = \frac{1}{2}v^2$. We consider two choices of Φ. The resulting manifold $\mathcal{M}$ is nicely parameterized by t, x_0. Since the augmenting costate variable is always given by $\sigma = -H(p)$, only t, x, p and S are included in the plots below.

4.1.1 An Example with Upstream Discontinuity

Consider $\Phi(x) = 2\arctan(x) - x$. Φ' is bounded, in accord with the hypotheses above. The resulting manifold $\mathcal{M}$ is illustrated in Figure 4, and in Figure 5 using different t cross-sections. The bottom row in each figure displays $S(t,\cdot)$ as a multivalued function of x, for the corresponding t values. As t increases the cross-section undergoes a "shearing" motion, with the upper half-plane moving to the right and the lower to the left. By the time t=1.2 the cross-section of $\mathcal{M}$ has formed an underloop, producing a nondifferentiable but Lipschitz function W, in accord with the implications of the variational interpretation. However for $t<0$ we see that an overloop develops, yielding a discontinuous W for $t<-1$.

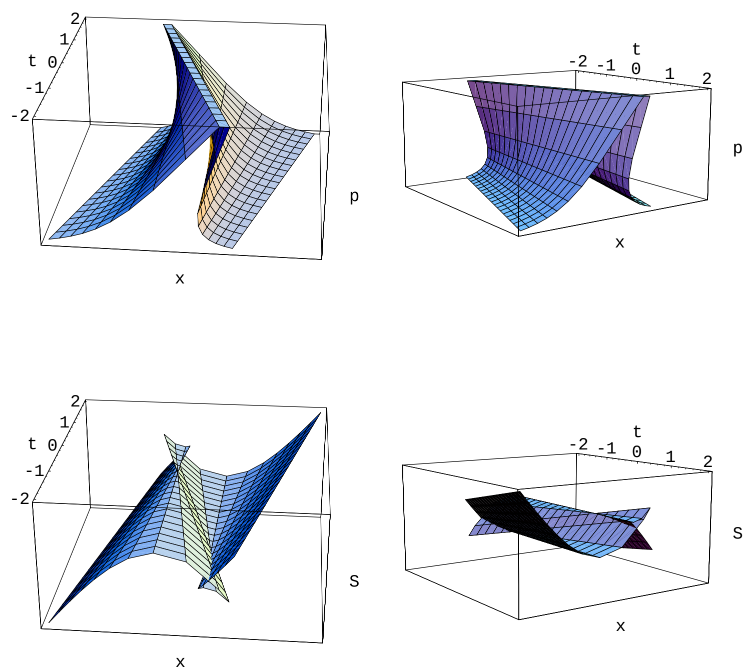

Figure 4: Manifold Perspectives for Example 4.1.1

Considering $t<0$ is obviously inappropriate in terms of the variational problem. In terms of the the basic hypotheses of Section 2 this example suggests that the direction of the bicharacteristic flow is significant for the continuity of $W(\cdot)$. We might hope that continuity of W (but not smoothness) will be maintained as we follow the bicharacteristics in their forward or "downstream" direction, but as we move "upstream" we are more likely to encounter discontinuities and thereby lose the viscosity subsolution property. The next example shows that in fact it is possible to lose continuity

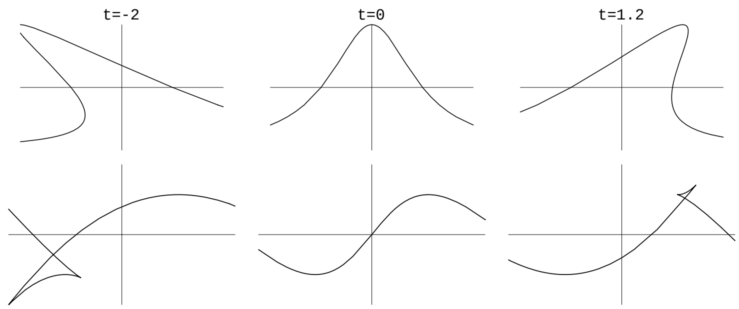

Figure 5: Cross-Sections for Example 4.1.1

in the "downstream" direction as well.

4.1.2 An Example without Local boundedness

Now take $\Phi(x) = \frac{1}{2}(\log(x^2+1) - x^2)$. We note immediately that $\Phi'(x) = \frac{-x^3}{x^2+1}$ is unbounded. The variational argument that $W(\cdot)$ is Lipschitz for $t > 0$ no longer applies. However the construction of $\mathcal{M}$, $S(\cdot)$ and then $W(\cdot)$ from (28) can still proceed. Figure 6 again displays t cross-sections of $\mathcal{M}$ on the top and the corresponding $S(t,\cdot)$ on the bottom row. We see that a pair of underloops has formed by $t = .95$, resulting in a nonsmooth but continuous W. However at $t = 1$ an "inversion" takes place, with the $p \to \pm\infty$ tails of $\mathcal{M}$ passing through the vertical axis. $\mathcal{M}$ violates the local boundedness hypothesis (A5) at this point. An overloop forms at this time, so that W develops a discontinuity. This discontinuity persists (although diminishing in size) as t continues to increase.

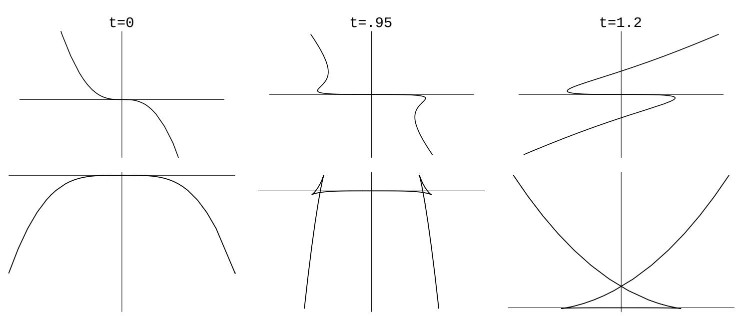

Figure 6: Example 4.1.2

It turns out that local boundedness does hold for the portion of $\mathcal{M}$ with $0 < t < 1$. This example illustrates that our local boundedness hypothesis is important for the continuity of $W(\cdot)$, but that it is not a property that propagates forward or "downstream" on $\mathcal{M}$. We pointed out in Section 2 that global boundedness does follow if the Hamiltonian has the property $H(x,p) \to 0$ as $|p| \to \infty$. In time dependent settings the augmented Hamiltonian (31) does not have this property. In the specific context of (29), the assumption of bounded $\Phi'(x)$ is a spatially global version (A5) which then does propagate downstream.

4.2 Comments on the Time-Dependent Equation

Byrnes [5] noted the connection between the case of $H(p) = \frac{1}{2}p^2$ above and Burgers' equation. His focus was on the Riccati PDE, which describes $\mathcal{M}$ under the assumption that it is a graph over (t,x): $p = p(t,x)$ (and $\sigma = -H(p(t,x))$). In the present setting the Riccati PDE would take the form

$$\frac{\partial p}{\partial t} + \frac{\partial p}{\partial x} \cdot H_p(p) = 0.$$

In higher dimensions the symmetry of $\partial p/\partial x$ follows from the Lagrange property of $\mathcal{M}$. This makes the Riccati PDE equivalent to the scalar conservation law

$$\frac{\partial p}{\partial t} + D_x H(p(t,x)) = 0. \tag{35}$$

For $H(p) = \frac{1}{2}p^2$ specifically we get Burgers' equation. The formation of shocks for (35) corresponds to the folding of $\mathcal{M}$ over state space, so that it is no longer a graph. In the context of (28) this corresponds to the formation of nonsmooth points, as illustrated in the examples. In the context of (35) it is natural to expect the minimizing p, as a function of (t,x), to provide a weak solution (see [22]). In fact the very construction (28) has been the basis of existence arguments for scalar conservation laws since the work of Hopf in the 1950's. Under appropriate technical conditions it produces the unique weak solution satisfying the usual entropy condition. (See Lax [16].) The Rankine-Hugoniot jump condition $[p]\,dx - [H]\,dt = 0$ and "equal area rule" are natural consequences.

5 Nonlinear L_2-Gain Calculations

We return in this final section to consider the implications of our results for the problem of L_2-gain calculation, introduced in Section 1. As pointed out there, van der Schaft [24], [25] has considered the use of Lagrange manifold constructions in this context, but only in terms of smooth solutions of (48) below in some domain Ω. We will obtain some of the same results under weaker hypotheses for which the construction yields a viscosity solution. The reader can refer to the papers of van der Schaft for more background on this topic and its relation to H_∞ control.

Recall from Section 1 that the problem concerns a control system

$$\dot{x}_t = f(x_t) + g(x_t)u_t \tag{36}$$

in which $u_t \in \mathbb{R}^m$ (locally L_2) is viewed as an input and the observed output is $y_t = h(x_t)$. We will call x_t solving (36) for some u_t a *controlled path*. The goal is to establish a bound γ on the L_2-norm of the map $u_\cdot \mapsto y_\cdot$. Two versions of this problem have been posed. One is to establish L_2-gain $\leq \gamma$ only for the zero-state response: for all $T > 0$ and controlled paths with $x_0 = 0$,

$$\int_0^T |h(x_t)|^2 \, dt \leq \gamma^2 \int_0^T |u_t|^2 \, dt. \tag{37}$$

The second is to generalize this to non-zero initial states x_0 allowing a nonnegative function $K(x_0)$ of the initial state to appear on the right:

$$\int_0^T |h(x_t)|^2 \, dt \leq \gamma^2 \int_0^T |u_t|^2 \, dt + K(x_0), \quad \text{all } T > 0. \tag{38}$$

The generalized version of the problem is to show there exists a (finite) function $K(\cdot) \geq K(0) = 0$ such that this holds for all initial conditions and all controlled paths. (This is the version of the problem referred to in Section 1.)

We define the *running cost* (or *supply rate* in the language of dissipative systems) to be

$$\ell(x, u) = \frac{1}{2}[\gamma^2 |u|^2 - |h(x)|^2], \tag{39}$$

and the associated Hamiltonian

$$\begin{aligned} H(x, p) &= \sup_u \{p \cdot (f(x) + g(x)u) - \ell(x, u)\} \\ &= \frac{1}{2}\gamma^{-2} p \cdot g(x)g(x)^T p + p \cdot f(x) + \frac{1}{2}|h(x)|^2. \end{aligned} \tag{40}$$

Note that the supremum is achieved for $u^*(p) = \gamma^{-2} g(x)^T p$, and $H_p(x, p) = f(x) + g(x)u^*(p)$. Thus the spatial part x_t of a bicharacteristic (x_t, p_t) is a controlled path

$$\dot{x}_t = H_p(x_t, p_t) = f(x_t) + g(x_t)u^*(p_t)$$

along which

$$H(x_t, p_t) = p_t \cdot \dot{x}_t - \ell(x_t, u^*(p_t)).$$

Both versions of the L_2-gain problem are expressed succinctly in terms of the *available storage* function, defined by

$$\phi_a(x_0) = -\inf \int_0^T \ell(x_t, u_t)\, dt, \tag{41}$$

where the infimum is over all $0 < T$ and controlled paths with the specified initial condition x_0. Considering $T \downarrow 0$ immediately implies

$$\phi_a(\cdot) \geq 0. \tag{42}$$

(However, without further assumptions on the control system it is possible for $\phi_a(x_0) = +\infty$.) L_2-gain $\leq \gamma$ for the zero-state response is equivalent to

$$\phi_a(0) = 0. \tag{43}$$

This is because (37) means $\phi_a(0) \leq 0$ which is equivalent to (43) by (42). L_2-gain $\leq \gamma$ holds in the generalized sense (38) if and only if

$$0 \leq \phi_a(\cdot) < \infty. \tag{44}$$

It is simple to check that ϕ_a obeys the Dissipation Inequality:

$$\phi(x_T) \leq \phi(x_0) + \int_0^T \ell(x_t, u_t)\, dt \quad \text{for all controlled paths.} \tag{45}$$

Moreover ϕ_a is minimal among all nonnegative functions $\phi(\cdot) \geq 0$ satisfying (45). James [15] shows that for a lower semi-continuous function ϕ, (45) is equivalent to the viscosity sense inequality

$$H(x, D^-\phi(x)) \leq 0. \tag{46}$$

(This is to say ϕ is a viscosity supersolution of $-H(x, D\phi(x)) = 0$, as many authors put it.) Ball and Helton [2] observe, by applying James' result in reverse time, that (45) is also equivalent to

$$H(x, D^+\phi(x)) \leq 0. \tag{47}$$

Soravia [23] and Ball and Helton [2] have shown that ϕ_a is actually a viscosity solution of

$$-H(x, D\phi(x)) = 0. \tag{48}$$

The question we consider here is whether we might be able to construct ϕ_a from an appropriate Lagrange manifold. Van der Schaft has recognized that the stable manifold at $(0,0)$ for the Hamiltonian system

$$\dot{x} = H_p(x,p), \quad \dot{p} = -H_x(x,p) \tag{49}$$

is a natural candidate. He makes appropriate controllability and observability assumptions which imply that this stable manifold is the graph of a smooth function over some neighborhood Ω_0 of 0 in state space. This provides (locally) the existence of a classical solution of (48), and thereby a local version of the L_2-gain property, holding for paths limited to Ω_0. We simply make our assumptions directly on the manifold $\mathcal{M}$ that we intend to work with. To be specific we assume the following:

(M1) $x = 0$, $p = 0$ is a hyperbolic equilibrium for (49) and $\mathcal{M}_s$ is the associated stable manifold.

(M2) Ω is an open region in state space containing 0 with the following properties:

a) for every $x \in \Omega$ there is some $(x, p) \in \mathcal{M}_s$ (i.e. $\mathcal{M}_s$ covers Ω);

b) for every $(x_0, p_0) \in \mathcal{M}_s$ with $x_0 \in \Omega$ the corresponding bicharacteristic (49) has $x_t \in \Omega$ for all $t \geq 0$.

(M3) $\mathcal{M}$ is the submanifold of $\mathcal{M}_s$ consisting of those $(x, p) \in \mathcal{M}_s$ with $x \in \Omega$. We assume $\mathcal{M}$ to satisfy (A1) – (A6) of Section 2.

(M4) $V_s(x) = \sup_{p:\,(x,p)\in\mathcal{M}} S(x, \cdot)$ is a continuous viscosity solution of (48), where S is the function on $\mathcal{M}$ determined $dS = p \cdot dx$, $S(0, 0) = 0$.

Several comments are in order. First, $\mathcal{M}_s$ and the hyperbolicity in (M1) are determined by $H(\cdot, \cdot)$. Second, Ω is not arbitrary but must be chosen to satisfy (M2) with respect to $\mathcal{M}_s$. $\mathcal{M}$ is then determined by the choice of Ω. It follows that Ω and $\mathcal{M}$ are forward invariant for (49): $x_0 \in \Omega$ and $(x_0, p_0) \in \mathcal{M}$ implies $x_t \in \Omega$ and $(x_t, p_t) \in \mathcal{M}$ for all $t \geq 0$. (This is important in Lemma 5.) Next, as noted by van der Schaft and others, $\mathcal{M}$ is a simply connected Lagrange manifold. Therefore the function $S(x, p)$ is well-defined on it. $H(0, 0) = 0$ (by (40)) and since H is constant along bicharacteristics in $\mathcal{M}$, all of which converge to $(0, 0)$ as $t \to +\infty$, we see that $H = 0$ everywhere on $\mathcal{M}$.

We accommodate the "-" sign in (48) by a simple change of variables. Note that (48) is equivalent to saying that $\psi = -\phi$ is a viscosity solution of

$$\tilde{H}(x, D\psi(x)) = 0, \quad x \in \Omega \tag{50}$$

with respect to the Hamiltonian

$$\tilde{H}(x, p) = H(x, -p), \tag{51}$$

Define a new Lagrange manifold $\tilde{\mathcal{M}}$ by

$$(x, p) \in \tilde{\mathcal{M}} \text{ if and only if } (x, -p) \in \mathcal{M}.$$

$\tilde{H}(x,\cdot)$ is convex since $H(x,\cdot)$ is, and $\tilde{H},\tilde{\mathcal{M}}$ satisfy our basic hypotheses since $H,\mathcal{M}$ do. The respective functions with $dS = p\cdot dx$ on $\mathcal{M}$ and $d\tilde{S} = p\cdot dx$ on $\tilde{\mathcal{M}}$ are related by

$$\tilde{S}(x,p) = -S(x,-p).$$

Our candidate to solve (50) is

$$W(x) = \inf_{p\colon (x,p)\in\tilde{\mathcal{M}}} \tilde{S}(x,p).$$

This is equivalent to

$$V_s(x) = -W(x) = -\inf_{p\colon (x,p)\in\tilde{\mathcal{M}}} \tilde{S}(x,p) = \sup_{p\colon (x,p)\in\mathcal{M}} S(x,p) \qquad (52)$$

being a viscosity solution of (48) in Ω.

One may check that the conversion from H to $\tilde{H}$ involves a time reversal. What was the "upstream" direction of a bicharacteristic on $\mathcal{M}$ becomes the "downstream" of its counterpart on $\tilde{\mathcal{M}}$. $\tilde{\mathcal{M}}$ is (a subset of) the unstable manifold associated with the equlibrium point at $(x,p)=(0,0)$ for the $\tilde{H}$ bicharacteristic system. Hence all of $\tilde{\mathcal{M}}$ is "downstream" from an arbitrarily small neighborhood of $(0,0)$ in $\tilde{\mathcal{M}}$. In accord with our observations following Example 4.1.1, this makes us optimistic that W will in fact be Lipschitz continuous and thus truly a viscosity solution of (50). We have simply assumed the validity of this in (M4).

It follows from James' result (46) that V_s satisfies the dissipation inequality, restricted to controlled paths with $x_t \in \Omega$. We abbreviate this by saying that "V_s satisfies (45) in Ω." Is is only reasonable to expect a comparison with the available storage function defined with a similar limitation. We let ϕ_a^Ω be defined as in (41) above, but with the additional restriction to paths $x_t \in \Omega$. The first of the following lemmas is clear.

Lemma 4 $0 \le \phi_a^\Omega(\cdot) \le \phi_a(\cdot)$ *in* Ω. ϕ_a^Ω *is minimal among* ϕ *which are nonnegative and satisfy (45) in* Ω.

Lemma 5 $V_s(0) \ge 0$ *and is minimal among those* ϕ *with* $\liminf_{x\to 0}\phi(x) \ge 0$ *and satisfying (45) in* Ω.

Proof. Since $(0,0)\in\mathcal{M}$ and $S(0,0)=0$, $V_s(0)\ge 0$ is immediate from the definition. Suppose ϕ satisfies (45) in Ω and $\liminf_{x\to 0}\phi(x)\ge 0$. Consider any $x_0\in\Omega$ and p_0 with $(x_0,p_0)\in\mathcal{M}$ and let (x_t,p_t) be the bicharacteristic through (x_0,p_0) for $t\ge 0$. We know $(x_t,p_t)\in\mathcal{M}$ and $x_t\in\Omega$ and that x_t is a controlled path with $u_t = u^*(p_t)$. Moreover, since $0 = H(x_t,p_t) = p_t\cdot\dot{x}_t - \ell(x_t,u_t)$ it follows that

$$S(x_T,p_T) - S(x_0,p_0) = \int_0^T p_t\cdot dx_t$$

$$
\begin{aligned}
&= \int_0^T \ell(x_t, u_t)\, dt \\
&\geq \phi(x_T) - \phi(x_0).
\end{aligned}
$$

As $T \to +\infty$ we know x_T, p_T and $S(x_T, p_T)$ all converge to 0. Thus the hypothesis $\liminf_{x\to 0} \phi(x) \geq 0$ implies that

$$
S(x_0, p_0) \leq \phi(x_0), \quad \text{all } x_0 \in \Omega, (x_0, p_0) \in \mathcal{M}.
$$

From this we conclude that $V_s(x_0) \leq \phi(x_0)$. Since V_s itself satisfies the dissipation inequality (45) in Ω, and $\liminf_{x\to 0} V_s(x) = V_s(0) \geq 0$, the proof is complete.

Theorem 4 *$V_s(\cdot) \leq \phi_a^\Omega(\cdot)$. $V_s = \phi_a^\Omega$ if and only if $V_s(\cdot) \geq 0$ in Ω.*

Proof. Since $\phi_a^\Omega(\cdot) \geq 0$, Lemma 5 implies $V_s(\cdot) \leq \phi_a^\Omega(\cdot)$. If $V_s(\cdot) \geq 0$ then Lemma 4 yields $V_s(\cdot) \geq \phi_a^\Omega(\cdot)$ and hence $V_s = \phi_a^\Omega$. If $V_s = \phi_a^\Omega$ then $V_s \geq 0$ is immediate from Lemma 4.

The idea is to compute $\mathcal{M}_s$, select the largest Ω possible satisfying our hypotheses and then observe the resulting V_s. If $V_s(\cdot) \geq 0$ in Ω then $V_s = \phi_a^\Omega$ and we have verified that L_2-gain $\leq \gamma$ in Ω. It is interesting to consider how this program might fail. If $V_s(x_0) < 0$ at some $x_0 \in \Omega$ (necessarily $x_0 \neq 0$), consider the null-controlled path $\dot{x}_t = f(x_t)$ through x_0. If x_t stays in Ω we must have

$$
V_s(x_T) \leq V_s(x_0) + \int_0^T \ell(x_t, 0)\, dt \leq V_s(x_0) < 0.
$$

In particular x_t is bounded away from 0. In the theory of nonlinear H_∞ control the L_2-gain criterion is usually taken together with a requirement of asymptotic stability of 0 for the null-controlled system. Thus $V_s(x_0) < 0$ implies that either Ω is not invariant for the null-controlled system, or that it is not asymptotically stable. This is in accord with the result that global asymptotic stability of the null-controlled system implies that any function satisfying the dissipation inequality (45) with $\phi(0) = 0$ must be nonnegative; see [25]. If we find $V_s(0) > 0$ then $\phi_a(0) \geq \phi_a^\Omega(0) \geq V_s(0) > 0$ and so the zero-state L_2-gain is not bounded by γ.

5.1 Linear-Quadratic Examples

Simple linear-quadratic examples in 2 dimensions will illustrate situations in which $V_s \neq \phi_a^\Omega$. We will use $\Omega = \mathbb{R}^2$. From the above remarks we know the null-controlled system must fail to be asymptotically stable. Consider the control system ($u \in \mathbb{R}^2$)

$$
\dot{x} = Fx + Gu
$$

with output given by

$$h(x) = Yx.$$

(F, G, Y are 2×2 real matrices.) The resulting Hamiltonian is

$$H(x,p) = \frac{1}{2\gamma^2} p \cdot GG^T p + p \cdot Fx + \frac{1}{2} x \cdot Y^T Y x.$$

The Hamiltonian system (49) is linear,

$$\begin{bmatrix} \dot{x} \\ \dot{p} \end{bmatrix} = \mathcal{H} \begin{bmatrix} x \\ p \end{bmatrix},$$

where

$$\mathcal{H} = \begin{bmatrix} F & \gamma^{-2} GG^T \\ -Y^T Y & -F^T \end{bmatrix}.$$

In both our examples $\mathcal{M}$ will be the stable subspace of $\mathcal{H}$, described by $p = Px$ where P is a particular symmetric 2×2 matrix. ($\Omega = \mathbb{R}^2$.) The resulting V_s is the quadratic function

$$V_s(x) = \frac{1}{2} x \cdot Px.$$

5.1.1 Example

Consider $\gamma = 1$ and

$$F = \begin{bmatrix} -2 & 0 \\ 0 & \frac{2}{3} \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

One calculates that

$$P = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad V_s(x) = \frac{1}{2}(x_1^2 - x_2^2).$$

Since V_s fails to be nonnegative we know $V_s \neq \phi_a$. In this example $\phi_a \equiv +\infty$, due to the instability of F. Indeed for a null-controlled response, x_t with $u_t \equiv 0$ the dissipation inequality (45) implies (because $\phi_a \geq 0$)

$$\int_0^T \frac{1}{2} x_t \cdot Y^T Y x_t \, dt \leq \phi_a(x_0),$$

which, since $Y^T Y$ is positive definite, converges to $+\infty$ as $T \to +\infty$, unless x_0 is in the stable subspace of F. Using an arbitrarily small control to move x_t out of this stable subspace one easily argues that $\phi_a(x_0) = +\infty$ on the stable subspace of F as well.

5.1.2 Example

Take $\gamma = 1$ again and

$$F = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 \end{bmatrix}.$$

One calculates that

$$P = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & -2 \end{bmatrix}, \quad V_s(x) = \frac{1}{4}x_1^2 - x_2^2.$$

Now however ϕ_a is finite, given by

$$\phi_a(x) = \frac{1}{4}x_1^2.$$

(Essentially, the instability is not observed.) Indeed, one can verify that ϕ_a is a classical solution of (48) and thus satisfies the dissipation inequality. By considering the feedback controls

$$u_t = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} x_t,$$

one sees that ϕ_a is the minimal nonnegative function satisfying (45).

5.2 A Nonlinear Example

Finally, we offer the following nonlinear example in 2 dimensions, indicating coordinates as $x_t = (x_{1,t}, x_{2,t})$. For (36) consider the linear system

$$\dot{x}_t = -x_t + u_t,$$

but with the following nonlinear output:

$$y_t = h(x_t) = (x_{1,t}, \frac{(1 + 5x_{1,t})x_{2,t}}{1 + x_{2,t}^2}).$$

Taking $\gamma = 4$, we calculate the unstable manifold $\mathcal{M}_s$ of the Hamiltonian system for (40) at $(0, 0)$. Figure 7 displays a selection of characteristics from $\mathcal{M}_s$ for $x_1 > 0$ with the corresponding S values plotted vertically. One can see that $V_s(\cdot)$ resulting from (52) develops a ''crease" over the x_1 axis as one moves away from the origin, resulting in a true viscosity solution to (48).

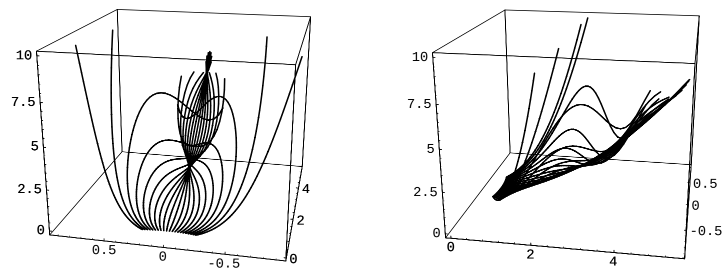

Figure 7: Nonlinear Example

References

[1] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (second edition) Springer-Verlag, New York, 1989.

[2] J. A. Ball and J. W. Helton, *Viscosity solutions of Hamilton-Jacobi equations arising in nonlinear H_∞-control* (Summary) J. Math. Systems Estimation and Control **6** (1996), pp. 109 – 112.

[3] M. Bardi and L. C. Evans, *On Hopf's formula for solutions of Hamilton-Jacobi equations* Nonlinear Anal., Th. Meth. and Appl. **8** (1984), pp. 1373 – 1381.

[4] P. Brunovsky, *On optimal stabilization of nonlinear systems*, Mathematical Theory of Control, A. V. Balakrishnan and Lucien W. Neustadt (ed.s), New York, 1967.

[5] C. Byrnes, *On the Riccati partial differential equation for nonlinear Bolza and Lagrange problems* (to appear).

[6] P. Cannarsa, H. M. Soner, *On the singularities of the viscosity solutions to Hamilton-Jacobi-Bellman equations* Indiana U. Math. J. **36** (1987), pp. 501 – 524.

[7] N. Caroff and H. Frankowska, *Conjugate points and shocks in nonlinear optimal control*, Trans. AMS **348** (1996), pp. 3133 – 3153.

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

[9] M. V. Day, *Regularity of boundary quasipotentials for planar systems*, Appl. Math. Optim. **30** (1994), pp. 79 – 101.

[10] M. V. Day and T. A. Darden, *Some regularity results on the Ventcel-Freidlin quasi-potential function*, Appl. Math. Optim. **13** (1985), pp. 259 – 282.

[11] W. Fleming and H. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer-Verlag, New York, 1993.

[12] H. Frankowska, *Hamilton-Jacobi equations: viscosity solutions and generalized solutions*, J. Math. Anal. Appl. **141** (1989), pp. 21 – 26.

[13] M. I. Freidlin and A. D. Wentzell, *Small Random Perturbations of Dynamical Systems*, Springer-Verlag, New York , 1984.

[14] E. Hopf *Generalized solutions of non-linear equations of first order*, J. Math. Mech. (1965), pp. 951 – 974.

[15] M. James, *A partial differential inequality for dissipative nonlinear systems*, Syst. and Control Letters **21** (1993), pp. 315–320.

[16] P. D. Lax, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, SIAM, Philadelphia, PA, 1973.

[17] D. L. Lukes, *Optimal regulation of nonlinear dynamical systems*, SIAM J. Control **7** (1969), pp. 75 – 100.

[18] V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics*, D. Reidel Dordrecht, Holland, 1981.

[19] R. S. Meier and D. L. Stein, *A scaling theory of bifurcations in the symmetric weak-noise escape problem*, J. Stat. Phys. **83** (1996), pp. 291 – 357.

[20] J. Milnor, *Topology from the Differentiable Viewpoint* Univ. Press of Virginia, Charlottesville, Virginia, 1965.

[21] B. Perthame, *Perturbed dynamical systems with an attracting singularity and weak viscosity limits in Hamilton-Jacobi equations* TAMS **317** (1990), pp. 723–748.

[22] J. Smoller, *Shock Waves and Reaction-Diffusion Equations*, Springer-Verlag, New York, 1983 .

[23] P. Soravia, *H_∞-control of nonlinear systems: differential games and viscosity solutions*, SIAM J. Control **34** (1996), pp. 1071 – 1097.

[24] A. J. van der Schaft, *L_2-gain analysis of nonlinear systems and nonlinear state feedback H_∞-control* IEEE Trans. Aut. Cont. **37** (1992), pp. 770 – 784.

[25] A. J. van der Schaft, *Nonlinear state space H_∞ control theory*, Essays on Control: perspectives in the Theory and its Applications, H. L. Trentelman and J. C. Willems (ed.s), Birkhäuser, Boston, 1993.