Viscosity Solutions of Hamilton-Jacobi Equations and Optimal Control Problems

Alberto Bressan

Penn State University

Contents

A first version of these lecture notes was written at NTNU, Trondheim, 2001.
This revised version was completed at Shanghai Jiao Tong University, Spring 2011.

1 - Preliminaries: the method of characteristics

A first order, scalar P.D.E. has the form

$$F(x, u, \nabla u) \;=\; 0 \qquad\qquad x \in \Omega \subseteq \mathbb{R}^n\,. \tag{1.1}$$

It is convenient to introduce the variable $p \doteq \nabla u$, so that $(p_1, \ldots, p_n) = (u_{x_1}, \ldots, u_{x_n})$. We assume that the $F = F(x, u, p)$ is a continuous function, mapping $\mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n$ into $\mathbb{R}$.

Given the boundary data

$$u(x) \;=\; \bar{u}(x) \qquad\qquad x \in \partial\Omega, \tag{1.2}$$

a solution can be constructed (at least locally, in a neighborhood of the boundary) by the classical method of characteristics. The idea is to obtain the values $u(x)$ along a curve $s \mapsto x(s)$ starting from the boundary of Ω, solving a suitable O.D.E. (figure 1.1).

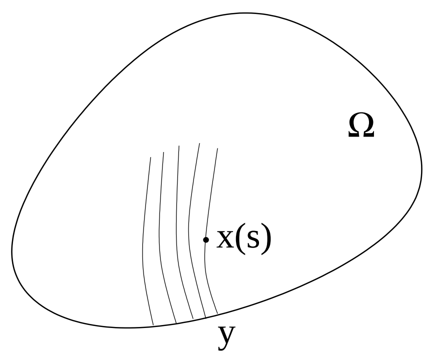

Figure 1.1. Characteristic curves starting from the boundary of Ω.

Fix a boundary point $y \in \partial\Omega$ and consider a curve $s \mapsto x(s)$ with $x(0) = y$. Call

$$u(s) \;\doteq\; u\big(x(s)\big), \qquad\qquad p(s) \;\doteq\; p\big(x(s)\big) = \nabla u\big(x(s)\big).$$

We seek an O.D.E. describing the evolution of u and $p = \nabla u$ along the curve. Denoting by an upper dot the derivative w.r.t. the parameter s, one has

$$\dot{u} \;=\; \sum_i u_{x_i}\, \dot{x}_i \;=\; \sum_i p_i\, \dot{x}_i\,, \tag{1.3}$$

$$\dot{p}_j \;=\; \dot{u}_{x_j} \;=\; \sum_i u_{x_j x_i}\, \dot{x}_i\,. \tag{1.4}$$

In general, $\dot{p}_j$ thus depends on the second derivatives of u. The key idea in the method of characteristics is that, by a careful choice of the curve $s \mapsto x(s)$, the terms involving second derivatives disappear from the equations. Differentiating (1.1) w.r.t. x_j we obtain

$$\frac{\partial F}{\partial x_j} + \frac{\partial F}{\partial u} u_{x_j} + \sum_i \frac{\partial F}{\partial p_i} u_{x_i x_j} \;=\; 0\,. \tag{1.5}$$

Hence

$$\sum_i \frac{\partial F}{\partial p_i} u_{x_j x_i} \;=\; -\frac{\partial F}{\partial x_j} - \frac{\partial F}{\partial u} p_j \,. \tag{1.6}$$

If we now make the choice $\dot{x}_i = \partial F/\partial p_i$, the right hand side of (1.4) is computed by (1.6). We thus obtain a closed system of equations, which do not involve second order derivatives:

$$\begin{cases} \dot{x}_i \;=\; \dfrac{\partial F}{\partial p_i} & i = 1, \ldots, n\,, \\ \dot{u} \;=\; \displaystyle\sum_i p_i \,\frac{\partial F}{\partial p_i}\,, & \\ \dot{p}_j \;=\; -\dfrac{\partial F}{\partial x_j} - \dfrac{\partial F}{\partial u} p_j & j = 1, \ldots, n\,. \end{cases} \tag{1.7}$$

This leads to a family of Cauchy problems, which in vector notation take the form

$$\begin{cases} \dot{x} \;=\; \dfrac{\partial F}{\partial p} \\ \dot{u} \;=\; p \cdot \dfrac{\partial F}{\partial p} \\ \dot{p} \;=\; -\dfrac{\partial F}{\partial x} - \dfrac{\partial F}{\partial u} \cdot p \end{cases} \qquad \begin{cases} x(0) \;=\; y \\ u(0) \;=\; u(y) \\ p(0) \;=\; \nabla u(y) \end{cases} \qquad y \in \partial\Omega\,. \tag{1.8}$$

The resolution of the first order boundary value problem (1.1)-(1.2) is thus reduced to the solution of a family of O.D.E's, depending on the initial point y. As y varies along the boundary of Ω, we expect that the union of the above curves $x(\cdot)$ will cover a neighborhood of $\partial\Omega$, where our solution u will be defined.

Remark 1.1. If F is affine w.r.t. p, the equation (1.1) takes the form

$$F(x, u, p) \;=\; p \cdot \alpha(x, u) + \beta(x, u) \;=\; 0\,.$$

In this case, the partial derivatives $\partial F/\partial p_i$ do not depend on p. The first two equations in (1.7) can thus be solved independently, without computing p from the third equation:

$$\dot{x} \;=\; \alpha(x, u)\,, \qquad\qquad \dot{u} \;=\; p \cdot \dot{x} \;=\; -\beta(x, u)\,.$$

Example 1.2. The equation

$$|\nabla u|^2 - 1 \;=\; 0 \qquad\qquad x \in \Omega \tag{1.9}$$

on $\mathbb{R}^2$ corresponds to (1.1) with $F(x, u, p) = p_1^2 + p_2^2 - 1$. Assigning the boundary data

$$u = 0 \qquad\qquad x \in \partial\Omega\,,$$

a solution is clearly given by the distance function

$$u(x) \;=\; \mathrm{dist}\,(x,\, \partial\Omega)\,.$$

The corresponding equations (1.8) are

$$\dot{x} = 2p\,, \qquad \dot{u} = p \cdot \dot{x} = 2\,, \qquad \dot{p} = 0\,.$$

Choosing the initial data at a point y we have

$$x(0) = y, \qquad u(0) = 0, \qquad p(0) = \mathbf{n}\,,$$

where $\mathbf{n}$ is the interior unit normal to the set Ω at the point y. In this case, the solution is constructed along the ray $x(s) = y + 2s\mathbf{n}$, and along this ray one has $u(x) = |x - y|$. Evan if the boundary $\partial\Omega$ is smooth, in general the distance function will be smooth only on a neighborhood of this boundary. If Ω is bounded, there will be a set γ of interior points $\bar{x}$ where the distance function is not differentiable (fig. 1.2). These are indeed the points such that

$$\text{dist}\,(\bar{x},\, \partial\Omega) \;=\; |\bar{x} - y_1| \;=\; |\bar{x} - y_2|$$

for two distinct points $y_1, y_2 \in \partial\Omega$.

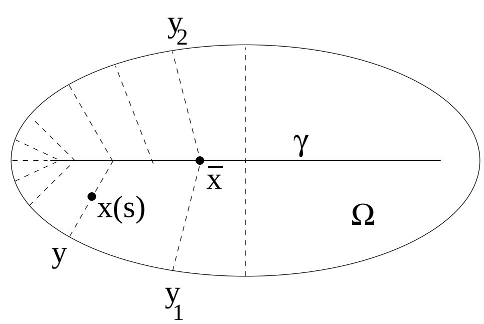

Figure 1.2

The previous example shows that, in general, the boundary value problem for a first order P.D.E. does not admit a global $\mathcal{C}^1$ solution. This suggests that we should relax our requirements, and consider solutions in a generalized sense. We recall that, by Rademacher's theorem, every Lipschitz continuous function $u : \Omega \mapsto I\!R$ is differentiable almost everywhere. It thus seems natural to introduce

Definition 1.3. A function u is a *generalized solution* of (1.1)-(1.2) if u is Lipschitz continuous on the closure $\overline{\Omega}$, takes the prescribed boundary values and satisfies the first order equation (1.1) at almost every point $x \in \Omega$.

Unfortunately, this concept of solution is far too weak, and does not lead to a useful uniqueness result.

Example 1.4. The boundary value problem

$$|u_x| - 1 \;=\; 0 \qquad x \in [-1, 1]\,, \qquad x(-1) \;=\; x(1) \;=\; 0, \tag{1.10}$$

admits infinitely many piecewise affine generalized solutions, as shown in fig. 1.3a.

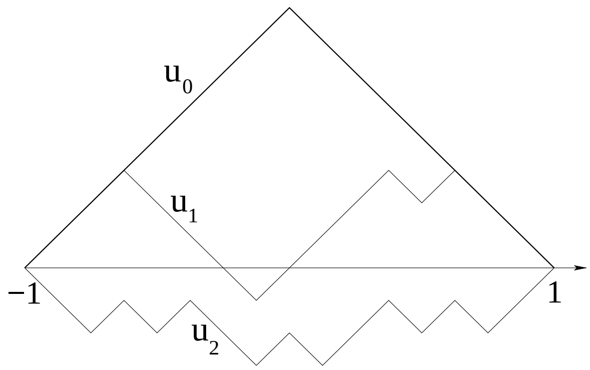

figure 1.3a

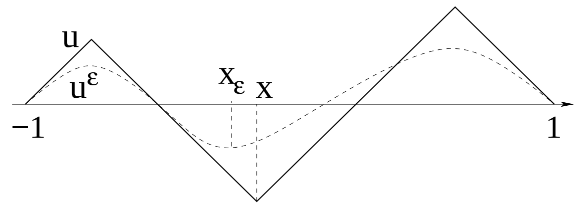

figure 1.3b

Observe that, among all these solutions, the distance function

$$u_0(x) \;=\; 1 - |x| \qquad\qquad x \in [-1,\,1]$$

is the only one that can be obtained as a vanishing viscosity limit. Indeed, any other generalized solution u with polygonal graph has at least one strict local minimum in the interior of the interval $[-1,1]$, say at a point x. Assume that $\lim\limits_{\varepsilon\to 0+} u^\varepsilon \to u$ uniformly on $[-1,1]$, for some family of smooth solutions to

$$|u^\varepsilon_x| - 1 \;=\; \varepsilon\, u^\varepsilon_{xx}\,.$$

Then for every $\varepsilon > 0$ sufficiently small the function u^ε will have a local minimum at a nearby point x_ε (fig. 1.3b). But this is impossible, because

$$\big|u^\varepsilon_x(x_\varepsilon)\big| - 1 \;=\; -1 \;\neq\; \varepsilon\, u^\varepsilon_{xx}(x_\varepsilon) \;\geq\; 0.$$

On the other hand, notice that if u^ε attains a local maximum at some interior point $x \in]-1,\,1[$, this does not lead to any contradiction.

In view of the previous example, one seeks a new concept of solution for the first order PDE (1.1), having the following properties:

1. For every boundary data (1.2), a unique solution exists, depending continuously on the boundary values and on the function F.

2. This solution u coincides with the limit of vanishing viscosity approximations. Namely, $u = \lim_{\varepsilon\to 0+} u^\varepsilon$, where the u^ε are solutions of

$$F(x, u^\varepsilon, \nabla u^\varepsilon) \;=\; \varepsilon\, \Delta u^\varepsilon\,.$$

3. In the case where (1.1) is the Hamilton-Jacobi equation describing the value function for some optimization problem, this new concept of solution should single out precisely the value function.

In the following sections we shall introduce the definition of *viscosity solution* and see how it fulfills the above requirements $\mathbf{1} - \mathbf{3}$.

2 - One-sided differentials

Let $u : \Omega \mapsto I\!R$ be a scalar function, defined on an open set $\Omega \subseteq I\!R^n$. The set of *super-differentials* of u at a point x is defined as

$$D^+u(x) \doteq \left\{ p \in I\!R^n ; \quad \limsup_{y \to x} \frac{u(y) - u(x) - p \cdot (y - x)}{|y - x|} \leq 0 \right\}. \tag{2.1}$$

In other words, a vector $p \in I\!R^n$ is a super-differential iff the hyperplane $y \mapsto u(x) + p \cdot (y - x)$ is tangent from above to the graph of u at the point x (fig. 2.1a). Similarly, the set of *sub-differentials* of u at a point x is defined as

$$D^-u(x) \doteq \left\{ p \in I\!R^n ; \quad \liminf_{y \to x} \frac{u(y) - u(x) - p \cdot (y - x)}{|y - x|} \geq 0 \right\}, \tag{2.2}$$

so that a vector $p \in I\!R^n$ is a sub-differential iff the hyperplane $y \mapsto u(x) + p \cdot (y - x)$ is tangent from below to the graph of u at the point x (fig. 2.1b).

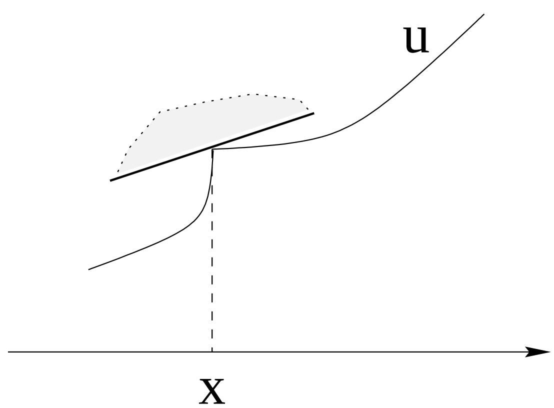

Figure 2.1a

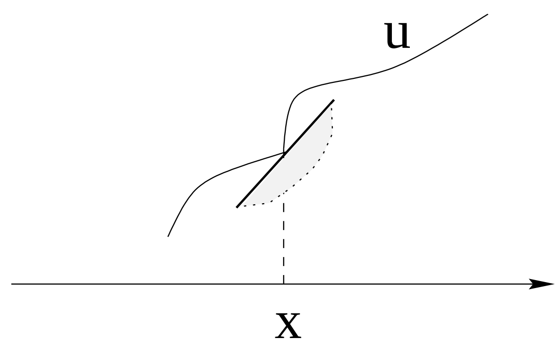

Figure 2.1b

Example 2.1. Consider the function (fig. 2.2)

$$u(x) \doteq \begin{cases} 0 & \text{if} \quad x < 0, \\ \sqrt{x} & \text{if} \quad x \in [0,1], \\ 1 & \text{if} \quad x > 1. \end{cases}$$

In this case we have

$$D^+u(0) = \emptyset, \qquad\qquad D^-u(0) = [0, \infty[\,,$$

$$D^+u(x) = D^-u(x) = \left\{ \frac{1}{2\sqrt{x}} \right\} \qquad\qquad x \in \,]0,1[\,,$$

$$D^+u(1) = \big[0, \ 1/2\big], \qquad\qquad D^-u(1) = \emptyset.$$

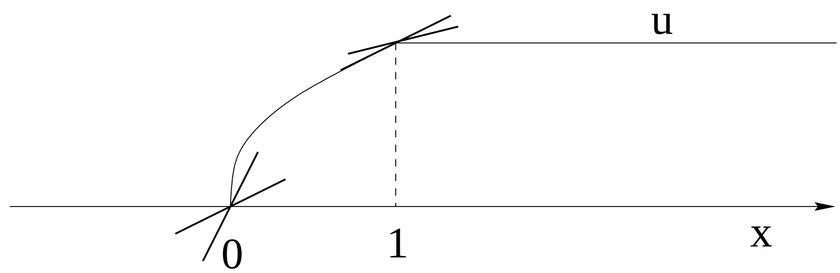

Figure 2.2

If $\varphi \in \mathcal{C}^1$, its differential at a point x is written as $\nabla\varphi(x)$. The following characterization of super- and sub-differentials is very useful.

Lemma 2.2. *Let $u \in \mathcal{C}(\Omega)$. Then*

(i) $p \in D^+u(x)$ if and only if there exists a function $\varphi \in \mathcal{C}^1(\Omega)$ such that $\nabla\varphi(x) = p$ and $u - \varphi$ has a local maximum at x.

(ii) $p \in D^-u(x)$ if and only if there exists a function $\varphi \in \mathcal{C}^1(\Omega)$ such that $\nabla\varphi(x) = p$ and $u - \varphi$ has a local minimum at x.

By adding a constant, it is not restrictive to assume that $\varphi(x) = u(x)$. In this case, we are saying that $p \in D^+u(x)$ iff there exists a smooth function $\varphi \geq u$ with $\nabla\varphi(x) = p$, $\varphi(x) = u(x)$. In other words, the graph of φ touches the graph of u from above at the point x (fig. 2.3a). A similar property holds for subdifferentials: $p \in D^-u(x)$ iff there exists a smooth function $\varphi \leq u$, with $\nabla\varphi(x) = p$, whose graph touches from below the graph of u at the point x (fig. 2.3b).

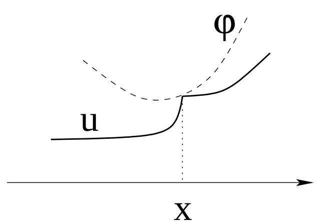

figure 2.3a

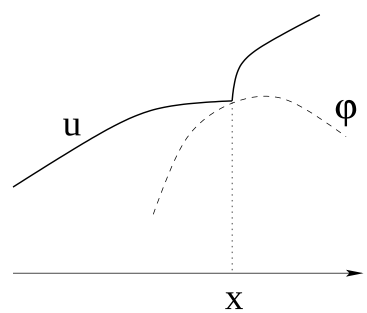

figure 2.3b

Proof of Lemma 2.2. Assume that $p \in D^+u(x)$. Then we can find $\delta > 0$ and a continuous, increasing function $\sigma : [0, \infty[\mapsto I\!R$, with $\sigma(0) = 0$, such that

$$u(y) \leq u(x) + p \cdot (y - x) + \sigma\big(|y - x|\big)|y - x|$$

for $|y - x| < \delta$. Define

$$\rho(r) \doteq \int_0^r \sigma(t)\,dt$$

and observe that

$$\rho(0) \;=\; \rho'(0) \;=\; 0, \qquad\qquad \rho(2r) \;\geq\; \sigma(r)\,r\,.$$

By the above properties, the function

$$\varphi(y) \;\doteq\; u(x) + p\cdot(y-x) + \rho\big(2|y-x|\big)$$

is in $\mathcal{C}^1(\Omega)$ and satisfies

$$\varphi(x) = u(x), \qquad\qquad \nabla\varphi(x) = p.$$

Moreover, for $|y-x| < \delta$ we have

$$u(y) - \varphi(y) \;\leq\; \sigma\big(|y-x|\big)|y-x| - \rho\big(2|y-x|\big) \;\leq\; 0.$$

Hence, the difference $u-\varphi$ attains a local maximum at the point x.

To prove the opposite implication, assume that $D\varphi(x) = p$ and $u-\varphi$ has a local maximum at x. Then

$$\limsup_{y\to x} \frac{u(y)-u(x)-p\cdot(y-x)}{|y-x|} \;\leq\; \limsup_{y\to x} \frac{\varphi(y)-\varphi(x)-p\cdot(y-x)}{|y-x|} \;=\; 0\,. \tag{2.3}$$

This completes the proof of (i). The proof of (ii) is entirely similar. □

Remark 2.3. By possibly replacing the function φ with $\tilde\varphi(y) = \varphi(y) \pm |y-x|^2$, it is clear that in the above lemma we can require that $u-\varphi$ attains a *strict* local minimum or a *strict* local maximum at the point x. This is important in view of the following stability result.

Lemma 2.4. *Let $u : \Omega \mapsto I\!R$ be continuous. Assume that, for some $\phi \in \mathcal{C}^1$, the function $u-\phi$ has a strict local minimum (a strict local maximum) at a point $x\in\Omega$. If $u_m \to u$ uniformly, then there exists a sequence of points $x_m \to x$ with $u_m(x_m) \to u(x)$ and such that $u_m - \phi$ has a local minimum (a local maximum) at x_m.*

Proof. Assume that $u-\phi$ has a strict local minimum at x. For every $\rho > 0$ sufficiently small, there exists $\varepsilon_\rho > 0$ such that

$$u(y) - \phi(y) \;>\; u(x) - \phi(x) + \varepsilon_\rho \qquad\quad \text{whenever} \;\; |y-x| = \rho\,.$$

By the uniform convergence $u_m \to u$, for all $m \geq N_\rho$ sufficiently large one has $u_m(y) - u(y) < \varepsilon_\rho/4$ for $|y-x| \leq \rho$. Hence

$$u_m(y) - \phi(y) \;>\; u_m(x) - \phi(x) + \frac{\varepsilon_\rho}{2} \qquad\qquad |y-x| = \rho\,,$$

This shows that $u_m - \phi$ has a local minimum at some point x_m, with $|x_m - x| < \rho$. Letting $\rho, \varepsilon_\rho \to 0$, we construct the desired sequence $\{x_m\}$. □

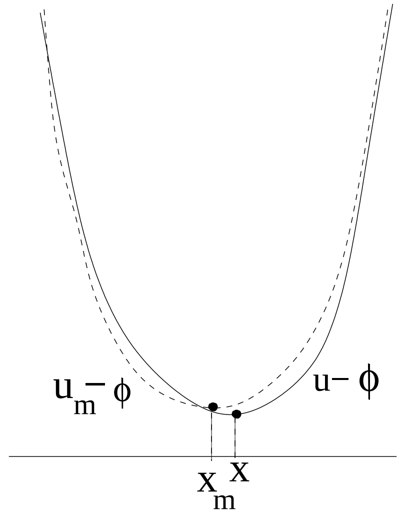

Figure 2.4a

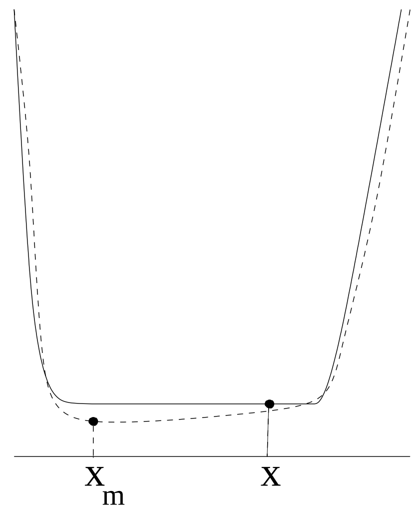

Figure 2.4b

This situation is illustrated in fig. 2.4a. On the other hand, if x is a point of non-strict local minimum for $u - \phi$, the slightly perturbed function $u_m - \phi$ may not have any local minimum x_m close to x (fig. 2.4b).

Some simple properties of super- and sub-differential are collected in the next lemma.

Lemma 2.5. *Let $u \in \mathcal{C}(\Omega)$. Then*

(i) If u is differentiable at x, then

$$D^+u(x) \;=\; D^-u(x) \;=\; \big\{\nabla u(x)\big\}\,. \tag{2.4}$$

(ii) If the sets $D^+u(x)$ and $D^-u(x)$ are both non-empty, then u is differentiable at x, hence (2.4) holds.

(iii) The sets of points where a one-sided differential exists:

$$\Omega^+ \doteq \big\{x \in \Omega;\;\; D^+u(x) \;\neq\; \emptyset\big\}, \qquad\qquad \Omega^- \doteq \big\{x \in \Omega;\;\; D^-u(x) \;\neq\; \emptyset\big\} \tag{2.5}$$

are both non-empty. Indeed, they are dense in Ω.

Proof. To prove (i), assume that u is differentiable at x. Trivially, $\nabla u(x) \in D^\pm u(x)$. On the other hand, if $\varphi \in \mathcal{C}^1(\Omega)$ is such that $u - \varphi$ has a local maximum at x, then $\nabla\varphi(x) = \nabla u(x)$. Hence $D^+u(x)$ cannot contain any vector other than $\nabla u(x)$.

To prove (ii), assume that the sets $D^+u(x)$ and $D^-u(x)$ are both non-empty. Then there we can find $\delta > 0$ and $\varphi_1, \varphi_2 \in \mathcal{C}^1(\Omega)$ such that (fig. 2.5)

$$\varphi_1(x) \;=\; u(x) \;=\; \varphi_2(x), \qquad\qquad \varphi_1(y) \;\leq\; u(y) \;\leq\; \varphi_2(y) \quad \text{whenever} \quad |y - x| < \delta.$$

By a standard comparison argument, this implies that u is differentiable at x and $\nabla u(x) = \nabla\varphi_1(x) = \nabla\varphi_2(x)$.

Concerning (iii), let $x_0 \in \Omega$ and $\varepsilon > 0$ be given. On the open ball $B(x_0, \varepsilon) \doteq \{x\,;\;\; |x - x_0| < \varepsilon\}$ centered at x_0 with radius ε, consider the smooth function (fig. 2.6)

$$\varphi(x) \;\doteq\; \frac{1}{\varepsilon^2 - |x - x_0|^2}\,.$$

Notice that $\varphi(x) \to +\infty$ as $|x - x_0| \to \rho-$. Therefore, the function $u - \varphi$ attains a global maximum at some interior point $y \in B(x_0, \varepsilon)$. By Lemma 2.2, the super-differential of u at y is non-empty. Indeed, $\nabla\varphi(y) = \frac{2(y-x_0)}{(\varepsilon^2 - |y-x_0|^2)^2} \in D^+u(x)$. The previous argument shows that, for every $x_0 \in \Omega$ and $\varepsilon > 0$, the set Ω^+ contains a point y such that $|y - x_0| < \varepsilon$. Therefore Ω^+ is dense in Ω. The case of sub-differentials is entirely similar. $\square$

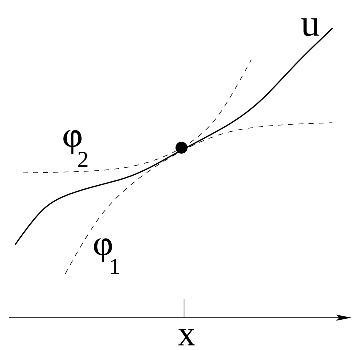

Figure 2.5

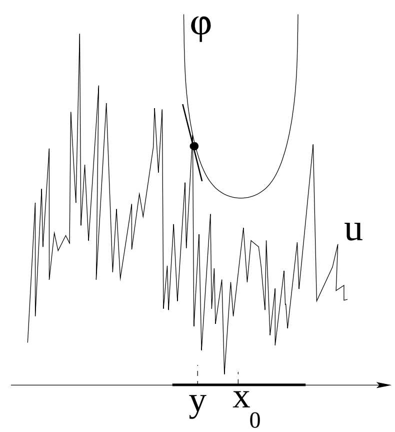

Figure 2.6

3 - Viscosity solutions

In the following, we consider the first order, partial differential equation

$$F\big(x,\, u(x),\, \nabla u(x)\big) \;=\; 0 \tag{3.1}$$

defined on an open set $\Omega \in I\!R^n$. Here $F : \Omega \times I\!R \times I\!R^n \mapsto I\!R$ is a continuous (nonlinear) function.

Definition 3.1. A function $u \in \mathcal{C}(\Omega)$ is a **viscosity subsolution** of (3.1) if

$$F\big(x, u(x), p\big) \;\leq\; 0 \qquad \text{for every } x \in \Omega,\;\; p \in D^+u(x). \tag{3.2}$$

Similarly, $u \in \mathcal{C}(\Omega)$ is a **viscosity supersolution** of (3.1) if

$$F\big(x, u(x), p\big) \;\geq\; 0 \qquad \text{for every } x \in \Omega,\;\; p \in D^-u(x). \tag{3.3}$$

We say that u is a **viscosity solution** of (3.1) if it is both a supersolution and a subsolution in the viscosity sense.

Similar definitions also apply to evolution equations of the form

$$u_t + H\big(t, x, u, \nabla u\big) = 0, \tag{3.4}$$

where ∇u denotes the gradient of u w.r.t. x. Recalling Lemma 2.2, we can reformulate these definitions in an equivalent form:

Definition 3.2. A function $u \in \mathcal{C}(\Omega)$ is a **viscosity subsolution** of (3.4) if, for every $\mathcal{C}^1$ function $\varphi = \varphi(t, x)$ such that $u - \varphi$ has a local maximum at (t, x), there holds

$$\varphi_t(t, x) + H(t, x, u, \nabla\varphi) \;\leq\; 0. \tag{3.5}$$

Similarly, $u \in \mathcal{C}(\Omega)$ is a **viscosity supersolution** of (3.4) if, for every $\mathcal{C}^1$ function $\varphi = \varphi(t, x)$ such that $u - \varphi$ has a local minimum at (t, x), there holds

$$\varphi_t(t, x) + H(t, x, u, \nabla\varphi) \;\geq\; 0. \tag{3.6}$$

Remark 3.3. In the definition of subsolution, we are imposing conditions on u only at points x where the super-differential is non-empty. Even if u is merely continuous, possibly nowhere differentiable, there are a lot of these points. Indeed, by Lemma 2.5, the set of points x where $D^+u(x) \neq \emptyset$ is dense on Ω. Similarly, for supersolutions we only impose conditions at points where $D^-u(x) \neq \emptyset$.

Remark 3.4. If u is a $\mathcal{C}^1$ function that satisfies (3.1) at every $x \in \Omega$, then u is also a solution in the viscosity sense. Viceversa, if u is a viscosity solution, then the equality (3.1) must hold at every point x where u is differentiable. In particular, if u is Lipschitz continuous, then by Rademacher's theorem it is a.e. differentiable. Hence (3.1) holds a.e. in Ω.

Remark 3.5. According to Definition 1.3, a Lipschitz continuous function is a *generalized solution* of (1.1) if the following implication holds true:

$$p \;=\; \nabla u(x) \qquad \Longrightarrow \qquad F(t, u(x), p) \;=\; 0\,. \tag{3.7}$$

This can be written in an equivalent way, splitting each equality into two inequalities:

$$\Big[p \in D^+u(x) \;\text{ and }\; p \in D^-u(x)\Big] \qquad \Longrightarrow \qquad \Big[F(t, u(x), p) \;\leq\; 0 \;\text{ and }\; F(t, u(x), p) \;\geq\; 0\Big]. \tag{3.8}$$

The definition of *viscosity solution*, on the other hand, requires two separate implications:

$$\begin{cases} p \in D^+u(x) \qquad \Longrightarrow \qquad F(t, u(x), p) \;\leq\; 0 \qquad (u \text{ is a viscosity subsolution})\,, \\ p \in D^-u(x) \qquad \Longrightarrow \qquad F(t, u(x), p) \;\geq\; 0 \qquad (u \text{ is a viscosity supersolution})\,. \end{cases} \tag{3.9}$$

Observe that, if $u(\cdot)$ satisfies the two implications in (3.9), then it also satisfies (3.8). In other words, if u is a viscosity solution, then u is also a generalized solution. However, the converse does not hold.

Example 3.6. Let $F(x,u,p) \doteq |p|-1$. Observe that the function $u(x)=1-|x|$ is a viscosity solution of

$$|u_x|-1 \;=\; 0 \tag{3.10}$$

on the open interval $]-1,\,1[$. Indeed, u is differentiable and satisfies the equation (3.10) at all points $x \neq 0$. Moreover, we have

$$D^+u(0) \;=\; [-1,\;1]\,, \qquad\qquad D^-u(0) \;=\; \emptyset. \tag{3.11}$$

To show that u is a supersolution, at the point $x=0$ there is nothing else to check. To show that u is a subsolution, take any $p \in [-1,\,1]$. Then $|p|-1 \leq 0$, as required.

It is interesting to observe that the same function $u(x)=1-|x|$ is NOT a viscosity solution of the equation

$$1-|u_x| \;=\; 0\,. \tag{3.12}$$

Indeed, at $x=0$, taking $p=0 \in D^+u(0)$ we find $1-|0|=1>0$. Since $D^-u(0)=\emptyset$, we conclude that the function $u(x)=1-|x|$ is a viscosity supersolution of (3.12), but not a subsolution.

4 - Stability properties

For nonlinear P.D.E's, the set of solutions may not be closed w.r.t. the topology of uniform convergence. In general, if $u_m \to u$ uniformly on a domain Ω, to conclude that u is itself a solution of the P.D.E. one should know, in addition, that all the derivatives $D^\alpha u_m$ that appear in the equation converge to the corresponding derivatives of u. This may not be the case in general.

Example 4.1. A sequence of generalized solutions to the equation

$$|u_x|-1 \;=\; 0\,, \qquad\qquad u(0) \;=\; u(1) \;=\; 0\,, \tag{4.1}$$

is provided by the saw-tooth functions (fig. 4.1)

$$u_m(x) \doteq \begin{cases} x-\dfrac{k-1}{m} & \text{if} \quad x \in \left[\dfrac{k-1}{m}\,,\;\; \dfrac{k-1}{m}+\dfrac{1}{2m}\right], \\[2ex] \dfrac{k}{m}-x & \text{if} \quad x \in \left[\dfrac{k}{m}-\dfrac{1}{2m}\,,\;\; \dfrac{k}{m}\right], \end{cases} \qquad k=1,\ldots,m\,. \tag{4.2}$$

Clearly $u_m \to 0$ uniformly on $[0,1]$, but the zero function is not a solution of (4.1). In this case, the convergence of the functions u_m is not accompanied by the convergence of their derivatives.

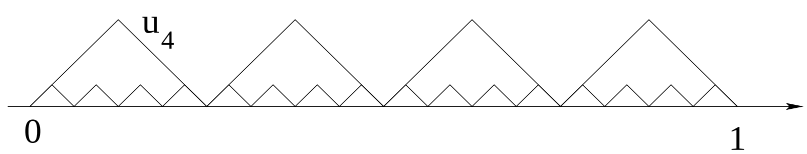

figure 4.1

The next lemma shows that, in the case of viscosity solutions, a general stability theorem holds, without any requirement about the convergence of derivatives.

Lemma 4.2. *Consider a sequence of continuous functions u_m, which provide viscosity sub-solutions (super-solutions) to*

$$F_m(x, u_m, \nabla u_m) = 0 \qquad x \in \Omega. \tag{4.3}$$

As $m \to \infty$, assume that $F_m \to F$ uniformly on compact subsets of $\Omega \times I\!R \times I\!R^n$ and $u_m \to u$ in $\mathcal{C}(\Omega)$. Then u is a subsolution (a supersolution) of (3.1).

Proof. To prove that u is a subsolution, let $\phi \in \mathcal{C}^1$ be such that $u - \phi$ has a strict local maximum at a point x. We need to show that

$$F\big(x, u(x), \nabla\phi(x)\big) \le 0. \tag{4.4}$$

By Lemma 2.4, there exists a sequence $x_m \to x$ such that $u_m - \phi$ has a local maximum at x_m, and $u_m(x_m) \to u(x)$ as $m \to \infty$. Since u_m is a subsolution,

$$F_m\big(x_m, u_m(x_m), \nabla\phi(x_m)\big) \le 0. \tag{4.5}$$

Letting $m \to \infty$ in (4.5), by continuity we obtain (4.4). $\square$

The above result should be compared with Example 4.1. Clearly, the functions u_m in (4.2) are *not* viscosity solutions.

The definition of viscosity solution is naturally motivated by the properties of vanishing viscosity limits.

Theorem 4.3. *Let u_ε be a family of smooth solutions to the viscous equation*

$$F\big(x,\, u_\varepsilon(x),\, \nabla u_\varepsilon(x)\big) = \varepsilon\, \Delta u_\varepsilon. \tag{4.6}$$

Assume that, as $\varepsilon \to 0+$, we have the convergence $u_\varepsilon \to u$ uniformly on an open set $\Omega \subseteq I\!R^n$. Then u is a viscosity solution of (3.1).

Proof. Fix $x \in \Omega$ and assume $p \in D^+u(x)$. To prove that u is a subsolution we need to show that $F(x,\, u(x),\, p) \le 0$.

1. By Lemma 2.2 and Remark 2.3, there exists $\varphi \in \mathcal{C}^1$ with $\nabla\varphi(x) = p$, such that $u - \varphi$ has a strict local maximum at x. For any $\delta > 0$ we can then find $0 < \rho \le \delta$ and a function $\psi \in \mathcal{C}^2$ such that

$$\big|\nabla\varphi(y) - \nabla\varphi(x)\big| \le \delta \qquad \text{if} \quad |y - x| \le \rho, \tag{4.7}$$

$$\|\psi - \varphi\|_{\mathcal{C}^1} \le \delta \tag{4.8}$$

and such that each function $u_\varepsilon - \psi$ has a local maximum inside the ball $B(x;\, \rho)$, for $\varepsilon > 0$ small enough.

2. Let x_ε be the location of this local maximum of $u_\varepsilon - \psi$. Since u_ε is smooth, this implies

$$\nabla\psi(x_\varepsilon) = \nabla u(x_\varepsilon), \qquad \Delta u(x_\varepsilon) \le \Delta\psi(x_\varepsilon). \tag{4.9}$$

Hence from (4.6) it follows

$$F\big(x,\, u_\varepsilon(x_\varepsilon),\;\; \nabla\psi(x_\varepsilon)\big) \;\le\; \varepsilon\,\Delta\psi(x_\varepsilon). \tag{4.10}$$

3. We can now select a sequence $\varepsilon_\nu \to 0+$ such that $\lim_{\nu\to\infty} x_{\varepsilon_\nu} = \tilde{x}$ for some limit point $\tilde{x}$. By the construction performed in step **1**, one has $|\tilde{x} - x| \le \rho$. Since $\psi \in \mathcal{C}^2$, we can pass to the limit in (4.10) and conclude

$$F\big(x,\, u(\tilde{x}),\;\; \nabla\psi(\tilde{x})\big) \;\le\; 0 \tag{4.11}$$

By (4.7)-(4.8) we have

$$\begin{aligned} \big|\nabla\psi(\tilde{x}) - p\big| \;&\le\; \big|\nabla\psi(\tilde{x}) - \nabla\varphi(\tilde{x})\big| + \big|\nabla\varphi(\tilde{x}) - \nabla\varphi(x)\big| \\ &\le\; \delta + \delta\,. \end{aligned} \tag{4.12}$$

Since $\delta > 0$ can be taken arbitrarily small, (4.11) and the continuity of F imply $F(x, u(x), p) \le 0$, showing that u is a subsolution. The fact that u is a supersolution is proved in an entirely similar way. □

Remark 4.4. In the light of the above result, it should not come as a surprise that the two equations

$$F(x, u, \nabla u) \;=\; 0 \qquad\qquad \text{and} \qquad\qquad -\,F(x, u, \nabla u) \;=\; 0$$

may have different viscosity solutions. Indeed, solutions to the first equation are obtained as limits of (4.6) as $\varepsilon \to 0+$, while solutions to the second equation are obtained as limits of (4.6) as $\varepsilon \to 0-$. These two limits can be substantially different.

5 - Comparison theorems

A remarkable feature of the notion of viscosity solutions is that on one hand it requires a minimum amount of regularity (just continuity), and on the other hand it is stringent enough to yield general comparison and uniqueness theorems.

The uniqueness proofs are based on a technique of doubling of variables, which reminds of Kruzhkov's uniqueness theorem for conservation laws [K]. We now illustrate this basic technique in a simple setting.

Theorem 5.1 (Comparison). *Let $\Omega \subset I\!R^n$ be a bounded open set. Let $u_1, u_2 \in \mathcal{C}(\overline{\Omega})$ be, respectively, viscosity sub- and supersolutions of*

$$u + H(x, \nabla u) \;=\; 0 \qquad\qquad x \in \Omega\,. \tag{5.1}$$

Assume that

$$u_1(x) \;\le\; u_2(x) \qquad\qquad \textit{for all} \;\; x \in \partial\Omega. \tag{5.2}$$

Moreover, assume that $H : \Omega \times I\!R^n \mapsto I\!R$ is uniformly continuous in the x-variable:

$$\big|H(x,p) - H(y,p)\big| \;\le\; \omega\Big(|x-y|\big(1+|p|\big)\Big), \tag{5.3}$$

for some continuous and non-decreasing function $\omega : [0,\infty[\mapsto [0,\infty[$ *with* $\omega(0) = 0$*. Then*

$$u_1(x) \;\le\; u_2(x) \qquad\qquad \textit{for all} \;\; x \in \overline{\Omega}. \tag{5.4}$$

Proof. To appreciate the main idea of the proof, consider first the case where u_1, u_2 are smooth. If the conclusion (5.4) fails, then the difference $u_1 - u_2$ attains a positive maximum at a point $x_0 \in \Omega$. This implies $p \doteq \nabla u_1(x_0) = \nabla u_2(x_0)$. By definition of sub- and supersolution, we now have

$$\begin{aligned} u_1(x_0) + H(x_0, p) \;&\le\; 0, \\ u_2(x_0) + H(x_0, p) \;&\ge\; 0. \end{aligned} \tag{5.5}$$

Subtracting the second from the first inequality in (5.5) we conclude $u_1(x_0) - u_2(x_0) \le 0$, reaching a contradiction.

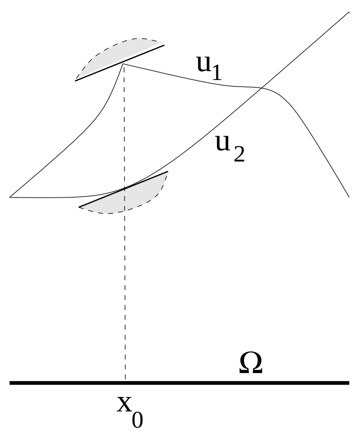

Figure 5.1a

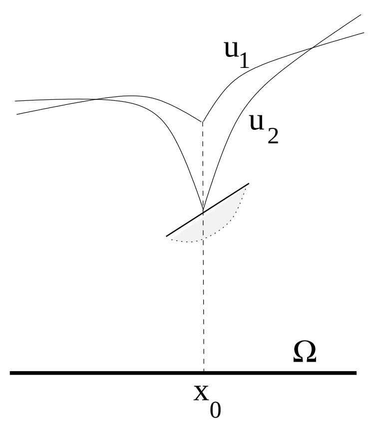

Figure 5.1b

Next, consider the non-smooth case. We can repeat the above argument and reach again a contradiction provided that we can find a point x_0 such that (fig. 5.1a)

(i) $u_1(x_0) > u_2(x_0)$,

(ii) some vector p lies at the same time in the upper differential $D^+u_1(x_0)$ and in the lower differential $D^-u_2(x_0)$.

A natural candidate for x_0 is a point where $u_1 - u_2$ attains a global maximum. Unfortunately, at such point one of the sets $D^+u_1(x_0)$ or $D^-u_2(x_0)$ may be empty, and the argument breaks down (fig. 5.1b). To proceed further, the key observation is that we do not need to compare values of u_1 and u_2 at exactly the same point. Indeed, to reach a contradiction, it suffices to find nearby points x_ε and y_ε such that (fig. 5.2)

(i′) $u_1(x_\varepsilon) > u_2(y_\varepsilon)$,

(ii′) some vector p lies at the same time in the upper differential $D^+u_1(x_\varepsilon)$ and in the lower differential $D^-u_2(y_\varepsilon)$.

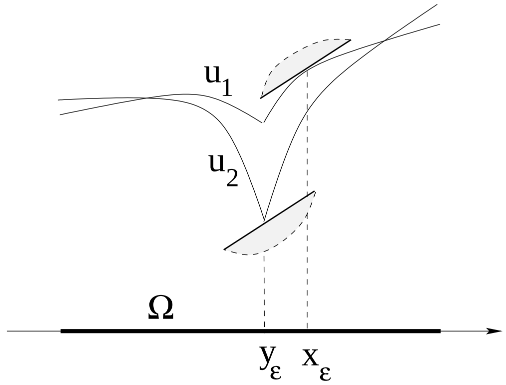

figure 5.2

Can we always find such points? It is here that the variable-doubling technique comes in. The key idea is to look at the function of two variables

$$\Phi_\varepsilon(x,y) \ \doteq\ u_1(x)-u_2(y)-\frac{|x-y|^2}{2\varepsilon}\,. \tag{5.6}$$

This clearly admits a global maximum over the compact set $\overline{\Omega}\times\overline{\Omega}$. If $u_1>u_2$ at some point x_0, this maximum will be strictly positive. Moreover, taking $\varepsilon>0$ sufficiently small, the boundary conditions imply that the maximum is attained at some interior point $(x_\varepsilon,y_\varepsilon)\in\Omega\times\Omega$. Notice that the points $x_\varepsilon,y_\varepsilon$ must be close to each other, otherwise the penalization term in (5.6) will be very large and negative.

We now observe that the function of a single variable

$$x\ \mapsto\ u_1(x)-\left(u_2(y_\varepsilon)+\frac{|x-y_\varepsilon|^2}{2\varepsilon}\right)\ =\ u_1(x)-\varphi_1(x) \tag{5.7}$$

attains its maximum at the point x_ε. Hence by Lemma 2.2

$$\frac{x_\varepsilon-y_\varepsilon}{\varepsilon}\ =\ \nabla\varphi_1(x_\varepsilon)\ \in\ D^+u_1(x_\varepsilon).$$

Moreover, the function of a single variable

$$y\ \mapsto\ u_2(y)-\left(u_1(x_\varepsilon)-\frac{|x_\varepsilon-y|^2}{2\varepsilon}\right)\ =\ u_2(y)-\varphi_2(y) \tag{5.8}$$

attains its minimum at the point y_ε. Hence

$$\frac{x_\varepsilon-y_\varepsilon}{\varepsilon}\ =\ \nabla\varphi_2(y_\varepsilon)\ \in\ D^-u_2(y_\varepsilon).$$

We have thus discovered two points x_ε, y_ε and a vector $p=(x_\varepsilon-y_\varepsilon)/\varepsilon$ which satisfy the conditions (i′)-(ii′).

We now work out the details of the proof, in several steps.

1. If the conclusion fails, then there exists $x_0 \in \Omega$ such that

$$u_1(x_0) - u_2(x_0) \;=\; \max_{x \in \overline{\Omega}} \big\{u_1(x) - u_2(x)\big\} \;\doteq\; \delta \;>\; 0. \tag{5.9}$$

For $\varepsilon > 0$, call $(x_\varepsilon, y_\varepsilon)$ a point where the function Φ_ε in (5.6) attains its global maximum on the compact set $\overline{\Omega} \times \overline{\Omega}$. By (5.9) one has

$$\Phi_\varepsilon(x_\varepsilon, y_\varepsilon) \;\geq\; \delta \;>\; 0. \tag{5.10}$$

2. Call M an upper bound for all values $\big|u_1(x)\big|$, $\big|u_2(x)\big|$, as $x \in \overline{\Omega}$. Then

$$\Phi_\varepsilon(x,y) \;\leq\; 2M - \frac{|x-y|^2}{2\varepsilon}\,,$$

$$\Phi_\varepsilon(x,y) \;\leq\; 0 \qquad \text{if} \quad |x-y|^2 \;\geq\; M\varepsilon.$$

Hence (5.10) implies

$$|x_\varepsilon - y_\varepsilon| \;\leq\; \sqrt{M\varepsilon}\,. \tag{5.11}$$

3. By the uniform continuity of the functions u_2 on the compact set $\overline{\Omega}$, for $\varepsilon' > 0$ sufficiently small we have

$$\big|u_2(x) - u_2(y)\big| \;<\; \frac{\delta}{2} \qquad\qquad \text{whenever} \quad |x-y| \;\leq\; \sqrt{M\varepsilon'}\,. \tag{5.12}$$

We now show that, choosing $\varepsilon < \varepsilon'$, the points $x_\varepsilon, y_\varepsilon$ cannot lie on the boundary of Ω. For example, if $x_\varepsilon \in \partial\Omega$, then by (5.2) and (5.11)-(5.12) it follows

$$\begin{aligned}\Phi_\varepsilon(x_\varepsilon, y_\varepsilon) \;&\leq\; \big(u_1(x_\varepsilon) - u_2(x_\varepsilon)\big) + \big|u_2(x_\varepsilon) - u_2(y_\varepsilon)\big| - \frac{|x_\varepsilon - y_\varepsilon|^2}{2\varepsilon}\\ &\leq\; 0 + \delta/2 + 0,\end{aligned}$$

in contradiction with (5.10).

4. Having shown that $x_\varepsilon, y_\varepsilon$ are interior points, we consider the functions of one single variable φ_1, φ_2 defined at (5.7)-(5.8). Since x_ε provides a local maximum for $u_1 - \varphi_1$ and y_ε provides a local minimum for $u_2 - \varphi_2$, we conclude that

$$p_\varepsilon \;\doteq\; \frac{x_\varepsilon - y_\varepsilon}{\varepsilon} \;\in\; D^+u_1(x_\varepsilon) \cap D^-u_2(y_\varepsilon). \tag{5.13}$$

From the definition of viscosity sub- and supersolution we now obtain

$$\begin{aligned}u_1(x_\varepsilon) + H(x_\varepsilon, p_\varepsilon) \;&\leq\; 0,\\ u_2(y_\varepsilon) + H(y_\varepsilon, p_\varepsilon) \;&\geq\; 0.\end{aligned} \tag{5.14}$$

5. From

$$u_1(x_\varepsilon) - u_2(x_\varepsilon) \;\leq\; \Phi_\varepsilon(x_\varepsilon, y_\varepsilon) \;\leq\; u_1(x_\varepsilon) - u_2(x_\varepsilon) + \big|u_2(x_\varepsilon) - u_2(y_\varepsilon)\big| - \frac{|x_\varepsilon - y_\varepsilon|^2}{2\varepsilon}$$

it follows

$$\big|u_2(x_\varepsilon)-u_2(y_\varepsilon)\big| - \frac{|x_\varepsilon-y_\varepsilon|^2}{2\varepsilon} \;\geq\; 0\,.$$

Using (5.11) and the uniform continuity of u_2, we thus obtain

$$\limsup_{\varepsilon\to 0+} \frac{|x_\varepsilon-y_\varepsilon|^2}{2\varepsilon} \;\leq\; \limsup_{\varepsilon\to 0+} \big|u_2(x_\varepsilon)-u_2(y_\varepsilon)\big| \;=\; 0\,. \tag{5.15}$$

6. By (5.10), subtracting the second from the first inequality in (5.14) and using (5.3), we obtain

$$\begin{aligned} \delta \;&\leq\; \Phi_\varepsilon(x_\varepsilon,y_\varepsilon) \\ &\leq\; u_1(x_\varepsilon)-u_2(y_\varepsilon) \\ &\leq\; \big|H(x_\varepsilon,p_\varepsilon)-H(y_\varepsilon,p_\varepsilon)\big| \\ &\leq\; \omega\Big((|x_\varepsilon-y_\varepsilon|\cdot\big(1+\varepsilon^{-1}|x_\varepsilon-y_\varepsilon|\big)\Big). \end{aligned} \tag{5.16}$$

This yields a contradiction, Indeed, by (5.15) the right hand side of (5.16) becomes arbitrarily small as $\varepsilon\to 0$. □

An easy consequence of the above result is the uniqueness of solutions to the boundary value problem

$$u+H(x,\nabla u) \;=\; 0 \qquad\qquad x\in\Omega, \tag{5.17}$$

$$u \;=\; \psi \qquad\qquad x\in\partial\Omega. \tag{5.18}$$

Corollary 5.2 (Uniqueness). *Let $\Omega\subset I\!R^n$ be a bounded open set. Let the Hamiltonian function H satisfy the equicontinuity assumption (5.3). Then the boundary value problem (5.17)-(5.18) admits at most one viscosity solution.*

Proof. Let u_1,u_2 be viscosity solutions. Since u_1 is a subsolution and u_2 is a supersolution, and $u_1=u_2$ on $\partial\Omega$, by Theorem 5.1 we conclude $u_1\leq u_2$ on $\overline{\Omega}$. Interchanging the roles of u_1 and u_2 one obtains $u_2\leq u_1$, completing the proof. □

By similar techniques, comparison and uniqueness results can be proved also for Hamilton-Jacobi equations of evolutionary type. Consider the Cauchy problem

$$u_t+H(t,x,\nabla u) \;=\; 0 \qquad\qquad (t,x)\in\,]0,T[\,\times I\!R^n, \tag{5.19}$$

$$u(0,x) \;=\; \bar u(x) \qquad\qquad x\in I\!R^n. \tag{5.20}$$

Here and in the sequel, it is understood that $\nabla u\doteq(u_{x_1},\ldots,u_{x_n})$ always refers to the gradient of u w.r.t. the space variables.

Theorem 5.3 (Comparison). *Let the function $H:[0,T]\times I\!R^n\times I\!R^n$ satisfy the Lipschitz continuity assumptions*

$$\big|H(t,x,p)-H(s,y,p)\big| \;\leq\; C\big(|t-s|+|x-y|\big)\big(1+|p|\big), \tag{5.21}$$

$$\big|H(t,x,p)-H(t,x,q)\big| \;\leq\; C\,|p-q|\,. \tag{5.22}$$

Let u,v be bounded, uniformly continuous sub- and super-solutions of (5.19) respectively. If $u(0,x)\leq v(0,x)$ for all $x\in I\!R^n$, then

$$u(t,x) \;\leq\; v(t,x) \qquad\qquad \textit{for all}\;\; (t,x)\in[0,T]\times I\!R^n. \tag{5.23}$$

Toward this result, as a preliminary we prove

Lemma 5.4. *Let u be a continuous function on $[0,T]\times I\!R^n$, which provides a subsolution of (5.19) for $t\in\,]0,T[$. If $\phi\in\mathcal{C}^1$ is such that $u-\phi$ attains a local maximum at a point (T,x_0), then*

$$\phi_t(T,x_0)+H\big(T,x_0,\nabla\phi(T,x_0)\big) \;\leq\; 0. \tag{5.24}$$

Proof. We can assume that (T,x_0) is a point of strict local maximum for $u-\phi$. For each $\varepsilon>0$ consider the function

$$\phi_\varepsilon(t,x) \;=\; \phi(t,x)+\frac{\varepsilon}{T-t}\,.$$

Each function $u-\phi_\varepsilon$ will then have a local maximum at a point $(t_\varepsilon,x_\varepsilon)$, with

$$t_\varepsilon \;<\; T, \qquad\qquad (t_\varepsilon,x_\varepsilon)\to(T,x_0) \quad \text{as}\;\; \varepsilon\to 0+\,.$$

Since u is a subsolution, one has

$$\phi_{\varepsilon,t}(t_\varepsilon,x_\varepsilon)+H\big(t_\varepsilon,x_\varepsilon,\nabla\phi_\varepsilon(t_\varepsilon,x_\varepsilon)\big) \;\leq\; 0\,,$$

$$\phi_t(t_\varepsilon,x_\varepsilon)+H\big(t_\varepsilon,x_\varepsilon,\nabla\phi(t_\varepsilon,x_\varepsilon)\big) \;\leq\; -\frac{\varepsilon}{(T-t_\varepsilon)^2}\,. \tag{5.25}$$

Letting $\varepsilon\to 0+$, from (5.25) we obtain (5.24). □

Proof of Theorem 5.3.

1. If (5.23) fails, then we can find $\lambda>0$ such that

$$\sup_{t,x}\Big\{u(t,x)-v(t,x)-2\lambda t\Big\} \;\doteq\; \sigma \;>\; 0. \tag{5.26}$$

Assume that the supremum in (5.26) is actually attained at a point (t_0,x_0), possibly with $t_0=T$. If both u and u are differentiable at such point, we easily obtain a contradiction, because

$$u_t(t_0,x_0)+H(t_0,x_0,\nabla u) \;\leq\; 0\,,$$
$$v_t(t_0,x_0)+H(t_0,x_0,\nabla v) \;\geq\; 0\,,$$

$$\nabla u(t_0,x_0) \;=\; \nabla v(t_0,x_0), \qquad\qquad u_t(t_0,x_0)-v_t(t_0,x_0)-2\lambda \;\geq\; 0\,.$$

2. To extend the above argument to the general case, we face two technical difficulties. First, the function in (5.26) may not attain its global maximum over the unbounded set $[0,T]\times I\!R^n$.

Moreover, at this point of maximum the functions u, v may not be differentiable. These problems are overcome by inserting a penalization term, and doubling the variables. As in the proof of Theorem 5.1, we introduce the function

$$\Phi_\varepsilon(t,x,s,y) \doteq u(t,x) - v(s,y) - \lambda(t+s) - \varepsilon\big(|x|^2+|y|^2\big) - \frac{1}{\varepsilon^2}\big(|t-s|^2+|x-y|^2\big).$$

Thanks to the penalization terms, the function Φ_ε clearly admits a global maximum at a point $(t_\varepsilon, x_\varepsilon, s_\varepsilon, y_\varepsilon) \in \big(]0,T] \times I\!\!R^n\big)^2$. Choosing $\varepsilon > 0$ sufficiently small, one has

$$\Phi_\varepsilon(t_\varepsilon, x_\varepsilon, s_\varepsilon, y_\varepsilon) \geq \max_{t,x} \Phi_\varepsilon(t,x,t,x) \geq \sigma/2.$$

3. We now observe that the function

$$(t,x) \mapsto u(t,x) - \Big[v(s_\varepsilon, y_\varepsilon) + \lambda(t+s_\varepsilon) + \varepsilon\big(|x|^2+|y_\varepsilon|^2\big) + \frac{1}{\varepsilon^2}\big(|t-s_\varepsilon|^2+|x-y_\varepsilon|^2\big)\Big] \doteq u(t,x) - \phi(t,x)$$

takes a maximum at the point $(t_\varepsilon, x_\varepsilon)$. Since u is a subsolution and ϕ is smooth, this implies

$$\lambda + \frac{2(t_\varepsilon - s_\varepsilon)}{\varepsilon^2} + H\left(t_\varepsilon,\, x_\varepsilon,\, \frac{2(x_\varepsilon - y_\varepsilon)}{\varepsilon^2} + 2\varepsilon x_\varepsilon\right) \leq 0. \tag{5.27}$$

Notice that, in the case where $t_\varepsilon = T$, (5.27) follows from Lemma 5.5.

Similarly, the function

$$(s,y) \mapsto v(s,y) - \Big[u(t_\varepsilon, x_\varepsilon) - \lambda(t_\varepsilon + s) - \varepsilon\big(|x_\varepsilon|^2+|y|^2\big) - \frac{1}{\varepsilon^2}\big(|t_\varepsilon - s|^2+|x_\varepsilon - y|^2\big)\Big] \doteq v(s,y) - \psi(s,y)$$

takes a maximum at the point $(t_\varepsilon, x_\varepsilon)$. Since v is a supersolution and ψ is smooth, this implies

$$-\lambda + \frac{2(t_\varepsilon - s_\varepsilon)}{\varepsilon^2} + H\left(s_\varepsilon,\, y_\varepsilon,\, \frac{2(x_\varepsilon - y_\varepsilon)}{\varepsilon^2} - 2\varepsilon y_\varepsilon\right) \geq 0. \tag{5.28}$$

4. Subtracting (5.28) from (5.27) and using (5.21)-(5.22) we obtain

$$\begin{aligned} 2\lambda &\leq H\left(s_\varepsilon,\, y_\varepsilon,\, \frac{2(x_\varepsilon - y_\varepsilon)}{\varepsilon^2} - 2\varepsilon y_\varepsilon\right) - H\left(t_\varepsilon,\, x_\varepsilon,\, \frac{2(x_\varepsilon - y_\varepsilon)}{\varepsilon^2} + 2\varepsilon x_\varepsilon\right) \\ &\leq C\varepsilon\big(|x_\varepsilon| + |y_\varepsilon|\big) + C\big(|t_\varepsilon - s_\varepsilon| + |x_\varepsilon - y_\varepsilon|\big)\left(1 + \frac{|x_\varepsilon - y_\varepsilon|}{\varepsilon^2} + \varepsilon\big(|x_\varepsilon| + |y_\varepsilon|\big)\right). \end{aligned} \tag{5.29}$$

To reach a contradiction we need to show that the right hand side of (5.29) approaches zero as $\varepsilon \to 0$.

5. Since u, v are globally bounded, the penalization terms must satisfy uniform bounds, independent of ε. Hence

$$|x_\varepsilon|,\, |y_\varepsilon| \leq \frac{C'}{\sqrt{\varepsilon}}, \qquad |t_\varepsilon - s_\varepsilon|,\, |x_\varepsilon - y_\varepsilon| \leq C'\varepsilon \tag{5.30}$$

for some constant C'. This implies

$$\varepsilon\big(|x_\varepsilon| + |y_\varepsilon|\big) \leq 2C'\sqrt{\varepsilon}. \tag{5.31}$$

To obtain a sharper estimate, we now observe that $\Phi_\varepsilon(t_\varepsilon, x_\varepsilon, s_\varepsilon, y_\varepsilon) \geq \Phi_\varepsilon(t_\varepsilon, x_\varepsilon, t_\varepsilon, x_\varepsilon)$, hence

$$\begin{aligned} u(t_\varepsilon, x_\varepsilon) - v(s_\varepsilon, y_\varepsilon) - &\lambda(t_\varepsilon + s_\varepsilon) - \varepsilon\big(|x_\varepsilon|^2 + |y_\varepsilon|^2\big) - \frac{1}{\varepsilon^2}\big(|t_\varepsilon - s_\varepsilon|^2 + |x_\varepsilon - y_\varepsilon|^2\big) \\ &\geq\ u(t_\varepsilon, x_\varepsilon) - v(t_\varepsilon, x_\varepsilon) - 2\lambda t_\varepsilon - 2\varepsilon|x_\varepsilon|^2, \end{aligned}$$

$$\frac{1}{\varepsilon^2}\big(|t_\varepsilon - s_\varepsilon|^2 + |x_\varepsilon - y_\varepsilon|^2\big) \ \leq\ v(t_\varepsilon, x_\varepsilon) - v(s_\varepsilon, y_\varepsilon) + \lambda(t_\varepsilon - s_\varepsilon) + \varepsilon\big(|x_\varepsilon|^2 - |y_\varepsilon|^2\big). \tag{5.32}$$

By the uniform continuity of v, the right hand side of (5.32) tends to zero as $\varepsilon \to 0$. Therefore

$$\frac{|t_\varepsilon - s_\varepsilon|^2 + |x_\varepsilon - y_\varepsilon|^2}{\varepsilon^2} \to 0 \qquad \text{as} \quad \varepsilon \to 0. \tag{5.33}$$

By (5.30), (5.31) and (5.33), the right hand side of (5.29) also approaches zero, This yields the desired contradiction. □

Corollary 5.5 (Uniqueness). *Let the function H satisfy the assumptions (5.21)-(5.22). Then the Cauchy problem (5.19)-(5.20) admits at most one bounded, uniformly continuous viscosity solution $u : [0, T] \times I\!R^n \mapsto I\!R$.*

6 - Control systems

The time evolution of a system, whose state is described by a finite number of parameters, can be usually modelled by an O.D.E.

$$\dot{x} \ =\ f(x) \qquad\qquad x \in I\!R^n.$$

Here and in the sequel the upper dot denotes a derivative w.r.t. time. In some cases, the system can be influenced also by the external input of a controller. An appropriate model is then provided by a **control system**, having the form

$$\dot{x} \ =\ f(x, u). \tag{6.1}$$

Here $x \in I\!R^n$, while the control $u : [0, T] \mapsto U$ is required to take values inside a given set $U \subseteq I\!R^m$. We denote by

$$\mathcal{U} \doteq \Big\{u : I\!R \mapsto I\!R^m \ \ \text{measurable}, \quad u(t) \in U \ \ \text{for a.e. } t\Big\}$$

the set of **admissible control functions**. To guarantee local existence and uniqueness of solutions, it is natural to assume that the map $f : I\!R^n \times I\!R^m \mapsto I\!R^n$ is Lipschitz continuous w.r.t. x and continuous w.r.t. u. The solution of the Cauchy problem (6.1) with initial condition

$$x(t_0) \ =\ x_0 \tag{6.2}$$

will be denoted as $t \mapsto x(t;\, t_0, x_0, u)$. It is clear that, as u ranges over the whole set of control functions, one obtains a family of possible trajectories for the system. Equivalently, these trajectories can be characterized as the solutions to the **differential inclusion**

$$\dot{x} \ \in\ F(x)\,, \qquad\qquad F(x) \ \doteq\ \Big\{f(x, \omega)\,; \quad \omega \in U\Big\}. \tag{6.3}$$

Example 6.1. Call $x(t) \in I\!R^2$ the position of a boat on a river, and let $\mathbf{v}(x)$ be the velocity of the water at the point x. If the boat simply drifts along with the currrent, its position is described by the differential equation

$$\dot{x} \;=\; \mathbf{v}(x).$$

If we assume that the boat is powered by an engine, and can move in any direction with speed $\leq \rho$ (relative to the water), the evolution can be modelled by the control system

$$\dot{x} \;=\; f(x,u) \;=\; \mathbf{v}(x) + u\,, \qquad\qquad |u| \leq \rho\,.$$

This is equivalent to a differential inclusion where the sets of velocities are balls with radius ρ (fig. 6.1):

$$\dot{x} \;\in\; F(x) \;=\; B\big(\mathbf{v}(x);\, \rho\big).$$

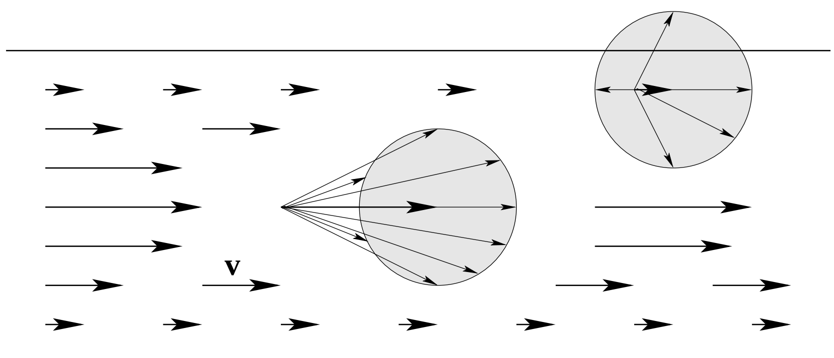

Figure 6.1. The possible velocities of a boat on a river.

Example 6.2 (cart on a rail). Consider a cart which can move without friction along a straight rail (Figure 6.1). For simplicity, assume that it has unit mass. Let $y(0) = \bar{y}$ be its initial position and $\dot{y}(0) = \bar{v}$ be its initial velocity. If no forces are present, its future position is simply given by

$$y(t) \;=\; \bar{y} + t\,\bar{v}\,.$$

Next, assume that a controller is able to push the cart, with an external force $u = u(t)$. The evolution of the system is then determined by the second order equation

$$\ddot{y}(t) \;=\; u(t)\,. \tag{6.4}$$

Calling $x_1(t) = y(t)$ and $x_2(t) = \dot{y}(t)$ respectively the position and the velocity of the cart at time t, the equation (6.4) can be written as a first order control system:

$$(\dot{x}_1, \dot{x}_2) \;=\; (x_2, u). \tag{6.5}$$

Given the initial condition $x_1(0) = \bar{y}$, $x_2(0) = \bar{v}$, the solution of (6.5) is provided by

$$\begin{aligned} x_1(t) &= \bar{y} + \bar{v}t + \int_0^t (t-s)u(s)\,ds, \\ x_2(t) &= \bar{v} + \int_0^t u(s)\,ds. \end{aligned}$$

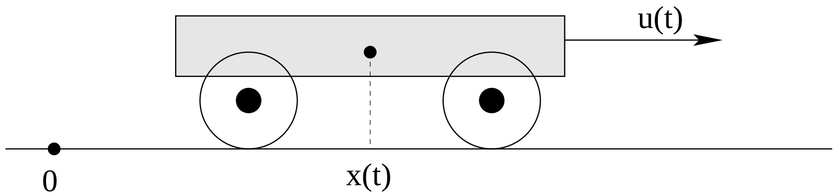

Figure 6.2. A cart moving along a straight, frictionless rail.

Assuming that the force satisfies the constraint

$$\big|u(t)\big| \ \leq \ 1\,,$$

the control system (6.5) is equivalent to the differential inclusion

$$(\dot x_1, \dot x_2) \ \in \ F(x_1,x_2) \ = \ \{(x_2, \omega)\,; \quad -1\leq \omega\leq 1\}\ .$$

We now consider the problem of steering the system to the origin $0\in I\!\!R^2$. In other words, we want the cart to be at the origin with zero speed. For example, if the initial condition is $(\bar y, \bar v) = (2,2)$, this goal is achieved by the open-loop control

$$\tilde u(t) = \begin{cases} -1 & \text{if} \ \ 0\leq t<4,\\ \ \ 1 & \text{if} \ \ 4\leq t<6,\\ \ \ 0 & \text{if} \ \ t\geq 6.\end{cases}$$

A direct computation shows that $(x_1(t), x_2(t)) = (0,0)$ for $t\geq 6$. Notice, however, that the above control would not accomplish the same task in connection with any other initial data $(\bar y, \bar v)$ different from $(2,2)$. This is a consequence of the backward uniqueness of solutions to the differential equation (6.5).

A related problem is that of asymptotic stabilization. In this case, we seek a feedback control function $u = u(x_1,x_2)$ such that, for every initial data $(\bar y,\bar v)$, the corresponding solution of the Cauchy problem

$$(\dot x_1, \dot x_2) \ = \ (x_2\,, \ u(x_1,x_2))\,, \qquad (x_1,x_2)(0) \ = \ (\bar y, \bar v)$$

approaches the origin as $t\to\infty$, i.e.

$$\lim_{t\to\infty} (x_1,x_2)(t) = (0,0).$$

There are several feedback controls which accomplish this task. For example, one can take $u(x_1,x_2) = -x_1-x_2$.

The above example is a special case of a control system having the form

$$\dot x \ = \ f(x) + g(x)\,u\,, \qquad\qquad u\in[-1,1]$$

where f,g are vector fields on $I\!\!R^n$. This is equivalent to a differential inclusion

$$\dot x \ \in \ F(x) \ = \ \big\{ f(x) + g(x)\,u\ ; \qquad u\in[-1,1]\big\}\,,$$

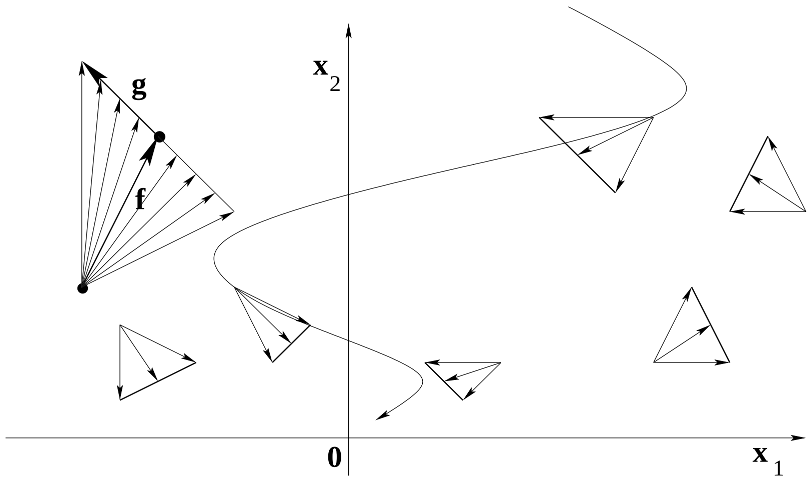

Figure 6.3. Velocity sets for a planar system which is linear w.r.t. the control.

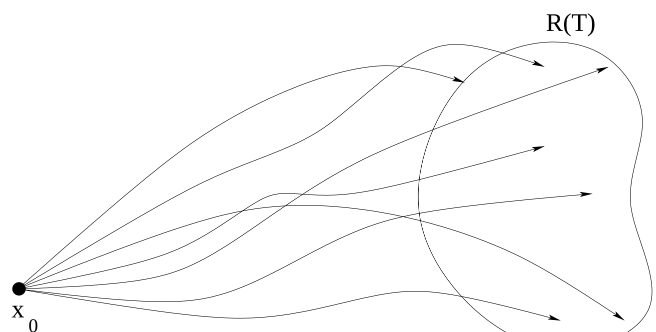

Figure 6.4

where each set $F(x)$ of possible velocities is a segment (fig. 6.3). Systems of this form, linear w.r.t. the control variable u, have been extensively studied using techniques from differential geometry.

Given a control system in the general form (6.1), the **reachable set** at time T starting from x_0 at time t_0 (fig. 6.4) will be denoted by

$$R(T) \doteq \Big\{x(T;\, t_0, x_0, u)\,; \quad u \in \mathcal{U}\Big\}. \tag{6.4}$$

The control u can be assigned as an **open loop control**, as a function of time: $t \mapsto u(t)$, or as a **feedback control**, as a function of the state: $x \mapsto u(x)$.

Among the major issues that one can study in connection with the control system (6.1) are the following.

1 - Dynamics. Starting from a point x_0, describe the set of all possible trajectories. Study the properties of the reachable set $R(T)$. In particular, one may determine whether $R(T)$ is closed, bounded, convex, with non-empty interior, etc…

2 - Stabilization. For each initial state x_0, find a control $u(\cdot)$ that steers the system toward the origin, so that

$$x(t; 0, x_0, u) \;\to\; 0 \qquad\qquad \text{as} \quad t \;\to\; +\infty\,.$$

Preferably, the stabilizing control should be found in feedback form. One thus looks for a function $u = u(x)$ such that all trajectories of the system

$$\dot x \;=\; f\big(x,\, u(x)\big)$$

approach the origin asymptotically as $t \to \infty$.

3 - Optimal Control. Find a control $u(\cdot) \in \mathcal{U}$ which is optimal w.r.t. a given cost criterion. For example, given the initial condition (6.2), one may seek to minimize the cost

$$J(u) \;\doteq\; \int_{t_0}^{T} L\big(x(t),\, u(t)\big)\, dt + \psi\big(x(T)\big)$$

over all control functions $u \in \mathcal{U}$. Here it is understood that $x(t) = x(t;\, t_0, x_0, u)$, while

$$L : \mathbb{R}^n \times U \mapsto \mathbb{R}, \qquad\qquad \psi : \mathbb{R}^n \mapsto \mathbb{R}$$

are continuous functions. We call L the **running cost** and ψ the **terminal cost**.

7 - The Pontryagin Maximum Principle

In connection with the system

$$\dot x = f(x,u), \qquad u(t) \in U, \qquad t \in [0,T], \qquad\qquad x(0) = x_0, \tag{7.1}$$

we consider the **Mayer problem:**

$$\max_{u\in\mathcal{U}}\; \psi\big(x(T,u)\big)\,. \tag{7.2}$$

Here there is no running cost, but we have a terminal payoff to be maximized over all admissible controls. Let $t \mapsto u^*(t)$ be an optimal control function, and let $t \mapsto x^*(t) = x(t;\, 0, x_0, u^*)$ be the corresponding optimal trajectory (fig. 7.1). We seek necessary conditions that will be satisfied by the control $u^*(\cdot)$.

As a preliminary, we recall some basic facts from O.D.E. theory. Let $t \mapsto x(t)$ be a solution of the O.D.E.

$$\dot x \;=\; g(t,x)\,. \tag{7.3}$$

Assume that $g : [0,T] \times \mathbb{R}^n \mapsto \mathbb{R}^n$ is measurable w.r.t. t and continuously differentiable w.r.t. x. Consider a family of nearby solutions (fig. 7.2), say $t \mapsto x_\varepsilon(t)$. Assume that at a given time s one has

$$\lim_{\varepsilon\to 0} \frac{x_\varepsilon(s) - x(s)}{\varepsilon} \;=\; v(s)\,.$$

Then the first order tangent vector

$$v(t) \;\doteq\; \lim_{\varepsilon\to 0} \frac{x_\varepsilon(t) - x(t)}{\varepsilon}$$

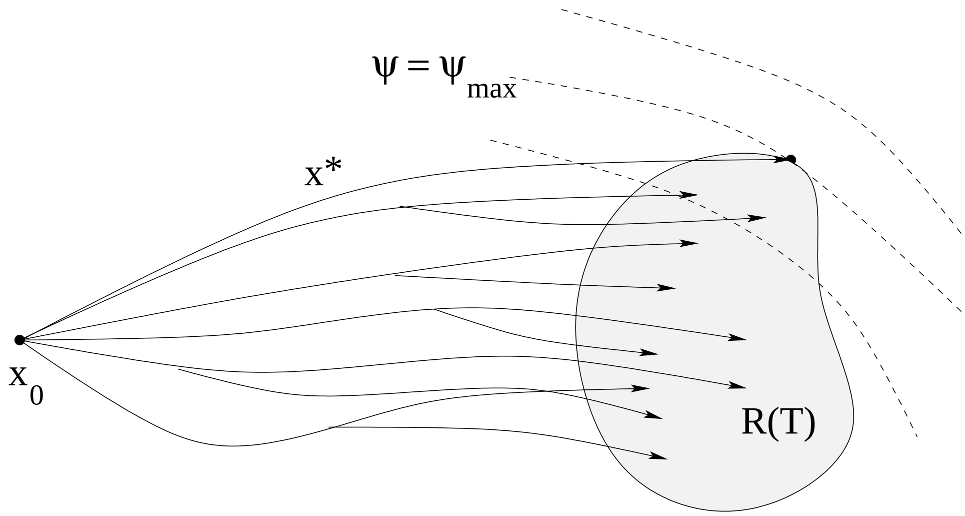

Figure 7.1. The optimal solution $x^*(\cdot)$ of the Mayer problem (7.2).

is well defined for every $t \in [0, T]$, and satisfies the linearized evolution equation

$$\dot{v}(t) \;=\; A(t)\, v(t)\,, \tag{7.4}$$

where

$$A(t) \;\doteq\; D_x g\big(t,\, x(t)\big) \tag{7.5}$$

is the $n \times n$ Jacobian matrix of first order partial derivatives of g w.r.t. x. Its entries are $A_{ij} = \partial g_i / \partial x_j$. Using the Landau notation, we can write $x_\varepsilon(t) = x(t) + \varepsilon v(t) + o(\varepsilon)$, where $o(\varepsilon)$ denotes an infinitesimal of higher order w.r.t. ε. The relations (7.4)-(7.5) are formally derived by equating terms of order ε in the first order approximation

$$\dot{x}_\varepsilon(t) \;=\; \dot{x}(t) + \varepsilon \dot{v}(t) + o(\varepsilon) \;=\; g(t, x_\varepsilon(t)) \;=\; \dot{x}(t) + D_x g(t, x(t))\, \varepsilon\, v(t) + o(\varepsilon)\,.$$

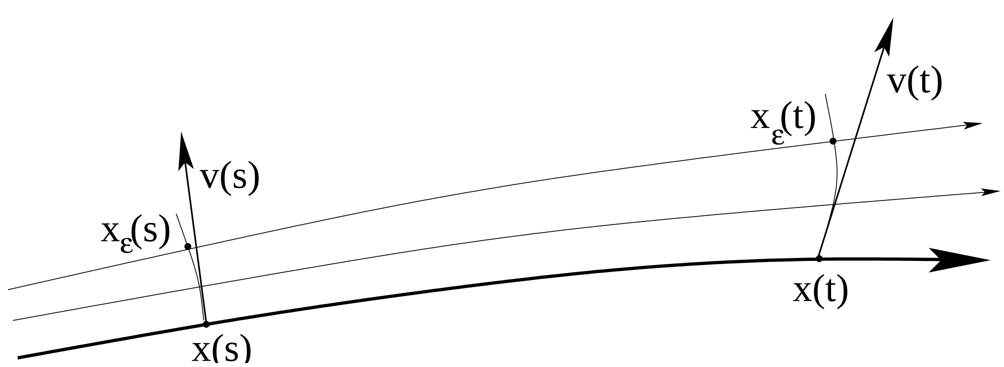

Figure 7.2. The evolution of a first order perturbation $v(t)$.

Together with (7.4), it is useful to consider the *adjoint system*

$$\dot{p}(t) \;=\; -p(t)A(t) \tag{7.6}$$

We regard $p \in I\!R^n$ as a row vector while $v \in I\!R^n$ is a column vector. Notice that, if $t \mapsto p(t)$ and $t \mapsto v(t)$ are any solutions of (7.6) and of (7.4) respectively, then the product $p(t)v(t)$ is constant in time. Indeed

$$\frac{d}{dt}\big(p(t)\,v(t)\big) \;=\; \dot{p}(t)v(t) + p(t)\dot{v}(t) \;=\; \big[-p(t)A(t)\big]v(t) + p(t)\big[A(t)v(t)\big] \;=\; 0. \tag{7.7}$$

After these preliminaries, we can now derive some necessary conditions for optimality. Since u^* is optimal, the payoff $\psi\big(x(T,u^*)\big)$ cannot be further increased by any perturbation of the control $u^*(\cdot)$. Fix a time $\tau \in]0,T]$ and a control value $\omega \in U$. For $\varepsilon > 0$ small, consider the **needle variation** $u_\varepsilon \in \mathcal{U}$ (fig. 7.3):

$$u_\varepsilon(t) \;=\; \begin{cases} \omega & \text{if} \quad t \in [\tau-\varepsilon,\,\tau]\,, \\ u^*(t) & \text{if} \quad t \notin [\tau-\varepsilon,\,\tau]\,. \end{cases} \tag{7.8}$$

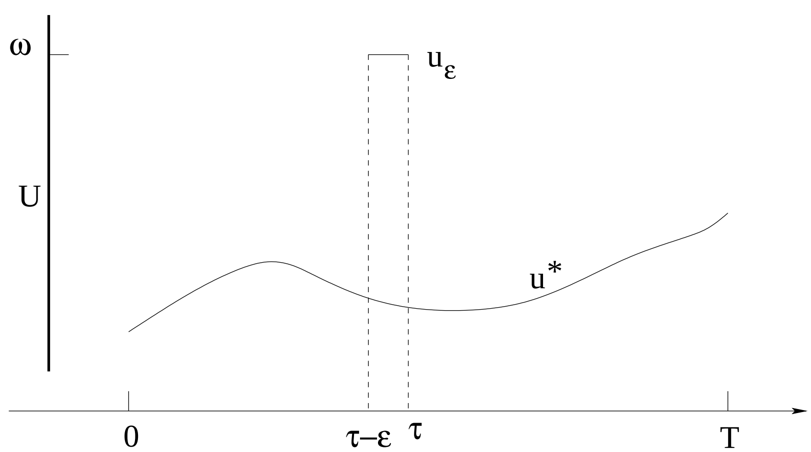

Figure 7.3. A needle variation of the control $u^*(\cdot)$.

Call $t \mapsto x_\varepsilon(t) = x(t, u_\varepsilon)$ the perturbed trajectory. We shall compute the terminal point $x_\varepsilon(T) = x(T, u_\varepsilon)$ and check that the value of ψ is not increased by this perturbation.

Assuming that the optimal control u^* is continuous at time $t = \tau$, we have

$$v(\tau) \;\doteq\; \lim_{\varepsilon\to 0} \frac{x_\varepsilon(\tau) - x^*(\tau)}{\varepsilon} \;=\; f\big(x^*(\tau),\,\omega\big) - f\big(x^*(\tau),\,u^*(\tau)\big). \tag{7.9}$$

Indeed, $x_\varepsilon(\tau-\varepsilon) = x^*(\tau-\varepsilon)$ and on the small interval $[\tau-\varepsilon,\;\tau]$ we have

$$\dot{x}_\varepsilon \;\approx\; f\big(x^*(\tau),\;\omega\big), \qquad\qquad \dot{x}^* \;\approx\; f\big(x^*(\tau),\;u^*(\tau)\big)\,.$$

Since $u_\varepsilon = u^*$ on the remaining interval $t \in [\tau, T]$, as in (7.4) the evolution of the tangent vector

$$v(t) \doteq \lim_{\varepsilon \to 0} \frac{x_\varepsilon(t) - x^*(t)}{\varepsilon} \qquad\qquad t \in [\tau, T]$$

is governed by the linear equation

$$\dot v(t) = A(t)\, v(t) \tag{7.10}$$

with $A(t) \doteq D_x f\big(x^*(t), u^*(t)\big)$. By maximality, $\psi\big(x_\varepsilon(T)\big) \le \psi\big(x^*(T)\big)$, therefore (fig. 7.4)

$$\nabla\psi\big(x^*(T)\big) \cdot v(T) \le 0. \tag{7.11}$$

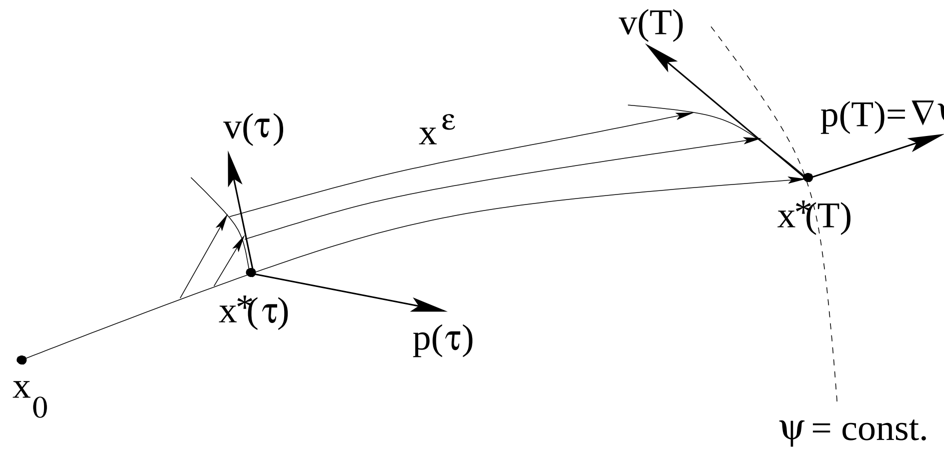

Figure 7.4. Transporting the vector $p(T)$ backward in time, along the optimal trajectory.

Summing up, the previous analysis has established the following:

For every time $\tau \in]0, T]$ where u^* is continuous and every admissible control value $\omega \in U$, we can generate the vector

$$v(\tau) \doteq f\big(x^*(\tau), \omega\big) - f\big(x^*(\tau), u^*(\tau)\big)$$

and propagate it forward in time, by solving the linearized equation (7.10). The inequality (7.11) is then a necessary condition for optimality.

Instead of propagating the (infinitely many) vectors $v(\tau)$ forward in time, it is more convenient to propagate the single vector $\nabla\psi$ backward. We thus define the row vector $t \mapsto p(t)$ as the solution of terminal value problem

$$\dot p(t) = -p(t)\, A(t), \qquad\qquad p(T) = \nabla\psi\big(x^*(T)\big). \tag{7.12}$$

By (7.7) one has $p(t)v(t) = p(T)v(T)$ for every t. In particular, (7.11) implies that

$$p(\tau) \cdot \Big[f\big(x^*(\tau), \omega\big) - f\big(x^*(\tau), u^*(\tau)\big) \Big] = \nabla\psi\big(x^*(T)\big) \cdot v(T) \le 0$$

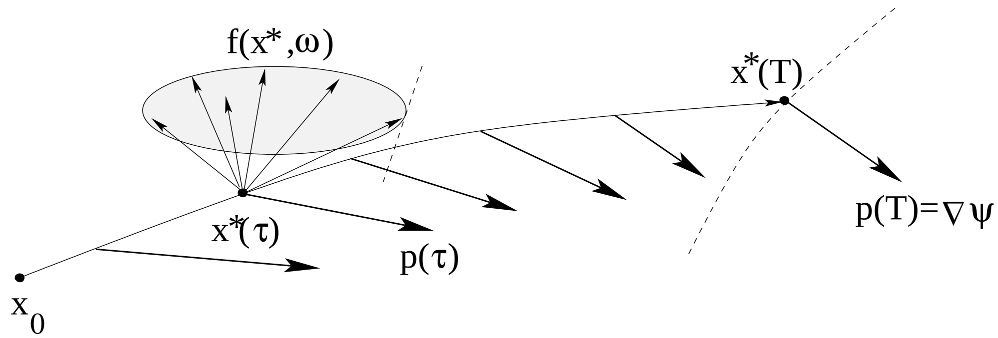

Figure 7.5. For a.e. time $\tau \in]0,T]$, the speed $\dot{x}^*(\tau)$ corresponding to the optimal control $u^*(\tau)$ is the one having inner product with $p(\tau)$ as large as possible

for every $\omega \in U$. Therefore (see fig.7.5)

$$p(\tau) \cdot \dot{x}^*(\tau) \;=\; p(\tau) \cdot f\big(x^*(\tau),\, u^*(\tau)\big) \;=\; \max_{\omega \in U} \Big\{ p(\tau) \cdot f\big(x^*(\tau),\, \omega\big) \Big\} . \tag{7.13}$$

With some additional care, one can show that the maximality condition (7.13) holds at every time τ which is a Lebesgue point of $u^*(\cdot)$, hence almost everywhere. The above result can be restated as

Theorem 7.1. Pontryagin Maximum Principle (Mayer Problem, free terminal point). *Consider the control system*

$$\dot{x} \;=\; f(x,u), \qquad u(t) \in U, \qquad t \in [0,T], \qquad x(0) = x_0 \,.$$

Let $t \mapsto u^*(t)$ be an optimal control and $t \mapsto x^*(t) = x(t,u^*)$ be the corresponding optimal trajectory for the maximization problem

$$\max_{u \in \mathcal{U}} \; \psi\big(x(T,u)\big) \,.$$

Define the vector $t \mapsto p(t)$ as the solution to the linear adjoint system

$$\dot{p}(t) \;=\; -p(t)\, A(t), \qquad\qquad A(t) \doteq D_x f\big(x^*(t),\, u^*(t)\big), \tag{7.14}$$

with terminal condition

$$p(T) = \nabla\psi\big(x^*(T)\big). \tag{7.15}$$

Then, for almost every $\tau \in [0,T]$ the following maximality condition holds:

$$p(\tau) \cdot f\big(x^*(\tau),\, u^*(\tau)\big) \;=\; \max_{\omega \in U} \Big\{ p(\tau) \cdot f\big(x^*(\tau),\, \omega\big) \Big\} . \tag{7.16}$$

In the above theorem, x, f, v represent column vectors, $D_x f$ is the $n \times n$ Jacobian matrix of first order partial derivatives of f w.r.t. x, while p is a row vector. In coordinates, the equations (7.14)-(7.15) can be rewritten as

$$\dot{p}_i(t) \;=\; -\sum_{j=1}^{n} p_j(t) \frac{\partial f_j}{\partial x_i}(t, x^*(t), u^*(t)), \qquad\qquad p_i(T) \;=\; \frac{\partial \psi}{\partial x_i}(x^*(T)),$$

while (7.16) takes the form

$$\sum_{i=1}^{n} p_i(t) \cdot f_i(t, x^*(t), u^*(t)) \;=\; \max_{\omega \in \mathbf{U}} \left\{ \sum_{i=1}^{n} p_i(t) \cdot f_i(t, x^*(t), \omega) \right\}.$$

Relying on the Maximum Principle, the computation of the optimal control requires two steps:

STEP 1: solve the pointwise maximixation problem (7.16), obtaining the optimal control u^* as a function of p, x, i.e.

$$u^*(x,p) \;=\; \operatorname*{argmax}_{\omega \in U} \big\{ p \cdot f(x, \omega) \big\}. \tag{7.17}$$

STEP 2: solve the two-point boundary value problem

$$\begin{cases} \dot{x} = f\big(x,\ u^*(x,p)\big), \\ \dot{p} = -p \cdot D_x f\big(x,\ u^*(x,p)\big), \end{cases} \qquad \begin{cases} x(0) = x_0\,, \\ p(T) = \nabla\psi\big(x(T)\big). \end{cases} \tag{7.18}$$

- In general, the function $u^* = u^*(p,x)$ in (7.17) is highly nonlinear. It may be multivalued or discontinuous.

- The two-point boundary value problem (7.18) can be solved by a **shooting method**: guess an initial value $p(0) = p_0$ and solve the corresponding Cauchy problem. Try to adjust the value of p_0 so that the terminal values $x(T),\ p(T)$ satisfy the given conditions.

Example 7.2 (Linear pendulum). Let $q(t) =$ be the position of a linearized pendulum with unit mass, controlled by an external force with magnitude $u(t) \in [-1, 1]$. Then $q(\cdot)$ satisfies the second order ODE

$$\ddot{q}(t) + q(t) \;=\; u(t), \qquad\qquad q(0) = \dot{q}(0) = 0, \qquad u(t) \in [-1, 1]\,.$$

We wish to maximize the terminal displacement $q(T)$.

Introducing the variables $x_1 = q$, $x_2 = \dot{q}$, we thus seek

$$\max_{u \in \mathcal{U}} \;\; x_1(T, u),$$

among all trajectories of the system

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = u - x_1 \end{cases} \qquad\qquad \begin{cases} x_1(0) = 0 \\ x_2(0) = 0. \end{cases}$$

Let $t \mapsto x^*(t) = x(t, u^*)$ be an optimal trajectory. The linearized equation for a tangent vector is

$$\begin{pmatrix} \dot{v}_1 \\ \dot{v}_2 \end{pmatrix} \;=\; \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

The corresponding adjoint vector $p = (p_1, p_2)$ satisfies

$$(\dot p_1, \dot p_2) \;=\; -(p_1, p_2)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad (p_1, p_2)(T) \;=\; \nabla\psi\big(x^*(T)\big) \;=\; (1, 0) \tag{7.19}$$

because $\psi(x) \doteq x_1$. In this special linear case, we can explicitly solve (7.19) without needing to know x^*, u^*. An easy computation yields

$$(p_1, p_2)(t) \;=\; \big(\cos(T-t)\,,\;\; \sin(T-t)\big)\,. \tag{7.20}$$

For each t, we must now choose the value $u^*(t) \in [-1,\, 1]$ so that

$$p_1 x_2 + p_2(-x_1 + u^*) \;=\; \max_{\omega\in[-1,1]} \;\big\{p_1 x_2 + p_2(-x_1 + \omega)\big\}\,.$$

By (7.20), the optimal control is

$$u^*(t) \;=\; \operatorname{sign} p_2(t) \;=\; \operatorname{sign}\big(\sin(T-t)\big)\,.$$

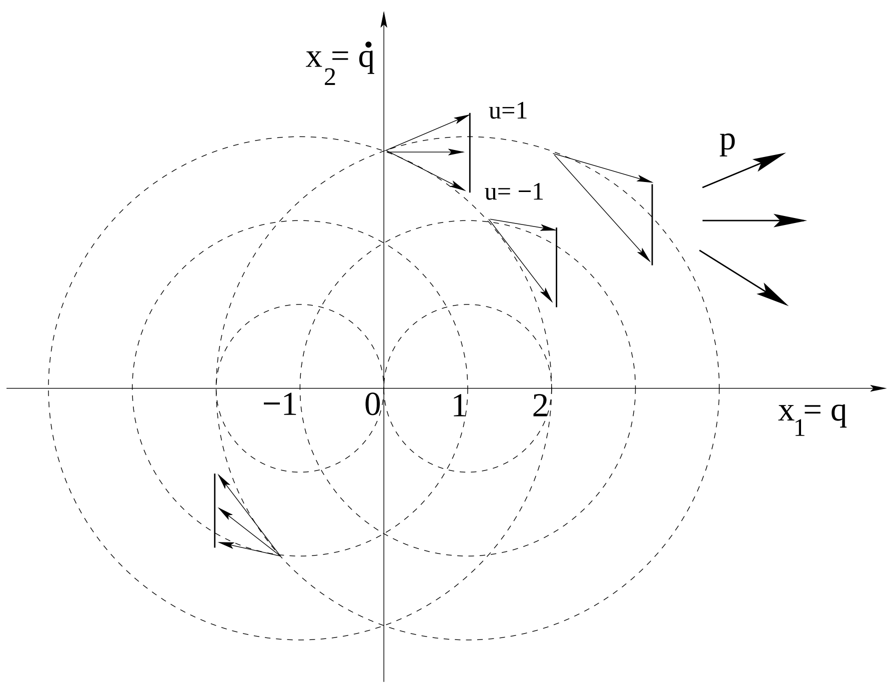

Figure 7.6

Example 7.3. Consider the problem on $I\!R^3$

$$\text{maximize } x_3(T) \qquad \text{over all controls } \; u : [0,T] \mapsto [-1,1]$$

for the system

$$\begin{cases} \dot{x}_1 = u \\ \dot{x}_2 = -x_1 \\ \dot{x}_3 = x_2 - x_1^2 \end{cases} \qquad \begin{cases} x_1(0) = 0 \\ x_2(0) = 0 \\ x_3(0) = 0. \end{cases} \tag{7.21}$$

The adjoint equations take the form

$$(\dot{p}_1, \dot{p}_2, \dot{p}_3) = (p_2 + 2x_1 p_3, \quad -p_3, \quad 0), \qquad (p_1, p_2, p_3)(T) = (0, 0, 1). \tag{7.22}$$

Maximizing the inner product $p \cdot \dot{x}$ we obtain the optimality conditions for the control u^*

$$p_1 u^* + p_2(-x_1) + p_3(x_2 - x_1^2) = \max_{\omega \in [-1,1]} \Big\{ p_1 \omega + p_2(-x_1) + p_3(x_2 - x_1^2) \Big\}, \tag{7.23}$$

$$\begin{cases} u^* = 1 & \text{if} \quad p_1 > 0, \\ u^* \in [-1, 1] & \text{if} \quad p_1 = 0, \\ u^* = -1 & \text{if} \quad p_1 < 0. \end{cases}$$

Solving the terminal value problem (7.22) for p_2, p_3 we find

$$p_3(t) \equiv 1, \qquad p_2(t) = T - t.$$

The function p_1 can now be found from the equations

$$\ddot{p}_1 = -1 + 2u^* = -1 + 2\,\mathrm{sign}(p_1), \qquad p_1(T) = 0, \quad \dot{p}_1(0) = p_2(0) = T,$$

with the convention: $\mathrm{sign}(0) = [-1, 1]$. The only solution is found to be

$$p_1(t) = \begin{cases} -\dfrac{3}{2}\left(\dfrac{T}{3} - t\right)^2 & \text{if} \quad 0 \le t \le T/3, \\ \\ 0 & \text{if} \quad T/3 \le t \le T. \end{cases}$$

The optimal control is

$$u^*(t) = \begin{cases} -1 & \text{if} \quad 0 \le t \le T/3, \\ 1/2 & \text{if} \quad T/3 \le t \le T. \end{cases}$$

Observe that on the interval $[T/3, \ T]$ the optimal control is derived not from the maximality condition (7.23) but from the equation $\ddot{p}_1 = (-1 + 2u) \equiv 0$. An optimal control with this property is called **singular**.

One should be aware that the Pontryagin Maximum Principle is a necessary condition, not sufficient for optimality.

Example 7.4. Consider the problem

$$\text{maximize:} \qquad x_2(T),$$

for the system with dynamics

$$\begin{cases} \dot{x}_1 = u, \\ \dot{x}_2 = x_1^2, \end{cases} \qquad \begin{cases} \dot{x}_1(0) = 0, \\ x_2(0) = 0, \end{cases} \qquad u(t) \in [-1, 1].$$

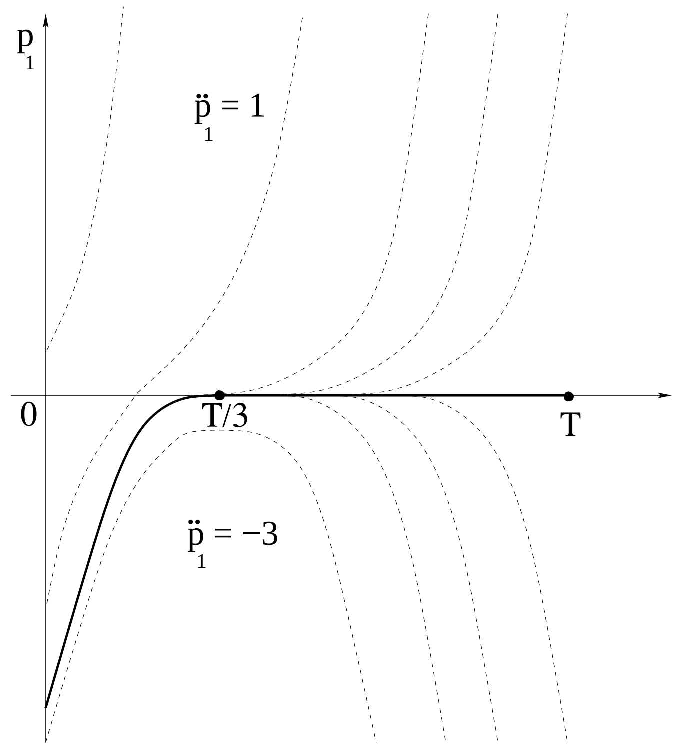

Figure 7.7

The control $u^*(t) \equiv 0$ yields the trajectory $(x_1^*(t),\, x_2^*(t)) \equiv (0,0)$. This solution satisfies the PMP. Indeed, the adjoint vector $(p_1(t), p_2(t)) \equiv (0,1)$ satisfies

$$\begin{cases} \dot p_1 &= -p_2\, x_2^* \equiv 0\,, \\ \dot p_2 &= 0\,, \end{cases} \qquad \begin{cases} p_1(T) = & 0\,, \\ p_2(T) & = 1\,, \end{cases}$$

$$p_1(t)u(t) + p_2(t)x_1^*(t) \;=\; 0 \;=\; \max_{\omega\in[-1,1]} \Big\{ p_1(t)\omega + p_2(t)x_1^*(t) \Big\}.$$

However, in this solution the terminal value $x_2^*(T) = 0$ provides the global minimum, not the global maximum! Any control $t \mapsto u(t) \in [-1,1]$ which is not identically zero yields a terminal value $x_2(T) > 0$. In this example, there are two optimal controls: $u_1(t) \equiv 1$ and $u_2(t) \equiv -1$.

8 - Extensions of the P.M.P.

In connection with the control system

$$\dot x \;=\; f(t,x,u) \qquad u(t) \in U, \qquad\qquad x(0) = x_0\,, \tag{8.1}$$

the more general optimization problem with terminal payoff and running cost

$$\max_{u\in\mathcal U} \left\{ \psi\big(x(T,u)\big) - \int_0^T L\big(t, x(t),\, u(t)\big)\, dt \right\}$$

can be easily reduced to a Mayer problem with only terminal payoff. Indeed, it suffices to introduce an additional variable x_{n+1} which evolves according to

$$\dot{x}_{n+1} = L\big(t, x(t), u(t)\big), \qquad x_{n+1}(0) = 0,$$

and consider the maximization problem

$$\max_{u \in \mathcal{U}} \Big\{ \psi\big(x(T,u)\big) - x_{n+1}(T,u) \Big\}.$$

Another important extension deals with the case where terminal constraints are given, say $x(T) \in S$, where the set S is defined as

$$S \doteq \big\{ x \in I\!R^n ; \ \ \phi_i(x) = 0, \qquad i = 1, \ldots, m \big\}.$$

Assume that, at a given point $x^* \in S$, the $m+1$ gradients $\nabla\psi$, $\nabla\phi_1, \ldots, \nabla\phi_m$ are linearly independent. Then the tangent space to S at x^* is

$$T_S = \big\{ v \in I\!R^n ; \quad \nabla\phi_i(x^*) \cdot v = 0 \qquad i = 1, \ldots, m \big\}, \tag{8.2}$$

while the tangent cone to the set

$$S^+ = \big\{ x \in S ; \ \ \psi(x) \geq \psi(x^*) \big\}$$

is

$$T_{S^+} = \Big\{ v \in I\!R^n ; \ \ \nabla\psi(x^*) \cdot v \geq 0, \qquad \nabla\phi_i(x^*) \cdot v = 0 \qquad i = 1, \ldots, m \Big\}. \tag{8.3}$$

When $x^* = x^*(T)$ is the terminal point of an admissible trajectory, we think of T_{S^+} as the **cone of profitable directions**, i.e. those directions in which we should like to move the terminal point, in order to increase the value of ψ and still satisfy the constraint $x(T) \in S$ (fig. 8.1).

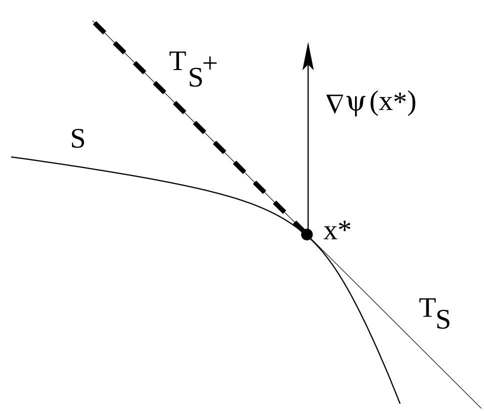

Figure 8.1. The tangent cone T_S and the cone of profitable directions T_{S^+}.

Lemma 8.1. *A vector $p \in I\!\!R^n$ satisfies*

$$p \cdot v \ \geq \ 0 \qquad\qquad \textit{for all} \ \ v \in T_{S^+} \tag{8.4}$$

if and only if it can be written as a linear combination

$$p \ = \ \lambda_0 \nabla\psi(x^*) + \sum_{i=1}^{m} \lambda_i \nabla\phi_i(x^*) \tag{8.5}$$

with $\lambda_0 \geq 0$.

Proof. Define the vectors

$$w_0 \ \doteq \ \nabla\psi(x^*)\,, \qquad\qquad w_i \ \doteq \ \nabla\phi_i(x^*) \qquad i = 1, \ldots, m\,.$$

By our previous assumption, these vectors are linearly independent. We can thus find additional vectors w_j, $j = m+1, \ldots, n-1$ so that

$$\big\{w_0,\, w_1,\, \cdots,\, w_m,\ \ w_{m+1},\, \ldots,\, w_{n-1}\big\}$$

is a basis of $I\!\!R^n$. Let

$$\big\{v_0,\, v_1,\, \cdots,\, v_m,\ \ v_{m+1},\, \ldots,\, v_{n-1}\big\}$$

be the dual basis, so that

$$v_i \cdot w_j = \begin{cases} 1 & \text{if} \quad i = j, \\ 0 & \text{if} \quad i \neq j. \end{cases}$$

We observe that

$$v \in T_{S^+} \qquad \text{if and only if} \qquad v \ = \ c_0 v_0 + \sum_{i=m+1}^{n-1} c_i v_i \tag{8.6}$$

for some $c_0 \geq 0$, $c_i \in I\!\!R$. An arbitrary vector $p \in I\!\!R^n$ can now be written as

$$p \ = \ \lambda_0 w_0 + \sum_{i=1}^{m} \lambda_i w_i + \sum_{i=m+1}^{n-1} \lambda_i w_j\,.$$

Moreover, every vector $v \in T_{S^+}$ can be decomposed as in (8.6). Therefore

$$p \cdot v \ = \ \lambda_0 c_0 + \sum_{i=m+1}^{n-1} \lambda_i c_i\,.$$

It is now clear that (8.4) holds if and only if $\lambda_0 \geq 0$ and $\lambda_i = 0$ for all $i = m+1, \ldots, n-1$. □

Next, consider a trajectory $t \mapsto x^*(t) = x(t, u^*)$, generated by the control $u^*(\cdot)$. To test its optimality, we need to perturb u^* in such a way that the terminal point $x(T)$ still lies in the admissible set S.

Example 8.2. Consider the problem

$$\text{maximize:} \quad x_3(T)$$

for a control system of the general form (8.1), with terminal constraint

$$x(T) \in S \doteq \{x = (x_1, x_2, x_3)\,; \quad x_2^2 + x_3^2 = 1\}.$$

Let $u^*(\cdot)$ be an admissible control, such that the trajectory $t \mapsto x^*(t) = x(t, u^*)$ reaches the terminal point $x^*(T) = (1, 0, 0)$.

Assume that we can construct a family of needle variations $u_\varepsilon(\cdot)$ generating the vector

$$\mathbf{v}_1(T) = \lim_{\varepsilon \to 0} \frac{x_\varepsilon(T) - x^*(T)}{\varepsilon} = (1, 0, 1).$$

This does not rule out the optimality of u^*, because none of the points $x_\varepsilon(T)$ for $\varepsilon > 0$ lies on the target set S.

Next, assume that we can construct a family of needle variations $u_\varepsilon(\cdot)$ generating the vector

$$\mathbf{v}_2(T) = \lim_{\varepsilon \to 0} \frac{x_\varepsilon(T) - x^*(T)}{\varepsilon} = (0, 0, 1).$$

Notice that the vector $\mathbf{v}_2$ is now tangent to the target set S at the point $x(T)$. In fact, $\mathbf{v}_2 \in T_{S+}$. Yet, this does not guarantee that any of the points $x_\varepsilon(T)$ should lie on S (see fig. 8.2, left). Again, we cannot rule out the optimality of $u^*(\cdot)$.

In order to obtain perturbations with terminal point $x_\varepsilon(T) \in S$, the key idea is to construct combinations of the needle variations (fig. 8.3, left). If a needle variation (τ_1, ω_1) produces a tangent vector $\mathbf{v}_1$ and the needle variation (τ_2, ω_2) produces the tangent vector $\mathbf{v}_2$, then by combining them we can produce any linear combination with positive coefficients $\theta_1\mathbf{v}_1 + \theta_2\mathbf{v}_2$ continuously depending on $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2_+$ (fig. 8.3, right). This guarantees that some intermediate value $x_\varepsilon^\theta(T)$ actually lies on the target set S (fig. 8.2, right).

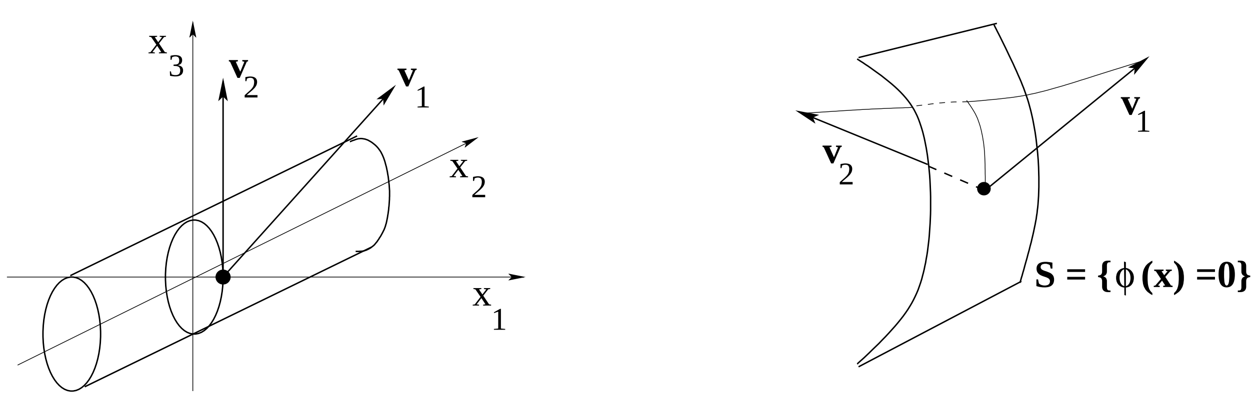

Figure 8.2.

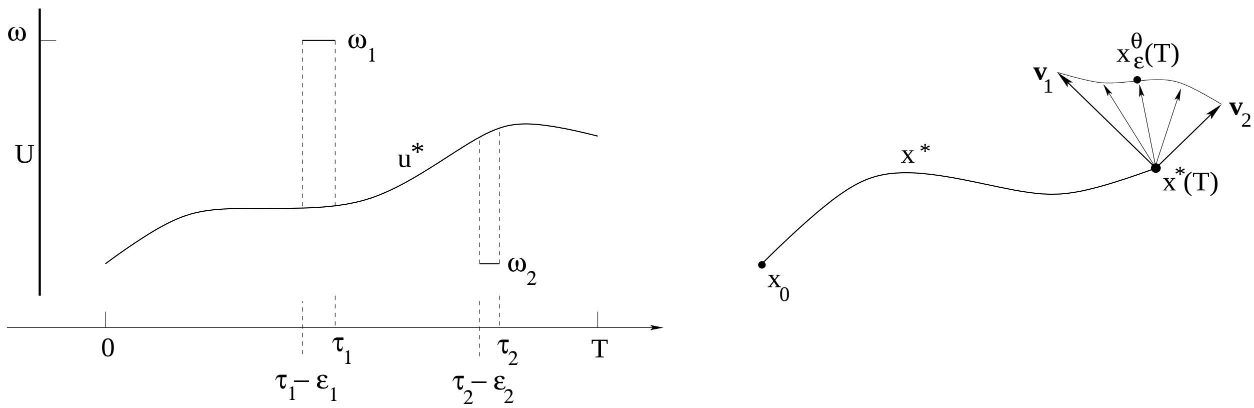

Figure 8.3.

As in the previous section, given $\tau \in]0,T]$ and $\omega \in U$, consider the family of needle variations

$$u_\varepsilon(t) \;=\; \begin{cases} \omega & \text{if} \qquad t \in [\tau-\varepsilon,\, \tau]\,, \\ u^*(t) & \text{if} \qquad t \notin [\tau-\varepsilon,\, \tau]\,. \end{cases}$$

Call

$$v^{\tau,\omega}(T) \;\doteq\; \lim_{\varepsilon\to 0} \frac{x(T,u_\varepsilon) - x(T,u^*)}{\varepsilon}$$

the first order variation of the terminal point of the corresponding trajectory. Define Γ to be the smallest convex cone containing all vectors $v^{\tau,\omega}$. In other words, Γ is the set of all finite linear combination of the vectors $v^{\tau,\omega}$ with positive coefficients.

We think of Γ as a **cone of feasible directions**, i.e. directions in which we can move the terminal point $x(T,u^*)$ by suitably perturbing the control u^* (fig. 8.4).

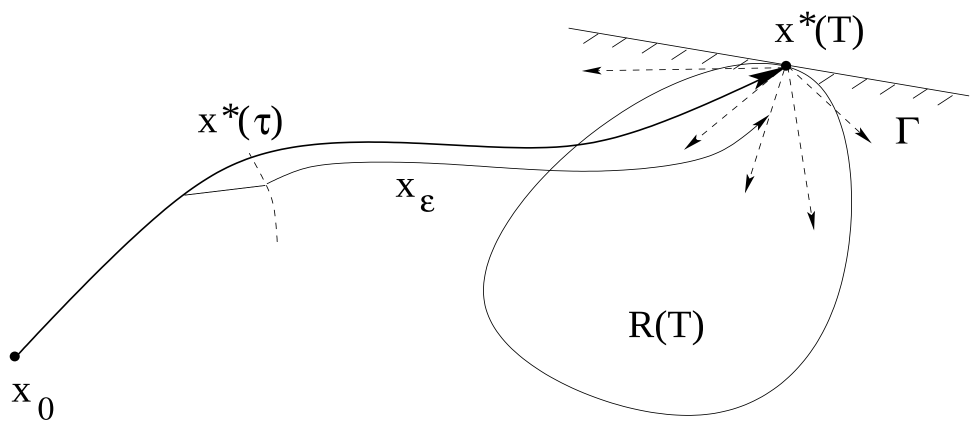

Figure 8.4. The cone Γ of feasible directions.

We can now state necessary conditions for optimality for the

Mayer Problem with terminal constraints:

$$\max_{u\in\mathcal{U}} \;\psi\big(x(T,u)\big), \tag{8.7}$$

for the control system

$$\dot{x} = f(t, x, u), \qquad u(t) \in U, \qquad t \in [0, T], \tag{8.8}$$

with initial and terminal constraints

$$x(0) = x_0, \qquad \phi_i\big(x(T)\big) = 0, \qquad i = 1, \ldots, m\,. \tag{8.9}$$

Theorem 8.3 (PMP, geometric version). *Let $t \mapsto x^*(t) = x(t, u^*)$ be an optimal trajectory for the problem (8.7)–(8.9), corresponding to the control $u^*(\cdot)$. Then the cones Γ and T_{S^+} are weakly separated, i.e. there exists a non-zero vector $p(T)$ such that*

$$p(T) \cdot v \;\geq\; 0 \qquad \textit{for all} \;\; v \in T_{S^+}\,, \tag{8.10}$$

$$p(T) \cdot v \;\leq\; 0 \qquad \textit{for all} \;\; v \in \Gamma\,. \tag{8.11}$$

This separation property is illustrated in fig. 8.5. An equivalent statement is:

Theorem 8.4 (PMP, analytic version). *Let $t \mapsto x^*(t) = x(t, u^*)$ be an optimal trajectory, corresponding to the control $u^*(\cdot)$. Then there exists a non-zero vector function $t \mapsto p(t)$ such that*

$$p(T) \;=\; \lambda_0\, \nabla\psi\big(x^*(T)\big) + \sum_{i=1}^{m} \lambda_i\, \nabla\phi_i\big(x^*(T)\big) \qquad \textit{with} \;\; \lambda_0 \geq 0\,, \tag{8.12}$$

$$\dot{p}(t) \;=\; -p(t)\, D_x f\big(t,\, x^*(t),\, u^*(t)\big) \qquad t \in [0, T]\,, \tag{8.13}$$

$$p(\tau) \cdot f\big(\tau,\, x^*(\tau),\, u^*(\tau)\big) \;=\; \max_{\omega \in U} \Big\{ p(\tau) \cdot f\big(\tau,\, x^*(\tau),\, \omega\big) \Big\} \qquad \textit{for a.e.}\; \tau \in [0, T]\,. \tag{8.14}$$

We show here the equivalence of the two formulations.
By Lemma 8.1, (8.10) is equivalent to (8.12).
Since every tangent vector $v^{\tau,\omega}$ satisfies the linear evolution equation

$$\dot{v}^{\tau,\omega}(t) \;=\; D_x f\big(t,\, x^*(t),\, u^*(t)\big)\, v^{\tau,\omega}(t),$$

if $t \mapsto p(t)$ satisfies (8.13) then the product $p(t) \cdot v^{\tau,\omega}(t)$ is constant. Therefore

$$p(T) \cdot v^{\tau,\omega}(T) \;\leq\; 0$$

if and only if

$$p(\tau) \cdot v^{\tau,\omega}(\tau) \;=\; p(\tau) \cdot \Big[f\big(\tau,\, x^*(\tau),\, \omega\big) - f\big(\tau,\, x^*(\tau),\, u^*(\tau)\big) \Big] \;\leq\; 0$$

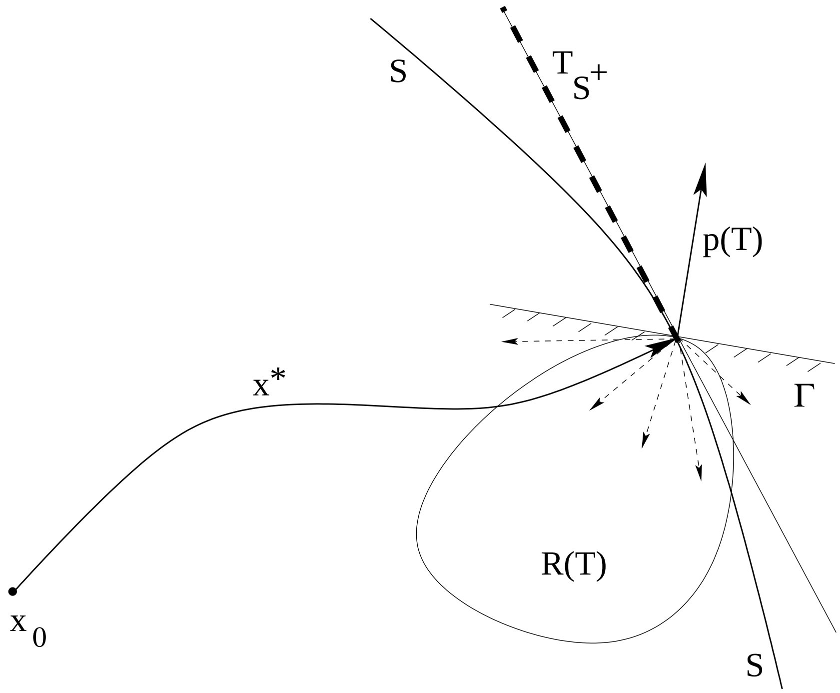

Figure 8.5. The hyperplane $\{w\,;\ \ p(T)\cdot w=0\}$ weakly separates the cones Γ and T_{S^+}.

if and only if (8.14) holds.

As a special case, consider the

Lagrange Minimization Problem with fixed initial and terminal points:

$$\min_{u\in\mathcal{U}}\int_0^T L(t,x,u)\,dt\,,\tag{8.15}$$

for the control system on $I\!R^n$

$$\dot x\ =\ f(t,x,u)\,,\qquad\qquad u(t)\in U\,,\tag{8.16}$$

with initial and terminal constraints

$$x(0)\ =\ a,\qquad\qquad x(T)\ =\ b\,.\tag{8.17}$$

An adaptation of the previous analysis yields

Theorem 8.5 (PMP, Lagrange problem). *Let $t\mapsto x^*(t)=x(t,u^*)$ be an optimal trajectory, corresponding to the optimal control $u^*(\cdot)$. Then there exist a constant $\lambda\geq 0$ and a row vector $t\mapsto p(t)$ (not both = 0) such that*

$$\dot p(t)\ =\ -p(t)\,D_xf\big(t,\,x^*(t),\,u^*(t)\big)-\lambda\,D_xL\big(t,\,x^*(t),\,u^*(t)\big)\,,\tag{8.18}$$

$$\begin{aligned}p(t)\cdot f\big(t,\,x^*(t),\,u^*(t)\big)+&\lambda\,L\big(t,\,x^*(t),\,u^*(t)\big)\\ &=\ \min_{\omega\in U}\Big\{p(t)\cdot f\big(t,\,x^*(t),\,\omega\big)+\lambda\,L\big(t,\,x^*(t),\,\omega\big)\Big\}\,.\end{aligned}\tag{8.19}$$

This follows by applying the previous results to the Mayer problem

$$\min_{u \in \mathcal{U}} \; x_{n+1}(T, u)$$

with

$$\dot{x} \;=\; f(t,x,u), \qquad\qquad \dot{x}_{n+1} \;=\; L(t,x,u), \qquad\qquad x_{n+1}(0) = 0\,.$$

Observe that the evolution of the adjoint vector $(p, p_{n+1}) = (p_1, \ldots, p_n, p_{n+1})$ is governed by the linear system

$$(\dot{p}_1, \ldots, \dot{p}_n, \dot{p}_{n+1}) \;=\; -(p_1, \ldots, p_n, p_{n+1}) \begin{pmatrix} \partial f_1/\partial x_1 & \cdots & \partial f_1/\partial x_n & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \partial f_n/\partial x_1 & \cdots & \partial f_n/\partial x_n & 0 \\ \partial L/\partial x_1 & \cdots & \partial L/\partial x_n & 0 \end{pmatrix}.$$

Because of the terminal constraints $(x_1, \ldots, x_n)(T) = (b_1, \ldots, b_n)$, the only requirement on the terminal value $(p_1, \ldots, p_n, p_{n+1})(T)$ is

$$p_{n+1}(T) \geq 0.$$

Since $\dot{p}_{n+1} = 0$, we have $p_{n+1}(t) \equiv \lambda$ for some constant $\lambda \geq 0$.

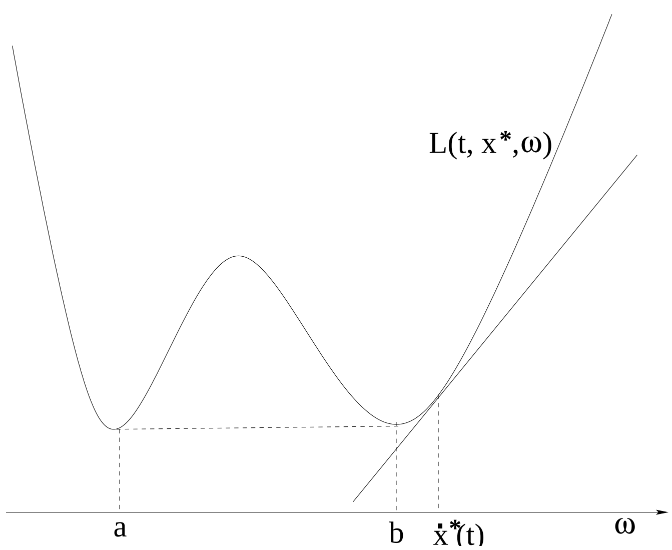

Figure 8.6. The Weierstrass necessary condition.

Theorem 8.5 can be further specialized to the

Standard Problem of the Calculus of Variations:

$$\text{minimize} \int_0^T L\big(t,\, x(t),\, \dot{x}(t)\big)\, dt \tag{8.20}$$

over all absolutely continuous functions $x : [0,T] \mapsto I\!R^n$ such that

$$x(0) = a, \qquad\qquad x(T) = b\,. \tag{8.21}$$

This corresponds to the optimal control problem (8.15), for the trivial control system

$$\dot x \;=\; u, \qquad\qquad u(t) \;\in\; U \;\doteq\; I\!R^n. \tag{8.22}$$

We assume that L is smooth, and that $x^*(\cdot)$ is an optimal solution. By Theorem 8.5 there exist a constant $\lambda \geq 0$ and a row vector $t \mapsto p(t)$ (not both $= 0$) such that

$$\dot p(t) \;=\; -\lambda \frac{\partial}{\partial x} L\big(t,\, x^*(t),\, \dot x^*(t)\big), \tag{8.23}$$

$$p(t)\cdot \dot x^*(t) + \lambda\, L\big(t,\, x^*(t),\, \dot x^*(t)\big) \;=\; \min_{\omega \in I\!R^n} \Big\{ p(t)\cdot \omega + \lambda\, L\big(t,\, x^*(t),\, \omega\big)\Big\}\,. \tag{8.24}$$

If $\lambda = 0$, then $p(t) \neq 0$. But in this case $\dot x^*$ cannot provide a minimum over the whole space $I\!R^n$. This contradiction shows that we must have $\lambda > 0$.

Since λ, p are determined up to a positive scalar multiple, we can assume $\lambda = 1$. According to (8.24), the global minimum on the right hand side is attained when $\omega = \dot x^*(t)$. Differentiating w.r.t. ω, a necessary condition is found to be

$$p(t) \;=\; -\frac{\partial}{\partial \dot x} L\big(t,\, x^*(t),\, \dot x^*(t)\big)\,. \tag{8.25}$$

Inserting (8.25) in the evolution equation (8.23) (with $\lambda = 1$), one obtains the famous **Euler-Lagrange equations**

$$\frac{d}{dt}\left[\frac{\partial}{\partial \dot x} L\big(t,\, x^*(t),\, \dot x^*(t)\big)\right] \;=\; \frac{\partial}{\partial x} L\big(t,\, x^*(t),\, \dot x^*(t)\big). \tag{8.26}$$

In addition, the minimality condition (8.24) (always with $\lambda = 1$) implies

$$p(t)\cdot \dot x^*(t) + L\big(t,\, x^*(t),\, \dot x^*(t)\big) \;\leq\; p(t)\cdot\omega + L\big(t,\, x^*(t),\, \omega\big) \qquad \text{for every } \;\omega \in I\!R^n.$$

Replacing $p(t)$ by its expression given at (8.25), one obtains the **Weierstrass necessary conditions**

$$L(t,\, x^*(t),\, \omega) \;\geq\; L\big(t,\, x^*(t),\, \dot x^*(t)\big) + \frac{\partial L\big(t,\, x^*(t),\, \dot x^*(t)\big)}{\partial \dot x}\cdot\big(\omega - \dot x^*(t)\big)\,, \tag{8.27}$$

valid for every $\omega \in I\!R^n$. In other words (fig. 8.6), for every time t, the graph of $\omega \mapsto L(t, x^*(t), \omega)$ lies entirely above its tangent hyperplane at the point $\big(t, x^*(t), \dot x^*(t)\big)$.

9 - Dynamic programming

Consider again a control system of the form

$$\dot x(t) \;=\; f\big(x(t),\, u(t)\big)\,, \qquad\qquad u(t) \in U\,, \qquad\quad t \in [0,T]\,. \tag{9.1}$$

We now assume that the set $U \subset I\!R^m$ of admissible control values is compact, while $f : I\!R^n \times U \mapsto I\!R^n$ is a continuous function such that

$$\big|f(x,u)\big| \;\leq\; C, \qquad \big|f(x,u) - f(y,u)\big| \;\leq\; C\,|x-y| \qquad \text{for all } x,y \in I\!R^n,\ u \in U\,, \tag{9.2}$$

for some constant C. Given an initial data

$$x(s) \;=\; y \in I\!R^n, \tag{9.3}$$

under the assumptions (9.2), for every choice of the measurable control function $u(\cdot) \in \mathcal{U}$ the Cauchy problem (9.1)-(9.2) has a unique solution, which we denote as $t \mapsto x(t;\, s, y, u)$ or sometimes simply as $t \mapsto x(t)$. We seek an admissible control function $u^* : [s,T] \mapsto U$, which minimizes the sum of a running and a terminal cost

$$J(s,y,u) \;\doteq\; \int_s^T L\big(x(t),\, u(t)\big)\, dt + \psi\big(x(T)\big). \tag{9.4}$$

Here it is understood that $x(t) = x(t;\, s, y, u)$, while

$$L : I\!R^n \times U \mapsto I\!R, \qquad \psi : I\!R^n \mapsto I\!R$$

are continuous functions. We shall assume that the functions L, ψ satisfy the bounds

$$\big|L(x,u)\big| \;\leq\; C, \qquad \big|\psi(x)\big| \;\leq\; C, \tag{9.5}$$

$$\big|L(x,u) - L(y,u)\big| \;\leq\; C\,|x-y|, \qquad \big|\psi(x) - \psi(y)\big| \;\leq\; C\,|x-y|, \tag{9.6}$$

for all $x, y \in I\!R^n$, $u \in U$. As in the previous sections, we call

$$\mathcal{U} \doteq \Big\{ u : I\!R \mapsto I\!R^m \;\; \text{measurable}, \quad u(t) \in U \;\; \text{for a.e. } t \Big\} \tag{9.7}$$

the family of admissible control functions.

Remark 9.1. The global bounds assumed in (9.2), (9.5), and (9.6) appear to be very restrictive. In practice, one can often obtain an a-priori bound on all trajectories of the control system which start from a bounded set S. Say, $|x(t)| \leq M$ for all $t \in [0,T]$. For all subsequent applications, it then suffices to assume that the bounds (9.2), (9.5), (9.6) hold as long as $|x|, |y| \leq M$.

According to the method of **dynamic programming**, an optimal control problem can be studied by looking at the **value function**:

$$V(s,y) \;\doteq\; \inf_{u(\cdot) \in \mathcal{U}} J(s,y,u). \tag{9.8}$$

We consider here a whole family of optimal control problem, all with the same dynamics (9.1) and cost functional (9.4). We are interested in how the minimum cost varies, depending on the initial data (s,y) in (9.3). As a preliminary, we prove

Lemma 9.2. *Let the functions f, ψ, L satisfy the assumptions (9.2), (9.5) and (9.6). Then the value function V in (9.8) is bounded and Lipschitz continuous. Namely, there exists a constant C' such that*

$$\big|V(s,y)\big| \;\leq\; C', \tag{9.9}$$

$$\big|V(s,y) - V(s',y')\big| \ \leq\ C'\big(|s-s'| + |y-y'|\big). \tag{9.10}$$

Proof. Let $(\bar{s},\bar{y})$ and $\varepsilon > 0$ be given. Choose a measurable control $u_\varepsilon : [0,T] \mapsto \mathcal{U}$ which is almost optimal for the optimization problem with initial data $x(\bar{s}) = \bar{y}$, namely

$$J(\bar{s},\bar{y},u_\varepsilon) \ \leq\ V(\bar{s},\bar{y}) + \varepsilon\,.$$

Call $t \mapsto x(t) = x(t;\bar{s},\bar{y},u_\varepsilon)$ the corresponding trajectory. Using the same control $u_\varepsilon(\cdot)$ in connection with a different initial data, say $x(s) = y$, we obtain a new trajectory $t \mapsto z(t) = x(t;s,y,u_\varepsilon)$. By the boundedness assumptions (9.2) it follows

$$|x(s) - z(s)| \ \leq\ |x(s) - x(\bar{s})| + |x(\bar{s}) - z(s)| \ \leq\ C|s-\bar{s}| + |\bar{y} - y|\,.$$

Since f is Lipschitz continuous, Gronwall's lemma yields

$$|x(t) - z(t)| \ \leq\ e^{C|t-s|}\,|x(s) - z(s)| \ \leq\ e^{CT}\,\Big(C|s-\bar{s}| + |\bar{y}-y|\Big)\,.$$

Using the bounds (9.5)-(9.6) we thus obtain

$$\begin{aligned}
J(s,y,u_\varepsilon) \ &=\ J(\bar{s},\bar{y},u_\varepsilon) + \int_s^{\bar{s}} L(z,u_\varepsilon)\,dt + \int_{\bar{s}}^T \Big(L(z,u_\varepsilon) - L(x,u_\varepsilon)\Big)\,dt + \psi(z(T)) - \psi(x(T)) \\
&\leq J(\bar{s},\bar{y},u_\varepsilon) + C|s-\bar{s}| + \int_{\bar{s}}^T C|z(t) - x(t)|\,dt + C|z(T) - x(T)| \\
&\leq J(\bar{s},\bar{y},u_\varepsilon) + C'\big(|s-\bar{s}| + |y-\bar{y}|\big)
\end{aligned}$$

for some constant C'. This implies

$$V(s,y) \ \leq\ J(s,y,u_\varepsilon) \ \leq\ V(\bar{s},\bar{y}) + \varepsilon + C'\big(|s-\bar{s}| + |y-\bar{y}|\big).$$

Letting $\varepsilon \to 0$ and interchanging the roles of (s,y) and $(\bar{s},\bar{y})$, one obtains the Lipschitz continuity of the value function V. $\square$

We will show that the value function V can be characterized as the unique viscosity solution to a Hamilton-Jacobi equation. Toward this goal, a basic step is provided by Bellman's principle of dynamic programming.

Theorem 9.3 (Dynamic Programming Principle). *For every $\tau \in [s,T]$ and $y \in I\!R^n$, one has*

$$V(s,y) \ =\ \inf_{u(\cdot)} \left\{ \int_s^\tau L\big(x(t;s,y,u),\ u(t)\big)\,dt \ +\ V\big(\tau,\ x(\tau;s,y,u)\big) \right\}. \tag{9.11}$$

In other words (fig. 9.1), the optimization problem on the time interval $[s,T]$ can be split into two separate problems:

- As a first step, we solve the optimization problem on the sub-interval $[\tau,T]$, with running cost L and terminal cost ψ. In this way, we determine the value function $V(\tau,\cdot)$, at time τ.

- As a second step, we solve the optimization problem on the sub-interval $[s,\tau]$, with running cost L and terminal cost $V(\tau,\cdot)$, determined by the first step.

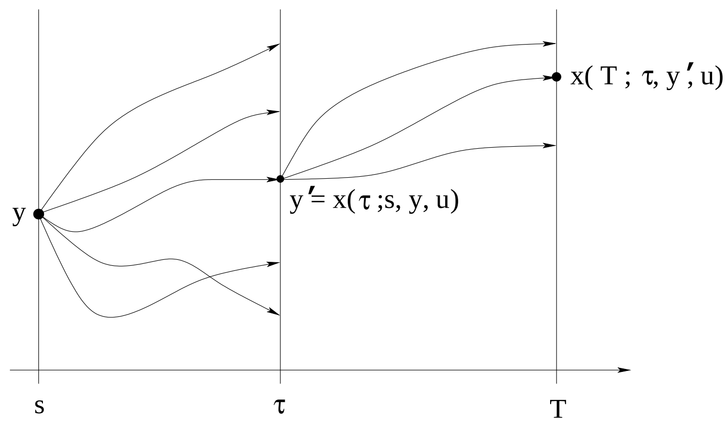

Figure 9.1. The optimization problem on $[s,T]$ can be decomposed in two sub-problems, on the time intervals $[s,\tau]$ and on $[\tau,T]$, respectively.

At the initial time s, by (9.11) we claim that the value function $V(s,\cdot)$ obtained in step 2 is the same as the value function corresponding to the global optimization problem over the whole interval $[s,T]$.

Proof. Call J^τ the right hand side of (9.11).

1. To prove that $J^\tau \le V(s,y)$, fix $\varepsilon > 0$ and choose a control $u : [s,T] \mapsto U$ such that

$$J(s,y,u) \;\le\; V(s,y) + \varepsilon.$$

Observing that

$$V\big(\tau, x(\tau;s,y,u)\big) \;\le\; \int_\tau^T L\big(x(t;s,y,u),\; u(t)\big)\,dt + \psi\big(x(T;s,y,u)\big),$$

we conclude

$$\begin{aligned} J^\tau \;&\le\; \int_s^\tau L\big(x(t;s,y,u),\; u(t)\big)\,dt + V\big(\tau, x(\tau;s,y,u)\big) \\ &\le\; J(s,y,u) \;\le\; V(s,y) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this first inequality is proved.

2. To prove that $V(s,y) \le J^\tau$, fix $\varepsilon > 0$. Then there exists a control $u_1 : [s,\tau] \mapsto U$ such that

$$\int_s^\tau L\big(x(t;s,y,u_1),\; u_1(t)\big)\,dt + V\big(\tau,\; x(\tau;s,y,u_1)\big) \;\le\; J^\tau + \varepsilon. \tag{9.12}$$

Moreover, there exists a control $u_2 : [\tau,T] \mapsto U$ such that

$$J\big(\tau,\; x(\tau;s,y,u_1),\; u_2\big) \;\le\; V\big(\tau,\; x(\tau;s,y,u_1)\big) + \varepsilon. \tag{9.13}$$

We now define a new control $u : [s,T] \mapsto U$ as the concatenation of u_1 and u_2:

$$u(t) \doteq \begin{cases} u_1(t) & \text{if} \quad t \in [s,\tau], \\ u_2(t) & \text{if} \quad t \in]\tau,T]. \end{cases}$$

By (9.12) and (9.13) it now follows

$$V(s,y) \;\le\; J(s,y,u) \;\le\; J^\tau + 2\varepsilon.$$

Since $\varepsilon > 0$ can be taken arbitrarily small, this second inequality is also proved. $\square$

10 - The Hamilton-Jacobi-Bellman Equation

The main goal of this section is to characterize the value function as the unique solution of a first order P.D.E., in the viscosity sense. In turn, this will provide a sufficient condition for the global optimality of a control function $u(\cdot)$. As in the previous section, we assume here that the set U is compact and that the functions f, L, ψ satisfy the bounds (9.2), (9.5) and (9.6).

Theorem 10.1. *In connection with the control system (9.1), consider the value function $V = V(s,y)$ defined by (9.8) and (9.4). Then V is the unique viscosity solution of the Hamilton-Jacobi-Bellman equation*

$$-\Big[V_t + H(x, \nabla V)\Big] \;=\; 0 \qquad\qquad (t,x) \in]0,T[\,\times I\!R^n, \tag{10.1}$$

with terminal condition

$$V(T,x) \;=\; \psi(x) \qquad\qquad x \in I\!R^n, \tag{10.2}$$

and Hamiltonian function

$$H(x,p) \;\doteq\; \min_{\omega\in U} \big\{p\cdot f(x,\omega) + L(x,\omega)\big\}. \tag{10.3}$$

Proof. By Lemma 9.2, the value function is bounded and uniformly Lipschitz continuous on $[0,T]\times I\!R^n$. The terminal condition (10.2) is obvious. To show that V is a viscosity solution, let $\varphi \in \mathcal{C}^1\big(]0,T[\,\times I\!R^n\big)$. Two separate statements need to be proved:

(P1) If $V-\varphi$ attains a local maximum at a point $(t_0,x_0) \in]0,T[\,\times I\!R^n$, then

$$\varphi_t(t_0,x_0) + \min_{\omega\in U}\big\{\nabla\varphi(t_0,x_0)\cdot f(x_0,\omega) + L(x_0,\omega)\big\} \;\geq\; 0. \tag{10.4}$$

(P2) If $V-\varphi$ attains a local minimum at a point $(t_0,x_0) \in]0,T[\,\times I\!R^n$, then

$$\varphi_t(t_0,x_0) + \min_{\omega\in U}\big\{\nabla\varphi(t_0,x_0)\cdot f(x_0,\omega) + L(x_0,\omega)\big\} \;\leq\; 0. \tag{10.5}$$

1. To prove (P1), we can assume that

$$V(t_0,x_0) \;=\; \varphi(t_0,x_0), \qquad\qquad V(t,x) \;\leq\; \varphi(t,x) \quad \text{for all } t,x\,.$$

If (10.4) does not hold, then there exists $\omega \in U$ and $\theta > 0$ such that

$$\varphi_t(t_0,x_0) + \nabla\varphi(t_0,x_0)\cdot f(x_0,\omega) + L(x_0,\omega) \;<\; -\theta. \tag{10.6}$$

We shall derive a contradiction by showing that this control value ω is "too good to be true". Namely, by choosing a control function $u(\cdot)$ with $u(t)\equiv\omega$ for $t\in[t_0,\ t_0+\delta]$ and such that u is nearly optimal on the remaining interval $[t_0+\delta,\ T]$, we obtain a total cost $J(t_0,x_0,u)$ strictly smaller than $V(t_0,x_0)$. Indeed, by continuity (10.6) implies

$$\varphi_t(t,x) + \nabla\varphi(t,x)\cdot f(x,\omega) \;<\; -L(x,\omega) - \theta. \tag{10.7}$$

whenever

$$|t-t_0| \;<\; \delta, \qquad\qquad |x-x_0| \;\leq\; C\delta, \tag{10.8}$$

for some $\delta > 0$ small enough and C the constant in (9.2). Let $x(t) \doteq x(t; t_0, x_0, \omega)$ be the solution of

$$\dot x(t) \;=\; f\big(x(t), \omega\big), \qquad\qquad x(t_0) \;=\; x_0,$$

i.e. the trajectory corresponding to the constant control $u(t) \equiv \omega$. We then have

$$\begin{aligned} V\big(t_0+\delta,\; x(t_0+\delta)\big) - V(t_0, x_0) \;&\leq\; \varphi\big(t_0+\delta,\; x(t_0+\delta)\big) - \varphi(t_0, x_0)\\ &=\; \int_{t_0}^{t_0+\delta} \frac{d}{dt}\varphi\big(t,\, x(t)\big)\,dt\\ &=\; \int_{t_0}^{t_0+\delta} \Big\{\varphi_t\big(t,\, x(t)\big) + \nabla\varphi\big(t,\, x(t)\big)\cdot f\big(x(t), \omega\big)\Big\}\,dt\\ &\leq\; -\int_{t_0}^{t_0+\delta} L\big(x(t),\, \omega\big)\,dt - \delta\theta, \end{aligned} \tag{10.9}$$

because of (10.7). On the other hand, the Dynamic Programming Principle (9.11) yields

$$V(t_0, x_0) \;\leq\; \int_{t_0}^{t_0+\delta} L\big(x(t), \omega\big)\,dt + V\big(t_0+\delta,\; x(t_0+\delta)\big)\,. \tag{10.10}$$

Together, (10.9) and (10.10) yield a contradiction, hence (P1) must hold.

2. To prove (P2), we can assume that

$$V(t_0, x_0) \;=\; \varphi(t_0, x_0), \qquad\qquad V(t,x) \;\geq\; \varphi(t,x) \quad \text{for all } t, x\,.$$

If (P2) fails, then there exists $\theta > 0$ such that

$$\varphi_t(t_0, x_0) + \nabla\varphi(t_0, x_0)\cdot f(x_0, \omega) + L(x_0, \omega) \;>\; \theta \qquad \text{for all } \omega \in U. \tag{10.11}$$

In this case, we shall reach a contradiction by showing that no control function $u(\cdot)$ is good enough. Namely, whatever control function $u(\cdot)$ we choose on the initial interval $[t_0,\, t_0+\delta]$, even if during the remaining time $[t_0+\delta,\, T]$ our control is optimal, the total cost will still be considerably larger than $V(t_0, x_0)$. Indeed, by continuity, (10.11) implies

$$\varphi_t(t,x) + \nabla\varphi(t,x)\cdot f(x, \omega) \;>\; \theta - L(x, \omega) \qquad \text{for all } \omega \in U, \tag{10.12}$$

for all t, x close to t_0, x_0, i.e. such that (10.8) holds. Choose an arbitrary control function $u : [t_0,\, t_0+\delta] \mapsto U$, and call $t \mapsto x(t) = x(t; t_0, x_0, u)$ the corresponding trajectory. We now have

$$\begin{aligned} V\big(t_0+\delta,\; x(t_0+\delta)\big) - V(t_0, x_0) \;&\geq\; \varphi\big(t_0+\delta,\; x(t_0+\delta)\big) - \varphi(t_0, x_0)\\ &=\; \int_{t_0}^{t_0+\delta} \frac{d}{dt}\varphi\big(t,\, x(t)\big)\,dt\\ &=\; \int_{t_0}^{t_0+\delta} \varphi_t\big(t,\, x(t)\big) + \nabla\varphi\big(t,\, x(t)\big)\cdot f\big(x(t), u(t)\big)\,dt\\ &\geq\; \int_{t_0}^{t_0+\delta} \theta - L\big(x(t), u(t)\big)\,dt, \end{aligned} \tag{10.13}$$

because of (10.12). Therefore, for every control function $u(\cdot)$ we have

$$V\big(t_0+\delta,\ x(t_0+\delta)\big) + \int_{t_0}^{t_0+\delta} L\big(x(t),u(t)\big)\,dt \ \geq\ V(t_0,x_0)+\delta\theta. \tag{10.14}$$

Taking the infimum of the left hand side of (10.14) over all control functions u, we see that this infimum is still $\geq V(t_0,x_0)+\delta\theta$. On the other hand, by the Dynamic Programming Principle (9.11), the infimum should be exactly $V(t_0,x_0)$. This contradiction shows that (P2) must hold, completing the proof. □

One can combine Theorems 5.3 and 10.1, and obtain sufficient conditions for the optimality of a control function. The usual setting is the following. Consider the problem of minimizing the cost functional (9.4). Assume that, for each initial condition (s,y), we can guess a "candidate" optimal control $u^{s,y} : [s,T] \mapsto U$. We then call

$$\widetilde{V}(s,y) \ \doteq\ J(s,y,u^{s,y}) \tag{10.15}$$

the corresponding cost. Typically, these control functions $u^{s,y}$ are found by applying the Pontryagin Maximum Principle, which provides a necessary condition for optimality. On the other hand, consider the true value function V, defined at (9.8) as the infimum of the cost over all admissible control functions $u(\cdot) \in \mathcal{U}$. By Theorem 10.1, this function V provides a viscosity solution to the Hamilton-Jacobi equation (10.1) with terminal condition $V(T,y)=\psi(y)$. If our function $\widetilde{V}$ at (10.15) also provides a viscosity solution to the same equations (10.1)-(10.2), then by the uniqueness of the viscosity solution stated in Theorem 5.3, we can conclude that $\widetilde{V}=V$. Therefore, all controls $u^{s,y}$ are optimal.

We conclude this section by exhibiting a basic relation between the O.D.E. satisfied by extremal trajectories according to Theorem 7.1, and the P.D.E. of dynamic programming (10.1). Namely:

The trajectories which satisfy the Pontryagin Maximum Principle provide characteristic curves for the Hamilton-Jacobi equation of Dynamic Programming.

We shall justify the above claim, assuming that all functions involved are sufficiently smooth. As a first step, we derive the equations of characteristics, in connection with the evolution equation

$$V_t + H(x,\nabla V) \ =\ 0. \tag{10.16}$$

Call $p \doteq \nabla V$ the spatial gradient of V, so that $p=(p_1,\ldots,p_n)=(V_{x_1},\ \ldots,\ V_{x_n})$. Observe that

$$\frac{\partial^2 V}{\partial x_i \partial x_j} \ =\ \frac{\partial p_i}{\partial x_j} = \frac{\partial p_j}{\partial x_i}\,.$$

Differentiating (10.16) w.r.t. x_i one obtains

$$\frac{\partial p_i}{\partial t} \ =\ \frac{\partial^2 V}{\partial x_i \partial t} \ =\ -\frac{\partial H}{\partial x_i} - \sum_j \frac{\partial H}{\partial p_j}\,\frac{\partial p_i}{\partial x_j}\,. \tag{10.17}$$

If now $t \mapsto x(t)$ is any smooth curve, the total derivative of p_i along x is computed by

$$\begin{aligned} \frac{d}{dt} p_i\big(t, x(t)\big) &= \frac{\partial p_i}{\partial t} + \sum_j \dot x_j \frac{\partial p_i}{\partial x_j} \\ &= -\frac{\partial H}{\partial x_i} + \sum_j \left(\dot x_j - \frac{\partial H}{\partial p_j}\right) \frac{\partial p_i}{\partial x_j}\,. \end{aligned} \tag{10.18}$$

In general, the right hand side of (10.18) contains the partial derivatives $\partial p_i/\partial x_j$. However, if we choose the curve $t \mapsto x(t)$ so that $\dot x = \partial H/\partial p$, the last term will disappear. This observation lies at the heart of the classical method of characteristics. To construct a smooth solution of the equation (10.16) with terminal data

$$V(T, x) = \psi(x), \tag{10.19}$$

we proceed as follows. For each point $\bar x$, we find the solution to the Hamiltonian system of O.D.E's

$$\begin{cases} \dot x_i = \dfrac{\partial H}{\partial p_i}(x, p)\,, \\ \dot p_i = -\dfrac{\partial H}{\partial x_i}(x, p)\,, \end{cases} \qquad \begin{cases} x_i(T) = \bar x_i\,, \\ p_i(T) = \dfrac{\partial \psi}{\partial x_i}(\bar x)\,. \end{cases} \tag{10.20}$$

This solution will be denoted as

$$t \mapsto x(t, \bar x), \qquad t \mapsto p(t, \bar x)\,. \tag{10.21}$$

For every t we have $\nabla V\big(t,\, x(t, \bar x)\big) = p(t, \bar x)$. To recover the function V, we observe that along each solution of (10.20) one has

$$\frac{d}{dt} V\big(t,\, x(t, \bar x)\big) = V_t + \dot x \cdot \nabla V = -H(x, p) + p \cdot \frac{\partial H}{\partial p}\,. \tag{10.22}$$

Therefore

$$V\big(t,\, x(t, \bar x)\big) = \psi(\bar x) + \int_t^T \left(H(x, p) - p \cdot \frac{\partial H}{\partial p}\right) ds\,, \tag{10.23}$$

where the integral is computed along the solution (10.21).

Next, assume that the hamiltonian function H comes from a minimization problem, and is thus given by (10.3). To simplify our derivation, in the following we shall assume that optimal controls exist, and take values in the interior of the admissible set U. This last assumption is certainly true if $U = \mathbb{R}^m$. By (10.3) we now have

$$H(x, p) = p \cdot f\big(x,\, u^*(x, p)\big) + L(x,\, u^*(x, p)) = \min_\omega \big\{ p \cdot f(x, \omega) + L(x, \omega)\big\}\,, \tag{10.24}$$

where

$$u^*(x, p) = \arg\min_\omega \big\{ p \cdot f(x, \omega) + L(x, \omega)\big\}\,. \tag{10.25}$$

At the point u^* where the minimum is attained, since u^* lies in the interior of U one has

$$p \cdot \frac{\partial f}{\partial u}\big(x,\, u^*(x, p)\big) + \frac{\partial}{\partial u} L(x,\, u^*(x, p)) = 0\,.$$

Hence

$$\begin{aligned}\frac{\partial H}{\partial p}\big(x,\, u^*(x,p)\big) \; &= \; f\big(x, u^*(x,p)\big),\\ \frac{\partial H}{\partial x}\big(x,\, u^*(x,p)\big) = \; &p\cdot\frac{\partial f}{\partial x}\big(x,\, u^*(x,p)\big) + \frac{\partial L}{\partial x}\big(x,\, u^*(x,p)\big)\,.\end{aligned}$$

The Hamiltonian system (10.20) thus takes the form

$$\begin{cases}\dot{x} \; = \; f\big(x,\, u^*(x,p)\big)\,,\\ \dot{p} \; = \; -p\cdot\dfrac{\partial f}{\partial x}\big(x,\, u^*(x,p)\big) - \dfrac{\partial L}{\partial x}\big(x,\, u^*(x,p)\big)\,,\end{cases} \qquad \begin{cases}x(T) = \; \bar{x}\,,\\ p(T) = \; \nabla\psi(\bar{x})\,.\end{cases} \tag{10.26}$$

We observe that the evolution equations in (10.26) and the optimality conditions (10.25) are precisely those given in the Pontryagin Maximum Principle. In other words, let $t \mapsto u^*(t)$ be a control for which the Pontryagin Maximum Principle is satisfied. Then the corresponding trajectory $x(\cdot)$ and the adjoint vector $p(\cdot)$ provide a solution to the equations of characteristics for the corresponding hamiltonian system (10.16).

11 - References

- **General theory of P.D.E.**

[E] L. C. Evans, *Partial Differential Equations*, Amer. Math. Soc. Graduate Studies in Mathematics, Vol. 19, Providence 1998.

- **Control systems: basic theory and the Pontryagin Maximum Principle**

[BP] A. Bressan and B. Piccoli, *Introduction to the Mathematical Theory of Control*, AIMS Series in Applied Mathematics, Springfield Mo. 2007.

[Ce] L. Cesari, *Optimization - Theory and Applications*, Springer, 1983.

[FR] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer, 1975.

[LM] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*, Wiley, 1967.

[PBGM] L. S. Pontryagin, V. Boltyanskii, R. V. Gamkrelidze and E. F. Mishenko, *The Mathematical Theory of Optimal Processes*, Wiley, 1962.

- **Viscosity solutions to Hamilton-Jacobi equations**

[Ba] G. Barles, *Solutions de viscosité des équations de Hamilton-Jacobi*, Springer, 1994.

[BC] M. Bardi and I. Capuzzo Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser, 1997.

[CL] M. Crandall and P. L. Lions, Viscosity solutions of Hamilton-Jacobi equations, *Trans. Amer. Math. Soc.* **277** (1983), 1-42.

[CILL] M. Crandall, H. Ishii and P. L. Lions, User's guide to viscosity solutions of second order partial differential equations, *Bull. Amer. Math. Soc.* **27** (1992), 1-67.

[TTN] D. V. Tran, M. Tsuji and D. T. S. Nguyen, *The Characteristic Method and its Generalizations for First Order Nonlinear Partial Differential Equations*, Chapman&Hall/CRC, 1999.

- **Scalar conservation laws**

[K] S. Kruzhkov, First order quasilinear equations with several space variables, *Math. USSR Sbornik* **10** (1970), 217-243.