Partial Differential Equations

SECOND EDITION

Lawrence C. Evans

Department of Mathematics
University of California, Berkeley

Graduate Studies in Mathematics

Volume 19

American Mathematical Society
Providence, Rhode Island

3.2. CHARACTERISTICS

3.2.1. Derivation of characteristic ODE.

We return to our basic nonlinear first-order PDE

$$F(Du, u, x) = 0 \quad \text{in } U, \tag{1}$$

subject now to the boundary condition

$$u = g \quad \text{on } \Gamma, \tag{2}$$

where $\Gamma \subseteq \partial U$ and $g : \Gamma \to \mathbb{R}$ are given. We hereafter suppose that F, g are smooth functions.

We develop next the method of *characteristics*, which solves (1), (2) by converting the PDE into an appropriate system of ODE. This is the plan. Suppose u solves (1), (2) and fix any point $x \in U$. We would like to calculate $u(x)$ by finding some curve lying within U, connecting x with a point $x^0 \in \Gamma$ and along which we can compute u. Since (2) says $u = g$ on Γ, we know the value of u at the one end x^0. We hope then to be able to calculate u all along the curve, and so in particular at x.

Finding the characteristic ODE. How can we choose a path in U so all this will work? Let us suppose the curve is described parametrically by the function $\mathbf{x}(s) = (x^1(s), \dots, x^n(s))$, the parameter s lying in some subinterval $I \subseteq \mathbb{R}$. Assuming u is a C^2 solution of (1), we define also

$$(3) \qquad z(s) := u(\mathbf{x}(s)).$$

In addition, set

$$(4) \qquad \mathbf{p}(s) := Du(\mathbf{x}(s));$$

that is, $\mathbf{p}(s) = (p^1(s), \dots, p^n(s))$, where

$$(5) \qquad p^i(s) = u_{x_i}(\mathbf{x}(s)) \quad (i = 1, \dots, n).$$

So $z(\cdot)$ gives the values of u along the curve and $\mathbf{p}(\cdot)$ records the values of the gradient Du. We must choose the function $\mathbf{x}(\cdot)$ in such a way that we can compute $z(\cdot)$ and $\mathbf{p}(\cdot)$.

For this, first differentiate (5):

$$(6) \qquad \dot{p}^i(s) = \sum_{j=1}^n u_{x_i x_j}(\mathbf{x}(s))\dot{x}^j(s) \qquad \left(\cdot = \frac{d}{ds}\right).$$

This expression is not too promising, since it involves the second derivatives of u. On the other hand, we can also differentiate the PDE (1) with respect to x_i:

$$(7) \qquad \sum_{j=1}^n F_{p_j}(Du, u, x)u_{x_j x_i} + F_z(Du, u, x)u_{x_i} + F_{x_i}(Du, u, x) = 0.$$

We are able to employ this identity to get rid of the second derivative terms in (6), provided we first set

$$(8) \qquad \dot{x}^j(s) = F_{p_j}(\mathbf{p}(s), z(s), \mathbf{x}(s)) \quad (j = 1, \dots, n).$$

Assuming now (8) holds, we evaluate (7) at $x = \mathbf{x}(s)$, obtaining thereby from (3), (4) the identity:

$$\sum_{j=1}^n F_{p_j}(\mathbf{p}(s), z(s), \mathbf{x}(s))u_{x_i x_j}(\mathbf{x}(s)) \\ + F_z(\mathbf{p}(s), z(s), \mathbf{x}(s))p^i(s) + F_{x_i}(\mathbf{p}(s), z(s), \mathbf{x}(s)) = 0.$$

Substitute this expression and (8) into (6):

$$(9) \qquad \begin{aligned} \dot{p}^i(s) = &-F_{x_i}(\mathbf{p}(s), z(s), \mathbf{x}(s)) \\ &- F_z(\mathbf{p}(s), z(s), \mathbf{x}(s))p^i(s) \quad (i = 1, \dots, n). \end{aligned}$$

Finally we differentiate (3):

$$(10) \qquad \dot{z}(s) = \sum_{j=1}^n u_{x_j}(\mathbf{x}(s))\dot{x}^j(s) = \sum_{j=1}^n p^j(s)F_{p_j}(\mathbf{p}(s), z(s), \mathbf{x}(s)),$$

the second equality holding by (5) and (8).

The characteristic equations. We summarize by rewriting equations (8)–(10) in vector notation:

$$(11) \qquad \begin{cases} \text{(a) } \dot{\mathbf{p}}(s) = -D_x F(\mathbf{p}(s), z(s), \mathbf{x}(s)) - D_z F(\mathbf{p}(s), z(s), \mathbf{x}(s))\mathbf{p}(s) \\ \text{(b) } \dot{z}(s) = D_p F(\mathbf{p}(s), z(s), \mathbf{x}(s)) \cdot \mathbf{p}(s) \\ \text{(c) } \dot{\mathbf{x}}(s) = D_p F(\mathbf{p}(s), z(s), \mathbf{x}(s)). \end{cases}$$

Furthermore,

$$(12) \qquad F(\mathbf{p}(s), z(s), \mathbf{x}(s)) \equiv 0.$$

These identities hold for $s \in I$.

The important system (11) of $2n + 1$ first-order ODE comprises the *characteristic equations* of the nonlinear first-order PDE (1). The functions $\mathbf{p}(\cdot) = (p^1(\cdot), \dots, p^n(\cdot))$, $z(\cdot)$, $\mathbf{x}(\cdot) = (x^1(\cdot), \dots, x^n(\cdot))$ are called the *characteristics*. We will sometimes refer to $\mathbf{x}(\cdot)$ as the *projected characteristic*: it is the projection of the full characteristics $(\mathbf{p}(\cdot), z(\cdot), \mathbf{x}(\cdot)) \subset \mathbb{R}^{2n+1}$ onto the physical region $U \subset \mathbb{R}^n$.

We have proved:

THEOREM 1 (Structure of characteristic ODE). *Let $u \in C^2(U)$ solve the nonlinear, first-order partial differential equation* (1) *in U. Assume $\mathbf{x}(\cdot)$ solves the* ODE (11)(c), *where $\mathbf{p}(\cdot) = Du(\mathbf{x}(\cdot))$, $z(\cdot) = u(\mathbf{x}(\cdot))$. Then $\mathbf{p}(\cdot)$ solves the* ODE (11)(a) *and $z(\cdot)$ solves the* ODE (11)(b), *for those s such that $\mathbf{x}(s) \in U$.*

We still need to discover appropriate initial conditions for the system of ODE (11), in order that this theorem be useful. We accomplish this in §3.2.3 below.

Remark. The characteristic ODE are truly remarkable in that they form an exact system of equations for $\mathbf{x}(\cdot)$, $z(\cdot) = u(\mathbf{x}(\cdot))$, and $\mathbf{p}(\cdot) = Du(\mathbf{x}(\cdot))$, whenever u is a smooth solution of the general nonlinear PDE (1). The key step in the derivation is our setting $\dot{\mathbf{x}} = D_pF$, so that—as explained above—the terms involving second derivatives drop out. We thereby obtain *closure* and in particular are not forced to introduce ODE for the second and higher derivatives of u.

3.2.2. Examples.

Before continuing our investigation of the characteristic equations (11), we pause to consider some special cases for which the structure of these equations is especially simple. We illustrate as well how we can sometimes actually solve the characteristic ODE and thereby explicitly compute solutions of certain first-order PDE, subject to appropriate boundary conditions.

a. F linear. Consider first the situation that our PDE (1) is linear and homogeneous and thus has the form

(13) $$F(Du, u, x) = \mathbf{b}(x) \cdot Du(x) + c(x)u(x) = 0 \quad (x \in U).$$

Then $F(p, z, x) = \mathbf{b}(x) \cdot p + c(x)z$, and so

$$D_pF = \mathbf{b}(x).$$

In this circumstance equation (11)(c) becomes

(14) $$\dot{\mathbf{x}}(s) = \mathbf{b}(\mathbf{x}(s)),$$

an ODE involving only the function $\mathbf{x}(\cdot)$. Furthermore equation (11)(b) becomes

(15) $$\dot{z}(s) = \mathbf{b}(\mathbf{x}(s)) \cdot \mathbf{p}(s).$$

Then equation (12) simplifies (15), yielding

(16) $$\dot{z}(s) = -c(\mathbf{x}(s))z(s).$$

This ODE is linear in $z(\cdot)$, once we know the function $\mathbf{x}(\cdot)$ by solving (14). In summary,

(17) $$\begin{cases} \text{(a)} \quad \dot{\mathbf{x}}(s) = \mathbf{b}(\mathbf{x}(s)) \\ \text{(b)} \quad \dot{z}(s) = -c(\mathbf{x}(s))z(s) \end{cases}$$

comprise the characteristic equations for the linear, first-order PDE (13). (We will see later that the equation for $\mathbf{p}(\cdot)$ is not needed.) □

Example 1. We demonstrate the utility of equations (17) by explicitly solving the problem

(18) $$\begin{cases} x_1 u_{x_2} - x_2 u_{x_1} = u & \text{in } U \\ u = g & \text{on } \Gamma, \end{cases}$$

where U is the quadrant $\{x_1 > 0, x_2 > 0\}$ and $\Gamma = \{x_1 > 0, x_2 = 0\} \subseteq \partial U$. The PDE in (18) is of the form (12), for $\mathbf{b} = (-x_2, x_1)$ and $c = -1$. Thus the equations (17) read

(19) $$\begin{cases} \dot{x}^1 = -x^2, \ \dot{x}^2 = x^1 \\ \dot{z} = z. \end{cases}$$

Accordingly we have

$$\begin{cases} x^1(s) = x^0 \cos s, \ x^2(s) = x^0 \sin s \\ z(s) = z^0 e^s = g(x^0)\, e^s, \end{cases}$$

where $x^0 \geq 0$, $0 \leq s \leq \frac{\pi}{2}$. Fix a point $(x_1, x_2) \in U$. We select $s > 0$, $x^0 > 0$ so that $(x_1, x_2) = (x^1(s), x^2(s)) = (x^0 \cos s, x^0 \sin s)$. That is, $x^0 = (x_1^2 + x_2^2)^{1/2}$, $s = \arctan\left(\frac{x_2}{x_1}\right)$. Therefore

$$u(x) = u(x^1(s), x^2(s)) = z(s) = g(x^0)\, e^s = g((x_1^2 + x_2^2)^{1/2})\, e^{\arctan\left(\frac{x_2}{x_1}\right)}.$$

□

b. F quasilinear. The partial differential equation (1) is quasilinear should it have the form

(20) $$F(Du, u, x) = \mathbf{b}(x, u(x)) \cdot Du(x) + c(x, u(x)) = 0.$$

In this circumstance $F(p, z, x) = \mathbf{b}(x, z) \cdot p + c(x, z)$, whence

$$D_pF = \mathbf{b}(x, z).$$

Hence equation (11)(c) reads

$$\dot{\mathbf{x}}(s) = \mathbf{b}(\mathbf{x}(s), z(s)),$$

and (11)(b) becomes

$$\dot{z}(s) = \mathbf{b}(\mathbf{x}(s), z(s)) \cdot \mathbf{p}(s) = -c(\mathbf{x}(s), z(s)), \quad \text{by (12).}$$

Consequently

(21) $$\begin{cases} \text{(a)} \quad \dot{\mathbf{x}}(s) = \mathbf{b}(\mathbf{x}(s), z(s)) \\ \text{(b)} \quad \dot{z}(s) = -c(\mathbf{x}(s), z(s)) \end{cases}$$

are the characteristic equations for the quasilinear first-order PDE (20). (Once again we do not require the equation for $\mathbf{p}(\cdot)$.) □

Example 2. The characteristic ODE (21) are in general difficult to solve, and so we work out in this example the simpler case of a boundary-value problem for a semilinear PDE:

$$\text{(22)} \qquad \begin{cases} u_{x_1} + u_{x_2} = u^2 & \text{in } U \\ u = g & \text{on } \Gamma. \end{cases}$$

Now U is the half-space $\{x_2 > 0\}$ and $\Gamma = \{x_2 = 0\} = \partial U$. Here $\mathbf{b} = (1,1)$ and $c = -z^2$. Then (21) becomes

$$\begin{cases} \dot{x}^1 = 1, \ \dot{x}^2 = 1 \\ \dot{z} = z^2. \end{cases}$$

Consequently

$$\begin{cases} x^1(s) = x^0 + s, \ x^2(s) = s \\ z(s) = \frac{z^0}{1-sz^0} = \frac{g(x^0)}{1-sg(x^0)}, \end{cases}$$

where $x^0 \in \mathbb{R}$, $s \geq 0$, provided the denominator is not zero.

Fix a point $(x_1, x_2) \in U$. We select $s > 0$ and $x^0 \in \mathbb{R}$ so that $(x_1, x_2) = (x^1(s), x^2(s)) = (x^0 + s, s)$; that is, $x^0 = x_1 - x_2$, $s = x_2$. Then

$$u(x) = u(x^1(s), x^2(s)) = z(s) = \frac{g(x^0)}{1 - sg(x^0)} = \frac{g(x_1 - x_2)}{1 - x_2 g(x_1 - x_2)}.$$

This solution of course makes sense only if $1 - x_2 g(x_1 - x_2) \neq 0$. □

c. F fully nonlinear. In the general case, we must integrate the full characteristic equations (11), if possible.

Example 3. Consider the fully nonlinear problem

$$\text{(23)} \qquad \begin{cases} u_{x_1} u_{x_2} = u & \text{in } U \\ u = x_2^2 & \text{on } \Gamma, \end{cases}$$

where $U = \{x_1 > 0\}$, $\Gamma = \{x_1 = 0\} = \partial U$. Here $F(p, z, x) = p_1 p_2 - z$, and hence the characteristic ODE (11) become

$$\begin{cases} \dot{p}^1 = p^1, \ \dot{p}^2 = p^2 \\ \dot{z} = 2p^1 p^2 \\ \dot{x}^1 = p^2, \ \dot{x}^2 = p^1. \end{cases}$$

We integrate these equations to find

$$\begin{cases} x^1(s) = p_2^0(e^s - 1), \ x^2(s) = x^0 + p_1^0(e^s - 1) \\ z(s) = z^0 + p_1^0 p_2^0 (e^{2s} - 1) \\ p^1(s) = p_1^0 e^s, \ p^2(s) = p_2^0 e^s, \end{cases}$$

where $x^0 \in \mathbb{R}$, $s \in \mathbb{R}$, and $z^0 = (x^0)^2$.

We must determine $p^0 = (p_1^0, p_2^0)$. Since $u = x_2^2$ on Γ, $p_2^0 = u_{x_2}(0, x^0) = 2x^0$. Furthermore the PDE $u_{x_1} u_{x_2} = u$ itself implies $p_1^0 p_2^0 = z^0 = (x^0)^2$, and so $p_1^0 = \frac{x^0}{2}$. Consequently the formulas above become

$$\begin{cases} x^1(s) = 2x^0(e^s - 1), x^2(s) = \frac{x^0}{2}(e^s + 1) \\ z(s) = (x^0)^2 e^{2s} \\ p^1(s) = \frac{x^0}{2} e^s, \ p^2(s) = 2x^0 e^s. \end{cases}$$

Fix a point $(x_1, x_2) \in U$. Select s and x^0 so that $(x_1, x_2) = (x^1(s), x^2(s)) = (2x^0(e^s - 1), \frac{x^0}{2}(e^s + 1))$. This equality implies $x^0 = \frac{4x_2 - x_1}{4}$, $e^s = \frac{x_1 + 4x_2}{4x_2 - x_1}$; and so

$$u(x) = u(x^1(s), x^2(s)) = z(s) = (x^0)^2 e^{2s} = \frac{(x_1 + 4x_2)^2}{16}.$$

□

3.2.3. Boundary conditions.

We return now to developing the general theory and intend in the section following to invoke the characteristic ODE (11) actually to solve the boundary-value problem (1), (2), at least in a small region near an appropriate portion Γ of ∂U.

a. Straightening the boundary. To simplify subsequent calculations, it is convenient first to change variables, so as to "flatten out" part of the boundary ∂U. To accomplish this, we first fix any point $x^0 \in \partial U$. Then utilizing the notation from §C.1, we find smooth mappings $\mathbf{\Phi}, \mathbf{\Psi} : \mathbb{R}^n \to \mathbb{R}^n$ such that $\mathbf{\Psi} = \mathbf{\Phi}^{-1}$ and $\mathbf{\Phi}$ straightens out ∂U near x^0. (See the illustration in §C.1.)

Given any function $u : U \to \mathbb{R}$, let us write $V := \mathbf{\Phi}(U)$ and set

$$\text{(24)} \qquad v(y) := u(\mathbf{\Psi}(y)) \quad (y \in V).$$

Then

$$\text{(25)} \qquad u(x) = v(\mathbf{\Phi}(x)) \quad (x \in U).$$

Now suppose that u is a C^1 solution of our boundary-value problem (1), (2) in U. What PDE does v then satisfy in V?

According to (25), we see

$$u_{x_i}(x) = \sum_{k=1}^{n} v_{y_k}(\mathbf{\Phi}(x)) \Phi_{x_i}^k(x) \quad (i = 1, \ldots, n);$$

that is,

$$Du(x) = Dv(y)D\mathbf{\Phi}(x).$$

Thus (1) implies

$$F(Dv(y)D\mathbf{\Phi}(\mathbf{\Psi}(y)), v(y), \mathbf{\Psi}(y)) = F(Du(x), u(x), x) = 0. \tag{26}$$

This is an expression having the form

$$G(Dv(y), v(y), y) = 0 \quad \text{in } V.$$

In addition $v = h$ on Δ, where $\Delta := \mathbf{\Phi}(\Gamma)$ and $h(y) := g(\mathbf{\Psi}(y))$.

In summary, our problem (1), (2) transforms to read

$$\begin{cases} G(Dv, v, y) = 0 & \text{in } V \\ v = h & \text{on } \Delta, \end{cases} \tag{27}$$

for G, h as above. The point is that if we change variables to straighten out the boundary near x^0, the boundary-value problem (1), (2) converts into a problem having the same form.

b. Compatibility conditions on boundary data. In view of the foregoing computations, if we are given a point $x^0 \in \Gamma$, *we may as well assume from the outset that* Γ *is flat near* x^0, lying in the plane $\{x_n = 0\}$.

We intend now to utilize the characteristic ODE to construct a solution (1), (2), at least near x^0, and for this we must discover appropriate initial conditions

$$\mathbf{p}(0) = p^0, \; z(0) = z^0, \; \mathbf{x}(0) = x^0. \tag{28}$$

Now clearly if the curve $\mathbf{x}(\cdot)$ passes through x^0, we should insist that

$$z^0 = g(x^0). \tag{29}$$

What should we require concerning $\mathbf{p}(0) = p^0$? Since (2) implies $u(x_1, \dots x_{n-1}, 0) = g(x_1, \dots, x_{n-1})$ near x^0, we may differentiate to find

$$u_{x_i}(x^0) = g_{x_i}(x^0) \quad (i = 1, \dots, n-1).$$

As we also want the PDE (1) to hold, we should therefore insist $p^0 = (p_1^0, \dots, p_n^0)$ satisfies these relations:

$$\begin{cases} p_i^0 = g_{x_i}(x^0) & (i = 1, \dots, n-1) \\ F(p^0, z^0, x^0) = 0. \end{cases} \tag{30}$$

These identities provide n equations for the n quantities $p^0 = (p_1^0, \dots, p_n^0)$.

We call (29) and (30) the *compatibility conditions.* A triple $(p^0, z^0, x^0) \in \mathbb{R}^{2n+1}$ verifying (29), (30) is *admissible.* Note z^0 is uniquely determined by the boundary condition and our choice of the point x^0, but a vector p^0 satisfying (30) may not exist or may not be unique.

c. Noncharacteristic boundary data. So now assume as above that $x^0 \in \Gamma$, that Γ near x^0 lies in the plane $\{x_n = 0\}$, and that the triple (p^0, z^0, x^0) is admissible. We are planning to construct a solution u of (1), (2) in U near x^0 by integrating the characteristic ODE (11). So far we have ascertained $\mathbf{x}(0) = x^0$, $z(0) = z^0$, $\mathbf{p}(0) = p^0$ are appropriate boundary conditions for the characteristic ODE, with $\mathbf{x}(\cdot)$ intersecting Γ at x^0. But we will need in fact to solve these ODE for *nearby* initial points as well and must consequently now ask if we can somehow appropriately perturb (p^0, z^0, x^0), keeping the compatibility conditions.

In other words, given a point $y = (y_1, \dots, y_{n-1}, 0) \in \Gamma$, with y close to x^0, we intend to solve the characteristic ODE

$$\begin{cases} \text{(a) } \dot{\mathbf{p}}(s) = -D_x F(\mathbf{p}(s), z(s), \mathbf{x}(s)) - D_z F(\mathbf{p}(s), z(s), \mathbf{x}(s))\mathbf{p}(s) \\ \text{(b) } \dot{z}(s) = D_p F(\mathbf{p}(s), z(s), \mathbf{x}(s)) \cdot \mathbf{p}(s) \\ \text{(c) } \dot{\mathbf{x}}(s) = D_p F(\mathbf{p}(s), z(s), \mathbf{x}(s)), \end{cases} \tag{31}$$

with the initial conditions

$$\mathbf{p}(0) = \mathbf{q}(y), \; z(0) = g(y), \; \mathbf{x}(0) = y. \tag{32}$$

Our task then is to find a function $\mathbf{q}(\cdot) = (q^1(\cdot), \dots, q^n(\cdot))$, so that

$$\mathbf{q}(x^0) = p^0 \tag{33}$$

and $(\mathbf{q}(y), g(y), y)$ is admissible; that is, the compatibility conditions

$$\begin{cases} q^i(y) = g_{x_i}(y) & (i = 1, \dots, n-1) \\ F(\mathbf{q}(y), g(y), y) = 0 \end{cases} \tag{34}$$

hold for all $y \in \Gamma$ close to x^0.

LEMMA 1 (Noncharacteristic boundary conditions). *There exists a unique solution* $\mathbf{q}(\cdot)$ *of* (33), (34) *for all* $y \in \Gamma$ *sufficiently close to* x^0, *provided*

$$F_{p_n}(p^0, z^0, x^0) \neq 0. \tag{35}$$

We say the admissible triple (p^0, z^0, x^0) is *noncharacteristic* if (35) holds. We henceforth assume this condition.

Proof. Our problem is to find $q^n(y)$ so that

$$F(\mathbf{q}(y), g(y), y) = 0,$$

where $q^i(y) = g_{x_i}(y)$ for $i = 1, \dots, n-1$. Since $F(p^0, z^0, x^0) = 0$, the Implicit Function Theorem (§C.7) implies we can indeed locally and uniquely solve for $q^n(y)$, provided that the noncharacteristic condition (35) is valid. □

General noncharacteristic condition. If Γ is not flat near x^0, the condition that Γ be noncharacteristic reads

$$D_pF(p^0, z^0, x^0) \cdot \boldsymbol{\nu}(x^0) \neq 0, \tag{36}$$

$\boldsymbol{\nu}(x^0)$ denoting the outward unit normal to ∂U at x^0. See Problem 7.

3.2.4. Local solution.

Remember that our aim is to use the characteristic ODE to build a solution u of (1), (2), at least near Γ. So as before we select a point $x^0 \in \Gamma$ and, as shown in §3.2.3, may as well assume that near x^0 the surface Γ is flat, lying in the plane $\{x_n = 0\}$. Suppose further that (p^0, z^0, x^0) is an admissible triple of boundary data, which is noncharacteristic. According to Lemma 1 there is a function $\mathbf{q}(\cdot)$ so that $p^0 = \mathbf{q}(x^0)$ and the triple $(\mathbf{q}(y), g(y), y)$ is admissible, for all y sufficiently close to x^0.

Given any such point $y = (y_1, \dots, y_{n-1}, 0)$, we solve the characteristic ODE (31), subject to initial conditions (32).

NOTATION. Let us write

$$\begin{cases} \mathbf{p}(s) = \mathbf{p}(y,s) = \mathbf{p}(y_1, \dots, y_{n-1}, s) \\ z(s) = z(y,s) = z(y_1, \dots, y_{n-1}, s) \\ \mathbf{x}(s) = \mathbf{x}(y,s) = \mathbf{x}(y_1, \dots, y_{n-1}, s) \end{cases}$$

to display the dependence of the solution of (31), (32) on s and y. Also, we will henceforth when convenient regard x^0 as lying in $\mathbb{R}^{n-1}$. □

LEMMA 2 (Local invertibility). *Assume we have the noncharacteristic condition $F_{p_n}(p^0, z^0, x^0) \neq 0$. Then there exist an open interval $I \subseteq \mathbb{R}$ containing 0, a neighborhood W of x^0 in $\Gamma \subset \mathbb{R}^{n-1}$, and a neighborhood V of x^0 in $\mathbb{R}^n$, such that for each $x \in V$ there exist unique $s \in I$, $y \in W$ such that*

$$x = \mathbf{x}(y,s).$$

The mappings $x \mapsto s, y$ are C^2.

Proof. We have $\mathbf{x}(x^0, 0) = x^0$. Consequently the Inverse Function Theorem (§C.6) gives the result, provided $\det D\mathbf{x}(x^0, 0) \neq 0$. Now

$$\mathbf{x}(y, 0) = (y, 0) \quad (y \in \Gamma);$$

and so if $i = 1, \dots, n-1$,

$$x^j_{y_i}(x^0, 0) = \begin{cases} \delta_{ij} & (j = 1, \dots, n-1) \\ 0 & (j = n). \end{cases}$$

Furthermore equation (31)(c) implies

$$x^j_s(x^0, 0) = F_{p_j}(p^0, z^0, x^0).$$

Thus

$$D\mathbf{x}(x^0, 0) = \begin{pmatrix} 1 & & 0 & F_{p_1}(p^0, z^0, x^0) \\ & \ddots & & \vdots \\ 0 & & 1 & \vdots \\ 0 & \cdots & 0 & F_{p_n}(p^0, z^0, x^0) \end{pmatrix}_{n \times n},$$

whence $\det D\mathbf{x}(x^0, 0) \neq 0$ follows from the noncharacteristic condition (35). □

In view of Lemma 2 for each $x \in V$, we can locally uniquely solve the equation

$$\begin{cases} x = \mathbf{x}(y, s), \\ \text{for } \ y = \mathbf{y}(x), \ s = s(x). \end{cases} \tag{37}$$

Finally, let us define

$$\begin{cases} u(x) := z(\mathbf{y}(x), s(x)) \\ \mathbf{p}(x) := \mathbf{p}(\mathbf{y}(x), s(x)) \end{cases} \tag{38}$$

for $x \in V$ and s, y as in (37).

We come finally to our principal assertion, namely, that we can locally weave together the solutions of the characteristic ODE into a solution of the PDE.

THEOREM 2 (Local Existence Theorem). *The function u defined above is C^2 and solves the* PDE

$$F(Du(x), u(x), x) = 0 \quad (x \in V),$$

with the boundary condition

$$u(x) = g(x) \quad (x \in \Gamma \cap V).$$

Proof. 1. First of all, fix $y \in \Gamma$ close to x^0 and, as above, solve the characteristic ODE (31), (32) for $\mathbf{p}(s) = \mathbf{p}(y,s)$, $z(s) = z(y,s)$, and $\mathbf{x}(s) = \mathbf{x}(y,s)$.

2. We assert that if $y \in \Gamma$ is sufficiently close to x^0, then

$$f(y,s) := F(\mathbf{p}(y,s), z(y,s), \mathbf{x}(y,s)) = 0 \quad (s \in I). \tag{39}$$

To see this, note

$$f(y,0) = F(\mathbf{p}(y,0), z(y,0), \mathbf{x}(y,0)) = F(\mathbf{q}(y), g(y), y) = 0, \tag{40}$$

by the compatibility condition (34). Furthermore

$$\begin{aligned} f_s(y,s) &= \sum_{j=1}^n F_{p_j}\dot{p}^j + F_z\dot{z} + \sum_{j=1}^n F_{x_j}\dot{x}^j \\ &= \sum_{j=1}^n F_{p_j}(-F_{x_j} - F_z p^j) + F_z(\sum_{j=1}^n F_{p_j}p^j) \\ &\quad + \sum_{j=1}^n F_{x_j}(Fp_j) \quad \text{according to (31)} \\ &= 0. \end{aligned}$$

This calculation and (40) prove (39).

3. In view of Lemma 2 and (37)–(39), we have

$$F(\mathbf{p}(x), u(x), x) = 0 \quad (x \in V).$$

To conclude, we must therefore show

$$\mathbf{p}(x) = Du(x) \quad (x \in V). \tag{41}$$

In order to prove (41), let us first demonstrate for $s \in I$, $y \in W$ that

$$z_s(y,s) = \sum_{j=1}^n p^j(y,s)x_s^j(y,s) \tag{42}$$

and

$$z_{y_i}(y,s) = \sum_{j=1}^n p^j(y,s)x_{y_i}^j(y,s) \quad (i = 1,\dots,n-1). \tag{43}$$

These formulas are obviously consistent with the equality (41) and will later help us prove it. The identity (42) results at once from the characteristic ODE (31)(b),(c). To establish (43), fix $y \in \Gamma$, $i \in \{1,\dots,n-1\}$, and set

$$r^i(s) := z_{y_i}(y,s) - \sum_{j=1}^n p^j(y,s)x_{y_i}^j(y,s). \tag{44}$$

We first note $r^i(0) = g_{x_i}(y) - q^i(y) = 0$ according to the compatibility condition (34). In addition, we can compute

$$\dot{r}^i(s) = z_{y_i s} - \sum_{j=1}^n p_s^j x_{y_i}^j + p^j x_{y_i s}^j. \tag{45}$$

To simplify this expression, let us first differentiate the identity (42) with respect to y_i:

$$z_{sy_i} = \sum_{j=1}^n p_{y_i}^j x_s^j + p^j x_{sy_i}^j. \tag{46}$$

Substituting (46) into (45), we discover

$$\dot{r}^i(s) = \sum_{j=1}^n p_{y_i}^j x_s^j - p_s^j x_{y_i}^j = \sum_{j=1}^n p_{y_i}^j F_{p_j} - (-F_{x_j} - F_z p^j)x_{y_i}^j, \tag{47}$$

according to (31)(a). Now differentiate (39) with respect to y_i:

$$\sum_{j=1}^n F_{p_j}p_{y_i}^j + F_z z_{y_i} + \sum_{j=1}^n F_{x_j}x_{y_i}^j = 0.$$

We employ this identity in (47), thereby obtaining

$$\dot{r}^i(s) = F_z\Big(\sum_{j=1}^n p^j x_{y_i}^j - z_{y_i}\Big) = -F_z r^i(s). \tag{48}$$

Hence $r^i(\cdot)$ solves the linear ODE (48), with the initial condition $r^i(0) = 0$. Consequently $r^i(s) = 0$ ($s \in I$, $i = 1,\dots,n-1$), and so identity (43) is verified.

4. We finally employ (42), (43) in proving (41). Indeed, if $j = 1,\dots,n$,

$$\begin{aligned} u_{x_j} &= z_s s_{x_j} + \sum_{i=1}^{n-1} z_{y_i} y_{x_j}^i \quad \text{by (38)} \\ &= \sum_{k=1}^n p^k x_s^k s_{x_j} + \sum_{i=1}^{n-1}\sum_{k=1}^n p^k x_{y_i}^k y_{x_j}^i \quad \text{by (42), (43)} \\ &= \sum_{k=1}^n p^k\Big(x_s^k s_{x_j} + \sum_{i=1}^{n-1} x_{y_i}^k y_{x_j}^i\Big) \\ &= \sum_{k=1}^n p^k x_{x_j}^k = \sum_{k=1}^n p^k \delta_{jk} = p^j. \end{aligned}$$

This assertion at last establishes (41) and so finishes up the proof. □

3.2.5. Applications.

We turn now to various special cases, to see how the local existence theory simplifies in these circumstances.

a. F linear. Recall that a linear, homogeneous, first-order PDE has the form

$$F(Du,u,x) = \mathbf{b}(x)\cdot Du(x) + c(x)u(x) = 0 \quad (x \in U). \tag{49}$$

Our noncharacteristic assumption (36) at a point $x^0 \in \Gamma$ as above becomes

$$\mathbf{b}(x^0)\cdot \boldsymbol{\nu}(x^0) \neq 0 \tag{50}$$

and thus does not involve z^0 or p^0 at all. Furthermore if we specify the boundary condition

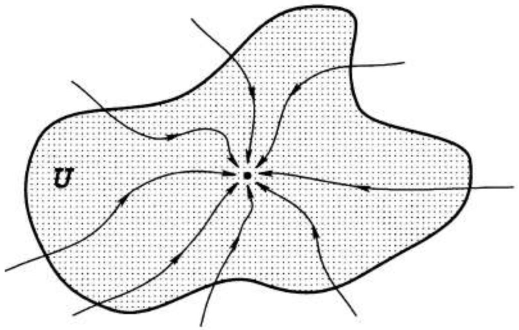
$$u = g \quad \text{on } \Gamma, \tag{51}$$

we can uniquely solve equation (34) for $\mathbf{q}(y)$ if $y \in \Gamma$ is near x^0. Thus we can apply the Local Existence Theorem 2 to construct a unique solution of (49), (51) in some neighborhood V containing x^0. Note carefully that although we have utilized the full characteristic equations (31) in the proof of Theorem 2, once we know the solution exists, we can use the reduced equations (17) (which do not involve $\mathbf{p}(\cdot)$) to compute the solution. Observe also that the projected characteristics $\mathbf{x}(\cdot)$ emanating from distinct points on Γ cannot cross, owing to uniqueness of solutions of the initial-value problem for the ODE (17)(a).

Example 4. Suppose the trajectories of the ODE

$$\dot{\mathbf{x}}(s) = \mathbf{b}(\mathbf{x}(s)) \tag{52}$$

are as drawn for Case 1. We are thus assuming the vector field $\mathbf{b}$ vanishes within U only at one point, which we will take to be the origin 0, and $\mathbf{b}\cdot\boldsymbol{\nu} < 0$ on $\Gamma := \partial U$.

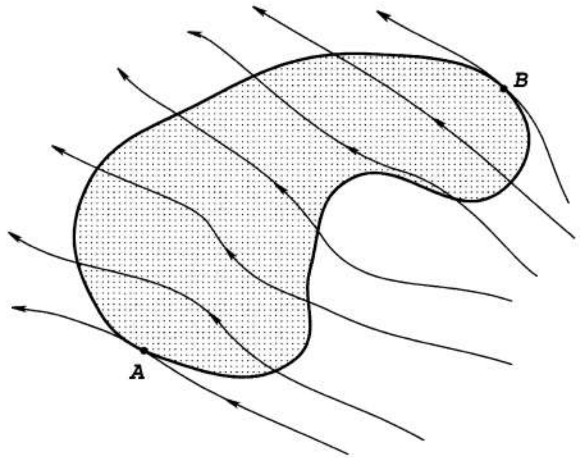


Case 1: flow to an attracting point

Can we solve the linear boundary-value problem

$$\begin{cases} \mathbf{b}\cdot Du = 0 & \text{in } U \\ u = g & \text{on } \Gamma\ ? \end{cases} \tag{53}$$

Invoking Theorem 2, we see that there exists a unique solution u defined *near* Γ and indeed that $u(\mathbf{x}(s)) \equiv u(\mathbf{x}(0)) = g(x^0)$ for each solution of the ODE (52), with the initial condition $\mathbf{x}(0) = x^0 \in \Gamma$. However, this solution cannot be smoothly continued to all of U (unless g is constant): any smooth solution of (53) is constant on trajectories of (52) and thus takes on different values near $x = 0$.



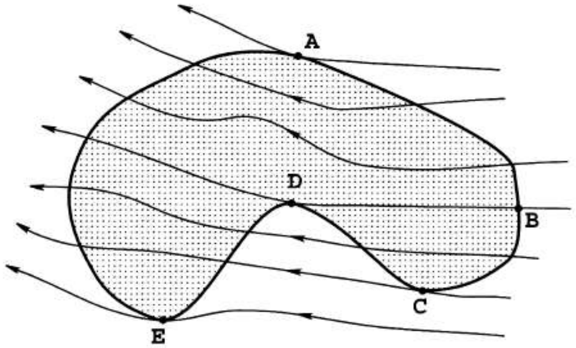
Case 2: flow across a domain

But now suppose the trajectories of the ODE (52) look like the illustration for Case 2. We are assuming that each trajectory of the ODE (except those through the characteristic points A, B) enters U precisely once, somewhere through the set

$$\Gamma := \{x \in \partial U \mid \mathbf{b}(x)\cdot\boldsymbol{\nu}(x) < 0\},$$

and exits U precisely once. In this circumstance we can find a smooth solution of (53) by setting u to be constant along each flow line.

Assume finally the flow looks like Case 3. We can now define u to be constant along trajectories, but then u will be discontinuous (unless $g(B) = g(D)$). Note that the point D is characteristic and that the local existence theory fails near D. □



Case 3: flow with characteristic points

b. F quasilinear. Should F be quasilinear, the PDE (1) is

$$F(Du, u, x) = \mathbf{b}(x, u) \cdot Du + c(x, u) = 0. \tag{54}$$

The noncharacteristic assumption (36) at a point $x^0 \in \Gamma$ reads $\mathbf{b}(x^0, z^0) \cdot \nu(x^0) \neq 0$, where $z^0 = g(x^0)$. As in the preceding example, if we specify the boundary condition

$$u = g \quad \text{on } \Gamma, \tag{55}$$

we can uniquely solve the equations (34) for $\mathbf{q}(y)$ if $y \in \Gamma$ near x^0. Thus Theorem 2 yields the existence of a unique solution of (54), (55) in some neighborhood V of x^0. We can compute this solution in V using the reduced characteristic equations (21), which do not explicitly involve $\mathbf{p}(\cdot)$.

In contrast to the linear case, however, *it is possible that the projected characteristics emanating from distinct points in* Γ *may intersect outside* V; such an occurrence usually signals the failure of our local solution to exist within all of U.

Example 5 (Characteristics for conservation laws). As an instance of a quasilinear first-order PDE, we turn now to the *scalar conservation law*

$$\begin{aligned} G(Du, u_t, u, x, t) &= u_t + \operatorname{div} \mathbf{F}(u) \\ &= u_t + \mathbf{F}'(u) \cdot Du = 0 \end{aligned} \tag{56}$$

in $U = \mathbb{R}^n \times (0, \infty)$, subject to the initial condition

$$u = g \quad \text{on } \Gamma = \mathbb{R}^n \times \{t = 0\}. \tag{57}$$

Here $\mathbf{F} : \mathbb{R} \to \mathbb{R}^n$, $\mathbf{F} = (F^1, \dots, F^n)$, and, as usual, we have set $t = x_{n+1}$. Also, "div" denotes the divergence with respect to the spatial variables $(x_1, \dots, x_n)$, and $Du = D_x u = (u_{x_1}, \dots, u_{x_n})$.

Since the direction $t = x_{n+1}$ plays a special role, we appropriately modify our notation. Writing now $q = (p, p_{n+1})$ and $y = (x, t)$, we have

$$G(q, z, y) = p_{n+1} + \mathbf{F}'(z) \cdot p,$$

and consequently

$$D_q G = (\mathbf{F}'(z), 1),\ D_y G = 0,\ D_z G = \mathbf{F}''(z) \cdot p.$$

Clearly the noncharacteristic condition (35) is satisfied at each point $y^0 = (x^0, 0) \in \Gamma$. Furthermore equation (21)(a) becomes

$$\begin{cases} \dot{x}^i(s) = F^{i\prime}(z(s)) \quad (i = 1, \dots, n) \\ \dot{x}^{n+1}(s) = 1. \end{cases} \tag{58}$$

Hence $x^{n+1}(s) = s$, in agreement with our having written $x_{n+1} = t$ above. In other words, we can identify the parameter s with the time t.

Equation (21)(b) reads $\dot{z}(s) = 0$. Consequently

$$z(s) = z^0 = g(x^0); \tag{59}$$

and (58) implies

$$\mathbf{x}(s) = \mathbf{F}'(g(x^0))s + x^0. \tag{60}$$

Thus the projected characteristic $\mathbf{y}(s) = (\mathbf{x}(s), s) = (\mathbf{F}'(g(x^0))s + x^0, s)$ $(s \geq 0)$ is a straight line, along which u is constant.

Crossing characteristics. But suppose now we apply the same reasoning to a different initial point $z^0 \in \Gamma$, where $g(x^0) \neq g(z^0)$. *The projected characteristics may possibly then intersect at some time* $t > 0$. Since Theorem 1 tells us $u \equiv g(x^0)$ on the projected characteristic through x^0 and $u \equiv g(z^0)$ on the projected characteristic through z^0, an apparent contradiction arises. The resolution is that *the initial-value problem* (56), (57) *does not in general have a smooth solution, existing for all times* $t > 0$. □

We will discuss in §3.4 the interesting possibility of extending the local solution (guaranteed to exist for short times by Theorem 2) to all times $t > 0$, as a kind of "weak" or "generalized" solution.

An implicit formula. We can eliminate s from equations (59), (60) to derive an implicit formula for u. Indeed given $x \in \mathbb{R}^n$ and $t > 0$, we see that since $s = t$,

$$u(\mathbf{x}(t), t) = z(t) = g(\mathbf{x}(t) - t\mathbf{F}'(z^0)) = g(\mathbf{x}(t) - t\mathbf{F}'(u(\mathbf{x}(t), t))).$$

Hence

$$u = g(x - t\mathbf{F}'(u)). \tag{61}$$

This implicit formula for u as a function of x and t is a nonlinear analogue of equation (3) in §2.1. It is easy to check that (61) does indeed give a solution, provided

$$1 + tDg(x - t\mathbf{F}'(u)) \cdot \mathbf{F}''(u) \neq 0.$$

In particular if $n = 1$, we require

$$1 + tg'(x - tF'(u))F''(u) \neq 0.$$

Note that if $F'' > 0$, but $g' < 0$, then this will definitely be false at some time $t > 0$. This failure of the implicit formula (61) reflects also the failure of the characteristic method. □

c. F fully nonlinear. The form of the full characteristic equations can be quite complicated for fully nonlinear first-order PDE, but sometimes a remarkable mathematical structure emerges.

Example 6 (Characteristics for the Hamilton–Jacobi equation). We look now at the general Hamilton–Jacobi PDE

$$G(Du, u_t, u, x, t) = u_t + H(Du, x) = 0, \tag{62}$$

where $Du = D_x u = (u_{x_1}, \dots, u_{x_n})$. Then writing $q = (p, p_{n+1})$, $y = (x, t)$, we have

$$G(q, z, y) = p_{n+1} + H(p, x);$$

and so

$$D_q G = (D_p H(p, x), 1),\ D_y G = (D_x H(p, x), 0),\ D_z G = 0.$$

Thus equation (11)(c) becomes

$$\begin{cases} \dot{x}^i(s) = H_{p_i}(\mathbf{p}(s), \mathbf{x}(s)) \quad (i = 1, \dots, n) \\ \dot{x}^{n+1}(s) = 1. \end{cases} \tag{63}$$

In particular we can identify the parameter s with the time t. Equation (11)(a) for the case at hand reads

$$\begin{cases} \dot{p}^i(s) = -H_{x_i}(\mathbf{p}(s), \mathbf{x}(s)) \quad (i = 1, \dots, n) \\ \dot{p}^{n+1}(s) = 0; \end{cases}$$

the equation (11)(b) is

$$\begin{aligned} \dot{z}(s) &= D_p H(\mathbf{p}(s), \mathbf{x}(s)) \cdot \mathbf{p}(s) + p^{n+1}(s) \\ &= D_p H(\mathbf{p}(s), \mathbf{x}(s)) \cdot \mathbf{p}(s) - H(\mathbf{p}(s), \mathbf{x}(s)). \end{aligned}$$

In summary, the characteristic equations for the Hamilton–Jacobi equation are

$$\begin{cases} \text{(a)} \quad \dot{\mathbf{p}}(s) = -D_x H(\mathbf{p}(s), \mathbf{x}(s)) \\ \text{(b)} \quad \dot{z}(s) = D_p H(\mathbf{p}(s), \mathbf{x}(s)) \cdot \mathbf{p}(s) - H(\mathbf{p}(s), \mathbf{x}(s)) \\ \text{(c)} \quad \dot{\mathbf{x}}(s) = D_p H(\mathbf{p}(s), \mathbf{x}(s)) \end{cases} \tag{64}$$

for $\mathbf{p}(\cdot) = (p^1(\cdot), \dots, p^n(\cdot))$, $z(\cdot)$, and $\mathbf{x}(\cdot) = (x^1(\cdot), \dots, x^n(\cdot))$.

The first and third of these equalities,

$$\begin{cases} \dot{\mathbf{x}} = D_p H(\mathbf{p}, \mathbf{x}) \\ \dot{\mathbf{p}} = -D_x H(\mathbf{p}, \mathbf{x}), \end{cases} \tag{65}$$

are called *Hamilton's equations*. We will discuss these ODE and their relationship to the Hamilton–Jacobi equation in much more detail, just below in §3.3. Observe that the equation for $z(\cdot)$ is trivial, once $\mathbf{x}(\cdot)$ and $\mathbf{p}(\cdot)$ have been found by solving Hamilton's equations. □

As for conservation laws (Example 5), the initial-value problem for the Hamilton–Jacobi equation does not in general have a smooth solution u lasting for all times $t > 0$.

3.3. INTRODUCTION TO HAMILTON–JACOBI EQUATIONS

In this section we study in some detail the initial-value problem for the Hamilton–Jacobi equation:

$$\begin{cases} u_t + H(Du) = 0 & \text{in } \mathbb{R}^n \times (0, \infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{1}$$

Here $u : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}$ is the unknown, $u = u(x, t)$, and $Du = D_x u = (u_{x_1}, \dots, u_{x_n})$. We are given the *Hamiltonian* $H : \mathbb{R}^n \to \mathbb{R}$ and the initial function $g : \mathbb{R}^n \to \mathbb{R}$.

Our goal is to find a formula for an appropriate weak or generalized solution, existing for all times $t > 0$, even after the method of characteristics has failed.

3.3.1. Calculus of variations, Hamilton's ODE.

Remember from §3.2.5 that two of the characteristic equations associated with the Hamilton–Jacobi PDE

$$u_t + H(Du, x) = 0$$

are Hamilton's ODE

$$\begin{cases} \dot{\mathbf{x}} = D_p H(\mathbf{p}, \mathbf{x}) \\ \dot{\mathbf{p}} = -D_x H(\mathbf{p}, \mathbf{x}), \end{cases}$$

which arise in the classical calculus of variations and in mechanics. (Note the x-dependence in H here.) In this section we recall the derivation of these ODE from a variational principle. We will then discover in §3.3.2 that this discussion contains a clue as to how to build a weak solution of the initial-value problem (1).

a. The calculus of variations. Assume that $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a given smooth function, hereafter called the *Lagrangian*.

NOTATION. We write

$$L = L(v, x) = L(v_1, \dots, v_n, x_1, \dots, x_n) \quad (v, x \in \mathbb{R}^n)$$

and

$$\begin{cases} D_v L = (L_{v_1} \cdots L_{v_n}) \\ D_x L = (L_{x_1} \cdots L_{x_n}). \end{cases}$$

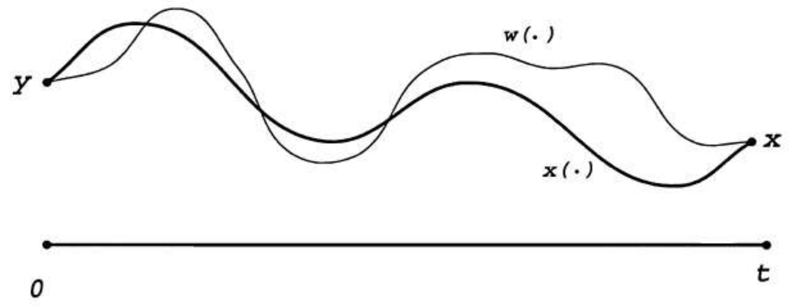
Thus in the formula (2) below "v" is the name of the variable for which we substitute $\dot{\mathbf{w}}(s)$, and "x" is the variable for which we substitute $\mathbf{w}(s)$. □

Now fix two points $x, y \in \mathbb{R}^n$ and a time $t > 0$. We introduce then the *action* functional

$$(2) \qquad I[\mathbf{w}(\cdot)] := \int_0^t L(\dot{\mathbf{w}}(s), \mathbf{w}(s))\, ds \quad \left(\cdot = \frac{d}{ds}\right),$$

defined for functions $\mathbf{w}(\cdot) = (w^1(\cdot), w^2(\cdot), \dots, w^n(\cdot))$ belonging to the *admissible class*

$$\mathcal{A} := \{\mathbf{w}(\cdot) \in C^2([0, t]; \mathbb{R}^n) \mid \mathbf{w}(0) = y, \mathbf{w}(t) = x\}.$$



A problem in the calculus of variations

Thus a C^2 curve $\mathbf{w}(\cdot)$ lies in $\mathcal{A}$ if it starts at the point y at time 0 and reaches the point x at time t.

A basic problem in the *calculus of variations* is to find a curve $\mathbf{x}(\cdot) \in \mathcal{A}$ satisfying

$$(3) \qquad I[\mathbf{x}(\cdot)] = \min_{\mathbf{w}(\cdot) \in \mathcal{A}} I[\mathbf{w}(\cdot)].$$

That is, we are asking for a function $\mathbf{x}(\cdot)$ which minimizes the functional $I[\cdot]$ among all admissible candidates $\mathbf{w}(\cdot) \in \mathcal{A}$.

We assume next that there in fact exists a function $\mathbf{x}(\cdot) \in \mathcal{A}$ satisfying our calculus of variations problem and will deduce some of its properties.

THEOREM 1 (Euler–Lagrange equations). *The function* $\mathbf{x}(\cdot)$ *solves the system of* Euler–Lagrange equations

$$(4) \qquad -\frac{d}{ds}\left(D_v L(\dot{\mathbf{x}}(s), \mathbf{x}(s))\right) + D_x L(\dot{\mathbf{x}}(s), \mathbf{x}(s)) = 0 \quad (0 \le s \le t).$$

This is a vector equation, consisting of n coupled second-order equations.

Proof. 1. Choose a smooth function $\mathbf{y} : [0, t] \to \mathbb{R}^n$, $\mathbf{y}(\cdot) = (y^1(\cdot), \dots, y^n(\cdot))$, satisfying

$$(5) \qquad \mathbf{y}(0) = \mathbf{y}(t) = 0,$$

and define for $\tau \in \mathbb{R}$

$$(6) \qquad \mathbf{w}(\cdot) := \mathbf{x}(\cdot) + \tau \mathbf{y}(\cdot).$$

Then $\mathbf{w}(\cdot) \in \mathcal{A}$ and so

$$I[\mathbf{x}(\cdot)] \le I[\mathbf{w}(\cdot)].$$

Thus the real-valued function

$$i(\tau) := I[\mathbf{x}(\cdot) + \tau\mathbf{y}(\cdot)]$$

has a minimum at $\tau = 0$, and consequently

$$i'(0) = 0 \quad \left(' = \frac{d}{d\tau}\right), \tag{7}$$

provided $i'(0)$ exists.

2. We explicitly compute this derivative. Observe

$$i(\tau) = \int_0^t L(\dot{\mathbf{x}}(s) + \tau\dot{\mathbf{y}}(s), \mathbf{x}(s) + \tau\mathbf{y}(s))\, ds,$$

and so

$$i'(\tau) = \int_0^t \sum_{i=1}^n L_{v_i}(\dot{\mathbf{x}} + \tau\dot{\mathbf{y}}, x + \tau\mathbf{y})\dot{y}^i + L_{x_i}(\dot{\mathbf{x}} + \tau\dot{\mathbf{y}}, x + \tau\mathbf{y})y^i\, ds.$$

Set $\tau = 0$ and remember (7):

$$0 = i'(0) = \int_0^t \sum_{i=1}^n L_{v_i}(\dot{\mathbf{x}}, \mathbf{x})\dot{y}^i + L_{x_i}(\dot{\mathbf{x}}, \mathbf{x})y^i\, ds.$$

We recall (5) and then integrate by parts in the first term inside the integral, to discover

$$0 = \sum_{i=1}^n \int_0^t \left[-\frac{d}{ds}\left(L_{v_i}(\dot{\mathbf{x}}, \mathbf{x})\right) + L_{x_i}(\dot{\mathbf{x}}, \mathbf{x})\right] y^i\, ds.$$

This identity is valid for all smooth functions $\mathbf{y}(\cdot) = (y^1(\cdot), \dots, y^n(\cdot))$ satisfying the boundary conditions (5), and so for $0 \le s \le t$

$$-\frac{d}{ds}\left(L_{v_i}(\dot{\mathbf{x}}, \mathbf{x})\right) + L_{x_i}(\dot{\mathbf{x}}, \mathbf{x}) = 0 \quad (i = 1, \dots, n).$$ □

Critical points. We have just demonstrated that any minimizer $\mathbf{x}(\cdot) \in \mathcal{A}$ of $I[\cdot]$ solves the Euler–Lagrange system of ODE. It is of course possible that a curve $\mathbf{x}(\cdot) \in \mathcal{A}$ may solve the Euler–Lagrange equations without necessarily being a minimizer: in this case we say $\mathbf{x}(\cdot)$ is a *critical point* of $I[\cdot]$. So every minimizer is a critical point, but a critical point need not be a minimizer.

Example. If $L(v, x) = \frac{1}{2}m|v|^2 - \phi(x)$, where $m > 0$, the corresponding Euler–Lagrange equation is

$$m\ddot{\mathbf{x}}(s) = \mathbf{f}(\mathbf{x}(s))$$

for $\mathbf{f} := -D\phi$. This is Newton's law for the motion of a particle of mass m moving in the force field $\mathbf{f}$ generated by the potential ϕ. (See Feynman–Leighton–Sands [**F-L-S,** Chapter 19].) □

b. Hamilton's equations. We now transform the Euler–Lagrange equations, a system of n second-order ODE, into Hamilton's equations, a system of $2n$ first-order ODE. We hereafter assume the C^2 function $\mathbf{x}(\cdot)$ is a critical point of the action functional and thus solves the Euler–Lagrange equations (4).

First we set

$$\mathbf{p}(s) := D_v L(\dot{\mathbf{x}}(s), \mathbf{x}(s)) \quad (0 \le s \le t); \tag{8}$$

$\mathbf{p}(\cdot)$ is called the *generalized momentum* corresponding to the *position* $\mathbf{x}(\cdot)$ and *velocity* $\dot{\mathbf{x}}(\cdot)$. We next make this important hypothesis:

$$\begin{cases} \text{Suppose for all } x, p \in \mathbb{R}^n \text{ that the equation} \\ \qquad p = D_v L(v, x) \\ \text{can be uniquely solved for } v \text{ as a smooth} \\ \text{function of } p \text{ and } x,\ v = \mathbf{v}(p, x). \end{cases} \tag{9}$$

We will examine this assumption in more detail later: see §3.3.2.

DEFINITION. *The* Hamiltonian H *associated with the Lagrangian* L *is*

$$H(p, x) := p \cdot \mathbf{v}(p, x) - L(\mathbf{v}(p, x), x) \quad (p, x \in \mathbb{R}^n),$$

where the function $\mathbf{v}(\cdot)$ *is defined implicitly by* (9).

Example (continued)**.** The Hamiltonian corresponding to the Lagrangian $L(v, x) = \frac{1}{2}m|v|^2 - \phi(x)$ is

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

The Hamiltonian is thus the total energy, the sum of the kinetic and potential energies (whereas the Lagrangian is the difference between the kinetic and potential energies). □

Next we rewrite the Euler–Lagrange equations in terms of $\mathbf{p}(\cdot), \mathbf{x}(\cdot)$:

THEOREM 2 (Derivation of Hamilton's ODE). *The functions* $\mathbf{x}(\cdot)$ *and* $\mathbf{p}(\cdot)$ *satisfy* Hamilton's equations*:*

$$\text{(10)} \qquad \begin{cases} \dot{\mathbf{x}}(s) = D_pH(\mathbf{p}(s), \mathbf{x}(s)) \\ \dot{\mathbf{p}}(s) = -D_xH(\mathbf{p}(s), \mathbf{x}(s)) \end{cases}$$

for $0 \leq s \leq t$. *Furthermore,*

$$\textit{the mapping } s \mapsto H(\mathbf{p}(s), \mathbf{x}(s)) \textit{ is constant.}$$

The equations (10) comprise a coupled system of $2n$ first-order ODE for $\mathbf{x}(\cdot) = (x^1(\cdot), \ldots, x^n(\cdot))$ and $\mathbf{p}(\cdot) = (p^1(\cdot), \ldots, p^n(\cdot))$ (defined by (8)).

Proof. First note from (8) and (9) that $\dot{\mathbf{x}}(s) = \mathbf{v}(\mathbf{p}(s), \mathbf{x}(s))$.

Let us hereafter write $\mathbf{v}(\cdot) = (v^1(\cdot), \ldots, v^n(\cdot))$. We compute for $i = 1, \ldots, n$ that

$$\begin{aligned} H_{x_i}(p, x) &= \sum_{k=1}^n p_k v^k_{x_i}(p, x) - L_{v_k}(\mathbf{v}(p, x), x) v^k_{x_i}(p, x) - L_{x_i}(\mathbf{v}(p, x), x) \\ &= -L_{x_i}(q, x) \quad \text{according to (9)} \end{aligned}$$

and

$$\begin{aligned} H_{p_i}(p, x) &= v^i(p, x) + \sum_{k=1}^n p_k v^k_{p_i}(p, x) - L_{v_k}(\mathbf{v}(p, x), x) v^k_{p_i}(p, x) \\ &= v^i(p, x), \quad \text{again by (9).} \end{aligned}$$

Thus

$$H_{p_i}(\mathbf{p}(s), \mathbf{x}(s)) = v^i(\mathbf{p}(s), \mathbf{x}(s)) = \dot{x}^i(s),$$

and likewise

$$\begin{aligned} H_{x_i}(\mathbf{p}(s), \mathbf{x}(s)) &= -L_{x_i}(\mathbf{v}(\mathbf{p}(s), \mathbf{x}(s)), \mathbf{x}(s)) = -L_{x_i}(\dot{\mathbf{x}}(s), \mathbf{x}(s)) \\ &= -\frac{d}{ds}\big(L_{v_i}(\dot{\mathbf{x}}(s), \mathbf{x}(s))\big) \text{ according to (4)} \\ &= -\dot{p}^i(s). \end{aligned}$$

Finally, observe

$$\begin{aligned} \frac{d}{ds}H(\mathbf{p}, \mathbf{x}) &= \sum_{i=1}^n H_{p_i}(\mathbf{p}, \mathbf{x})\dot{p}^i + H_{x_i}(\mathbf{p}, \mathbf{x})\dot{x}^i \\ &= \sum_{i=1}^n H_{p_i}(\mathbf{p}, \mathbf{x})(-H_{x_i}(\mathbf{p}, \mathbf{x})) + H_{x_i}(\mathbf{p}, \mathbf{x})H_{p_i}(\mathbf{p}, \mathbf{x}) = 0. \end{aligned}$$

□

See Arnold [**Ar1,** Chapter 9] for more on Hamilton's ODE and Hamilton–Jacobi PDE in classical mechanics. We are employing here different notation than is customary in mechanics: our notation is better overall for PDE theory.

3.3.2. Legendre transform, Hopf–Lax formula.

Now let us try to find a connection between the Hamilton–Jacobi PDE and the calculus of variations problem (2)–(4). To simplify further, we also drop the x-dependence in the Hamiltonian, so that afterwards $H = H(p)$. We start by reexamining the definition of the Hamiltonian in §3.3.1.

a. Legendre transform. We hereafter suppose the Lagrangian $L : \mathbb{R}^n \to \mathbb{R}$ satisfies these conditions:

$$\text{(11)} \qquad \text{the mapping } v \mapsto L(v) \text{ is convex}$$

and

$$\text{(12)} \qquad \lim_{|v| \to \infty} \frac{L(v)}{|v|} = +\infty.$$

The convexity implies L is continuous.

DEFINITION. *The* Legendre transform *of* L *is*

$$\text{(13)} \qquad L^*(p) := \sup_{q \in \mathbb{R}^n} \{p \cdot v - L(v)\} \quad (p \in \mathbb{R}^n).$$

This is also referred to as the *Fenchel transform.*

Motivation for Legendre transform. Why do we make this definition? For some insight let us note in view of (12) that the "sup" in (13) is really a "max"; that is, there exists some $v^* \in \mathbb{R}^n$ for which

$$L^*(p) = p \cdot v^* - L(v^*)$$

and the mapping $v \mapsto p \cdot v - L(v)$ has a maximum at $v = v^*$. But then $p = DL(v^*)$, provided L is differentiable at v^*. Hence the equation $p = DL(v)$ is solvable (although perhaps not uniquely) for v in terms of p, $v^* = \mathbf{v}(p)$. Therefore

$$L^*(p) = p \cdot \mathbf{v}(p) - L(\mathbf{v}(p)).$$

However, this is almost exactly the definition of the Hamiltonian H associated with L in §3.3.1 (where, recall, we are now assuming the variable x does not appear). We consequently henceforth write

$$\text{(14)} \qquad H = L^*.$$

Thus (13) tells us how to obtain the Hamiltonian H from the Lagrangian L.

Now we ask the converse question: given H, how do we compute L?

THEOREM 3 (Convex duality of Hamiltonian and Lagrangian). *Assume L satisfies* (11), (12) *and define H by* (13), (14).

(i) *Then*

$$\text{the mapping } p \mapsto H(p) \text{ is convex}$$

and

$$\lim_{|p|\to\infty} \frac{H(p)}{|p|} = +\infty.$$

(ii) *Furthermore*

$$L = H^*. \tag{15}$$

Thus H is the Legendre transform of L, and vice versa:

$$L = H^*, \ H = L^*.$$

We say H and L are *dual* convex functions. The identity (15) implies that the three statements

$$\begin{cases} p \cdot v = L(v) + H(p) \\ p = DL(v) \\ v = DH(p) \end{cases} \tag{16}$$

are equivalent provided H is differentiable at p and L is differentiable at v: see Problem 11.

Proof. 1. For each fixed v, the function $p \mapsto p \cdot v - L(v)$ is linear; and consequently the mapping

$$p \mapsto H(p) = L^*(p) = \sup_{v\in\mathbb{R}^n} \{p \cdot v - L(v)\}$$

is convex. Indeed, if $0 \le \tau \le 1$, $p, \hat{p} \in \mathbb{R}^n$, we have

$$\begin{aligned} H(\tau p + (1-\tau)\hat{p}) &= \sup_v \{(\tau p + (1-\tau)\hat{p}) \cdot v - L(v)\} \\ &\le \tau \sup_v \{p \cdot v - L(v)\} \\ &\quad + (1-\tau) \sup_v \{\hat{p} \cdot v - L(v)\} \\ &= \tau H(p) + (1-\tau) H(\hat{p}). \end{aligned}$$

2. Fix any $\lambda > 0$, $p \neq 0$. Then

$$\begin{aligned} H(p) &= \sup_{v\in\mathbb{R}^n} \{p \cdot v - L(v)\} \\ &\ge \lambda|p| - L(\lambda \frac{p}{|p|}) \quad (v = \lambda \frac{p}{|p|}) \\ &\ge \lambda|p| - \max_{B(0,\lambda)} L. \end{aligned}$$

Thus $\liminf_{|p|\to\infty} \frac{H(p)}{|p|} \ge \lambda$ for all $\lambda > 0$.

3. In view of (14)

$$H(p) + L(v) \ge p \cdot v$$

for all $p, v \in \mathbb{R}^n$, and consequently

$$L(v) \ge \sup_{p\in\mathbb{R}^n} \{p \cdot v - H(p)\} = H^*(v).$$

On the other hand

$$\begin{aligned} H^*(v) &= \sup_{p\in\mathbb{R}^n} \{p \cdot v - \sup_{r\in\mathbb{R}^n} \{p \cdot r - L(r)\}\} \\ &= \sup_{p\in\mathbb{R}^n} \inf_{r\in\mathbb{R}^n} \{p \cdot (v - r) + L(r)\}. \end{aligned}$$

Now since $v \mapsto L(v)$ is convex, according to §B.1 there exists $s \in \mathbb{R}^n$ such that

$$L(r) \ge L(v) + s \cdot (r - v) \quad (r \in \mathbb{R}^n).$$

(If L is differentiable at q, take $s = DL(v)$.) Putting $p = s$ above, we compute

$$H^*(v) \ge \inf_{r\in\mathbb{R}^n} \{s \cdot (v - r) + L(r)\} = L(v). \qquad \square$$

b. Hopf–Lax formula. Let us now return to the initial-value problem (1) for the Hamilton–Jacobi equation and conclude from (64) in §3.2.5 that the corresponding characteristic equations are

$$\begin{cases} \dot{\mathbf{p}} = 0 \\ \dot{z} = DH(\mathbf{p}) \cdot \mathbf{p} - H(\mathbf{p}) \\ \dot{\mathbf{x}} = DH(\mathbf{p}). \end{cases}$$

The first and third of these are Hamilton's ODE, which we in §3.3.1 derived from a minimization problem for associated Lagrangian $L = H^*$. Remembering (16), we can therefore understand the second of the characteristic equations as asserting

$$\dot{z} = DH(\mathbf{p}) \cdot \mathbf{p} - H(\mathbf{p}) = L(\dot{\mathbf{x}}).$$

But at least for such short times that (1) has a smooth solution u, we have $z(t) = u(\mathbf{x}(t), t)$ and consequently

$$u(x, t) = \int_0^t L(\dot{\mathbf{x}}(s))\, ds + g(\mathbf{x}(0)).$$

Our intention is to modify this expression, to make sense even for large times $t > 0$ when (1) does not have a smooth solution. The variational principle for the action discussed in §3.3.1 provides the clue. Given $x \in \mathbb{R}^n$ and $t > 0$, we therefore propose to *minimize* among curves $\mathbf{w}(\cdot)$ satisfying $\mathbf{w}(t) = x$ the expression

$$\int_0^t L(\dot{\mathbf{w}}(s))\,ds + g(\mathbf{w}(0)),$$

which is the action augmented with the value of the initial data. We accordingly now *define*

$$u(x,t) := \inf\left\{\int_0^t L(\dot{\mathbf{w}}(s))\,ds + g(\mathbf{w}(0)) \mid \mathbf{w}(t) = x\right\}, \tag{17}$$

the infimum taken over all C^1 functions $\mathbf{w}(\cdot)$. (Better justification for this guess will be provided much later, in Chapter 10.)

We must investigate the sense in which the function u given by (17) actually solves the initial-value problem for the Hamilton–Jacobi PDE:

$$\begin{cases} u_t + H(Du) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{18}$$

Recall we are assuming H is smooth,

$$\begin{cases} H \text{ is convex and} \\ \lim_{|p|\to\infty} \frac{H(p)}{|p|} = +\infty. \end{cases} \tag{19}$$

We henceforth suppose also

$$g : \mathbb{R}^n \to \mathbb{R} \quad \text{is Lipschitz continuous;} \tag{20}$$

this means $\operatorname{Lip}(g) := \sup_{\substack{x,y\in\mathbb{R}^n \\ x\neq y}} \left\{\frac{|g(x)-g(y)|}{|x-y|}\right\} < \infty$.

First we note that formula (17) can be simplified:

THEOREM 4 (Hopf–Lax formula). *If $x \in \mathbb{R}^n$ and $t > 0$, then the solution $u = u(x,t)$ of the minimization problem (17) is*

$$u(x,t) = \min_{y\in\mathbb{R}^n}\left\{tL\left(\frac{x-y}{t}\right) + g(y)\right\}. \tag{21}$$

DEFINITION. *We call the expression on the right-hand side of* (21) *the* Hopf–Lax formula.

Proof. 1. Fix any $y \in \mathbb{R}^n$ and define $\mathbf{w}(s) := y + \frac{s}{t}(x-y)$ $(0 \le s \le t)$. Then the definition (17) of u implies

$$u(x,t) \le \int_0^t L(\dot{\mathbf{w}}(s))\,ds + g(y) = tL\left(\frac{x-y}{t}\right) + g(y),$$

and so

$$u(x,t) \le \inf_{y\in\mathbb{R}^n}\left\{tL\left(\frac{x-y}{t}\right) + g(y)\right\}.$$

2. On the other hand, if $\mathbf{w}(\cdot)$ is any C^1 function satisfying $\mathbf{w}(t) = x$, we have

$$L\left(\frac{1}{t}\int_0^t \dot{\mathbf{w}}(s)\,ds\right) \le \frac{1}{t}\int_0^t L(\dot{\mathbf{w}}(s))\,ds$$

by Jensen's inequality (§B.1). Thus if we write $y = \mathbf{w}(0)$, we find

$$tL\left(\frac{x-y}{t}\right) + g(y) \le \int_0^t L(\dot{\mathbf{w}}(s))\,ds + g(y);$$

and consequently

$$\inf_{y\in\mathbb{R}^n}\left\{tL\left(\frac{x-y}{t}\right) + g(y)\right\} \le u(x,t).$$

3. We have so far shown

$$u(x,t) = \inf_{y\in\mathbb{R}^n}\left\{tL\left(\frac{x-y}{t}\right) + g(y)\right\},$$

and leave it as an exercise to prove that the infimum above is really a minimum. □

We now commence a study of various properties of the function u defined by the Hopf–Lax formula (21). Our ultimate goal is showing this formula provides a reasonable weak solution of the initial-value problem (18) for the Hamilton–Jacobi equation.

First, we record some preliminary observations.

LEMMA 1 (A functional identity). *For each $x \in \mathbb{R}^n$ and $0 \le s < t$, we have*

$$u(x,t) = \min_{y\in\mathbb{R}^n}\left\{(t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s)\right\}. \tag{22}$$

In other words, to compute $u(\cdot,t)$, we can calculate u at time s and then use $u(\cdot,s)$ as the initial condition on the remaining time interval $[s,t]$.

Proof. 1. Fix $y \in \mathbb{R}^n$, $0 < s < t$ and choose $z \in \mathbb{R}^n$ so that

$$u(y,s) = sL\left(\frac{y-z}{s}\right) + g(z). \tag{23}$$

Now since L is convex and $\frac{x-z}{t} = \left(1-\frac{s}{t}\right)\frac{x-y}{t-s} + \frac{s}{t}\frac{y-z}{s}$, we have

$$L\left(\frac{x-z}{t}\right) \le \left(1-\frac{s}{t}\right) L\left(\frac{x-y}{t-s}\right) + \frac{s}{t} L\left(\frac{y-z}{s}\right).$$

Thus

$$\begin{aligned} u(x,t) \le tL\left(\frac{x-z}{t}\right) + g(z) &\le (t-s)L\left(\frac{x-y}{t-s}\right) + sL\left(\frac{y-z}{s}\right) + g(z) \\ &= (t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s), \end{aligned}$$

by (23). This inequality is true for each $y \in \mathbb{R}^n$. Therefore, since $y \mapsto u(y,s)$ is continuous (according to the first part of the proof Lemma 2 below), we have

$$u(x,t) \le \min_{y \in \mathbb{R}^n} \left\{ (t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) \right\}. \tag{24}$$

2. Now choose w such that

$$u(x,t) = tL\left(\frac{x-w}{t}\right) + g(w), \tag{25}$$

and set $y := \frac{s}{t}x + \left(1-\frac{s}{t}\right)w$. Then $\frac{x-y}{t-s} = \frac{x-w}{t} = \frac{y-w}{s}$. Consequently

$$\begin{aligned} &(t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) \\ &\qquad \le (t-s)L\left(\frac{x-w}{t}\right) + sL\left(\frac{y-w}{s}\right) + g(w) \\ &\qquad = tL\left(\frac{x-w}{t}\right) + g(w) = u(x,t), \end{aligned}$$

by (25). Hence

$$\min_{y \in \mathbb{R}^n} \left\{ (t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) \right\} \le u(x,t). \tag{26}$$

□

LEMMA 2 (Lipschitz continuity). *The function u is Lipschitz continuous in $\mathbb{R}^n \times [0,\infty)$, and*

$$u = g \quad \text{on } \mathbb{R}^n \times \{t=0\}.$$

Proof. 1. Fix $t > 0$, $x, \hat{x} \in \mathbb{R}^n$. Choose $y \in \mathbb{R}^n$ such that

$$tL\left(\frac{x-y}{t}\right) + g(y) = u(x,t). \tag{27}$$

Then

$$\begin{aligned} u(\hat{x},t) - u(x,t) &= \min_z \left\{ tL\left(\frac{\hat{x}-z}{t}\right) + g(z) \right\} - tL\left(\frac{x-y}{t}\right) - g(y) \\ &\le g(\hat{x}-x+y) - g(y) \le \operatorname{Lip}(g)|\hat{x}-x|. \end{aligned}$$

Hence

$$u(\hat{x},t) - u(x,t) \le \operatorname{Lip}(g)|\hat{x}-x|;$$

and, interchanging the roles of $\hat{x}$ and x, we find

$$|u(x,t) - u(\hat{x},t)| \le \operatorname{Lip}(g)|x-\hat{x}|. \tag{28}$$

2. Now select $x \in \mathbb{R}^n$, $t > 0$. Choosing $y = x$ in (21), we discover

$$u(x,t) \le tL(0) + g(x). \tag{29}$$

Furthermore,

$$\begin{aligned} u(x,t) &= \min_{y \in \mathbb{R}^n} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\} \\ &\ge g(x) + \min_{y \in \mathbb{R}^n} \left\{ -\operatorname{Lip}(g)|x-y| + tL\left(\frac{x-y}{t}\right) \right\} \\ &= g(x) - t \max_{z \in \mathbb{R}^n} \{\operatorname{Lip}(g)|z| - L(z)\} \quad \left(z = \frac{x-y}{t}\right) \\ &= g(x) - t \max_{w \in B(0,\operatorname{Lip}(g))} \max_{z \in \mathbb{R}^n} \{w \cdot z - L(z)\} \\ &= g(x) - t \max_{B(0,\operatorname{Lip}(g))} H. \end{aligned}$$

This inequality and (29) imply

$$|u(x,t) - g(x)| \le Ct$$

for

$$C := \max(|L(0)|, \max_{B(0,\operatorname{Lip}(g))} |H|). \tag{30}$$

3. Finally select $x \in \mathbb{R}^n$, $0 < \hat{t} < t$. Then $\operatorname{Lip}(u(\cdot,t)) \leq \operatorname{Lip}(g)$ by (28) above. Consequently Lemma 1 and calculations like those employed in step 2 above imply

$$|u(x,t) - u(x,\hat{t})| \leq C|t - \hat{t}|$$

for the constant C defined by (30). □

Now Rademacher's Theorem (which we will prove later, in §5.8.3) asserts that a Lipschitz function is differentiable almost everywhere. Consequently in view of Lemma 2 our function u defined by the Hopf–Lax formula (21) is differentiable for a.e. $(x,t) \in \mathbb{R}^n \times (0,\infty)$. The next theorem asserts u in fact solves the Hamilton–Jacobi PDE wherever u is differentiable.

THEOREM 5 (Solving the Hamilton–Jacobi equation). *Suppose $x \in \mathbb{R}^n$, $t > 0$, and u defined by the Hopf–Lax formula (21) is differentiable at a point $(x,t) \in \mathbb{R}^n \times (0,\infty)$. Then*

$$u_t(x,t) + H(Du(x,t)) = 0.$$

Proof. 1. Fix $v \in \mathbb{R}^n$, $h > 0$. Owing to Lemma 1,

$$\begin{aligned} u(x+hv, t+h) &= \min_{y \in \mathbb{R}^n} \left\{ hL\left(\frac{x+hv-y}{h}\right) + u(y,t) \right\} \\ &\leq hL(v) + u(x,t). \end{aligned}$$

Hence

$$\frac{u(x+hv,t+h) - u(x,t)}{h} \leq L(v).$$

Let $h \to 0^+$, to compute

$$v \cdot Du(x,t) + u_t(x,t) \leq L(v).$$

This inequality is valid for all $v \in \mathbb{R}^n$, and so

(31) $$u_t(x,t) + H(Du(x,t)) = u_t(x,t) + \max_{v \in \mathbb{R}^n}\{v \cdot Du(x,t) - L(v)\} \leq 0.$$

The first equality holds since $H = L^*$.

2. Now choose z such that $u(x,t) = tL\left(\frac{x-z}{t}\right) + g(z)$. Fix $h > 0$ and set $s = t - h$, $y = \frac{s}{t}x + \left(1 - \frac{s}{t}\right)z$. Then $\frac{x-z}{t} = \frac{y-z}{s}$, and thus

$$\begin{aligned} u(x,t) - u(y,s) &\geq tL\left(\frac{x-z}{t}\right) + g(z) - \left[sL\left(\frac{y-z}{s}\right) + g(z)\right] \\ &= (t-s)L\left(\frac{x-z}{t}\right). \end{aligned}$$

That is,

$$\frac{u(x,t) - u\left(\left(1 - \frac{h}{t}\right)x + \frac{h}{t}z, t-h\right)}{h} \geq L\left(\frac{x-z}{t}\right).$$

Let $h \to 0^+$, to see that

$$\frac{x-z}{t} \cdot Du(x,t) + u_t(x,t) \geq L\left(\frac{x-z}{t}\right).$$

Consequently

$$\begin{aligned} u_t(x,t) + H(Du(x,t)) &= u_t(x,t) + \max_{v \in \mathbb{R}^n}\{v \cdot Du(x,t) - L(v)\} \\ &\geq u_t(x,t) + \frac{x-z}{t} \cdot Du(x,t) - L\left(\frac{x-z}{t}\right) \\ &\geq 0. \end{aligned}$$

This inequality and (31) complete the proof. □

We summarize:

THEOREM 6 (Hopf–Lax formula as solution). *The function u defined by the Hopf–Lax formula (21) is Lipschitz continuous, is differentiable a.e. in $\mathbb{R}^n \times (0,\infty)$, and solves the initial-value problem*

(32) $$\begin{cases} u_t + H(Du) = 0 & \text{a.e. in } \mathbb{R}^n \times (0,\infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

3.3.3. Weak solutions, uniqueness.

a. Semiconcavity. In view of Theorem 6 above it may seem reasonable to define a weak solution of the initial-value problem (18) to be a Lipschitz function which agrees with g on $\mathbb{R}^n \times \{t = 0\}$ and solves the PDE a.e. on $\mathbb{R}^n \times (0,\infty)$. However this turns out to be an inadequate definition, *as such weak solutions would not in general be unique.*

Example. Consider the initial-value problem

(33) $$\begin{cases} u_t + |u_x|^2 = 0 & \text{in } \mathbb{R} \times (0,\infty) \\ u = 0 & \text{on } \mathbb{R} \times \{t = 0\}. \end{cases}$$

One obvious solution is

$$u_1(x,t) \equiv 0.$$

However the function

$$u_2(x,t) := \begin{cases} 0 & \text{if } |x| \geq t \\ x - t & \text{if } 0 \leq x \leq t \\ -x - t & \text{if } -t \leq x \leq 0 \end{cases}$$

is Lipschitz continuous and also solves the PDE a.e. (everywhere, in fact, except on the lines $x = 0, \pm t$). It is easy to see that actually there are infinitely many Lipschitz functions satisfying (33). □

This example shows we must presumably require more of a weak solution than merely that it satisfy the PDE a.e. We will look to the Hopf–Lax formula (21) for a further clue as to what is needed to ensure uniqueness. The following lemma demonstrates that u inherits a kind of "one-sided" second-derivative estimate from the initial function g.

LEMMA 3 (Semiconcavity). *Suppose there exists a constant C such that*

$$g(x+z) - 2g(x) + g(x-z) \le C|z|^2 \tag{34}$$

for all $x, z \in \mathbb{R}^n$. Define u by the Hopf–Lax formula (21). Then

$$u(x+z,t) - 2u(x,t) + u(x-z,t) \le C|z|^2$$

for all $x, z \in \mathbb{R}^n$, $t > 0$.

We say g is *semiconcave* provided (34) holds. It is easy to check that (34) is valid if g is C^2 and $\sup_{\mathbb{R}^n} |D^2 g| < \infty$. Note that g is semiconcave if and only if the mapping $x \mapsto g(x) - \frac{C}{2}|x|^2$ is concave for some constant C.

Proof. Choose $y \in \mathbb{R}^n$ so that $u(x,t) = tL\left(\frac{x-y}{t}\right) + g(y)$. Then, putting $y+z$ and $y-z$ in the Hopf–Lax formulas for $u(x+z,t)$ and $u(x-z,t)$, we find

$$\begin{aligned} u(x+z,t) &- 2u(x,t) + u(x-z,t) \\ &\le \left[tL\left(\frac{x-y}{t}\right) + g(y+z)\right] - 2\left[tL\left(\frac{x-y}{t}\right) + g(y)\right] \\ &\quad + \left[tL\left(\frac{x-y}{t}\right) + g(y-z)\right] \\ &= g(y+z) - 2g(y) + g(y-z) \\ &\le C|z|^2, \quad \text{by (34).} \end{aligned}$$

□

As a semiconcavity condition for u will turn out to be important, we pause to identify some other circumstances under which it is valid. We will no longer assume g to be semiconcave but will suppose the Hamiltonian H to be uniformly convex.

DEFINITION. *A C^2 convex function $H : \mathbb{R}^n \to \mathbb{R}$ is called* uniformly convex *(with constant $\theta > 0$) if*

$$\sum_{i,j=1}^{n} H_{p_i p_j}(p)\xi_i\xi_j \ge \theta|\xi|^2 \quad \textit{for all } p, \xi \in \mathbb{R}^n. \tag{35}$$

We now prove that even if g is not semiconcave, the uniform convexity of H forces u to become semiconcave for times $t > 0$: this is a kind of mild regularizing effect for the Hopf–Lax solution of the initial-value problem (18).

LEMMA 4 (Semiconcavity again). *Suppose that H is uniformly convex (with constant θ) and u is defined by the Hopf–Lax formula (21). Then*

$$u(x+z,t) - 2u(x,t) + u(x-z,t) \le \frac{1}{\theta t}|z|^2$$

for all $x, z \in \mathbb{R}^n$, $t > 0$.

Proof. 1. We note first using Taylor's formula that (35) implies

$$H\left(\frac{p_1+p_2}{2}\right) \le \frac{1}{2}H(p_1) + \frac{1}{2}H(p_2) - \frac{\theta}{8}|p_1 - p_2|^2. \tag{36}$$

Next we claim that for the Lagrangian L we have the estimate

$$\frac{1}{2}L(v_1) + \frac{1}{2}L(v_2) \le L\left(\frac{v_1+v_2}{2}\right) + \frac{1}{8\theta}|v_1 - v_2|^2 \tag{37}$$

for all $v_1, v_2 \in \mathbb{R}^n$. Verification is left as an exercise.

2. Now choose y so that $u(x,t) = tL\left(\frac{x-y}{t}\right) + g(y)$. Then using the same value of y in the Hopf–Lax formulas for $u(x+z,t)$ and $u(x-z,t)$, we calculate

$$\begin{aligned} u(x+z,t) &- 2u(x,t) + u(x-z,t) \\ &\le \left[tL\left(\frac{x+z-y}{t}\right) + g(y)\right] - 2\left[tL\left(\frac{x-y}{t}\right) + g(y)\right] \\ &\quad + \left[tL\left(\frac{x-z-y}{t}\right) + g(y)\right] \\ &= 2t\left[\frac{1}{2}L\left(\frac{x+z-y}{t}\right) + \frac{1}{2}L\left(\frac{x-z-y}{t}\right) - L\left(\frac{x-y}{t}\right)\right] \\ &\le 2t\frac{1}{8\theta}\left|\frac{2z}{t}\right|^2 \le \frac{1}{\theta t}|z|^2, \end{aligned}$$

the next-to-last inequality following from (37). □

b. Weak solutions, uniqueness. In this section we show that semiconcavity conditions of the sorts discovered for the Hopf–Lax solution u in Lemmas 3 and 4 can be utilized as uniqueness criteria.

DEFINITION. *We say that a Lipschitz continuous function* $u : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}$ *is a* weak solution *of the initial-value problem:*

$$\begin{cases} u_t + H(Du) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t=0\} \end{cases} \tag{38}$$

provided

(a) $u(x,0) = g(x)$ $(x \in \mathbb{R}^n)$,

(b) $u_t(x,t) + H(Du(x,t)) = 0$ *for a.e.* $(x,t) \in \mathbb{R}^n \times (0,\infty)$,

and

(c) $u(x+z,t) - 2u(x,t) + u(x-z,t) \le C\left(1 + \frac{1}{t}\right)|z|^2$

for some constant $C \ge 0$ *and all* $x, z \in \mathbb{R}^n$, $t > 0$.

Next we prove that a weak solution of (38) is unique, the key point being that this uniqueness assertion follows from the *inequality* condition (c).

THEOREM 7 (Uniqueness of weak solutions). *Assume H is C^2 and satisfies* (19) *and g satisfies* (20). *Then there exists at most one weak solution of the initial-value problem* (38).

Proof*. 1. Suppose that u and $\tilde{u}$ are two weak solutions of (38) and write $w := u - \tilde{u}$.

Observe now that at any point (y,s) where both u and $\tilde{u}$ are differentiable and solve our PDE, we have

$$\begin{aligned} w_t(y,s) &= u_t(y,s) - \tilde{u}_t(y,s) \\ &= -H(Du(y,s)) + H(D\tilde{u}(y,s)) \\ &= -\int_0^1 \frac{d}{dr} H(rDu(y,s) + (1-r)D\tilde{u}(y,s))\,dr \\ &= -\int_0^1 DH(rDu(y,s) + (1-r)D\tilde{u}(y,s))\,dr \cdot (Du(y,s) - D\tilde{u}(y,s)) \\ &=: -\mathbf{b}(y,s) \cdot Dw(y,s). \end{aligned}$$

Consequently

$$w_t + \mathbf{b} \cdot Dw = 0 \quad \text{a.e.} \tag{39}$$

*Omit on first reading.

2. Write $v := \phi(w) \ge 0$, where $\phi : \mathbb{R} \to [0,\infty)$ is a smooth function to be selected later. We multiply (39) by $\phi'(w)$ to discover

$$v_t + \mathbf{b} \cdot Dv = 0 \quad \text{a.e.} \tag{40}$$

3. Now choose $\varepsilon > 0$ and define $u^\varepsilon := \eta_\varepsilon * u$, $\tilde{u}^\varepsilon := \eta_\varepsilon * \tilde{u}$, where η_ε is the standard mollifier in the x and t variables. Then according to §C.4

$$|Du^\varepsilon| \le \operatorname{Lip}(u), \ |D\tilde{u}^\varepsilon| \le \operatorname{Lip}(\tilde{u}), \tag{41}$$

and

$$Du^\varepsilon \to Du, \ D\tilde{u}^\varepsilon \to D\tilde{u} \quad \text{a.e., as } \epsilon \to 0. \tag{42}$$

Furthermore inequality (c) in the definition of weak solution implies

$$D^2u^\varepsilon, D^2\tilde{u}^\varepsilon \le C\left(1 + \frac{1}{s}\right) I \tag{43}$$

for an appropriate constant C and all $\varepsilon > 0$, $y \in \mathbb{R}^n$, $s > 2\varepsilon$. Verification is left as an exercise.

4. Write

$$\mathbf{b}_\varepsilon(y,s) := \int_0^1 DH(rDu^\varepsilon(y,s) + (1-r)D\tilde{u}^\varepsilon(y,s))\,dr. \tag{44}$$

Then (40) becomes

$$v_t + \mathbf{b}_\varepsilon \cdot Dv = (\mathbf{b}_\varepsilon - \mathbf{b}) \cdot Dv \quad \text{a.e.};$$

hence

$$v_t + \operatorname{div}(v\mathbf{b}_\varepsilon) = (\operatorname{div} \mathbf{b}_\varepsilon)v + (\mathbf{b}_\varepsilon - \mathbf{b}) \cdot Dv \quad \text{a.e.} \tag{45}$$

5. Now

$$\begin{aligned} \operatorname{div} \mathbf{b}_\varepsilon &= \int_0^1 \sum_{k,l=1}^n H_{p_k p_l}(rDu^\varepsilon + (1-r)D\tilde{u}^\varepsilon)(ru^\varepsilon_{x_l x_k} + (1-r)\tilde{u}^\varepsilon_{x_l x_k})\,dr \\ &\le C\left(1 + \frac{1}{s}\right) \end{aligned} \tag{46}$$

for some constant C, in view of (41), (43). Here we note that H convex implies $D^2H \ge 0$.

6. Fix $x_0 \in \mathbb{R}^n$, $t_0 > 0$, and set

$$R := \max\{|DH(p)| \mid |p| \le \max(\mathrm{Lip}(u), \mathrm{Lip}(\tilde{u}))\}. \tag{47}$$

Define also the cone

$$C := \{(x,t) \mid 0 \le t \le t_0, |x - x_0| \le R(t_0 - t)\}.$$

Next write

$$e(t) = \int_{B(x_0, R(t_0-t))} v(x,t)\,dx$$

and compute for a.e. $t > 0$:

$$\begin{aligned}
\dot{e}(t) &= \int_{B(x_0,R(t_0-t))} v_t\,dx - R\int_{\partial B(x_0,R(t_0-t))} v\,dS \\
&= \int_{B(x_0,R(t_0-t))} -\operatorname{div}(v\mathbf{b}_\varepsilon) + (\operatorname{div}\mathbf{b}_\varepsilon)v + (\mathbf{b}_\varepsilon - \mathbf{b})\cdot Dv\,dx \\
&\quad - R\int_{\partial B(x_0,R(t_0-t))} v\,dS \quad \text{by (45)} \\
&= -\int_{\partial B(x_0,R(t_0-t))} v(\mathbf{b}_\varepsilon \cdot \nu + R)\,dS \\
&\quad + \int_{B(x_0,R(t_0-t))} (\operatorname{div}\mathbf{b}_\varepsilon)v + (\mathbf{b}_\varepsilon - \mathbf{b})\cdot Dv\,dx \\
&\le \int_{B(x_0,R(t_0-t))} (\operatorname{div}\mathbf{b}_\varepsilon)v + (\mathbf{b}_\varepsilon - \mathbf{b})\cdot Dv\,dx \quad \text{by (41), (44)} \\
&\le C\left(1 + \frac{1}{t}\right)e(t) + \int_{B(x_0,R(t_0-t))} (\mathbf{b}_\varepsilon - \mathbf{b})\cdot Dv\,dx
\end{aligned}$$

by (46). The last term on the right-hand side goes to zero as $\varepsilon \to 0$, for a.e. $t > 0$, according to (41), (42) and the Dominated Convergence Theorem. Thus

$$\dot{e}(t) \le C\left(1 + \frac{1}{t}\right)e(t) \quad \text{for a.e. } 0 < t < t_0. \tag{48}$$

7. Fix $0 < \varepsilon < r < t_0$ and choose the function $\phi(z)$ to equal zero if

$$|z| \le \varepsilon[\mathrm{Lip}(u) + \mathrm{Lip}(\tilde{u})]$$

and to be positive otherwise. Since $u \equiv \tilde{u}$ on $\mathbb{R}^n \times \{t = 0\}$,

$$v = \phi(w) = \phi(u - \tilde{u}) = 0 \quad \text{at } \{t = \varepsilon\}.$$

Thus $e(\varepsilon) = 0$. Consequently Gronwall's inequality (§B.2) and (48) imply

$$e(r) \le e(\varepsilon)e^{\int_\varepsilon^r C(1+\frac{1}{s})ds} = 0.$$

Hence

$$|u - \tilde{u}| \le \varepsilon[\mathrm{Lip}(u) + \mathrm{Lip}(\tilde{u})] \quad \text{on } B(x_0, R(t_0 - r)).$$

This inequality is valid for all $\varepsilon > 0$, and so $u \equiv \tilde{u}$ in $B(x_0, R(t_0 - r))$. Therefore, in particular, $u(x_0, t_0) = \tilde{u}(x_0, t_0)$. □

In light of Lemmas 3, 4 and Theorem 7, we have

THEOREM 8 (Hopf–Lax formula as weak solution). *Suppose H is C^2 and satisfies* (19) *and g satisfies* (20). *If either g is semiconcave or H is uniformly convex, then*

$$u(x,t) = \min_{y \in \mathbb{R}^n}\left\{tL\left(\frac{x-y}{t}\right) + g(y)\right\}$$

is the unique weak solution of the initial-value problem (38) *for the Hamilton–Jacobi equation.*

Examples. (i) Consider the initial-value problem:

$$\begin{cases} u_t + \frac{1}{2}|Du|^2 = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = |x| & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{49}$$

Here $H(p) = \frac{1}{2}|p|^2$ and so $L(v) = \frac{1}{2}|v|^2$. The Hopf–Lax formula for the unique, weak solution of (49) is

$$u(x,t) = \min_{y \in \mathbb{R}^n}\left\{\frac{|x-y|^2}{2t} + |y|\right\}. \tag{50}$$

Assume $|x| > t$. Then

$$D_y\left(\frac{|x-y|^2}{2t} + |y|\right) = \frac{y-x}{t} + \frac{y}{|y|} \quad (y \ne 0);$$

and this expression equals zero if $x = y + \frac{y}{|y|}t$, $y = (|x| - t)\frac{x}{|x|} \ne 0$. Thus $u(x,t) = |x| - \frac{t}{2}$ if $|x| > t$. If $|x| \le t$, the minimum in (50) is attained at $y = 0$. Consequently

$$u(x,t) = \begin{cases} |x| - t/2 & \text{if } |x| \ge t \\ \frac{|x|^2}{2t} & \text{if } |x| \le t. \end{cases}$$

Observe that the solution becomes semiconcave at times $t > 0$, even though the initial function $g(x) = |x|$ is not semiconcave. This accords with Lemma 4.

(ii) We next examine the problem with reversed initial conditions:

(51)
$$\begin{cases} u_t + \frac{1}{2}|Du|^2 = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = -|x| & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

Then
$$u(x,t) = \min_{y \in \mathbb{R}^n} \left\{ \frac{|x-y|^2}{2t} - |y| \right\}.$$

Now
$$D_y\left(\frac{|x-y|^2}{2t} - |y|\right) = \frac{y-x}{t} - \frac{y}{|y|} \quad (y \neq 0),$$

and this equals zero if $x = y - \frac{y}{|y|}t$, $y = (|x| + t)\frac{x}{|x|}$. Thus

$$u(x,t) = -|x| - \frac{t}{2} \quad (x \in \mathbb{R}^n,\ t \geq 0).$$

The initial function $g(x) = -|x|$ is semiconcave, and the solution remains so for times $t > 0$. □

In Chapter 10 we will again study Hamilton–Jacobi PDE and discover another and better notion of weak solution, applicable even if H is not convex.

3.5. PROBLEMS

In the following exercises, all given functions are assumed smooth, unless otherwise stated.

1. Prove
$$u(x,t,a,b) = a \cdot x - tH(a) + b \quad (a \in \mathbb{R}^n, b \in \mathbb{R})$$
is a complete integral of the Hamilton–Jacobi equation
$$u_t + H(Du) = 0.$$

2. Compute the envelopes of the family of lines
$$x_1 + a^2 x_2 - 2a = 0 \quad (a \in \mathbb{R})$$
in $\mathbb{R}^2$ and of the family of planes
$$2a_1 x_1 + 2a_2 x_2 - x_3 + a_1^2 + a_2^2 = 0 \quad (a_1, a_2 \in \mathbb{R})$$

in $\mathbb{R}^3$. Draw pictures illustrating the geometric meaning of the envelopes.

3. Suppose that the formula $G(x,z,a)=0$ implicitly defines the function $z=u(x,a)$, where $x,a\in\mathbb{R}^n$. Assume further that we can eliminate the variables a from the identities

$$\begin{cases} G(x,u,a)=0 \\ G_{x_i}(x,u,a)+G_z(x,u,a)u_{x_i}=0 \quad (i=1,\dots,n), \end{cases}$$

to solve for $u=u(x)$.

(a) Find a PDE that u solves if $G=\sum_{i=1}^n a_i x_i^2+z^3$.

(b) What is the PDE characterizing all spheres in $\mathbb{R}^{n+1}$ with unit radius and center in $\mathbb{R}^n\times\{z=0\}$?

4. (a) Write down the characteristic equations for the PDE

$$(*) \qquad u_t+b\cdot Du=f \quad \text{in } \mathbb{R}^n\times(0,\infty),$$

where $b\in\mathbb{R}^n$, $f=f(x,t)$.

(b) Use the characteristic ODE to solve $(*)$ subject to the initial condition

$$u=g \quad \text{on } \mathbb{R}^n\times\{t=0\}.$$

Make sure your answer agrees with formula (5) in §2.1.2.

5. Solve using characteristics:

(a) $x_1u_{x_1}+x_2u_{x_2}=2u, \quad u(x_1,1)=g(x_1)$.

(b) $x_1u_{x_1}+2x_2u_{x_2}+u_{x_3}=3u, \quad u(x_1,x_2,0)=g(x_1,x_2)$.

(c) $uu_{x_1}+u_{x_2}=1, \quad u(x_1,x_1)=\frac{1}{2}x_1$.

6. Given a smooth vector field $\mathbf{b}$ on $\mathbb{R}^n$, let $\mathbf{x}(s)=\mathbf{x}(s,x,t)$ solve the ODE

$$\begin{cases} \dot{\mathbf{x}}=\mathbf{b}(\mathbf{x}) \quad (s\in\mathbb{R}) \\ \mathbf{x}(t)=x. \end{cases}$$

(a) Define the Jacobian

$$J(s,x,t):=\det D_x\mathbf{x}(s,x,t)$$

and derive *Euler's formula*:

$$J_s=\operatorname{div}\mathbf{b}(\mathbf{x})J.$$

(b) Demonstrate that

$$u(x,t):=g(\mathbf{x}(0,x,t))J(0,x,t)$$

solves

$$\begin{cases} u_t+\operatorname{div}(u\mathbf{b})=0 & \text{in } \mathbb{R}^n\times(0,\infty) \\ u=g & \text{on } \mathbb{R}^n\times\{t=0\}. \end{cases}$$

(Hint: Show $\frac{\partial}{\partial s}(u(\mathbf{x},s)J)=0$.)

7. Verify assertion (36) in §3.2.3, that when Γ is not flat near x^0, the noncharacteristic condition is

$$D_pF(p^0,z^0,x^0)\cdot\boldsymbol{\nu}(x^0)\neq 0.$$

8. Confirm that the formula $u=g(x-t\mathbf{F}'(u))$ from §3.2.5 provides an implicit solution for the conservation law

$$u_t+F(u)_x=0.$$

9. Consider the problem of minimizing the action $\int_0^t L(\dot{\mathbf{w}}(s),\mathbf{w}(s))\,ds$ over the new admissible class

$$\mathcal{A}:=\{\mathbf{w}(\cdot)\in C^2([0,t];\mathbb{R}^n)\mid \mathbf{w}(t)=x\},$$

where we do not require that $\mathbf{w}(0)=y$.

(a) Show that a minimizer $\mathbf{x}(\cdot)\in\mathcal{A}$ solves the Euler–Lagrange equations

$$-\frac{d}{ds}\left(D_vL(\dot{\mathbf{x}}(s),\mathbf{x}(s))\right)+D_xL(\dot{\mathbf{x}}(s),\mathbf{x}(s))=0 \quad (0\le s\le t).$$

(b) Prove that

$$D_vL(\dot{\mathbf{x}}(0),\mathbf{x}(0))=0.$$

(c) Suppose now that $\mathbf{x}(\cdot)\in\mathcal{A}$ minimizes the modified action

$$\int_0^t L(\dot{\mathbf{w}}(s),\mathbf{w}(s))\,ds+g(\mathbf{w}(0)).$$

Show that $\mathbf{x}(\cdot)$ solves the usual Euler–Lagrange equations and determine the boundary condition at $s=0$.

10. If $H:\mathbb{R}^n\to\mathbb{R}$ is convex, we write $L=H^*$.

(a) Let $H(p)=\frac{1}{r}|p|^r$, for $1<r<\infty$. Show

$$L(v)=\frac{1}{s}|v|^s, \quad \text{where } \frac{1}{r}+\frac{1}{s}=1.$$

(b) Let $H(p)=\frac{1}{2}\sum_{i,j=1}^n a_{ij}p_ip_j+\sum_{i=1}^n b_ip_i$, where $A=((a_{ij}))$ is a symmetric, positive definite matrix, $b\in\mathbb{R}^n$. Compute $L(v)$.

11. Let $H : \mathbb{R}^n \to \mathbb{R}$ be convex. We say v belongs to the *subdifferential* of H at p, written
$$v \in \partial H(p),$$
if
$$H(r) \geq H(p) + v \cdot (r - p) \quad \text{for all } r \in \mathbb{R}^n.$$
Prove $v \in \partial H(p)$ if and only if $p \in \partial L(v)$ if and only if $p \cdot v = H(p) + L(v)$, where $L = H^*$.

12. Assume $L_1, L_2 : \mathbb{R}^n \to \mathbb{R}$ are convex, smooth and superlinear. Show that
$$\min_{v \in \mathbb{R}^n} (L_1(v) + L_2(v)) = \max_{p \in \mathbb{R}^n} (-H_1(p) - H_2(-p)),$$
where $H_1 = L_1^*, H_2 = L_2^*$.

13. Prove that the Hopf–Lax formula reads
$$\begin{aligned} u(x,t) &= \min_{y \in \mathbb{R}^n} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\} \\ &= \min_{y \in B(x,Rt)} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\} \end{aligned}$$
for $R = \sup_{\mathbb{R}^n} |DH(Dg)|$, $H = L^*$. (This proves *finite propagation speed* for a Hamilton–Jacobi PDE with convex Hamiltonian and Lipschitz continuous initial function g.)

14. Let E be a closed subset of $\mathbb{R}^n$. Show that if the Hopf–Lax formula could be applied to the initial-value problem
$$\begin{cases} u_t + |Du|^2 = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = \begin{cases} 0 & x \in E \\ +\infty & x \notin E \end{cases} & \text{on } \mathbb{R}^n \times \{t = 0\}, \end{cases}$$
it would give the solution
$$u(x,t) = \frac{1}{4t} \operatorname{dist}(x,E)^2.$$

15. Provide all details for the proof of Lemma 4 in §3.3.3.

16. Assume u^1, u^2 are two solutions of the initial-value problems
$$\begin{cases} u_t^i + H(Du^i) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u^i = g^i & \text{on } \mathbb{R}^n \times \{t = 0\} \ (i = 1,2), \end{cases}$$
given by the Hopf–Lax formula. Prove the L^∞-contraction inequality
$$\sup_{\mathbb{R}^n} |u^1(\cdot,t) - u^2(\cdot,t)| \leq \sup_{\mathbb{R}^n} |g^1 - g^2| \quad (t > 0).$$

17. Show that
$$u(x,t) := \begin{cases} -\frac{2}{3}\left(t + \sqrt{3x + t^2}\right) & \text{if } 4x + t^2 > 0 \\ 0 & \text{if } 4x + t^2 < 0 \end{cases}$$
is an (unbounded) entropy solution of $u_t + \left(\frac{u^2}{2}\right)_x = 0$.

18. Assume $u(x+z) - u(x) \leq Ez$ for all $z > 0$. Let $u^\epsilon = \eta_\epsilon * u$, and show
$$u_x^\epsilon \leq E.$$

19. Assume $F(0) = 0$, u is a continuous integral solution of the conservation law
$$\begin{cases} u_t + F(u)_x = 0 & \text{in } \mathbb{R} \times (0,\infty) \\ u = g & \text{on } \mathbb{R} \times \{t = 0\}, \end{cases}$$
and u has compact support in $\mathbb{R} \times [0,T]$ for each time $T > 0$. Prove
$$\int_{-\infty}^{\infty} u(\cdot,t)\,dx = \int_{-\infty}^{\infty} g\,dx$$
for all $t > 0$.

20. Compute explicitly the unique entropy solution of
$$\begin{cases} u_t + \left(\frac{u^2}{2}\right)_x = 0 & \text{in } \mathbb{R} \times (0,\infty) \\ u = g & \text{on } \mathbb{R} \times \{t = 0\}, \end{cases}$$
for
$$g(x) = \begin{cases} 1 & \text{if } x < -1 \\ 0 & \text{if } -1 < x < 0 \\ 2 & \text{if } 0 < x < 1 \\ 0 & \text{if } x > 1. \end{cases}$$
Draw a picture documenting your answer, being sure to illustrate what happens for all times $t > 0$.

3.6. REFERENCES

Section 3.1 A nice source for this material is Courant–Hilbert [**C-H,** Chapter 2].

Section 3.2 This derivation of the characteristic differential equations is found in Carathéodory [**C**]. The proof of Theorem 2 follows Garabedian [**G,** Chapter 2], John [**J2,** Chapter 1], etc. Chester [**Ch**] and Sneddon [**Sn**] are also good texts for more

Chapter 8

THE CALCULUS OF VARIATIONS

8.1. INTRODUCTION

8.1.1. Basic ideas.

We introduce some new ideas by supposing that we wish to solve a particular partial differential equation, which for simplicity we write in the abstract form

$$A[u] = 0. \tag{1}$$

In this formula $A[\cdot]$ denotes a given, possibly nonlinear partial differential operator and u is the unknown. There is, of course, no general theory for solving all such PDE.

The *calculus of variations* identifies an important class of such nonlinear problems that can be solved using relatively simple techniques from nonlinear functional analysis. This is the class of *variational problems*, that is,

PDE of the form (1), where the nonlinear operator $A[\cdot]$ is the "derivative" of an appropriate "energy" functional $I[\cdot]$. Symbolically we write

$$A[\cdot] = I'[\cdot]. \tag{2}$$

Then problem (1) reads

$$I'[u] = 0. \tag{3}$$

The advantage of this new formulation is that we now can recognize solutions of (1) as being critical points of $I[\cdot]$. These in certain circumstances may be relatively easy to find: if, for instance, the functional $I[\cdot]$ has a minimum at u, then presumably (3) is valid and thus u is a solution of the original PDE (1). *The point is that whereas it is usually extremely difficult to solve (1) directly, it may be much easier to discover minimum (or maximum or other critical) points of the functional* $I[\cdot]$.

In addition of course, many of the laws of physics and other scientific disciplines arise directly as variational principles.

8.1.2. First variation, Euler–Lagrange equation.

Suppose now $U \subset \mathbb{R}^n$ is a bounded, open set with smooth boundary ∂U and we are given a smooth function

$$L : \mathbb{R}^n \times \mathbb{R} \times \bar{U} \to \mathbb{R}.$$

We call L the *Lagrangian*.

NOTATION. We will write

$$L = L(p, z, x) = L(p_1, \dots, p_n, z, x_1, \dots, x_n)$$

for $p \in \mathbb{R}^n$, $z \in \mathbb{R}$, and $x \in U$. Thus "p" is the name of the variable for which we substitute $Dw(x)$ below, and "z" is the variable for which we substitute $w(x)$. We also set

$$\begin{cases} D_p L = (L_{p_1}, \dots, L_{p_n}) \\ D_z L = L_z \\ D_x L = (L_{x_1}, \dots, L_{x_n}). \end{cases}$$

This notation will clarify the theory to follow.

We make the vague ideas in §8.1.1 more precise by now assuming $I[\cdot]$ to have the explicit form

$$I[w] := \int_U L(Dw(x), w(x), x)\, dx, \tag{4}$$

for smooth functions $w : \bar{U} \to \mathbb{R}$ satisfying, say, the boundary condition

$$w = g \quad \text{on } \partial U. \tag{5}$$

Let us now additionally suppose some particular smooth function u, satisfying the requisite boundary condition $u = g$ on ∂U, happens to be a minimizer of $I[\cdot]$ among all functions w satisfying (5). *We will demonstrate that u is then automatically a solution of a certain nonlinear partial differential equation.*

To confirm this, first choose any smooth function $v \in C_c^\infty(U)$ and consider the real-valued function

$$i(\tau) := I[u + \tau v] \quad (\tau \in \mathbb{R}). \tag{6}$$

Since u is a minimizer of $I[\cdot]$ and $u + \tau v = u = g$ on ∂U, we observe that $i(\cdot)$ has a minimum at $\tau = 0$. Therefore

$$i'(0) = 0. \tag{7}$$

We explicitly compute this derivative (called the *first variation*) by writing out

$$i(\tau) = \int_U L(Du + \tau Dv, u + \tau v, x)\, dx. \tag{8}$$

Thus

$$i'(\tau) = \int_U \sum_{i=1}^n L_{p_i}(Du + \tau Dv, u + \tau v, x) v_{x_i} + L_z(Du + \tau Dv, u + \tau v, x) v\, dx.$$

Let $\tau = 0$, to deduce from (7) that

$$0 = i'(0) = \int_U \sum_{i=1}^n L_{p_i}(Du, u, x) v_{x_i} + L_z(Du, u, x) v\, dx.$$

Finally, since v has compact support, we can integrate by parts and obtain

$$0 = \int_U \left[-\sum_{i=1}^n \left(L_{p_i}(Du, u, x)\right)_{x_i} + L_z(Du, u, x) \right] v\, dx.$$

As this equality holds for all test functions v, we conclude u solves the

nonlinear PDE

$$-\sum_{i=1}^n \left(L_{p_i}(Du, u, x)\right)_{x_i} + L_z(Du, u, x) = 0 \quad \text{in } U. \tag{9}$$

This is the *Euler–Lagrange equation* associated with the energy functional $I[\cdot]$ defined by (4). Observe that (9) is a quasilinear, second-order PDE in divergence form.

In summary, any smooth minimizer of $I[\cdot]$ is a solution of the Euler–Lagrange partial differential equation (9), and thus—conversely—we can try to find a solution of (9) by searching for minimizers of (4).

Example 1 (Dirichlet's principle). Take

$$L(p, z, x) = \tfrac{1}{2}|p|^2.$$

Then $L_{p_i} = p_i$ $(i = 1, \dots, n)$, $L_z = 0$; and so the Euler–Lagrange equation associated with the functional

$$I[w] := \tfrac{1}{2} \int_U |Dw|^2\, dx$$

is

$$\Delta u = 0 \quad \text{in } U.$$

This fact is *Dirichlet's principle*, previously introduced in §2.2.5. □

Example 2 (Generalized Dirichlet's principle). Write

$$L(p, z, x) = \tfrac{1}{2} \sum_{i,j=1}^n a^{ij}(x) p_i p_j - z f(x),$$

where $a^{ij} = a^{ji}$ $(i, j = 1, \dots, n)$. Then $L_{p_i} = \sum_{j=1}^n a^{ij}(x) p_j$ $(i = 1, \dots, n)$, $L_z = -f(x)$. Hence the Euler–Lagrange equation associated with the functional

$$I[w] := \int_U \tfrac{1}{2} \sum_{i,j=1}^n a^{ij} w_{x_i} w_{x_j} - w f\, dx$$

is the divergence structure linear equation

$$-\sum_{i,j=1}^n \left(a^{ij} u_{x_j}\right)_{x_i} = f \quad \text{in } U.$$

We will see later (in §8.1.3 and §8.2) that the uniform ellipticity condition on the a^{ij} $(i,j=1,\dots,n)$ is a natural further assumption, required to prove the existence of a minimizer. Consequently from the nonlinear viewpoint of the calculus of variations, the divergence structure form of a linear second-order elliptic PDE is completely natural. This observation provides some much belated motivation for the bilinear form techniques utilized in Chapter 6. □

Example 3 (Nonlinear Poisson equation). Assume we are given a smooth function $f:\mathbb{R}\to\mathbb{R}$, and define its antiderivative $F(z)=\int_0^z f(y)\,dy$. Then the Euler–Lagrange equation associated with the functional

$$I[w] := \int_U \tfrac{1}{2}|Dw|^2 - F(w)\,dx$$

is the nonlinear Poisson equation

$$-\Delta u = f(u) \quad \text{in } U. \qquad \square$$

Example 4 (Minimal surfaces). Let

$$L(p,z,x) = (1+|p|^2)^{1/2},$$

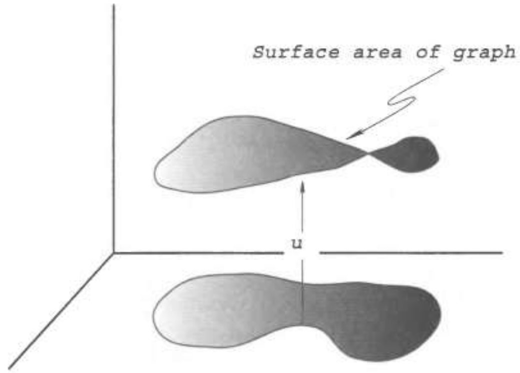
so that

$$I[w] = \int_U (1+|Dw|^2)^{1/2}\,dx$$

is the area of the graph of the function $w:U\to\mathbb{R}$. The associated Euler–Lagrange equation is

$$\sum_{i=1}^n \left(\frac{u_{x_i}}{(1+|Du|^2)^{1/2}}\right)_{x_i} = 0 \quad \text{in } U. \tag{10}$$

This partial differential equation is the *minimal surface equation*. The expression $\operatorname{div}\left(\frac{Du}{(1+|Du|^2)^{1/2}}\right)$ on the left side of (10) is n times the *mean curvature* of the graph of u. Thus *a minimal surface has zero mean curvature.* □



A minimal surface

8.1.3. Second variation.

We continue in the spirit of the calculations from §8.1.2 by computing now the *second variation* of $I[\cdot]$ at the function u. This we find by observing that since u gives a minimum for $I[\cdot]$, we must have

$$i''(0) \geq 0,$$

$i(\cdot)$ defined as above by (6). In view of (8) we can calculate

$$\begin{aligned} i''(\tau) = \int_U \sum_{i,j=1}^n & L_{p_ip_j}(Du+\tau Dv, u+\tau v, x) v_{x_i} v_{x_j} \\ &+ 2\sum_{i=1}^n L_{p_iz}(Du+\tau Dv, u+\tau v, x) v_{x_i} v \\ &+ L_{zz}(Du+\tau Dv, u+\tau v, x) v^2\,dx. \end{aligned}$$

Setting $\tau=0$, we derive the inequality

$$\begin{aligned} 0 \leq i''(0) = \int_U \sum_{i,j=1}^n & L_{p_ip_j}(Du,u,x) v_{x_i} v_{x_j} \\ &+ 2\sum_{i=1}^n L_{p_iz}(Du,u,x) v_{x_i} v + L_{zz}(Du,u,x) v^2\,dx, \end{aligned} \tag{11}$$

holding for all test functions $v\in C_c^\infty(U)$.

We can extract useful information from inequality (11), as follows. First, note after a routine approximation argument that estimate (11) is valid for any Lipschitz continuous function v vanishing on ∂U. We then fix $\xi \in \mathbb{R}^n$ and define

$$v(x) := \epsilon\rho\left(\frac{x\cdot\xi}{\epsilon}\right)\zeta(x) \quad (x \in U), \tag{12}$$

where $\zeta \in C_c^\infty(U)$ and $\rho : \mathbb{R} \to \mathbb{R}$ is the periodic "zig-zag" function defined by

$$\rho(x) = \begin{cases} x & \text{if } 0 \le x \le \frac{1}{2} \\ 1 - x & \text{if } \frac{1}{2} \le x \le 1 \end{cases} \quad \rho(x+1) = \rho(x) \quad (x \in \mathbb{R}).$$

Thus

$$|\rho'| = 1 \quad \text{a.e.} \tag{13}$$

Observe further that $v_{x_i}(x) = \rho'\left(\frac{x\cdot\xi}{\epsilon}\right)\xi_i\zeta + O(\epsilon)$ as $\epsilon \to 0$, and so our substituting (12) into (11) yields

$$0 \le \int_U \sum_{i,j=1}^n L_{p_ip_j}(Du, u, x)(\rho')^2\xi_i\xi_j\zeta^2\,dx + O(\epsilon).$$

We recall (13) and send $\epsilon \to 0$, thereby obtaining the inequality

$$0 \le \int_U \sum_{i,j=1}^n L_{p_ip_j}(Du, u, x)\xi_i\xi_j\zeta^2\,dx.$$

Since this estimate holds for all $\zeta \in C_c^\infty(U)$, we deduce

$$\sum_{i,j=1}^n L_{p_ip_j}(Du, u, x)\xi_i\xi_j \ge 0 \quad (\xi \in \mathbb{R}^n,\ x \in U). \tag{14}$$

We will see later in §8.2 that this necessary condition contains a clue as to the basic convexity assumption on the Lagrangian L required for the existence theory.

8.1.4. Systems.

a. Euler–Lagrange equations. The foregoing considerations generalize quite easily to the case of systems, the only new complications being largely notational. Recall from §A.1 that $\mathbb{M}^{m\times n}$ denotes the space of real $m \times n$ matrices, and assume the smooth Lagrangian function

$$L : \mathbb{M}^{m\times n} \times \mathbb{R}^m \times \bar{U} \to \mathbb{R}$$

is given.

NOTATION. We will write

$$L = L(P, z, x) = L(p_1^1, \dots, p_n^m, z^1, \dots, z^m, x_1, \dots, x_n)$$

for $P \in \mathbb{M}^{m\times n}$, $z \in \mathbb{R}^m$, and $x \in U$, where

$$P = \begin{pmatrix} p_1^1 & \dots & p_n^1 \\ & \ddots & \\ p_1^m & \dots & p_n^m \end{pmatrix}_{m\times n}.$$

(We are now employing superscripts to denote rows, since this notational convention simplifies the following formulas.)

As in §8.1.2 we associate with L the functional

$$I[\mathbf{w}] := \int_U L(D\mathbf{w}(x), \mathbf{w}(x), x)\,dx, \tag{15}$$

defined for smooth functions $\mathbf{w} : \bar{U} \to \mathbb{R}^m$, $\mathbf{w} = (w^1, \dots, w^m)$, satisfying the boundary condition $\mathbf{w} = \mathbf{g}$ on ∂U, $\mathbf{g} : \partial U \to \mathbb{R}^m$ being given. Here

$$D\mathbf{w}(x) = \begin{pmatrix} w_{x_1}^1 & \dots & w_{x_n}^1 \\ & \ddots & \\ w_{x_1}^m & \dots & w_{x_n}^m \end{pmatrix}_{m\times n}$$

is the gradient matrix of $\mathbf{w}$ at x.

Let us now show that any smooth minimizer $\mathbf{u} = (u^1, \dots, u^m)$ of $I[\cdot]$, taken among functions equal to $\mathbf{g}$ on ∂U, must solve a certain *system* of nonlinear partial differential equations. We therefore fix $\mathbf{v} = (v^1, \dots, v^m) \in C_c^\infty(U; \mathbb{R}^m)$ and write

$$i(\tau) := I[\mathbf{u} + \tau\mathbf{v}].$$

As before,

$$i'(0) = 0.$$

From this we readily deduce as above the equality

$$0 = i'(0) = \int_U \sum_{i=1}^n \sum_{k=1}^m L_{p_i^k}(D\mathbf{u}, \mathbf{u}, x)v_{x_i}^k + \sum_{k=1}^m L_{z^k}(D\mathbf{u}, \mathbf{u}, x)v^k\,dx\,.$$

As this identity is valid for all choices of $v^1, \dots, v^m$, we conclude after integrating by parts that

$$-\sum_{i=1}^n \left(L_{p_i^k}(D\mathbf{u}, \mathbf{u}, x)\right)_{x_i} + L_{z^k}(D\mathbf{u}, \mathbf{u}, x) = 0 \quad \text{in } U \ (k = 1, \dots, m). \tag{16}$$

This coupled, quasilinear *system* of PDE comprises the *Euler–Lagrange equations* for the functional $I[\cdot]$ defined by (15).

b. Null Lagrangians. Surprisingly, it turns out to be interesting to study certain systems of nonlinear partial differential equations for which *every* smooth function is a solution.

DEFINITION. *The function L is called a* null Lagrangian *if the system of Euler–Lagrange equations*

$$(17)\quad -\sum_{i=1}^{n}\left(L_{p_i^k}(D\mathbf{u},\mathbf{u},x)\right)_{x_i}+L_{z^k}(D\mathbf{u},\mathbf{u},x)=0 \quad \text{in } U\ (k=1,\dots,m)$$

is automatically solved by all smooth functions $\mathbf{u}:U\to\mathbb{R}^m$.

The importance of null Lagrangians is that the corresponding energy

$$I[\mathbf{w}]=\int_U L(D\mathbf{w},\mathbf{w},x)\,dx$$

depends only on the boundary conditions:

THEOREM 1 (Null Lagrangians and boundary conditions). *Let L be a null Lagrangian. Assume $\mathbf{u},\tilde{\mathbf{u}}$ are two functions in $C^2(\bar U,\mathbb{R}^m)$ such that*

$$(18)\qquad \mathbf{u}\equiv\tilde{\mathbf{u}} \quad \text{on } \partial U.$$

Then

$$(19)\qquad I[\mathbf{u}]=I[\tilde{\mathbf{u}}].$$

Proof. Define

$$i(\tau):=I[\tau\mathbf{u}+(1-\tau)\tilde{\mathbf{u}}] \quad (0\le\tau\le1)\ .$$

Then

$$\begin{aligned}
i'(\tau)&=\int_U\sum_{i=1}^{n}\sum_{k=1}^{m}L_{p_i^k}(\tau D\mathbf{u}+(1-\tau)D\tilde{\mathbf{u}},\tau\mathbf{u}+(1-\tau)\tilde{\mathbf{u}},x)(u^k_{x_i}-\tilde u^k_{x_i})\\
&\qquad+\sum_{k=1}^{m}L_{z^k}(\tau D\mathbf{u}+(1-\tau)D\tilde{\mathbf{u}},\tau\mathbf{u}+(1-\tau)\tilde{\mathbf{u}},x)(u^k-\tilde u^k)\,dx\\
&=\sum_{k=1}^{m}\int_U\Big[-\sum_{i=1}^{n}\left(L_{p_i^k}(\tau D\mathbf{u}+(1-\tau)D\tilde{\mathbf{u}},\tau\mathbf{u}+(1-\tau)\tilde{\mathbf{u}},x)\right)_{x_i}\\
&\qquad+L_{z^k}(\tau D\mathbf{u}+(1-\tau)D\tilde{\mathbf{u}},\tau\mathbf{u}+(1-\tau)\tilde{\mathbf{u}},x)\Big](u^k-\tilde u^k)\,dx\\
&=0,
\end{aligned}$$

the last equality holding since the Euler–Lagrange system of PDE is satisfied by $\tau\mathbf{u}+(1-\tau)\tilde{\mathbf{u}}$. The identity (19) follows. □

In the scalar case that $m=1$ the only null Lagrangians are the boring examples where L is linear in the variable p. For the case of systems ($m>1$), however, there are certain nontrivial examples, which will turn out to be important for us later.

NOTATION. If A is an $n\times n$ matrix, we denote by

$$\operatorname{cof} A$$

the *cofactor* matrix, whose $(k,i)^{\text{th}}$ entry is $(\operatorname{cof} A)^k_i=(-1)^{i+k}d(A)^k_i$, where $d(A)^k_i$ is the determinant of the $(n-1)\times(n-1)$ matrix obtained by deleting the k^{th} row and i^{th} column from A.

LEMMA (Divergence-free rows). *Let $\mathbf{u}:\mathbb{R}^n\to\mathbb{R}^n$ be a smooth function. Then*

$$(20)\qquad \sum_{i=1}^{n}(\operatorname{cof} D\mathbf{u})^k_{i,x_i}=0 \quad (k=1,\dots,n).$$

Proof. 1. From linear algebra we recall the identity

$$(21)\qquad (\det P)I=P^T(\operatorname{cof} P) \quad (P\in\mathbb{M}^{n\times n});$$

that is,

$$(22)\qquad (\det P)\delta_{ij}=\sum_{k=1}^{n}p^k_i(\operatorname{cof} P)^k_j \quad (i,j=1,\dots,n).$$

Thus in particular

$$(23)\qquad \frac{\partial\det P}{\partial p^k_m}=(\operatorname{cof} P)^k_m \quad (k,m=1,\dots,n).$$

2. Now set $P=Du$ in (22), differentiate with respect to x_j, and sum $j=1,\dots,n$, to find

$$\sum_{j,k,m=1}^{n}\delta_{ij}(\operatorname{cof} D\mathbf{u})^k_m u^k_{x_m x_j}=\sum_{k,j=1}^{n}u^k_{x_i x_j}(\operatorname{cof} D\mathbf{u})^k_j+u^k_{x_i}(\operatorname{cof} D\mathbf{u})^k_{j,x_j}$$

for $i = 1, \dots, n$. This identity simplifies to read

$$(24) \qquad \sum_{k=1}^{n} u^k_{x_i} \left(\sum_{j=1}^{n} (\operatorname{cof} D\mathbf{u})^k_{j,x_j} \right) = 0 \quad (i = 1, \dots, n).$$

3. Now if $\det D\mathbf{u}(x_0) \neq 0$, we deduce from (24) that

$$\sum_{j=1}^{n} (\operatorname{cof} D\mathbf{u})^k_{j,x_j} = 0 \quad (k = 1, \dots, n)$$

at x_0. If instead $\det D\mathbf{u}(x_0) = 0$, we choose a number $\epsilon > 0$ so small that $\det(D\mathbf{u}(x_0) + \epsilon I) \neq 0$, apply steps 1–2 to $\tilde{\mathbf{u}} := \mathbf{u} + \varepsilon x$, and send $\epsilon \to 0$. □

THEOREM 2 (Determinants as null Lagrangians). *The determinant function*

$$L(P) = \det P \quad (P \in \mathbb{M}^{n \times n})$$

is a null Lagrangian.

Proof. We must show that for any smooth function $\mathbf{u} : U \to \mathbb{R}^n$,

$$\sum_{i=1}^{n} \left(L_{p^k_i}(D\mathbf{u}) \right)_{x_i} = 0 \quad (k = 1, \dots, n).$$

According to (23) we have $L_{p^k_i} = (\operatorname{cof} P)^k_i$ $(i, k = 1, \dots, n)$. But then employing the notation and conclusion of the lemma, we see

$$\sum_{i=1}^{n} \left(L_{p^k_i}(D\mathbf{u}) \right)_{x_i} = \sum_{i=1}^{n} (\operatorname{cof} D\mathbf{u})^k_{i,x_i} = 0 \quad (k = 1, \dots, n). \qquad \square$$

Some other interesting null Lagrangians are introduced in the exercises.

c. Application. A nice application is a quick analytic proof of a topological fixed point theorem.

THEOREM 3 (Brouwer's Fixed Point Theorem). *Assume*

$$\mathbf{u} : B(0,1) \to B(0,1)$$

is continuous, where $B(0,1)$ denotes the closed unit ball in $\mathbb{R}^n$. Then $\mathbf{u}$ has a fixed point; that is, there exists a point $x \in B(0,1)$ with

$$\mathbf{u}(x) = x.$$

Proof. 1. Write $B = B(0,1)$. We first of all claim that there does not exist a smooth function

$$(25) \qquad \mathbf{w} : B \to \partial B$$

such that

$$(26) \qquad \mathbf{w}(x) = x \quad \text{for all } x \in \partial B.$$

2. Suppose to the contrary that such a function $\mathbf{w}$ exists. Let us temporarily write $\tilde{\mathbf{w}}$ for the identity function, so that $\tilde{\mathbf{w}}(x) = x$ for all $x \in B$. In view of (26), $\mathbf{w} \equiv \tilde{\mathbf{w}}$ on ∂B. Since the determinant is a null Lagrangian, Theorem 1 implies

$$(27) \qquad \int_B \det D\mathbf{w}\, dx = \int_B \det D\tilde{\mathbf{w}}\, dx = |B| \neq 0.$$

On the other hand, (25) implies $|\mathbf{w}|^2 \equiv 1$; and so differentiating, we find

$$(28) \qquad (D\mathbf{w})^T \mathbf{w} = \mathbf{0}.$$

Since $|\mathbf{w}| = 1$, (28) says 0 is an eigenvalue of $D\mathbf{w}^T$ for each $x \in B$. Therefore $\det D\mathbf{w} \equiv 0$ in B. This contradicts (27) and thereby proves no smooth function $\mathbf{w}$ satisfying (25), (26) can exist.

3. Next we show there does not exist any continuous function $\mathbf{w}$ verifying (25), (26). Indeed if $\mathbf{w}$ were such a function, we continuously extend $\mathbf{w}$ by setting $\mathbf{w}(x) = x$ if $x \in \mathbb{R}^n - B$. Observe that $\mathbf{w}(x) \neq 0$ $(x \in \mathbb{R}^n)$. Fix $\epsilon > 0$ so small that $\mathbf{w}_1 := \eta_\epsilon * \mathbf{w}$ satisfies $\mathbf{w}_1(x) \neq 0$ $(x \in \mathbb{R}^n)$. Note also that since η_ϵ is radial, we have $\mathbf{w}_1(x) = x$ if $x \in \mathbb{R}^n - B(0,2)$, for $\epsilon > 0$ sufficiently small. Then

$$\mathbf{w}_2 := \frac{2\mathbf{w}_1}{|\mathbf{w}_1|}$$

would be a smooth mapping satisfying (25), (26) (with the ball $B(0,2)$ replacing $B = B(0,1)$), in contradiction to step 1.

4. Finally suppose $\mathbf{u} : B \to B$ is continuous but has no fixed point. Define now the mapping $\mathbf{w} : B \to \partial B$ by setting $\mathbf{w}(x)$ to be the point on ∂B hit by the ray emanating from $\mathbf{u}(x)$ and passing through x. This mapping is well defined since $\mathbf{u}(x) \neq x$ for all $x \in B$. In addition $\mathbf{w}$ is continuous and satisfies (25), (26).

But this in turn is a contradiction to step 2. □

We will employ Brouwer's Fixed Point Theorem several times in Chapter 9.

8.7. PROBLEMS

In the exercises U always denotes a bounded, open subset of $\mathbb{R}^n$, with smooth boundary. All given functions are assumed smooth, unless otherwise stated.

1. This problem illustrates that a weakly convergent sequence can be rather badly behaved.
   (a) Prove $u_k(x) = \sin(kx) \rightharpoonup 0$ as $k \to \infty$ in $L^2(0,1)$.
   (b) Fix $a, b \in \mathbb{R}$, $0 < \lambda < 1$. Define
$$u_k(x) := \begin{cases} a & \text{if } j/k \le x < (j+\lambda)/k \\ b & \text{if } (j+\lambda)/k \le x < (j+1)/k \end{cases} \quad (j = 0, \dots, k-1).$$

   Prove $u_k \rightharpoonup \lambda a + (1-\lambda) b$ in $L^2(0,1)$.
2. Find $L = L(p,z,x)$ so that the PDE
$$-\Delta u + D\phi \cdot Du = f \quad \text{in } U$$
is the Euler–Lagrange equation corresponding to the functional $I[w] := \int_U L(Dw, w, x)\,dx$.
(Hint: Look for a Lagrangian with an exponential term.)
3. The *elliptic regularization* of the heat equation is the PDE
$$(*) \qquad u_t - \Delta u - \epsilon u_{tt} = 0 \quad \text{in } U_T,$$
where $\epsilon > 0$ and $U_T = U \times (0,T]$. Show that $(*)$ is the Euler–Lagrange equation corresponding to an energy functional $I_\varepsilon[w] := \int\int_{U_T} L_\varepsilon(Dw, w_t, w, x, t)\,dxdt$.
(Hint: Look for a Lagrangian with an exponential term involving t.)
4. Assume $\eta : \mathbb{R}^n \to \mathbb{R}$ is C^1.
   (a) Show $L(P,z,x) = \eta(z)\det P$ $(P \in \mathbb{M}^{n\times n}, z \in \mathbb{R}^n)$ is a null Lagrangian.
   (b) Deduce that if $\mathbf{u} : \mathbb{R}^n \to \mathbb{R}^n$ is C^2, then
$$\int_U \eta(\mathbf{u}) \det D\mathbf{u}\,dx$$
   depends only on $\mathbf{u}|_{\partial U}$.
5. (Continuation) Fix $x_0 \notin \mathbf{u}(\partial U)$, and choose a function η as above so that $\int_{\mathbb{R}^n} \eta\,dz = 1$, $\operatorname{spt}\eta \subset B(x_0,r)$, r taken so small that $B(x_0,r) \cap \mathbf{u}(\partial U) = \emptyset$. Define
$$\deg(\mathbf{u}, x_0) = \int_U \eta(\mathbf{u}) \det D\mathbf{u}\,dx,$$
the *degree* of $\mathbf{u}$ relative to x_0. Prove the degree is an integer.
6. Let $\Sigma \subset \mathbb{R}^3$ denote the graph of the smooth function $u : U \to \mathbb{R}$, $U \subset \mathbb{R}^2$. Then
$$(*) \qquad \int_U (1 + |Du|^2)^{-\frac{3}{2}} \det D^2 u \;\, dx$$
represents the integral of the Gauss curvature over Σ. Prove that this expression depends only upon Du restricted to ∂U. (The Gauss–Bonnet Theorem in differential geometry computes $(*)$ in terms of the geodesic curvature of $\partial\Sigma$.)

7. Let $m = n$. Prove

$$L(P) = \operatorname{tr}(P^2) - \operatorname{tr}(P)^2 \quad (P \in \mathbb{M}^{n\times n})$$

is a null Lagrangian.

8. Explain why the methods in §8.2 will not work to prove the existence of a minimizer of the functional

$$I[w] := \int_U (1 + |Dw|^2)^{1/2}\, dx$$

over $\mathcal{A} := \{w \in W^{1,q}(U) \mid w = g \text{ on } \partial U\}$, for any $1 \le q < \infty$.

9. (Second variation for systems) Assume $\mathbf{u} : U \to \mathbb{R}^m$ is a smooth minimizer of the functional

$$I[\mathbf{w}] := \int_U L(D\mathbf{w}, \mathbf{w}, x)\, dx.$$

(a) Show

$$\sum_{i,j=1}^{n} \sum_{k,l=1}^{m} \frac{\partial^2 L}{\partial p_i^k \partial p_j^l}(D\mathbf{u}, \mathbf{u}, x)\eta_k \eta_l \xi_i \xi_j \ge 0$$

for all $x \in U$, $\xi \in \mathbb{R}^n$, $\eta \in \mathbb{R}^m$.

(b) Provide an example of a nonconvex function $L : \mathbb{M}^{m\times n} \to \mathbb{R}$ satisfying

$$\sum_{i,j=1}^{n} \sum_{k,l=1}^{m} \frac{\partial^2 L(P)}{\partial p_i^k \partial p_j^l}\eta_k \eta_l \xi_i \xi_j \ge 0$$

for all $P \in \mathbb{M}^{m\times n}$, $\xi \in \mathbb{R}^n$, $\eta \in \mathbb{R}^m$.

10. Use the methods of §8.4.1 to show the existence of a nontrivial weak solution $u \in H_0^1(U)$, $u \not\equiv 0$, of

$$\begin{cases} -\Delta u = |u|^{q-1}u & \text{in } U \\ u = 0 & \text{on } \partial U \end{cases}$$

for $1 < q < \frac{n+2}{n-2}$, $n > 2$.

11. Assume $\beta : \mathbb{R} \to \mathbb{R}$ is smooth, with

$$0 < a \le \beta'(z) \le b \quad (z \in \mathbb{R})$$

for constants a, b. Let $f \in L^2(U)$. Formulate what it means for $u \in H^1(U)$ to be a weak solution of the nonlinear boundary-value problem

$$\begin{cases} -\Delta u = f & \text{in } U \\ \frac{\partial u}{\partial \nu} + \beta(u) = 0 & \text{on } \partial U. \end{cases}$$

Prove there exists a unique weak solution.

12. Assume u is a smooth minimizer of the area integral

$$I[w] = \int_U (1 + |Dw|^2)^{1/2}\, dx,$$

subject to given boundary conditions $w = g$ on ∂U and the constraint

$$J[w] = \int_U w\, dx = 1.$$

Prove the graph of u is a surface of constant mean curvature. (Hint: Recall Example 4 in §8.1.2.)

13. Assume $f \in L^2(U)$. Prove the *dual variational principle* that

$$\min_{w \in H_0^1(U)} \int_U \frac{1}{2}|Dw|^2 - fw\, dx = \max_{\substack{\xi \in L^2(U;\mathbb{R}^n) \\ \operatorname{div} \xi = f}} -\frac{1}{2}\int_U |\xi|^2\, dx.$$

14. (Multivalued PDE) Show that the variational inequality (26) for the obstacle problem in §8.4.2 can be rewritten as

$$-\Delta u + \beta(u - h) \ni f$$

for the multivalued function

$$\beta(z) := \begin{cases} 0 & \text{if } z > 0 \\ (-\infty, 0] & \text{if } z = 0 \\ \emptyset & \text{if } z < 0. \end{cases}$$

(See also Problem 3 in Chapter 9.)

15. (Pointwise gradient constraint)

(a) Show there exists a unique minimizer $u \in \mathcal{A}$ of

$$I[w] := \int_U \tfrac{1}{2}|Dw|^2 - fw\, dx,$$

where $f \in L^2(U)$ and

$$\mathcal{A} := \{w \in H_0^1(U) \mid |Dw| \le 1 \text{ a.e.}\}.$$

(b) Prove

$$\int_U Du \cdot D(w - u)\, dx \ge \int_U f(w - u)\, dx$$

for all $w \in \mathcal{A}$.

16. Assume $n \geq 3$ and U is a bounded open set containing 0. Show that $\mathbf{u} := \frac{x}{|x|}$ belongs to $H^1(U;\mathbb{R}^n)$ and is a harmonic mapping into the sphere S^{n-1}. That is, show $\mathbf{u}$ is a weak solution of

$$\begin{cases} -\Delta \mathbf{u} = |D\mathbf{u}|^2\mathbf{u} \\ |\mathbf{u}| = 1 \end{cases} \quad \text{in } U.$$

17. Let $u, \hat{u} \in H_0^1(U)$ both be positive minimizers of the Dirichlet energy

$$I[w] := \int_U |Dw|^2.$$

Suppose also that $u, \hat{u} > 0$ within U. Follow the hints to give a new proof that

$$u \equiv \hat{u} \quad \text{in } U.$$

(Hint: Define $w := \left(\frac{u^2+\hat{u}^2}{2}\right)^{1/2}$, $s := \frac{u^2}{u^2+\hat{u}^2}$ and $\eta := \frac{u^2+\hat{u}^2}{2}$; and show that

$$|Dw|^2 = \eta \left| s\frac{Du}{u} + (1-s)\frac{D\hat{u}}{\hat{u}} \right|^2.$$

Deduce

$$|Dw|^2 \leq \eta \left(s\left|\frac{Du}{u}\right|^2 + (1-s)\left|\frac{D\hat{u}}{\hat{u}}\right|^2 \right) = \frac{1}{2}|Du|^2 + \frac{1}{2}|D\hat{u}|^2$$

and therefore $\frac{Du}{u} = \frac{D\hat{u}}{\hat{u}}$ almost everywhere.)

(Belloni–Kawohl, Manuscripta Math. 109 (2002), 229–231)

18. Assume that a_1, a_2 are smooth, positive functions on $\bar{U}$ such that $a_1 \leq a_2$. Let u_1, u_2 be smooth solutions of

$$\operatorname{div}(a_1 Du_1) = 0, \quad \operatorname{div}(a_2 Du_2) = 0 \quad \text{in } U$$

with $Du_2 \neq 0$ a. e. Suppose finally that

$$u_1 = u_2, \ \frac{\partial u_1}{\partial \nu} = \frac{\partial u_2}{\partial \nu}, \ a_1 = a_2 \quad \text{on } \partial U.$$

Prove that

$$a_1 \equiv a_2, \ u_1 \equiv u_2 \quad \text{within } U.$$

(Hint: Observe that $\int_U a_1|Du_1|^2\,dx = \int_U a_2|Du_2|^2\,dx$.)

19. (Momentum conservation) Given a solution u of the nonlinear wave equation $\Box u + f(u) = 0$, apply Noether's Theorem to the transformations $\mathbf{x}(x,t,\tau) = (x + \tau e_k, t)$, $w(x,t,\tau) = u(x + \tau e_k, t)$ to calculate the momentum density p_k and the momentum flux $\mathbf{j}_k$ satisfying the conservation laws

$$(p_k)_t - \operatorname{div}(\mathbf{j}_k) = 0 \quad (k = 1, \dots, n).$$

20. Let u be harmonic in some region $U \subseteq \mathbb{R}^n$ and assume $B(0,R) \subset U$, $u(0) = 0$, $u \not\equiv 0$. Define for $0 < r < R$ the functions

$$a(r) := \frac{1}{r^{n-1}} \int_{\partial B(0,r)} u^2\,dS, \ b(r) := \frac{1}{r^{n-2}} \int_{B(0,r)} |Du|^2\,dx.$$

We derived in §8.6.2 the monotonicity formula

$$\dot{b} = \frac{2}{r^{n-2}} \int_{\partial B(0,r)} u_r^2\,dS.$$

(a) Prove that

$$\dot{a} = \frac{2}{r^{n-1}} \int_{\partial B(0,r)} u u_r\,dS = \frac{2}{r} b.$$

(b) Show

$$b^2 \leq \frac{r}{2} a\dot{b}.$$

(c) Define the *frequency function*

$$f := \frac{b}{a}$$

and derive *Almgren's monotonicity formula*: $\dot{f} \geq 0$.

(d) Demonstrate next that $\frac{\dot{a}}{a} \leq \frac{\beta}{r}$ and consequently

$$a(r) \geq \gamma r^{\beta} \quad (0 < r < R)$$

for $\beta := \frac{2b(R)}{a(R)}$ and $\gamma := \frac{a(R)}{R^{\beta}}$. This is an estimate from below on how fast a nonconstant harmonic function must grow near a point where it vanishes.

8.8. REFERENCES

Section 8.1 See Giaquinta [**Gi**] and Giaquinta–Hildebrandt [**G-H**] for more about the calculus of variations. Struwe [**Str**] and Zeidler [**Zd,** Vol. 3] are good references for variational methods.

Chapter 10

HAMILTON–JACOBI EQUATIONS

10.1. INTRODUCTION, VISCOSITY SOLUTIONS

This chapter investigates the existence, uniqueness and other properties of appropriately defined weak solutions of the initial-value problem for the *Hamilton-Jacobi equation:*

$$(1) \qquad \begin{cases} u_t + H(Du, x) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

Here the Hamiltonian $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is given, as is the initial function $g : \mathbb{R}^n \to \mathbb{R}$. The unknown is $u : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}$, $u = u(x,t)$, and $Du = D_x u = (u_{x_1}, \dots, u_{x_n})$. We will write $H = H(p,x)$, so that "p" is the name of the variable for which we substitute the gradient Du in the PDE.

We recall from our study of characteristics in §3.2 that in general there can be no smooth solution of (1) lasting for all times $t \geq 0$. We recall further that if H depends only on p and is convex, then the Hopf–Lax formula (expression (21) in §3.3.2) provides us with a type of generalized solution.

In this chapter we consider the general case that H depends also on x and, more importantly, is no longer necessarily convex in the variable p. We

will discover in these new circumstances a different way to define a weak solution of (1).

An approximation. Our approach is to consider first this problem:

$$(2) \qquad \begin{cases} u^\epsilon_t + H(Du^\epsilon, x) - \epsilon\Delta u^\epsilon = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u^\epsilon = g & \text{on } \mathbb{R}^n \times \{t = 0\}, \end{cases}$$

for $\epsilon > 0$. The idea is that whereas (1) involves a fully nonlinear first-order PDE, (2) is an initial-value problem for a quasilinear parabolic PDE, which turns out to have a smooth solution. The term $\epsilon\Delta$ in (2) in effect regularizes the Hamilton–Jacobi equation. Then of course we hope that as $\epsilon \to 0$ the solutions u^ϵ of (2) will converge to some sort of weak solution of (1). This technique is the method of *vanishing viscosity.*

However, as $\epsilon \to 0$ we can expect to lose control over the various estimates of the function u^ϵ and its derivatives: these estimates depend strongly on the regularizing effect of $\epsilon\Delta$ and blow up as $\epsilon \to 0$. However, it turns out that we can often in practice at least be sure that the family $\{u^\epsilon\}_{\epsilon>0}$ is bounded and equicontinuous on compact subsets of $\mathbb{R}^n \times [0,\infty)$. Consequently the Arzela–Ascoli compactness criterion, §C.7, ensures that

$$(3) \qquad u^{\epsilon_j} \to u \quad \text{locally uniformly in } \mathbb{R}^n \times [0,\infty),$$

for some subsequence $\{u^{\epsilon_j}\}_{j=1}^\infty$ and some limit function

$$(4) \qquad u \in C(\mathbb{R}^n \times [0,\infty)).$$

Now we can surely expect that u is some kind of solution of our initial-value problem (1), but as we only know that u is continuous and have absolutely no information as to whether Du and u_t exist in any sense, such an interpretation is difficult.

Similar problems have arisen before in Chapters 8 and 9, where we had to deal with the weak convergence of various would-be approximate solutions to other nonlinear partial differential equations. Remember in particular that in §9.1 we solved a divergence structure quasilinear elliptic PDE by passing to limits using the method of Browder and Minty. Roughly speaking, we there integrated by parts to throw "hard–to–control" derivatives onto a fixed test function and only then tried to go to limits to discover a solution. We will for the Hamilton–Jacobi equation (1) attempt something similar. We will fix a smooth test function v and will pass from (2) to (1) as $\epsilon \to 0$ by first "putting the derivatives onto v".

But since (1) is fully nonlinear and in particular is not of divergence structure, we cannot just integrate by parts to switch to differentiations on v, as we did in §9.1. Instead we will exploit the maximum principle to accomplish this transition, at least at certain points.

We will call the solution we build a *viscosity solution*, in honor of the vanishing viscosity technique. Our main goal will then be to discover an intrinsic characterization of such generalized solutions of (1).

10.1.1. Definitions.

Motivation for definition of viscosity solution. We henceforth assume that H, g are continuous and will as necessary later add further hypotheses.

The technique alluded to above works as follows. Fix any smooth test function $v \in C^\infty(\mathbb{R}^n \times (0,\infty))$ and suppose

$$\text{(5)} \qquad \begin{cases} u - v \text{ has a } \textit{strict} \text{ local maximum at some point} \\ (x_0, t_0) \in \mathbb{R}^n \times (0,\infty). \end{cases}$$

This means

$$(u-v)(x_0,t_0) > (u-v)(x,t)$$

for all points (x,t) sufficiently close to (x_0,t_0), with $(x,t) \neq (x_0,t_0)$.

Now recall (3). We claim for each sufficiently small $\epsilon_j > 0$, there exists a point $(x_{\epsilon_j}, t_{\epsilon_j})$ such that

$$\text{(6)} \qquad u^{\epsilon_j} - v \text{ has a local maximum at } (x_{\epsilon_j}, t_{\epsilon_j})$$

and

$$\text{(7)} \qquad (x_{\epsilon_j}, t_{\epsilon_j}) \to (x_0, t_0) \quad \text{as } j \to \infty.$$

To confirm this, note that for each sufficiently small $r > 0$, (5) implies $\max_{\partial B}(u-v) < (u-v)(x_0,t_0)$, B denoting the closed ball in $\mathbb{R}^{n+1}$ with center (x_0,t_0) and radius r. In view of (3), $u^{\epsilon_j} \to u$ uniformly on B, and so $\max_{\partial B}(u^{\epsilon_j} - v) < (u^{\epsilon_j} - v)(x_0,t_0)$ provided ϵ_j is small enough. Consequently $u^{\epsilon_j} - v$ attains a local maximum at some point in the interior of B. We can next replace r by a sequence of radii tending to zero to obtain (6), (7).

Now owing to (6), we see that the equations

$$\text{(8)} \qquad Du^{\epsilon_j}(x_{\epsilon_j}, t_{\epsilon_j}) = Dv(x_{\epsilon_j}, t_{\epsilon_j}),$$

$$\text{(9)} \qquad u_t^{\epsilon_j}(x_{\epsilon_j}, t_{\epsilon_j}) = v_t(x_{\epsilon_j}, t_{\epsilon_j})$$

and the inequality

$$\text{(10)} \qquad -\Delta u^{\epsilon_j}(x_{\epsilon_j}, t_{\epsilon_j}) \geq -\Delta v(x_{\epsilon_j}, t_{\epsilon_j})$$

hold. We consequently can calculate

$$\text{(11)} \qquad \begin{aligned} &v_t(x_{\epsilon_j}, t_{\epsilon_j}) + H(Dv(x_{\epsilon_j}, t_{\epsilon_j}), x_{\epsilon_j}) \\ &\quad = u_t^{\epsilon_j}(x_{\epsilon_j}, t_{\epsilon_j}) + H(Du^{\epsilon_j}(x_{\epsilon_j}, t_{\epsilon_j}), x_{\epsilon_j}) \quad \text{by (8),(9)} \\ &\quad = \epsilon_j \Delta u^{\epsilon_j}(x_{\epsilon_j}, t_{\epsilon_j}) \quad \text{by (2)} \\ &\quad \leq \epsilon_j \Delta v(x_{\epsilon_j}, t_{\epsilon_j}) \quad \text{by (10).} \end{aligned}$$

Now let $\epsilon_j \to 0$ and remember (7). Since v is smooth and H is continuous, we deduce

$$\text{(12)} \qquad v_t(x_0,t_0) + H(Dv(x_0,t_0), x_0) \leq 0.$$

We have established this inequality assuming (5). Suppose now instead that

$$\text{(13)} \qquad u - v \text{ has a local maximum at } (x_0,t_0)$$

but that this maximum is not necessarily strict. Then $u - \tilde{v}$ has a strict local maximum at (x_0,t_0), for $\tilde{v}(x,t) := v(x,t) + \delta(|x-x_0|^2 + (t-t_0)^2)$ $(\delta > 0)$. We thus conclude as above that $\tilde{v}_t(x_0,t_0) + H(D\tilde{v}(x_0,t_0), x_0) \leq 0$, whereupon (12) again follows.

Consequently (13) implies inequality (12). Similarly, we deduce the reverse inequality

$$\text{(14)} \qquad v_t(x_0,t_0) + H(Dv(x_0,t_0), x_0) \geq 0,$$

provided

$$\text{(15)} \qquad u - v \text{ has a local minimum at } (x_0,t_0).$$

The proof is exactly like that above, except that the inequalities in (10), and thus in (11), are reversed.

In summary, we have discovered for any smooth function v that inequality (12) follows from (13), and (14) from (15). We have in effect put the derivatives onto v, at the expense of certain inequalities holding. □

Our intention now is to *define* a weak solution of (1) in terms of (12), (13) and (14), (15).

DEFINITION. *Assume that u is bounded and uniformly continuous on $\mathbb{R}^n \times [0,T]$, for each $T > 0$. We say that u is a* viscosity solution *of the initial-value problem (1) for the Hamilton–Jacobi equation provided*

(i) $u = g$ *on* $\mathbb{R}^n \times \{t = 0\}$, *and*

(ii) *for each* $v \in C^\infty(\mathbb{R}^n \times (0,\infty))$,

$$(16) \quad \begin{cases} \textit{if } u - v \textit{ has a local maximum at a point } (x_0,t_0) \in \mathbb{R}^n \times (0,\infty), \\ \textit{then} \\ \qquad v_t(x_0,t_0) + H(Dv(x_0,t_0),x_0) \leq 0, \end{cases}$$

and

$$(17) \quad \begin{cases} \textit{if } u - v \textit{ has a local minimum at a point } (x_0,t_0) \in \mathbb{R}^n \times (0,\infty), \\ \textit{then} \\ \qquad v_t(x_0,t_0) + H(Dv(x_0,t_0),x_0) \geq 0. \end{cases}$$

Remark. Note carefully that by definition a viscosity solution satisfies (16), (17), and so all subsequent deductions must be based on these inequalities. The previous discussion was purely motivational.

For emphasis, we repeat the same point, which has caused some confusion among students. To verify that a given function u is a viscosity solution of the Hamilton–Jacobi equation $u_t + H(Du,x) = 0$, we must confirm that (16), (17) hold for all smooth functions v. Now the argument above shows that *if* u is constructed using the vanishing viscosity method, it is indeed a viscosity solution. But we will also see later in §10.3 that viscosity solutions can be built in entirely different ways, which have nothing whatsoever to do with vanishing viscosity.

The point is that the inequalities (16), (17) provide an intrinsic characterization, and indeed the very definition, of our generalized solutions.

We devote the rest of this chapter to demonstrating that viscosity solutions provide an appropriate and useful notion of weak solutions for our Hamilton–Jacobi PDE.

10.1.2. Consistency.

Let us begin by checking that the notion of viscosity solution is consistent with that of a classical solution. First of all, note that if $u \in C^1(\mathbb{R}^n \times [0,\infty))$ solves (1) and if u is bounded and uniformly continuous, then u is a viscosity solution. That is, we assert that any *classical solution of* $u_t + H(Du,x) = 0$ *is also a viscosity solution.* The proof is easy. If v is smooth and $u - v$ obtains a local maximum at (x_0,t_0), then

$$\begin{cases} Du(x_0,t_0) = Dv(x_0,t_0) \\ u_t(x_0,t_0) = v_t(x_0,t_0). \end{cases}$$

Consequently

$$\begin{aligned} & v_t(x_0,t_0) + H(Dv(x_0,t_0),x_0) \\ & \quad = u_t(x_0,t_0) + H(Du(x_0,t_0),x_0) = 0, \end{aligned}$$

since u solves (1). A similar equality holds at any point (x_0,t_0) where $u - v$ has a local minimum.

Next we assert that *any sufficiently smooth viscosity solution is a classical solution* and, even more, that if a viscosity solution is differentiable at some point, then it solves the Hamilton–Jacobi PDE there. We will need the following calculus fact:

LEMMA (Touching by a C^1 function). *Assume* $u : \mathbb{R}^n \to \mathbb{R}$ *is continuous and is also differentiable at the point* x_0. *Then there exists a function* $v \in C^1(\mathbb{R}^n)$ *such that*

$$(18) \qquad u(x_0) = v(x_0)$$

and

$$(19) \qquad u - v \textit{ has a strict local maximum at } x_0.$$

Proof. 1. We may as well assume

$$(20) \qquad x_0 = 0,\ u(0) = Du(0) = 0,$$

for otherwise we could consider $\tilde{u}(x) := u(x + x_0) - u(x_0) - Du(x_0) \cdot x$ in place of u.

2. In view of (20) and our hypothesis, we have

$$(21) \qquad u(x) = |x|\rho_1(x),$$

where

$$(22) \qquad \rho_1 : \mathbb{R}^n \to \mathbb{R} \text{ is continuous}, \ \ \rho_1(0) = 0.$$

Set

$$(23) \qquad \rho_2(r) := \max_{x \in B(0,r)} \{|\rho_1(x)|\} \quad (r \geq 0).$$

Then

$$(24) \qquad \rho_2 : [0,\infty) \to [0,\infty) \text{ is continuous}, \ \ \rho_2(0) = 0,$$

and

(25) ρ_2 is nondecreasing.

3. Now write

$$v(x) := \int_{|x|}^{2|x|} \rho_2(r)\, dr + |x|^2 \quad (x \in \mathbb{R}^n).$$

Since $|v(x)| \leq |x|\rho_2(2|x|) + |x|^2$, we observe

(26) $$v(0) = Dv(0) = 0.$$

Furthermore if $x \neq 0$, we have

$$Dv(x) = \frac{2x}{|x|}\rho_2(2|x|) - \frac{x}{|x|}\rho_2(|x|) + 2x,$$

and so $v \in C^1(\mathbb{R}^n)$.

4. Finally note that if $x \neq 0$,

$$\begin{aligned} u(x) - v(x) &= |x|\rho_1(x) - \int_{|x|}^{2|x|} \rho_2(r)\, dr - |x|^2 \\ &\leq |x|\rho_2(|x|) - \int_{|x|}^{2|x|} \rho_2(r)\, dr - |x|^2 \\ &\leq -|x|^2 \quad \text{by (25)} \\ &< 0 = u(0) - v(0). \end{aligned}$$

Thus $u - v$ has a strict local maximum at 0, as required. □

THEOREM 1 (Consistency of viscosity solutions). *Let u be a viscosity solution of* (1), *and suppose u is differentiable at some point $(x_0, t_0) \in \mathbb{R}^n \times (0, \infty)$. Then*

$$u_t(x_0, t_0) + H(Du(x_0, t_0), x_0) = 0.$$

Proof. 1. Applying the lemma above to u, with $\mathbb{R}^{n+1}$ replacing $\mathbb{R}^n$ and (x_0, t_0) replacing x_0, we deduce there exists a C^1 function v such that

(27) $u - v$ has a strict maximum at (x_0, t_0).

2. Now set $v^\epsilon := \eta_\epsilon * v$, η_ϵ denoting the usual mollifier in the $n + 1$ variables (x, t). Then

(28) $$\begin{cases} v^\epsilon \to v \\ Dv^\epsilon \to Dv \quad \text{uniformly near } (x_0, t_0) \\ v_t^\epsilon \to v_t; \end{cases}$$

and so (27) implies

(29) $u - v^\epsilon$ has a maximum at some point (x_ϵ, t_ϵ),

with

(30) $$(x_\epsilon, t_\epsilon) \to (x_0, t_0) \quad \text{as } \epsilon \to 0.$$

Applying then the definition of viscosity solution, we see

$$v_t^\epsilon(x_\epsilon, t_\epsilon) + H(Dv^\epsilon(x_\epsilon, t_\epsilon), x_\epsilon) \leq 0.$$

Let $\epsilon \to 0$ and use (28), (30) to deduce

(31) $$v_t(x_0, t_0) + H(Dv(x_0, t_0), x_0) \leq 0.$$

But in view of (27), we see that since u is differentiable at (x_0, t_0),

$$Du(x_0, t_0) = Dv(x_0, t_0),\ \ u_t(x_0, t_0) = v_t(x_0, t_0).$$

Substitute above, to conclude from (31) that

(32) $$u_t(x_0, t_0) + H(Du(x_0, t_0), x_0) \leq 0.$$

3. Now apply the lemma above to $-u$ in $\mathbb{R}^{n+1}$, to find a C^1 function v such that $u - v$ has a strict minimum at (x_0, t_0). Then, arguing as above, we likewise deduce

$$u_t(x_0, t_0) + H(Du(x_0, t_0), x_0) \geq 0.$$

This inequality and (32) complete the proof. □

10.4. PROBLEMS

1. Assume u is a viscosity solution of

$$u_t + H(Du, x) = 0 \quad \text{in } \mathbb{R}^n \times (0,\infty).$$

Show that $\hat{u} := -u$ is a viscosity solution of

$$\hat{u}_t + \hat{H}(D\hat{u}, x) = 0 \quad \text{in } \mathbb{R}^n \times (0,\infty),$$

for $\hat{H}(p,x) := -H(-p,x)$.

2. Let $\{u^k\}_{k=1}^{\infty}$ be viscosity solutions of the Hamilton–Jacobi equations

$$u^k_t + H(Du^k, x) = 0 \quad \text{in } \mathbb{R}^n \times (0,\infty)$$

$(k = 1, \dots)$, and suppose $u^k \to u$ uniformly. Assume as well that H is continuous. Show u is a viscosity solution of

$$u_t + H(Du, x) = 0 \quad \text{in } \mathbb{R}^n \times (0,\infty).$$

Hence the uniform limits of viscosity solutions are viscosity solutions.

3. Suppose for each $\epsilon > 0$ that u^ϵ is a smooth solution of the parabolic equation

$$u^\epsilon_t + H(Du^\epsilon, x) - \epsilon \sum_{i,j=1}^{n} a^{ij} u^\epsilon_{x_i x_j} = 0$$

in $\mathbb{R}^n \times (0,\infty)$, where the smooth coefficients a^{ij} $(i,j = 1,\dots,n)$ satisfy the uniform ellipticity condition from Chapter 6. Suppose also that H is continuous and that $u^\epsilon \to u$ uniformly as $\epsilon \to 0$.

Prove that u is a viscosity solution of $u_t + H(Du, x) = 0$. (This exercise shows that viscosity solutions do not depend upon the precise structure of the parabolic smoothing.)

4. Let u^i $(i = 1, 2)$ be viscosity solutions of

$$\begin{cases} u^i_t + H(Du^i, x) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u^i = g^i & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

Assume H satisfies condition (3) in §10.2. Prove the *contraction* property

$$\sup_{\mathbb{R}^n} |u^1(\cdot,t) - u^2(\cdot,t)| \le \sup_{\mathbb{R}^n} |g^1 - g^2| \quad (t \ge 0).$$

5. (a) Show that $u(x) := 1 - |x|$ is a viscosity solution of

$$(*) \qquad \begin{cases} |u'| = 1 & \text{in } (-1,1) \\ u(-1) = u(1) = 0. \end{cases}$$

This means that for each $v \in C^\infty(-1,1)$, if $u - v$ has a maximum (minimum) at a point $x_0 \in (-1,1)$, then $|v'(x_0)| \le 1$ (≥ 1).

(b) Show that $\tilde{u}(x) := |x| - 1$ is *not* a viscosity solution of $(*)$.

(c) Show that $\tilde{u}$ is a viscosity solution of

$$(**) \qquad \begin{cases} -|\tilde{u}'| = -1 & \text{in } (-1,1) \\ \tilde{u}(-1) = \tilde{u}(1) = 0. \end{cases}$$

(Hint: What is the meaning of a viscosity solution of $(**)$?)

(d) Why do problems $(*)$, $(**)$ have different viscosity solutions?

6. Let $U \subset \mathbb{R}^n$ be open, bounded. Set $u(x) := \operatorname{dist}(x, \partial U)$ $(x \in U)$. Prove that u is Lipschitz continuous and that it is a viscosity solution of the eikonal equation

$$|Du| = 1 \quad \text{in } U.$$

This means that for each $v \in C^\infty(U)$, if $u - v$ has a maximum (minimum) at a point $x_0 \in U$, then $|Dv(x_0)| \le 1$ (≥ 1).

7. Suppose an open set $U \subset \mathbb{R}^n$ is subdivided by a smooth hypersurface Γ into the subregions V^+ and V^-. Let ν denote the unit normal to Γ, pointing into V^+. Assume that u is a viscosity solution of

$$H(Du) = 0 \quad \text{in } U$$

and that u is smooth in $\bar{V}^+$ and $\bar{V}^-$. Write u^+_ν for the limit of $Du \cdot \nu$ along Γ from within V^+, and write u^-_ν for the limit from within V^-. Prove that along Γ we have the inequalities

$$H(\lambda u^-_\nu + (1-\lambda) u^+_\nu) \ge 0 \quad \text{if } u^-_\nu \le u^+_\nu$$

and

$$H(\lambda u^-_\nu + (1-\lambda) u^+_\nu) \le 0 \quad \text{if } u^+_\nu \le u^-_\nu,$$

for each $0 \le \lambda \le 1$.

8. A surface described by the graph of $u : \mathbb{R}^2 \to \mathbb{R}$ is illuminated by parallel light rays from the vertical e_3 direction. We assume the surface has constant albedo and in addition is *Lambertian*, meaning that incoming right rays are scattered equally in all directions. Then the intensity $i = i(x)$ of the reflected light above the point $x \in \mathbb{R}^2$ is given by the formula $i = e_3 \cdot \nu$, where ν is the upward pointing unit normal to the surface.

Show that u solves a PDE of the form

$$|Du| = n$$

for a given function $n = n(x)$, computed in terms of the intensity i. (Finding the surface by solving this PDE for u is the *shape from shading* problem.)

9. A yacht starts at the point $(x_1, 0)$ on the positive x_1-axis and sails to the right at speed $b_1 > 0$. Another yacht is initially at the point $(0, x_2)$ along the positive x_2-axis and starts in pursuit, sailing always towards the first yacht at speed $b_2 > b_1$.

   Find the PDE solved by

$$u(x_1, x_2) := \text{time it takes the second yacht to intercept the first.}$$

   (Think of this as a dynamic programming problem, but with no controls.)

10. (Infinite horizon control problem) Assume $\mathbf{f}$ and r satisfy the conditions given in §10.3. Given a point $x \in \mathbb{R}^n$ and a control belonging to $\mathcal{A} := \{\boldsymbol{\alpha} : [0, \infty) \to A \mid \boldsymbol{\alpha}(\cdot) \text{ is measurable}\}$, let $\mathbf{x}(\cdot)$ be the unique solution of the ODE

$$\begin{cases} \dot{\mathbf{x}}(s) = \mathbf{f}(\mathbf{x}(s), \boldsymbol{\alpha}(s)) & (s > 0) \\ \mathbf{x}(0) = x. \end{cases}$$

   Fix $\lambda > 0$ and define the discounted cost

$$C_x[\boldsymbol{\alpha}(\cdot)] := \int_0^\infty e^{-\lambda s} r(\mathbf{x}(s), \boldsymbol{\alpha}(s))\, ds.$$

   Define the value function

$$u(x) := \inf_{\boldsymbol{\alpha}(\cdot) \in \mathcal{A}} C_x[\boldsymbol{\alpha}(\cdot)].$$

   (a) Show that u is bounded and that if $\lambda > \operatorname{Lip}[\mathbf{f}]$, then u is Lipschitz continuous.

   (b) Show that if $0 < \lambda \le \operatorname{Lip}[\mathbf{f}]$, then u is Hölder continuous for some exponent $0 < \alpha < 1$.

11. (Continuation) Prove that the value function u is a viscosity solution of the PDE

$$\lambda u - \min_{a \in A} \{\mathbf{f}(x, a) \cdot Du + r(x, a)\} = 0 \quad \text{in } \mathbb{R}^n.$$

   (This means that if v is smooth and $u - v$ has a local maximum at a point x_0, then

$$\lambda u - \min_{a \in A} \{\mathbf{f}(x, a) \cdot Dv + r(x, a)\} \le 0$$

   at x_0, and that the opposite inequality holds if $u - v$ has a local minimum at x_0.)

The next sequence of exercises develops some of the theory of viscosity solutions for fully nonlinear elliptic PDE of second order.

12. Remember from §A.1 that if $R, S \in \mathbb{S}^n$, we write $R \ge S$ if $R - S$ is nonnegative definite. A function $F : \mathbb{S}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $F = F(R, p, x)$, is *elliptic* provided

$$R \ge S \text{ implies } F(R, p, x) \le F(S, p, x).$$

   Here $\mathbb{S}^n$ denotes the space of real, $n \times n$ symmetric matrices.

   (a) Show that $F(R) = -\operatorname{tr} R$ is elliptic.

   (b) More generally, show that if $A \in \mathbb{S}^n$ and $A \ge 0$, then $F(R) = -A : R = -\operatorname{tr}(AR^T)$ is elliptic.

   (c) Show that if for each $k = 1, \dots, m$, F_k is elliptic, then so are $\max_k F_k$ and $\min_k F_k$.

13. Let F be continuous and elliptic. We say that a function $u \in C(U)$ is a *viscosity solution* of the fully nonlinear elliptic PDE

$$(*) \qquad F(D^2u, Du, x) = 0 \quad \text{in } U,$$

   provided for each $v \in C^\infty(U)$, (i) if $u - v$ has a local maximum at a point $x_0 \in U$, then $F(D^2u(x_0), Du(x_0), x_0) \le 0$ and (ii) if $u - v$ has a local minimum at a point $x_0 \in U$, then $F(D^2u(x_0), Du(x_0), x_0) \ge 0$.

   Show that if u is a C^2 solution of $(*)$, then u is a viscosity solution.

14. Assume that u_k is a viscosity solution of

$$F(D^2u_k, Du_k, x) = 0 \quad \text{in } U$$

   for $k = 1, \dots$. Suppose $u_k \to u$ uniformly and show u is a viscosity solution of

$$F(D^2u, Du, x) = 0 \quad \text{in } U.$$

10.5. REFERENCES

Section 10.1 The definition of viscosity solutions presented here is due to Crandall, Evans and Lions (Trans. AMS 282 (1984), 487–502), who recast an earlier definition set forth in the basic