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

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

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

$$D_qG = (\mathbf{F}'(z), 1),\ D_yG = 0,\ D_zG = \mathbf{F}''(z)\cdot p.$$

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

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

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

(59) $$z(s) = z^0 = g(x^0);$$

and (58) implies

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

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.

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

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

Hence

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

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 (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

(62) $$G(Du, u_t, u, x, t) = u_t + H(Du, x) = 0,$$

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

$$D_qG = (D_pH(p,x), 1),\ D_yG = (D_xH(p,x), 0),\ D_zG = 0.$$

Thus equation (11)(c) becomes

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

In particular we can identify the parameter s with the time t.

Equation (11)(a) for the case at hand reads

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

the equation (11)(b) is

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

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

$$(64)\qquad \begin{cases} \text{(a)} & \dot{\mathbf{p}}(s) = -D_xH(\mathbf{p}(s),\mathbf{x}(s)) \\ \text{(b)} & \dot{z}(s) = D_pH(\mathbf{p}(s),\mathbf{x}(s))\cdot\mathbf{p}(s) - H(\mathbf{p}(s),\mathbf{x}(s)) \\ \text{(c)} & \dot{\mathbf{x}}(s) = D_pH(\mathbf{p}(s),\mathbf{x}(s)) \end{cases}$$

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,

$$(65)\qquad \begin{cases} \dot{\mathbf{x}} = D_pH(\mathbf{p},\mathbf{x}) \\ \dot{\mathbf{p}} = -D_xH(\mathbf{p},\mathbf{x}), \end{cases}$$

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:

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

Here $u : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}$ is the unknown, $u = u(x,t)$, and $Du = D_xu = (u_{x_1},\dots,u_{x_n})$. 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

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

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

$$L = L(q,x) = L(q_1,\dots,q_n,x_1,\dots,x_n) \quad (q,x \in \mathbb{R}^n)$$

and

$$\begin{cases} D_qL = (L_{q_1}\cdots L_{q_n}) \\ D_xL = (L_{x_1}\cdots L_{x_n}). \end{cases}$$

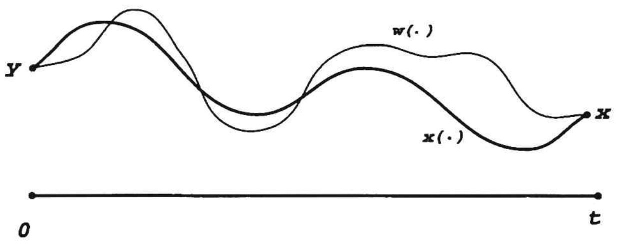
Thus in the formula (2) below "q" is the name of the variable for which we substitute $\dot{\mathbf{w}}(s)$, and "x" is the variable for which we substitute $\mathbf{w}(s)$. □

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

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

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 then to find a curve $\mathbf{x}(\cdot) \in \mathcal{A}$ satisfying

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

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

$$-\frac{d}{ds}\left(D_qL(\dot{\mathbf{x}}(s),\mathbf{x}(s))\right) + D_xL(\dot{\mathbf{x}}(s),\mathbf{x}(s)) = 0 \quad (0 \le s \le t). \tag{4}$$

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

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

$$\mathbf{v}(0) = \mathbf{v}(t) = 0, \tag{5}$$

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

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

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

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

Thus the real-valued function

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

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

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

and so

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

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

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

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

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

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

$$-\frac{d}{ds}\left(L_{q_i}(\dot{\mathbf{x}},\mathbf{x})\right) + L_{x_i}(\dot{\mathbf{x}},\mathbf{x}) = 0$$

for $0 \le s \le t$, $i = 1,\dots,n$. □

Remark. 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 $L(q,x) = \frac{1}{2}m|q|^2 - \phi(x)$, 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 ODE.

We now convert 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

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

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

(9) Suppose for all $x, p \in \mathbb{R}^n$ that the equation
$$p = D_qL(q,x)$$
can be uniquely solved for q as a smooth function of p and x, $q = \mathbf{q}(p,x)$.

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

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

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

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

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

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

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

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

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

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

Let us hereafter write $\mathbf{q}(\cdot) = (q^1(\cdot), \dots, q^n(\cdot))$. We then compute for $i = 1, \dots, n$:

$$\begin{aligned} \frac{\partial H}{\partial x_i}(p,x) &= \sum_{k=1}^n p_k \frac{\partial q^k}{\partial x_i}(p,x) - \frac{\partial L}{\partial q_k}(q,x)\frac{\partial q^k}{\partial x_i}(p,x) - \frac{\partial L}{\partial x_i}(q,x) \\ &= -\frac{\partial L}{\partial x_i}(q,x) \quad \text{by (9),} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial H}{\partial p_i}(p,x) &= q^i(p,x) + \sum_{k=1}^n p_k \frac{\partial q^k}{\partial p_i}(p,x) - \frac{\partial L}{\partial q_k}(q,x)\frac{\partial q^k}{\partial p_i}(p,x) \\ &= q^i(p,x), \quad \text{again by (9).} \end{aligned}$$

Thus

$$\frac{\partial H}{\partial p_i}(\mathbf{p}(s), \mathbf{x}(s)) = q^i(\mathbf{p}(s), \mathbf{x}(s)) = \dot{x}^i(s);$$

and likewise

$$\begin{aligned} \frac{\partial H}{\partial x_i}(\mathbf{p}(s), \mathbf{x}(s)) &= -\frac{\partial L}{\partial x_i}(\mathbf{q}(\mathbf{p}(s), \mathbf{x}(s)), \mathbf{x}(s)) = -\frac{\partial L}{\partial x_i}(\dot{\mathbf{x}}(s), \mathbf{x}(s)) \\ &= -\frac{d}{ds}\left(\frac{\partial L}{\partial q_i}(\dot{\mathbf{x}}(s), \mathbf{x}(s))\right) \quad \text{according to (4)} \\ &= -\dot{p}^i(s). \end{aligned}$$

Finally, observe

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

□

Remark. See Arnold [**AR,** 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:

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

and

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

The convexity implies L is continuous.

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

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

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 $q^* \in \mathbb{R}^n$ for which

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

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

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

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

(14) $$H = L^*.$$

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*

(15) $$L = H^*.$$

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

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

We say H and L are *dual* convex functions. □

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

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

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

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

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

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

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

3. In view of (14)

$$H(p) + L(q) \geq p \cdot q$$

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

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

On the other hand

$$\text{(16)} \qquad \begin{aligned} H^*(q) &= \sup_{p \in \mathbb{R}^n} \{p \cdot q - \sup_{r \in \mathbb{R}^n} \{p \cdot r - L(r)\}\} \\ &= \sup_{p \in \mathbb{R}^n} \inf_{r \in \mathbb{R}^n} \{p \cdot (q - r) + L(r)\}. \end{aligned}$$

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

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

(If L is differentiable at q, take $s = DL(q)$.) Taking $p = s$ in (16), we compute

$$H^*(q) \geq \inf_{r \in \mathbb{R}^n} \{s \cdot (q - r) + L(r)\} = L(q).$$

□

b. Hopf–Lax formula.

Let us now return to the initial-value problem (1). Recall that the calculus of variations problem with Lagrangian L, discussed in §3.3.1, led to Hamilton's ODE for the associated Hamiltonian H. Since these ODE are also the characteristic equations of the Hamilton–Jacobi PDE, we conjecture there is probably a direct connection between this PDE and the calculus of variations.

So if $x \in \mathbb{R}^n$ and $t > 0$ are given, we should presumably try to minimize the action

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

over functions $\mathbf{w} : [0, t] \to \mathbb{R}^n$ satisfying $\mathbf{w}(t) = x$. But what should we take for $\mathbf{w}(0)$? As we must somehow take into account the initial condition for our PDE, let us try modifying the action to include the function g evaluated at $\mathbf{w}(0)$:

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

Next let us construct a candidate for a solution to the initial-value problem (1), in terms of a variational principle entailing this modified action. We accordingly set

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

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

We propose now to investigate the sense in which u so defined by (17) actually solves the initial-value problem for the Hamilton–Jacobi PDE:

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

Recall we are assuming H is smooth,

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

We henceforth suppose also

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

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 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*

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

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 $\mathbf{w}(s) := y + \frac{s}{t}(x - y)$ $(0 \leq s \leq t)$. Then the definition (17) of u implies

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

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 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 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

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

Hence

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

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

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

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

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

Furthermore,

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

This inequality and (29) imply

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

for

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

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

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

Hence

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

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

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

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

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

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,

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