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

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

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

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

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

The first and third of these equalities,

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

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:

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

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

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

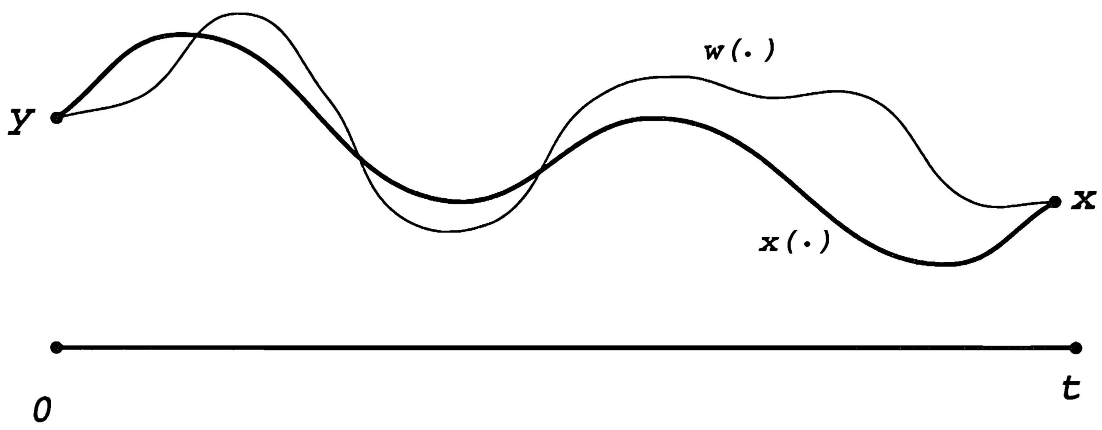
Thus in the formula (2) below "v" is the name of the variable for which we substitute $\dot{\mathbf{w}}(s)$, and "x" is the variable for which we substitute $\mathbf{w}(s)$. □

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

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

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

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

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

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

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

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

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

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

Thus the real-valued function

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

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 $0 \le s \le t$

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

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

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

$\mathbf{p}(\cdot)$ 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*:*

$$\left\{\begin{aligned} \dot{\mathbf{x}}(s) &= D_pH(\mathbf{p}(s),\mathbf{x}(s)) \\ \dot{\mathbf{p}}(s) &= -D_xH(\mathbf{p}(s),\mathbf{x}(s)) \end{aligned}\right. \tag{10}$$

for $0 \le s \le t$. *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 $\mathbf{x}(\cdot) = (x^1(\cdot),\dots,x^n(\cdot))$ and $\mathbf{p}(\cdot) = (p^1(\cdot),\dots,p^n(\cdot))$ (defined by (8)).

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

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

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

and

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

The convexity implies L is continuous.

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

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

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

$$H = L^*. \tag{14}$$

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*

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

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

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

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 $0 \leq \tau \leq 1$, $p, \hat{p} \in \mathbb{R}^n$, we have

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

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

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

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

3. In view of (14)

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

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

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

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

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

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

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

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 $\mathbf{w}(s) := y + \frac{s}{t}(x-y)$ $(0 \leq s \leq t)$. 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

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

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

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

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

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

and consequently

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

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 $0 \leq s < t$, 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 $[s,t]$.