VII. Hamiltonian Methods

UBC M402 Lecture Notes ©2015 by Philip D. Loewen

A. The Hamiltonian

Throughout this writeup, we consider a fixed open subset \( \Omega \) of the \((t, x)\)-space \( \mathbb{R} \times \mathbb{R}^n \) and a given Lagrangian \( L(t, x, v) : \Omega \times \mathbb{R}^n \to \mathbb{R} \) of class \( C^1 \). We assume that for each \((t, x)\) in \( \Omega \), the function \( v \mapsto L(t, x, v) \) is strictly convex. Since \( L \) is smooth, this is equivalent to the requirement that the vector-valued function \( v \mapsto L_v(t, x, v) \) be strictly monotonic. In particular, any extremal arc for \( L \) whose graph lies in the set \( \Omega \) will be continuously differentiable by (WE1).

Associated with the Lagrangian \( L \) is the Hamiltonian, defined by taking the Legendre-Fenchel transform of \( L \) with respect to the velocity variables:

\[
H(t, x, p) := \sup \{ pv - L(t, x, v) : v \in \mathbb{R}^n \}.
\]

(1)

Formally, \( H \) maps \( \Omega \times (\mathbb{R}^n)^* \) into the extended real line \( \mathbb{R} \cup \{+\infty\} \): whereas the arguments \( x \) and \( v \) of the Lagrangian \( L \) are (column) vectors in \( \mathbb{R}^n \), the argument \( p \) in (1) is a (row) vector in \((\mathbb{R}^n)^*\). Thus the juxtaposition \( pv \) in (1) indicates the product of a \( 1 \times n \) matrix with an \( n \times 1 \) matrix, which produces a scalar result.

Since the function \( v \mapsto L(t, x, v) \) is strictly convex, the function being maximized on the right side of (1) is strictly concave. If the maximum is attained at all, the maximizing \( v \) must be unique, and characterized by a vanishing gradient. Thus the point \( v = \Phi(t, x, p) \) is the maximizer if and only if

\[
v = \Phi(t, x, p) \iff H(t, x, p) = pv - L(t, x, v) \iff p = L_v(t, x, v). \tag{2}
\]

An important consequence of (2) is the pair of identities

\[
\Phi(t, x, L_v(t, x, v)) = v, \quad L_v(t, x, \Phi(t, x, p)) = p. \tag{3}
\]

If \( L \) has superlinear growth as \(|v| \to \infty\), these hold for all \((t, x)\) in \( \Omega \), \( v \) in \( \mathbb{R}^n \), and \( p \) in \((\mathbb{R}^n)^*\). If \( L \) has slow growth, they hold in suitable subsets. The calculations below apply on the interior of these sets.

In favourable circumstances, the mapping \( \Phi \) will be differentiable. We may then use (3) to compute the partial derivatives of \( H \) as follows.

\[
H_t(t, x, p) = \frac{\partial}{\partial t} [p\Phi(t, x, p) - L(t, x, \Phi(t, x, p))] = p\Phi_t - L_t - L_v\Phi_t = -L_t(t, x, \Phi(t, x, p)) \tag{4}
\]

\[
H_x(t, x, p) = \nabla_x [p\Phi(t, x, p) - L(t, x, \Phi(t, x, p))] = p\Phi_x - L_x - L_v\Phi_x = -L_x(t, x, \Phi(t, x, p)) \tag{5}
\]

\[
H_p(t, x, p) = \nabla_p [p\Phi(t, x, p) - L(t, x, \Phi(t, x, p))] = \Phi + p\Phi_p - L_v\Phi_p = \Phi(t, x, p) \tag{6}
\]

The identity \( \Phi \equiv H_p \) in (6) is of particular importance: it frees us to replace \( \Phi \) by \( H_p \) throughout the discussion below. We must never lose sight of its meaning, however:
When these statements hold, we also have

\[ L(t, x, v) = pv - H(t, x, p) \iff v = H_p(t, x, p) \iff 0 = \nabla_p [pv - H(t, x, p)]. \]

Now it is easy to see that definition (1) produces a function \( H \) that is convex with respect to the variable \( p \), so the last condition here holds if and only if the vector \( p \) actually maximizes the function in square brackets. Thus the Lagrangian can be recovered from the Hamiltonian by the Legendre-Fenchel transform as well:

\[ L(t, x, v) = \sup_{p \in (\mathbb{R}^n)^*} \{ pv - H(t, x, p) \}. \tag{7} \]

(“The Legendre-Fenchel transform is an involution: apply it twice and you get back where you started.”)

The calculations above allow us to rewrite much of the theory developed so far in Hamiltonian terms. The following result deals with the Euler-Lagrange equation, whose equivalent formulation in part (c) is known as “the canonical equations of Hamilton”.

A.1. Theorem (Hamiltonian Necessary Conditions). Let \( \hat{x} \) and \( p \) be arcs in \( \mathbb{R}^n \) and \( (\mathbb{R}^n)^* \). Then the following are equivalent:

(a) \( \frac{d}{dt} \hat{L}_v(t) = \hat{L}_x(t) \), and \( p(t) = \hat{L}_v(t) \);

(b) \( [\hat{p}(t) p(t)] = \nabla_{x,v} L(t, \hat{x}(t), \hat{x}(t)) \);

(c) \( -\hat{p}(t) = H_x(t, \hat{x}(t), p(t)) \), \( \hat{x}(t) = H_p(t, \hat{x}(t), p(t)) \).

When these statements hold, we also have

\[ H(t, \hat{x}(t), p(t)) = p(t)\hat{x}(t) - L(t, \hat{x}(t), \hat{x}(t)) \]

or equivalently,

\[ v = \hat{x}(t) \text{ maximizes } v \mapsto p(t)v - L(t, \hat{x}(t), v) \text{ uniquely over } v \in \mathbb{R}^n. \tag{8b} \]

Proof. (a \iff b) Obvious.

(b \implies c) From (2), the condition \( p(t) = L_v(t, \hat{x}(t), \hat{x}(t)) \) is equivalent to \( \hat{x}(t) = \Phi(t, \hat{x}(t), p(t)) \). Thus a simple application of (5) and (6) leads from (b) to (c).

(c \implies b) From (6), we get \( \hat{x}(t) = \Phi(t, \hat{x}(t), p(t)) \). That is, by (3), \( p(t) = L_v(t, \hat{x}(t), \hat{x}(t)) \). From (5), we have \( -\hat{p}(t) = -L_x(t, \hat{x}(t), \hat{x}(t)) \).

Notice that the maximization statement in (8b) can be rewritten as the inequality

\[ L(t, \hat{x}(t), v) - L(t, \hat{x}(t), \hat{x}(t)) - \hat{L}_v(t) \left[ v - \hat{x}(t) \right] \geq 0 \quad \forall v \in \mathbb{R}^n. \]

This is precisely the necessary condition of Weierstrass for \( \hat{x} \). (As we have seen in class, this condition holds for every extremal under the convexity assumption we have made on \( L \).)

Finally, consider the second Weierstrass-Erdmann condition for \( \hat{x} \). This states that \( \frac{d}{dt} \left[ \hat{L} - \hat{L}_v \hat{x} \right] = \hat{L}_t \). Thanks to (4) and the definition \( p(t) = \hat{L}_v(t) \), condition (8a)
allows us to rewrite (WE2) as
\[ \frac{d}{dt} H(t, \hat{x}(t), p(t)) = H_t(t, \hat{x}(t), p(t)). \]
In particular, for problems where the Lagrangian—and hence the Hamiltonian—has no explicit time-dependence, the Hamiltonian will be constant along every extremal trajectory.

A.2. Example. Consider, for fixed \( q > 1 \), the function \( L(t, x, v) = \frac{|v|^q}{q} \). (Here we use the Euclidean norm.) We write \( \text{sgn}(v) \) for the vector “signum” of \( v \), namely
\[ \text{sgn}(v) = \begin{cases} v^T |v|, & \text{if } v \neq 0, \\ 0^T, & \text{if } v = 0. \end{cases} \]
This gives \( \text{sgn}(v)v = |v| \) for each \( v \in \mathbb{R}^n \), and \( p\text{sgn}(p) = |p| \) for each \( p \in (\mathbb{R}^n)^* \). We may now compute \( L_v(t, x, v) = |v| q - 1 \text{sgn}(v) \). Thus (2) specifies that \( \Phi(t, x, p) = v \) exactly when \( p = |v|^{q-1} \text{sgn}(v) \), i.e., when \( |p| = |v|^{q-1} \) and \( p\text{sgn}(p)^T = \text{sgn}(v) \). These relations give \( v = |p|^{1/(q-1)} \text{sgn}(p) \), so
\[ H(t, x, p) = p|p|^{1/(q-1)} \text{sgn}(p) - \frac{|p|^{q/(q-1)} \text{sgn}(p)}{q} = |p|^{q'} q, \]
where \( q' \) is the exponent conjugate to \( q \), defined by \( \frac{1}{q} + \frac{1}{q'} = 1 \).

B. The Principle of Optimality

Relative Optimality. Recall that any piecewise smooth function on a closed interval is called an arc. We will say that an arc \( x \), with associated interval of definition \( [a, b] \), has graph in \( \Omega \) whenever
\[ (t, x(t)) \in \Omega \quad \forall t \in (a, b). \]
(Strictly speaking, the graph of \( x \) is a set including the two endpoints \( (a, x(a)) \) and \( (b, x(b)) \), and our terminology allows these two points to lie outside \( \Omega \). This is not a bug; it’s a feature—and a useful one, too, as we shall see.)

B.1. Definition. Let an arc \( \hat{x} \), with interval of definition \( [a, b] \), have graph in \( \Omega \). To say that \( \hat{x} \) is optimal relative to its endpoints (for the Lagrangian \( L \)) means that \( \hat{x} \) solves the following fixed-endpoint problem in the calculus of variations:
\[
\text{minimize } \int_a^b L(t, x(t), \dot{x}(t)) \, dt \\
\text{subject to } x \in \text{PWS}[a, b], \ x(a) = \hat{x}(a), \ x(b) = \hat{x}(b).
\]
In words, among all arcs \( x \) joining the endpoints of \( \hat{x} \), the choice \( x = \hat{x} \) gives the smallest integral value.

Optimality relative to endpoints is a hereditary property: if an arc has it, then so do all its sub-arcs. This is the point of the next paragraph.
B.2. The Principle of Optimality. Suppose \( \hat{x} \), an arc with associated interval \([a, b]\), is optimal relative to its endpoints for the Lagrangian \( L \). Then for any subinterval \([\alpha, \beta]\) of \([a, b]\), the restricted arc \( \hat{x}_{[\alpha, \beta]} \) is optimal relative to its endpoints.

The reason for this is simple. Let \( \tilde{\Lambda}[y] := \int_\alpha^\beta L(t, x(t), \dot{x}(t)) \, dt \). If there were an arc \( y \) on \([\alpha, \beta]\) with graph in \( \Omega \), endpoints \( y(\alpha) = \hat{x}(\alpha) \), \( y(\beta) = \hat{x}(\beta) \), and objective value \( \tilde{\Lambda}[y] < \Lambda[\hat{x}] \), then we could piece together \( \hat{x} \) and \( y \) to define the arc

\[
z(t) := \begin{cases} \hat{x}(t), & a < t \leq \alpha, \\ y(t), & \alpha < t \leq \beta, \\ \hat{x}(t), & \beta < t \leq b, \end{cases}
\]

for which \( \Lambda[z] < \Lambda[\hat{x}] \). This would contradict the optimality of \( \hat{x} \), so no such \( y \) can exist. In other words, \( \hat{x} \) is optimal relative to its endpoints. (This argument depends on our ability to manage piecewise smooth competing functions: even if both \( \hat{x} \) and \( y \) are smooth arcs, the concatenated arc \( z \) may be only piecewise smooth.)

C. The Hamilton-Jacobi Verification Technique

It is sometimes possible to simplify a variational problem by adding a constant to the objective value—especially when that constant comes in the form of the integral of an exact differential. We have met this idea before; here is another example.

Example. The arc \( \hat{x}(t) = e^t \) is a global minimizer in the problem

\[
\min \left\{ \Lambda[x] \right\}^{\text{def}} \left[ \int_0^1 \left( \dot{x}(t)^2 + x(t)^2 \right) \, dt : x(0) = 1, \; x(1) = e \right].
\]

This is evident because for any admissible arc \( x \),

\[
\Lambda[x] = \int_0^1 \left( \dot{x}(t)^2 - 2x(t)\dot{x}(t) + x(t)^2 \right) \, dt + \int_0^1 2x(t)\dot{x}(t) \, dt
\]

\[
= \int_0^1 (\dot{x}(t) - x(t))^2 \, dt + x(t)^2 \bigg|_{t=0}^{t=1}
\]

\[
= e^2 - 1 + \int_0^1 (\dot{x}(t) - x(t))^2 \, dt.
\]

The integral on the RHS vanishes when \( x = \hat{x} \), but gives a positive value for every other admissible arc.

We can formalize this trick.

C.0. Theorem (Basic Verification—Lagrangian). If some \( W \in C^1(\Omega) \) obeys

(i) \( W_t(t, x) + W_x(t, x)v - L(t, x, v) \leq 0 \; \forall (t, x, v) \in \Omega \times \mathbb{R}^n \).

then any arc \( \hat{x} \in PWS[a, b] \) with \( (t, \hat{x}(t)) \in \Omega \; \forall t \in [a, b] \) along which

(ii) \( W_t(t, \hat{x}(t)) + W_x(t, \hat{x}(t))\dot{x}(t) - L(t, \hat{x}(t), \dot{x}(t)) = 0 \; \forall t \in [a, b] \)

is optimal relative to its endpoints.
C.2. Theorem (Verification). Suppose some $W \in C^1(\Omega)$ obeys

(i) $W(t, x(t)) = 0 \forall (t, x)$. Let $\hat{x} \in PWS[a, b]$ be any arc with $(t, \hat{x}(t)) \in \Omega \forall t \in (a, b)$. If

(ii) $W(t, \hat{x}(t)) + H(t, \hat{x}(t), W_x(t, \hat{x}(t))) = 0$ for all $t \in (a, b)$,

(iii) $\hat{x}(t) = H_p(t, \hat{x}(t), W_x(t, \hat{x}(t)))$ for all $t \in (a, b)$, and

(iv) these two limits exist: $w_0 := \lim_{t \to a^+} W(t, \hat{x}(t))$, $w_1 := \lim_{t \to b^-} W(t, \hat{x}(t))$.

Proof. For any arc $x$ with graph in $\Omega$ and $x(a) = \hat{x}(a)$, $x(b) = \hat{x}(b)$,

$$
\int_a^b L(t, x(t), \dot{x}(t)) dt \geq \int_a^b \frac{d}{dt} W(t, x(t)) dt
$$

$$
= W(b, x(b)) - W(a, x(a))
$$

$$
= W(b, \hat{x}(b)) - W(a, \hat{x}(a))
$$

$$
= \int_a^b \frac{d}{dt} W(t, \hat{x}(t)) dt = \int_a^b L(t, \hat{x}(t), \dot{x}(t)) dt.
$$

To express this in Hamiltonian form, notice that hypothesis (i) is equivalent to

$$
\sup_v \{W_t(t, x) + W_x(t, x)v - L(t, x, v)\} \leq 0 \forall (t, x) \in \Omega.
$$

while line (ii) says that when $(t, x) = (t, \hat{x}(t))$, the maximizing $v$ is precisely $\dot{x}(t)$.

C.1. Theorem (Basic Verification—Hamiltonian). If some $W$ in $C^1(\Omega)$ obeys

(i) $W(t, x) + H(t, x, W_x(t, x)) \leq 0 \forall (t, x) \in \Omega$,

then any arc $\hat{x} \in PWS[a, b]$ with $(t, \hat{x}(t)) \in \Omega$ for all $t \in [a, b]$ along which

(ii) $W(t, \hat{x}(t)) + H(t, \hat{x}(t), W_x(t, \hat{x}(t))) = 0 \forall t \in (a, b), and

(iii) $\hat{x}(t) = H_p(t, \hat{x}(t), W_x(t, \hat{x}(t))) \forall t \in (a, b)$,

is optimal relative to its endpoints. Furthermore,

$$
\int_a^b L(t, \hat{x}(t), \dot{x}(t)) dt = W(b, \hat{x}(b)) - W(a, \hat{x}(a)).
$$

Proof. Line (i) is equivalent to hypothesis (i) of Theorem C.0. Lines (ii)–(iii) together are equivalent to hypothesis (ii) of that result.

Endpoints on the Boundary of $\Omega$. In cases where the arc whose optimality is in question has endpoints lying outside the set $\Omega$, there is more work to do. Here is a suitable modification of Theorem C.1. Instead of simply asserting that the trial arc $\hat{x}$ is optimal relative to its endpoints, this result offers the apparently weaker conclusion that among all arcs joining the endpoints of $\hat{x}$ and satisfying certain other conditions, it is $\hat{x}$ that gives the minimum. We will discuss these other conditions after proving the theorem.

VII. Hamiltonian Methods
then the arc \( \hat{x} \) with associated interval \([a, b]\) solves the following problem:

\[
\min_{x, [\alpha, \beta]} \left\{ \int_{\alpha}^{\beta} L(t, x(t), \dot{x}(t)) \, dt : (t, x(t)) \in \Omega \quad \forall t \in (\alpha, \beta), \right. \\
\left. \quad \lim_{r \to \alpha^+} W(r, x(r)) = w_0, \\
\quad \lim_{s \to \beta^-} W(s, x(s)) = w_1 \right\}.
\]

Moreover, the numerical value of the minimum in \((\hat{P})\) is \(w_1 - w_0\).

**Proof.** Thanks to the definition of \(H\), hypothesis (i) implies that

\[ L(t, x, v) \geq W(t, x) + W_x(t, x)v \quad \forall v \in \mathbb{R}^n, \forall (t, x) \in \Omega. \]

Now consider any arc \(x \in PWS[\alpha, \beta]\) admissible for \((\hat{P})\). Fix any closed subinterval \([r, s]\) of \((\alpha, \beta)\). On \([r, s]\), choosing \((x, v) = (x(t), \dot{x}(t))\) in (*) above gives

\[
\int_r^s L(t, x(t), \dot{x}(t)) \, dt = \left[ \int_r^s L - \frac{d}{dt} W(t, x(t)) \right] \, dt + W(t, x(t)) \bigg|_r^s \\
\geq 0 + W(s, x(s)) - W(r, x(r)).
\]

Now in the limit as \(r \to \alpha^+\) and \(s \to \beta^-\) in this inequality, we obtain—by the admissibility condition for \((\hat{P})\)—

\[
\int_{\alpha}^{\beta} L(t, x(t), \dot{x}(t)) \, dt \geq \lim_{s \to \beta^-} W(s, x(s)) - \lim_{r \to \alpha^+} W(r, x(r)) = w_1 - w_0. \quad (\dagger)
\]

Now we repeat the steps above for the arc \(x = \hat{x}\) on \([a, b]\), noting that instead of an inequality based on (i), this choice gives an identity based on (ii)–(iii). Thus

\[
\int_{\alpha}^{\beta} L(t, \hat{x}(t), \dot{x}(t)) \, dt = w_1 - w_0. \quad (\ddagger)
\]

Comparing (\dagger) and (\ddagger) gives the result. // // //

**Discussion.** Any function \(W(t, x)\) satisfying the conditions of Theorem C.2 on some region \(\Omega\), relative to some arc \(\hat{x}\), is called a *verification function for \(\hat{x}\)*, since it “verifies” the optimality of this arc relative to its endpoints.

To appreciate the significance of Theorem C.2, it helps to see how it generalizes Theorem C.1. Suppose, therefore, that the two endpoints \((a, \hat{x}(a))\) and \((b, \hat{x}(b))\) of the arc in question actually lie in \(\Omega\). Then the continuity of \(W\) throughout \(\Omega\) guarantees the existence of the limits \(w_0 = W(a, \hat{x}(a))\) and \(w_1 = W(b, \hat{x}(b))\) mentioned in condition (iv). Indeed, any arc \(x\) with the same interval of definition and endpoints as \(\hat{x}\) will automatically satisfy \(W(a, x(a)) = w_0\) and \(W(b, x(b)) = w_1\), and consequently be admissible in problem \((\hat{P})\). The result of Theorem C.2 implies that \(\hat{x}\) yields the lowest integral value among all such arcs, i.e., that \(\hat{x}\) is optimal relative to its endpoints.

But Theorem C.2 says more: while \(\hat{x}\) compares favourably against all other arcs with the same endpoints, it can actually compete against a wider class of challengers.
VII. Hamiltonian Methods 7

For the endpoint conditions in problem (\(\hat{P}\)) can conceivably be satisfied not just by arcs with the same endpoints as \(\hat{x}\), but by any other arcs whose initial and terminal points lie on the corresponding level sets of \(W\) passing through the initial and terminal points of \(\hat{x}\). Theorem C.1 shows that even in this wider class of competitors, it is \(\hat{x}\) that gives the minimum.

The final refinement in Theorem C.2 is its provision for initial and final points that lie on the boundary of the set \(\Omega\) where the integrand \(L\) and the verification function \(W\) are defined. As the proof shows, a modification of the proof of Theorem C.1 goes through if we understand all our integrals as (possibly) improper Riemann integrals. This will turn out to be important in our later study of the value function.

The Hamilton-Jacobi Equation. A particularly attractive way to satisfy conditions (i) and (ii) in the verification theorems C.1 and C.2 is to choose the function \(W\) as a smooth solution of the Hamilton-Jacobi Equation:

\[
W_t(t, x) + H(t, x, W_x(t, x)) = 0, \quad \forall (t, x) \in \Omega.
\] (11)

This equation typically has many solutions, as the next example illustrates. But for each solution \(W\) one finds, condition (iii) in the cited theorems holds along any arc \(x\) for which \(\dot{x}(t) = H_p(t, x, W_x(t, x))\) holds. In other words, when a smooth solution \(W\) of (11) is available, the definition \(\phi(t, x) = H_p(t, x, W_x(t, x))\) leads to a slope field whose trajectories (\(\dot{x} = \phi(t, x)\)) are all optimal relative to their endpoints in \(\Omega\).

C.3. Example (Three in One). Consider the simple Lagrangian \(L(t, x, v) = \frac{1}{2}v^2\).

Since \(L\) is convex, we know that any extremal arc will be optimal relative to its endpoints. But to illustrate our theory, we consider several possible verification functions.

The Hamiltonian corresponding to \(L = \frac{1}{2}v^2\) is easy to compute (Example A.2):

\[
H(t, x, p) = \sup \{pv - \frac{1}{2}v^2\} = \frac{1}{2}p^2; \quad H_p(t, x, p) = p.
\]

(i) Consider \(W(t, x) = x - \frac{1}{2}t\) with \(\Omega = \mathbb{R}^2\). Here we find

\[
W_t(t, x) + \frac{1}{2}W_x(t, x)^2 = -\frac{1}{2} + \frac{1}{2}(1)^2 = 0,
\]

so (11) holds everywhere, and thus conditions (i)–(ii) of Theorem C.1 hold. According to that result, any arc \(\hat{x}\) obeying condition (iii), i.e.,

\[
\dot{x}(t) = H_p(t, x(t), W_x(t, x(t))) = W_x(t, x(t)) = 1
\]

will be optimal relative to its endpoints. This criterion identifies the family of all straight lines with slope 1.

(ii) Fix \(\varepsilon > 0\) and consider \(W(t, x) = \frac{(x + \varepsilon)^2}{2(t + \varepsilon)}\). This function is well-behaved on the set \(\Omega = \{(t, x) : t > -\varepsilon, x \in \mathbb{R}\}\). Again the left side of (11) amounts to

\[
W_t(t, x) + \frac{1}{2}W_x(t, x)^2 = \frac{(x + \varepsilon)^2}{2(t + \varepsilon)^2} + \frac{1}{2} \left[\frac{x + \varepsilon}{t + \varepsilon}\right]^2 = 0,
\]

so again (11) holds, and conditions C.1(i)–(ii) follow. Condition C.1(iii) requires that the candidate arc \(\hat{x}\) obey \(\dot{x}(t) = \frac{1}{2}W_x(t, x(t)) = (x + \varepsilon)/(t + \varepsilon)\). The general solution of
this linear equation is \( \hat{x}(t) = -\varepsilon + m(t + \varepsilon), m \in \mathbb{R} \). From Theorem C.1, we deduce that any of these straight lines is optimal relative to its endpoints, provided both endpoints lie in the open set \( \Omega \). To push the left endpoint back to \( t = -\varepsilon \), however, we cannot rely on Theorem C.1: we must switch to Theorem C.2, and pay some attention to the limiting endpoint conditions in problem \( (\hat{P}) \). Details are discussed below for the particular case \( \varepsilon = 0 \).

(iii) Finally, take \( W(t, x) = x^2/(2t) \) on the open set \( \Omega = \{(t, x) : t > 0, x \in \mathbb{R}\} \). Since this is simply the case \( \varepsilon = 0 \) of (ii), \( W \) satisfies (11) and conditions C.2(i)–(ii) hold. Condition C.2(iii) requires that the derivative of the candidate arc \( \hat{x} \) coincide with \( H_p(t, \hat{x}(t), W_x(t, \hat{x}(t))) = W_x(t, \hat{x}(t)) \). In our case this requirement reduces to \( \dot{\hat{x}}(t) = \hat{x}(t)/t \), which is true for any arc \( \hat{x}(t) = mt, m \in \mathbb{R} \).

Let us use Theorem C.2 to show that any straight line from the origin to a point in the right half plane is optimal relative to its endpoints. It suffices to confirm that any arc \( x \) with graph in \( \Omega \) and \( x(0) = 0 \) satisfies the left endpoint condition in problem \( (\hat{P}) \). To do this, we note that since \( x \) is an arc, its derivative exists and is bounded in some interval of the form \( (0, h) \), \( h > 0 \). Consequently L’Hospital’s rule can be applied to evaluate the limit

\[
\lim_{t \to 0^+} W(t, x(t)) = \lim_{t \to 0^+} \frac{x(t)^2}{2t} = \lim_{t \to 0^+} \frac{2x(t)\dot{x}(t)}{2} = 0.
\]

This confirms the first part of hypothesis C.2(iv), with \( w_0 = 0 \), simultaneously. The desired result follows.

This last choice of \( W \) illustrates why endpoints on the boundary of \( \Omega \) force us to consider generalized endpoint conditions in \( (\hat{P}) \). A simpler condition like

\[
\lim_{(t,x) \to (a, \hat{x}(a)) \atop (t,x) \in \Omega} W(t, x) = w_0
\]

would be inadequate because the limit on the left-hand side fails to exist when \( (a, \hat{x}(a)) = (0, 0) \): approaching the origin along any curve \( x = k\sqrt{t} \) will produce a limit value of \( k^2 \). Since different curves give different limits, the limit in (12) fails to exist. Notice, however, that all the curves giving nonzero limits are not arcs: for arcs, local boundedness of the derivative is often enough to derive the endpoint behaviour necessary to confirm condition C.2(iv) and apply problem \( (\hat{P}) \).

If one’s primary goal were to confirm that the arc \( \hat{x}(t) = t \) on the interval \([0, 1]\) is optimal relative to its endpoints, then any one of the three functions described above would do the job. This is a nice feature of verification theory: to handle a single specific arc, there are probably lots of suitable verification functions, and only one is required. To explain where all these functions come from, at least partially, notice that each of them actually verifies the optimality of a whole family of extremals for the given Lagrangian, with different families corresponding to different choices for \( W \). The arc \( \hat{x}(t) = t \) is the only one whose optimality can be verified by each of the three \( W \)-functions above.

// // //

**Optimal Feedback.** Example C.3 illustrates an important feature of the Hamilton-Jacobi theory: once we have a smooth solution \( W(t, x) \) of (11), it will verify the
optimality of a whole family of arcs. This family is characterized by the differential equation implicit in condition C.1(iii), namely,
\[ \dot{x}(t) = \phi(t, x(t)), \quad \text{where} \quad \phi(t, x) := H_p(t, x, W_x(t, x)). \]  
(13)

Here is the striking conclusion:

Every smooth solution \( W \) of the Hamilton-Jacobi Equation (11) induces a differential equation (13). All trajectories of (13) are optimal relative to their endpoints.

Engineers use the Hamilton-Jacobi theory to produce feedback control laws. The idea is to choose the right solution \( W \) of (11) (which one is “the right one” will be discussed shortly), and calculate the corresponding function \( \phi(t, x) \). Then just run your physical system (manufacturing plant, etc.) according to the differential equation (13): even if a disturbance comes along and perturbs the state \( x \) in some way beyond your control, using the same law to generate the system’s subsequent motion will ensure that the evolution is best possible from the time of disturbance. This ability to reject disturbances and always follow the best possible course is the key to reliable systems operation.

Practice. With \( L(t, x, v) = \frac{1}{2}v^2 \), consider \( W(t, x) = \frac{(x + \varepsilon)^2}{2(t + \eta)} \). Find the set \( \Omega \) on which \( W \) is \( C^2 \), and show that \( W \) satisfies (11) in \( \Omega \). Then identify the family of arcs \( x \) whose optimality is verified by this function \( W \).

Extremality. The theory of necessary conditions implies that every arc optimal relative to its endpoints must satisfy the Euler-Lagrange equation. Thus any arc \( \hat{x} \) satisfying the conditions of Theorem C.1 or Theorem C.2 must be extremal. But in contrast to the hard work required to prove the Euler-Lagrange equations for general minimizers, the task is very simple when there exists a verification function \( W \) of class \( C^2 \).

Let us put ourselves in the situation described in the statement of Theorem C.1, and add the hypothesis that \( W \in C^2(\Omega) \). Conditions C.1(i)–(iii) imply that \( \hat{x} \) is \( C^1 \), and satisfies both
\[ (a) \quad W_t(t, x) + W_x(t, x)v - L(t, x, v) \leq 0 \quad \forall v \in \mathbb{R}^n, \forall (t, x) \in \Omega; \quad \text{and} \]
\[ (b) \quad W_t(t, \hat{x}(t)) + W_x(t, \hat{x}(t))\dot{\hat{x}}(t) - L(t, \hat{x}(t), \dot{\hat{x}}(t)) = 0 \quad \text{for all} \quad t \in (a, b). \]

For each fixed \( t \) in \( (a, b) \), these two statements imply that the point \( (x, v) = (\hat{x}(t), \dot{\hat{x}}(t)) \) gives a local maximum to the function
\[ (x, v) \mapsto W_t(t, x) + W_x(t, x)v - L(t, x, v). \]

In particular, the \( x \)- and \( v \)-partial derivatives of this function must both vanish at \( (\hat{x}(t), \dot{\hat{x}}(t)) \): this gives
\[ 0 = W_{tx}(t, \hat{x}(t)) + W_{xx}(t, \hat{x}(t))\dot{\hat{x}}(t) - \dot{L}_x(t, \hat{x}(t), \dot{\hat{x}}(t)) \]
\[ (14) \]
\[ 0 = W_x(t, \hat{x}(t)) - \dot{L}_v(t, \hat{x}(t), \dot{\hat{x}}(t)). \]
\[ (15) \]

Here (15) reveals \( \dot{L}_v(t) = \dot{W}_x(t) \), and in view of this, (14) produces \( \frac{d}{dt}\dot{\hat{x}}(t) = \dot{L}_x(t) \). Thus \( \hat{x} \) satisfies the differentiated Euler-Lagrange equation at each time \( t \) in \( (a, b) \).
This derivation of the Euler-Lagrange equation makes a nice little exercise, but it raises a serious question: where is the hard work in the verification theory? If it is not in the proof of the verification theorem, and not in the proof of extremality, then it must be in the construction of the verification function $W$ itself. This is certainly the case: we will investigate sources of verification functions in later sections.

The preceding arguments about extremality are particularly attractive when we have in hand a $C^2$ function $W$ that actually solves the Hamilton-Jacobi Equation (11). As noted above, the hypotheses of Theorem C.2 are then fulfilled by any arc satisfying the differential equation $\dot{x}(t) = \phi(t, x(t))$, where $\phi(t, x) := H_p(t, x, W_x(t, x))$. The arguments just given show that each of these arcs will necessarily be an extremal for $L$. Thus we can view the function $\phi(t, x)$ as the “slope function” for a whole family of optimal trajectories in the set $\Omega$. This is the point of contact between the Hamilton-Jacobi verification technique and the classical theory of “fields of extremals”. Our approach so far has been to solve (11) and then derive a slope function; in the theory of fields, we start with a slope function and generate a solution to (11).

**Problems with Fixed Endpoints.** For a calculus of variations problem in which the fundamental interval $[a, b]$ and the endpoints $(a, A)$, $(b, B)$ are fixed in advance, the theory above is tailor-made. Indeed, in this case, any admissible arc on $[a, b]$ that happens to be optimal relative to its endpoints will solve the given problem. Thus Theorem C.2 applies directly in the fixed-endpoint case.

**Problems with Free Endpoints.** The modified endpoint conditions introduced in Theorem C.2 give useful flexibility to the theory. Consider, for simplicity, the case where both endpoints of the candidate arc $\hat{x}$ lie in the open set $\Omega$. In this case the endpoint conditions of problem $(\bar{P})$ require only that every competing arc $x$ should have its initial point on the curve $S_0 = \{(t, x) \in \Omega : W(t, x) - w_0 = 0\}$ and its final point on the curve $S_1 = \{(t, x) \in \Omega : W(t, x) - w_1 = 0\}$. Thus the conclusion of Theorem C.2 is actually that the arc $\hat{x}$ on the interval $[a, b]$ solves a certain curve-to-curve problem in the calculus of variations, where the curves are $S_0$ and $S_1$.

Let us illustrate this with the problem from Example C.3, taking as our arc $\hat{x}(t) = t$ on the interval $[1, 3]$. The function $W(t, x) = x^2/(2t)$ from part (iii) has $W(1, 1) = 1/2$ and $W(3, 3) = 3/2$, so the level curves of $W$ through these points are $S_0 : x^2 = 2t$ and $S_1 : x^2 = 3t$.

Theorem C.2 shows that $\hat{x}$ gives the minimum in the problem of joining $S_0$ to $S_1$ while minimizing the integral of $\frac{1}{2}\dot{x}^2$. Indeed, it shows that any one of the extremals identified by $W$, i.e., any straight line through the origin, accumulates an equal value for the integral of $\frac{1}{2}\dot{x}^2$ as it passes between these two curves: all of these arcs are minimizers for this curve-to-curve problem. See Figure C.1.

Similarly, the verification function $W(t, x) = x - \frac{1}{2}t$ from part (i) of Example C.3 shows that all of the straight lines of slope 1 from the curve $S_0$ to $S_1$, where

$$S_0 : x - \frac{1}{2}t = W(t, x) = W(1, 1) = \frac{1}{2}, \quad S_1 : x - \frac{1}{2}t = W(t, x) = W(3, 3) = \frac{3}{2},$$

produce the same integral value of $\frac{1}{2}\dot{x}^2$, and this is the lowest possible integral value for any curve from $S_0$ to $S_1$. See Figure C.2.
To harness this observation and turn Theorem C.2 into a tool for verifying the optimality of particular arcs in point-to-curve or even curve-to-curve problems, it will be necessary to choose a verification function $W$ whose level curves coincide exactly with the given initial and terminal curves in the problem’s statement. This is usually possible in point-to-curve and curve-to-point problems, but rarely possible in curve-to-curve problems. The key to arranging it lies in making sure that the trajectories of the differential equation $\dot{x} = \phi(t,x)$ characterizing the optimal arcs for $W$ all cross the given target curves transversally, in the sense of the usual transversality conditions for initial and terminal points (case $\ell^{(1)} \equiv 0$, $\ell^{(2)} \equiv 0$). Details are given in Appendices A3.11-13 in Sagan’s book; Example D.4 below illustrates some of the basic ideas.

**D. Value Functions**

For all the potential power and elegance of the verification technique, we still need a reliable source of verification functions. In this section we describe one possibility, the value function.

Let a Lagrangian $L(t,x,v)$ be given. Fix an initial time $a_0$ and a starting vector
$A_0$, and then define, for each $T > a_0$ and $X \in \mathbb{R}^n$, the value function

$$V(T, X) := \inf \left\{ \int_{a_0}^{T} L(t, x(t), \dot{x}(t)) \, dt : x \in \text{PWS}[a_0, T], \right.$$ 

$$x(a_0) = A_0, \ x(T) = X, \quad P(T, X)$$

$$(t, x(t)) \in \Omega \ \forall \ t \in (a_0, T) \}.$$ 

Let $\Sigma(T, X)$ denote the subset of $\text{PWS}[a_0, T]$ consisting of the minimizing arcs in problem $P(T, X)$. The principle of optimality implies the following property of $V$:

**D.1. Lemma.** For any arc $x$ with associated interval $[a, b]$, and graph in $\Omega$, the function $v$ defined below is nonincreasing on $[a, b]$:

$$v(t) := V(t, x(t)) - \int_{a}^{t} L(r, x(r), \dot{x}(r)) \, dr \quad (16)$$

Moreover, for any $(T, X)$ in $\Omega$, the related function $\hat{v}$ is constant along any arc $\hat{x}$ in $\Sigma(T, X)$.

**Proof.** Consider first any arc $x$ with associated interval $[a, b]$, and graph in $\Omega$. Choose any nondegenerate subinterval $[t_1, t_2]$ of $[a, b]$, and note that

$$v(t_2) - v(t_1) = V(t_2, x(t_2)) - \left[ V(t_1, x(t_1)) + \int_{t_1}^{t_2} L(r, x(r), \dot{x}(r)) \, dr \right]. \quad (17)$$
Figure D.1. Sketch for the proof of Lemma D.1.

To prove that this quantity is nonpositive, consider Figure D.1.

Fix any \( \varepsilon > 0 \). By the definition of \( V(t_1, x(t_1)) \), there must be some arc \( z \) with graph in \( \Omega \) such that \( z(a_0) = A_0, \ z(t_1) = x(t_1) \), and

\[
\int_{a_0}^{t_1} L(t, z(t), \dot{z}(t)) \, dt \leq V(t_1, x(t_1)) + \varepsilon.
\]

If we define

\[
y(t) := \begin{cases} 
z(t), & a_0 \leq t < t_1, \\
x(t), & t_1 \leq t \leq t_2,
\end{cases}
\]

then \( y \) is a piecewise smooth function joining \( (a_0, A_0) \) to \( (t_2, x(t_2)) \). As such, we can be sure that

\[
V(t_2, x(t_2)) \leq \int_{a_0}^{t_2} L(t, y(t), \dot{y}(t)) \, dt \leq \varepsilon + V(t_1, x(t_1)) + \int_{t_1}^{t_2} L(r, x(r), \dot{x}(r)) \, dr.
\]

Since \( \varepsilon > 0 \) is arbitrary, however, (18) must remain valid in the limit as \( \varepsilon \to 0^+ \). This completes the proof that the quantity in (17) is at most zero.

Now if \( \hat{x} \) lies in \( \Sigma(T, X) \) for some \( (T, X) \in \Omega \), and we define \( \hat{v} \) as in (16), then the principle of optimality implies that

\[
V(t_2, \hat{x}(t_2)) = \int_{a_0}^{t_2} L(t, \hat{x}, \dot{\hat{x}}) \, dt = V(t_1, \hat{x}(t_1)) + \int_{t_1}^{t_2} L(t, \hat{x}, \dot{\hat{x}}) \, dt
\]

\[
= V(t_1, \hat{x}(t_1)) + \left[ \int_{a_0}^{t_2} L(t, \hat{x}, \dot{\hat{x}}) \, dt - \int_{a_0}^{t_1} L(t, \hat{x}, \dot{\hat{x}}) \, dt \right].
\]

Rearranging the result shows that \( \hat{v}(t_2) = \hat{v}(t_1) \); since \( a_0 \leq t_1 < t_2 \leq T \) are arbitrary, the function \( \hat{v} \) must be constant, as claimed. /////
D.2. Corollary. Suppose $V$ is differentiable at some point $(t, x)$ in $\Omega$. Then

$$V_t(t, x) + H(t, x, V_x(t, x)) \leq 0. \quad (19)$$

If, in addition, problem $P(t, x)$ has a solution $\hat{x}$, then equality holds in (19), and

$$\dot{\hat{x}}(t-) = H_p(t, \hat{x}(t), V_x(t, \hat{x}(t))). \quad (20)$$

Proof. For any fixed $v$ in $\mathbb{R}^n$, define an arc $y: [t, t + h] \to \mathbb{R}^n$ by $y(s) = x + (s - t)v$. By Lemma D.1, the function

$$w(s) := V(s, y(s)) - \int_t^s L(r, y(r), \dot{y}(r)) \, dr$$

is nonincreasing on $[t, t + h]$. In particular, the right-hand derivative of $w$ at $s = t$ must be nonpositive. Calculating this derivative gives

$$0 \geq V_t(t, x) + V_x(t, x)v - L(t, x, v).$$

Since $v$ is arbitrary, inequality (19) follows from the definition of the Hamiltonian (1).

Now if $\hat{x}$ solves problem $P(t, x)$, the function $\hat{v}$ defined in (16) is constant on $[a_0, t]$. In particular, the left-hand derivative of $\hat{v}(s)$ at $s = t$ must vanish. Calculating this derivative just as above, and noting that $\hat{x}(t) = x$, gives

$$0 = V_t(t, x) + V_x(t, x)\dot{\hat{x}}(t-) - L(t, x, \dot{\hat{x}}(t-)).$$

In conjunction with the inequality just established, this proves not only that equality holds in (a), but also that the maximizing $v$ in the definition of $H(t, x, V_x(t, x))$ is $v = \hat{x}(t-)$. This gives (b). /////

Corollary D.2(b) implies that when the value function $V$ smooth on some open set $\Omega$, it satisfies the Hamilton-Jacobi equation (11) there. So one way to apply the verification technique is to compute the value function.

Interpretations. In the previous section we saw that every smooth solution of the Hamilton-Jacobi equation (11) not only demonstrates the optimality of certain arcs, but also gives an easy proof that these arcs satisfy the Euler-Lagrange equations. The derivation given there remains valid when the solution in question is the value function, and in this case it gives an attractive interpretation to each ingredient of the Hamiltonian formulation of the necessary conditions (Theorem A.1). Let $\hat{x}$ be a minimizing arc obeying (20) above, and define $p(t) = \hat{L}_v(t)$ to set up the Euler-Lagrange equations in canonical form A.1(c). Then (15) implies that $p(t) = V_x(t, \hat{x}(t))$, while the Hamilton-Jacobi equation itself says $-V_t = H$ along the optimal arc. Thus we have an interpretable result:

$$\nabla V(t, \hat{x}(t)) = (-H(t, \hat{x}(t), p(t)), p(t)). \quad (21)$$

In other words, $p(t)$ measures the marginal value of a perturbation in the state-value of the arc $\hat{x}$ at the instant $t$, while $-H(t, \hat{x}(t), p(t))$ tells the marginal value of an infinitesimal time change at the position $(t, \hat{x}(t))$. 
Computing the Value Function. Using the value function to generate sufficient conditions for optimality may seem like circular reasoning at first, since one can only be sure of the value function in cases where one is already certain about the optimality of certain arcs. This objection is legitimate, but it fails to account for the independent validity of the verification theorem C.2. This acts as the referee in a game of “guess and check” that we now explain.

Fix some initial point \((a_0, A_0)\), and select a unique extremal for \(L\) through each point \((T, X)\) in \(\Omega\). Use the notation \(x(t; T, X)\) to indicate this extremal: ‘guess’ that the minimum value \(V(T, X)\) equals the integral cost

\[
W(T, X) := \int_{a_0}^{T} L(t, x(t; T, X), x_t(t; T, X)) \, dt.
\]

Then simply check whether the function \(W\) satisfies the Hamilton-Jacobi equation (11). If it does, then Theorem C.1 will confirm that the extremal \(x(t; T, X)\) is optimal relative to its endpoints on any closed subinterval of \((a_0, T]\), provided that condition C.1(iii) holds, i.e.,

\[
x_t(t; T, X) = H_p(t, x(t; T, X), W_X(t, x(t; T, X))) \quad \forall t \in (a, T).
\]

This turns out to be an automatic consequence of the construction just outlined, as we now demonstrate. (We assume that \(x(t; T, X)\) is sufficiently smooth in all three variables.)

The definition of the function \(x(t; T, X)\) leads to the identities

(a) \(x(t; r, x(r; T, X)) = x(t; T, X)\) whenever \(a < r < T\),
(b) \(x(T; T, X) = X\) for all \((T, X)\) in \(\Omega\), so in particular, \(x_X(T; T, X) = 1\),
(c) \(x(a_0; T, X) = A_0\) for all \((T, X)\) in \(\Omega\), so in particular, \(x_X(a_0; T, X) = 0\).

The definition of \(W\) gives, for any \((T, X)\) in \(\Omega\),

\[
W_X(T, X) = \int_{a_0}^{T} \left[ L_x(t, x(t; T, X), x_t(t; T, X)) x_X(t; T, X) \\
+ L_v(t, x(t; T, X), x_t(t; T, X)) x_t X(t; T, X) \right] \, dt \\
= \left. L_v(t, x(t; T, X), x_t(t; T, X)) x_X(t; T, X) \right|_{a_0}^{T} \\
+ \int_{a_0}^{T} \left( L_x - \frac{d}{dt} L_v \right) x_X(t; T, X) \, dt \\
= L_v(T, x(T; T, X), x_t(T; T, X)).
\]

Here we have used integration by parts to produce the middle line, and then applied the Euler-Lagrange equation of the extremal family and identities (b) and (c) to get the bottom line. Now the last line shows that

\[
H_p(T, x(T; T, X), W_X(T, X)) = x_t(T; T, X).
\]

This is an identity valid for all \((T, X)\) in \(\Omega\). In particular, replacing \((T, X)\) by \((t, x(t; T, X))\) for some fixed \((T, X)\) in \(\Omega\) and \(t\) in \((a, T)\), we find

\[
x_t(t; t, x(t; T, X)) = H_p(t, x(t; t, x(t; T, X)), W_X(t, x(t; T, X))).
\]
Recalling identity (a), we see that this is precisely the desired result.

To obtain the additional conclusion that $W(T, X) = V(T, X)$, further analysis of the left endpoint condition is required. This is because the construction typically produces a function $W$ that is discontinuous at the point $(a_0, A_0)$, so shifting the conclusion about optimality of extremals back to the initial point requires special care. The next example illustrates this procedure.

D.3. Example. Consider the following variational problem:

$$\min \left\{ \int_0^T (\dot{x}(t)^2 - x(t)^2) \, dt : x(0) = 0, \ x(T) = X \right\}. \quad (22)$$

We already know that any extremal arc for this problem must be a $C^2$ solution of the Euler-Lagrange equation, $\ddot{x} + x = 0$. The initial condition $x(0) = 0$ identifies the family of extremals $x(t) = c \sin t$, $c \in \mathbb{R}$. This family “bunches up” when $t = \pi$, so this point is conjugate to 0 along any extremal obeying the given initial condition. It follows from Jacobi’s necessary condition that the problem has no weak local minimizers, and hence no solution at all, in cases where $T > \pi$. We therefore concentrate on the situations when $0 < T < \pi$, where any value of $X$ is allowed, and when $T = \pi$, where only $X = 0$ corresponds to an admissible extremal.

If $0 < T < \pi$, there is a unique admissible extremal through $(T, X)$, namely

$$x(t; T, X) = \frac{X}{\sin T} \sin t, \ 0 \leq t \leq T.$$ 

On the set $\Omega = \{(T, X) : 0 < T < \pi, X \in \mathbb{R}\}$, this depends smoothly on $(T, X)$. As suggested above, we may conjecture that the value function equals

$$W(T, X) := \int_0^T (x_t(t; T, X)^2 - x(t; T, X)^2) \, dt$$

$$= \frac{X^2}{\sin^2 T} \int_0^T (\cos^2 t - \sin^2 t) \, dt = X^2 \cot T.$$ 

Thus we hope that $W(t, x) = x^2 \cot t$ will satisfy (11). This is easy to check: for $L(x, v) = v^2 - x^2$, we have $H(x, p) = p^2/4 + x^2$ with $H_p(x, p) = p/2$, so

$$W_t(t, x) + H(x, W_x(t, x)) = -x^2 \csc^2 t + \frac{1}{4} (2x \cot t)^2 + x^2$$

$$= x^2 [\csc^2 t + \cot^2 t + 1] = 0,$$

as required. The theory suggests that each of the extremals in our construction will satisfy condition C.2(iii) automatically, and this is indeed the case: for $x(t) = (X/\sin T) \sin t$, we have

$$H_p(t, x(t), W_x(t, x(t))) = \frac{1}{2} W_x(t, x(t)) = x(t) \cot t = \frac{X}{\sin T} \sin t \cot t = \dot{x}(t),$$

as required. According to the basic Verification Theorem C.1, each of our extremals $x(t) = c \sin t$ is optimal relative to its endpoints, provided both endpoints lie in the set $\Omega$. To extend this conclusion to extremal arcs starting from $(0, 0)$, we must apply
Theorem C.2, since \((0, 0) \not\in \Omega\). Conditions (i)–(iii) are the same as for Theorem C.1: what remains is to consider the limiting behaviour of \(W\) at the origin. For any arc \(x\) with graph in \(\Omega\) and \(x(0) = 0\), we know that \(\dot{x}(t)\) exists and is bounded in some interval of the form \((0, h), h > 0\). Consequently L’Hospital’s rule gives

\[
\lim_{t \to 0^+} W(t, x(t)) = \lim_{t \to 0^+} \frac{x(t)^2 \cos t}{\sin t} = \lim_{t \to 0^+} \frac{2x(t)\dot{x}(t) \cos t - x(t)^2 \sin t}{\cos t} = 0.
\]

This shows that \(w_0 = 0\) in problem \((\hat{P})\), and that every arc \(x\) with \(x(0) = 0\) satisfies the left-hand boundary condition in Theorem C.2. It follows that for every \((T, X)\) in \(\Omega\), the arc \(x(t) = X(\sin t / \sin T)\) solves problem (22), and the minimum value is really \(V(T, X) = X^2 \cot T\).

Again, we note that the curves of constant \(W\), defined by \(X^2 \cot T = c, c \in \mathbb{R}\), can be interpreted as follows: for any two such curves \(S_0\) and \(S_1\), all sinusoidal arcs from the family discussed here that join \(S_0\) to \(S_1\) accumulate the same integral value in making the journey, and this value is the absolute minimum associated with any possible transition between the curves in question. Some curves of constant \(W\) are sketched lightly in Figure D.2, together with a few of the sinusoidal extremal arcs, shown in bold.

\[
\text{Figure D.2. Curves of Constant } W \text{ and Extremals in Example D.3.}
\]

Let us now pass to the case where \(T = \pi\). As noted above, every arc in the extremal family \(x(t) = c \sin t, c \in \mathbb{R}\), satisfies the condition \(x(\pi) = 0\), yet \((\pi, 0) \not\in \Omega\).
Theorem C.2 still applies, however, since the right endpoint condition at \( b = \pi \) is satisfied by any arc for which \( x(\pi) = 0 \). Again, this is because the definition of an arc implies that there is some \( h > 0 \) for which \( \dot{x}(t) \) exists and is bounded on \((\pi - h, \pi)\), so L'Hospital’s rule can be used to evaluate

\[
\lim_{t \to \pi^-} W(t, x(t)) = \lim_{t \to \pi^-} \frac{x(t)^2 \cos t}{\sin t} = \lim_{t \to \pi^-} \frac{2x(t)\dot{x}(t) \cos t - x(t)^2 \sin t}{\cos t} = 0.
\]

Consequently \( w_1 = 0 \) in condition C.2(iv), and the right-hand endpoint condition in problem \((\hat{P})\) holds for any arc with \( x(\pi) = 0 \). Thus Theorem C.2 implies that each and every arc \( x(t) = c \sin t, c \in \mathbb{R} \), gives a global minimum in problem (22) when \((T, X) = (\pi, 0)\).

Next we treat a related curve-to-point problem handled in a similar way.

**D.4. Example.** Consider the following free-initial-point problem:

\[
\min \left\{ \int_0^T (\dot{x}(t)^2 - x(t)^2) \, dt : x(T) = X \right\}. \tag{23}
\]

The Euler-Lagrange equation and the natural boundary condition together reveal the unique admissible extremal as

\[
x(t) = \frac{X}{\cos T} \cos t, \quad 0 \leq t \leq T.
\]

It is easy to evaluate the associated integral cost:

\[
\int_0^T (\dot{x}(t)^2 - x(t)^2) \, dt = -X^2 \tan T.
\]

This is our conjecture for the value function in the original problem, so we test \( W(t, x) = -x^2 \tan t \) in Theorem C.1. The largest reasonable domain for \( W \) containing the points of interest is \( \Omega = \{(t, x) : -\pi/2 < t < \pi/2, \, x \in \mathbb{R}\} \). For points \((t, x)\) in \( \Omega \), we compute

\[
W_t(t, x) + H(t, x, W_x(t, x)) = -x^2 \sec^2 t + x^2 + \frac{1}{4}(-2x \tan t)^2 = 0,
\]

so \( W \) does indeed satisfy the Hamilton-Jacobi Equation (11). Thus conditions (i) and (ii) of the basic verification theorem hold throughout \( \Omega \), for any arc \( \hat{x} \). Furthermore, the arguments above guarantee that each of the extremals above satisfies condition (iii), namely \( \dot{x}(t) = H_p(t, x(t), W_x(t, x(t))) \)—this can also be checked by hand. Thus Theorem C.1 implies that each of these extremals is optimal relative to its endpoints, provided these endpoints lie in the set \( \Omega \). In particular, we deduce that whenever \( 0 < T < \pi/2 \), the minimum value in the original problem is precisely equal to \(-X^2 \tan T\), and that this value is attained by the extremal \( x(t) = X \cos t/\cos T \).

The elementary analysis above shows that the minimum value tends to \(-\infty\) as the endpoint \((T, X)\) approaches any point of the right-hand boundary \( T = \pi/2 \) of the set \( \Omega \) except \((\pi/2, 0)\). The reader is invited to prove that for any arc \( x \) on the interval \([0, \pi/2]\) satisfying \( x(\pi/2) = 0 \), the boundedness of \( \dot{x} \) on an interval near \( \pi/2 \) implies, thanks to L’Hospital’s rule, that

\[
\lim_{t \to \pi/2^-} -x(t)^2 \tan(t) = 0,
\]
so that the right-hand limit \( w_1 = 0 \) in condition (iv) of Theorem C.2. It follows that the problem with right endpoint \((T, X) = (\pi/2, 0)\) has a global minimum value of 0, attained by each of the arcs \( x(t) = c \cos t \ (c \in \mathbb{R}) \).

Notice that in this problem the transversality conditions at the left-hand boundary lead to a verification function \( W \) for which one of the level curves, namely the set where \( W(t, x) = 0 \), coincides exactly with the given initial curve \( t = 0 \). A sketch of the curves of constant \( W \) and the associated extremals can be obtained from Figure D.2 above, simply by shifting the axes: taking the vertical line through \( t = \pi/2 \) in Fig. D.2 as the new \( x \)-axis to produce a figure corresponding to the current problem.

\[
\textbf{Free Endpoint Problems.} \quad \text{The discussion so far leads to a quick-and-dirty approach to free-endpoint problems of the form}
\]

\[
\min \{ \ell^{(2)}(b, x(b)) + \int_a^b L(t, x(t), \dot{x}(t)) \, dt : b > a, \ x \in PWS[a, b],
\]

\[
x(a) = A, \ (b, x(b)) \in S,
\]

\[
(t, x(t)) \in \Omega \ \forall t \in (a, b) \},
\]

for some given endpoint cost function \( \ell^{(2)} \) and target set \( S \). Knowing the optimal value function \( V(T, X) \) (and how it is attained) for just the integral part of this problem is a huge help. Indeed, in this case, it suffices to minimize the function \( \ell^{(2)} + V : \Omega \to \mathbb{R} \) over the set \( S \cap \Omega \ldots \) and this is just a calculus problem.

To see why this works, suppose the point \((\hat{b}, \hat{B})\) minimizes the function \( V(b, B) + \ell^{(2)}(b, B) \) over the set \( S \cap \Omega \). Associated with \( V(\hat{b}, \hat{B}) \) is an arc \( \hat{x} \) joining \((a, A)\) to \((\hat{b}, \hat{B})\) and minimizing the associated integral relative to these endpoints. It follows that for any arc \( x \) with graph in \( \Omega \) satisfying the problem’s constraints on some interval \([a, b] \),

\[
\ell^{(2)}(b, x(b)) + \int_a^b L(t, x, \dot{x}) \, dt \geq \ell^{(2)}(b, x(b)) + V(b, x(b)) \quad \text{(def. of } V) \]

\[
\geq \ell^{(2)}(\hat{b}, \hat{B}) + V(\hat{b}, \hat{B}) \quad \text{(optimality of } (\hat{b}, \hat{B})) \]

\[
= \ell^{(2)}(\hat{b}, \hat{x}(\hat{b})) + \int_a^b \hat{L}(t) \, dt \quad \text{(optimality of } \hat{x}).
\]

This does it.

\[ \textbf{D.5. Example.} \quad \text{Here is a problem where the terminal point is completely free:} \]

\[
\min \left\{ 4(x(b) - 1)^2 + b + \int_0^b \dot{x}(t)^2 \, dt : b > 0, \ x \in PWS[0, b], \ x(0) = 0 \right\}.
\]

We have seen (Example C.3) that the lowest integral cost accumulated in passing from \((0, 0)\) to \((t, x)\) equals \( V(t, x) = x^2/t \). Thus the lowest total cost for a given arc \( x \) on \([0, b]\) with right endpoint \( x(b) = B \) will equal

\[
f(b, B) := 4(B - 1)^2 + b + B^2/b.
\]
The methods of calculus show that \( f \) has a global minimum over the right half plane at the point \((\hat{b}, \hat{B}) = (3/4, 3/4)\). Thus the arc \( \hat{x}(t) = t \) on \([0, 3/4]\) gives the global minimum in this free-endpoint problem.

\[\]

E. Dynamic Programming

The Principle of Optimality is built on such a simple idea that it can easily be applied to problems other than choosing piecewise smooth functions. Any sort of problem involving “paths” or “trajectories” that can be joined up end-to-end to produce a new “path” will succumb to the same analysis. Consider, for example, the problem of finding the best path through a rectangular lattice of \((m+1)(n+1)\) points, arranged into \(n+1\) vertical columns with \(m+1\) rows in each. Let \( P(i, j) \) be the point in row \( i \) and column \( j \), where the row index \( i = 0, \ldots, m \) and the column index \( j = 0, \ldots, n \). An admissible path in this model consists of a polygonal line from left to right across the lattice, passing through exactly one point in each column. This can be described by a function \( x(j) \), which takes the column number \( j \) as input and returns the corresponding row number as output. See Figure E.1 below, in which (for example) \( x(0) = 3 \), \( x(1) = 2 \), \( x(2) = 3 \), etc.

![Figure E.1: Discrete Dynamic Programming Setup](image)

To assign a cost to every such path, suppose there is a function \( I(i_1, i_2; j) \) which tells the cost of a straight line from the point \( P(i_1, j - 1) \) to the point \( P(i_2, j) \), and the total cost is the sum of the costs for each segment. Then the total cost of the path \( x \) is

\[
M[x] := \sum_{j=1}^{n} I(x(j - 1), x(j); j),
\]
and our optimization problem is to minimize $M[x]$ over all possible paths $x$.

The principle of optimality in this case would involve a value function $V(i,j)$ defined at each lattice point as the smallest possible sum accumulated by paths from the leftmost column ($j = 0$) to the point $P(i,j)$. The principle of optimality would say that for every $j = 1, 2, \ldots, n$ and every $i = 0, \ldots, m$,

$$V(i,j) = \min \{ V(i',j-1) + I(i',i;j) : i' = 1, \ldots, m \}.$$  \hspace{1cm} (24)

Again, the interpretation is simple: to find the cheapest path to the point $P(i,j)$ look at the cheapest path to every point in column $j - 1$, and add the cost of the link from there to $P(i,j)$. The minimum of these sums over all intermediate points in column $j - 1$ gives the lowest travel cost; and the intermediate point itself tells where to look back to to realize the minimum.

Equation (24) leads directly to a simple algorithm for computing the best path through the lattice:

1. Initialize: Use known information to assign a value to each point in the first column. After this step, $V(i,0)$ is known for all $i = 0, \ldots, m$.
2. Propagate knowledge of $V$ to the right, one column at a time. For each fixed column $j$, use (24) to work out $V(i,j)$ for each row $i = 0, 1, \ldots, m$. During this process, record the index $i' = R(i,j)$ that actually gives the minimum.
3. Read off the answer, from right to left. In the last column ($j = n$), the smallest number $V(i,n)$ gives the cost of the cheapest path through the lattice. The node where this occurs, say $i$, is the optimal right endpoint: thus $x(n) = i$. The associated index $R(x(n),n) = i'$ tells you the best previous node, so $x(n-1) = R(x(n),n)$. In general $x(j-1) = R(x(j),j)$ for each $j = 1, 2, \ldots, n$: to make this work on the computer, you have to apply it as $j$ counts backwards from $n$ down to 1.
4. Print your output; relax.

This little algorithm for minimization in a lattice problem is interesting in its own right, but the real point of introducing it here is that it provides a way to approximate the solutions of calculus of variations problems. Suppose the problem, defined on the fixed interval $[a,b]$, is to minimize

$$\Lambda[x] := \ell(1)(x(a)) + \ell(2)(x(b)) + \int_a^b L(t, x(t), \dot{x}(t)) \, dt.$$ 

Make some reasonable guess about a box in the $(t,x)$-plane that is likely to contain the optimal arc: that is, choose $\alpha < \beta$ such that the optimal arc will have graph in

$$\Omega = \{(t,x) : a < t < b, \alpha < x < \beta \}.$$ 

Then, discretize in both time and space. Build a lattice of time values $a = t_0 < t_1 < \ldots < t_n = b$ and space values $\alpha = x_0 < x_1 < \ldots < x_m = \beta$. The picture looks exactly like that in Figure E.1, only now the lattice point $P(i,j)$ corresponds to the point $(t_j,x_i)$ in the variational problem. The cost of getting from one lattice point to the next is just the integral of $L$ along a straight line segment. The segment joining $(t_{j-1},x_{i'})$ to $(t_j,x_i)$ has equation $x(t) = x_{i'} + (x_i - x_{i'})(t - t_{j-1})/(t_j - t_{j-1})$—we
approximate the resulting integral as follows:

$$\int_{t_{j-1}}^{t_j} L(t, x, \dot{x}) \, dt \approx (t_j - t_{j-1})L\left(\frac{t_j - t_{j-1}}{2}, \frac{x_i' + x_i}{2}, \frac{x_i - x_i'}{t_j - t_{j-1}}\right) =: I(i', i; j).$$

To make the algorithm work, one would set $V(i, 0) = \ell^{(1)}(x_i)$ in Step 1; use the cost function $I$ just defined to loop through Step 2; and take care in Step 3 to minimize not just $V(i, n)$, but the total cost $V(i, n) + \ell^{(2)}(x_i)$, comprised of the path cost and the endpoint cost. This brief description should make it possible to experiment with some computer programs.

In practical computations involving fixed endpoint conditions like $x(a) = A$, the general method set out above can be used with an endpoint cost function designed to vigorously discourage all other starting points. The obvious choice is

$$\ell^{(1)}(x) := \begin{cases} 0, & \text{if } x = A, \\ +\infty, & \text{if } x \neq A. \end{cases}$$

This works perfectly in theory, but on the computer it will be necessary to replace $+\infty$ by some astronomically large real number.

**F. Other Solutions of the Hamilton-Jacobi Equation**

As Example C.3 clearly shows, there are often several different solutions of the Hamilton-Jacobi equation that will serve to verify the optimality of a given trajectory. In particular problems, it may be easier to find one of these than to calculate the value function. Here are some alternatives.

**Separation of Variables.** There are many problems where the Hamilton-Jacobi equation splits up into the form

$$F(t, W_t(t, x)) - G(x, W_x(t, x)) = 0. \tag{25}$$

Here the guess $W(t, x) = u(t) + w(x)$ for some $u$ and $w$ is effective: plugging this into (25) gives

$$F(t, \dot{u}(t)) = G(x, \nabla w(x)).$$

Since the two sides depend on different variables, they must both be constant—say $\alpha$. Then we have two simpler equations to solve: the ordinary differential equation $F(t, \dot{u}(t)) = \alpha$ for $u$, and the partial differential equation $G(x, \nabla w(x)) = \alpha$ ($x \in \mathbb{R}^n$) for $w$. Integrating these to produce solutions $u(t; \alpha)$ and $w(x; \alpha)$, and then recombining these, together with an arbitrary additive constant $\beta$, produces $W(t, x; \alpha, \beta) = u(t; \alpha) + w(x; \alpha) + \beta$, a so-called “complete solution” of the Hamilton-Jacobi equation.

**F.2. Example.** Consider again the Lagrangian $L(t, x, v) = \frac{1}{2}v^2$ of Example C.3. The associated Hamilton-Jacobi equation reads

$$W_t(t, x) + \frac{1}{2} (W_x(t, x))^2 = 0,$$

which fits the pattern of (25) with $F(t, u) = u$ and $G(x, y) = -\frac{1}{2}y^2$. Here it makes sense to choose the separation constant in the form $-\frac{1}{2}\alpha^2$ for some real $\alpha$, so that
straight line of slope \( \alpha \). In particular, if one has in mind a particular straight line, a suitable choice of \( u(t) = -\frac{1}{2} \alpha^2 \), whence \( u(t) = -\frac{1}{2} \alpha^2 t + \beta_1 \), some \( \beta_1 \in \mathbb{R} \).

Combining these solutions produces the complete solution \( W(t, x; \alpha, \beta) = 2 \alpha 1 - \frac{1}{2} \alpha^2 t + \beta \), in which both constants \( \alpha \) and \( \beta \) are arbitrary. To see which arcs have their optimality confirmed by such a function, we need only consider the slope function \( \phi(t, x) = H_p(t, x, W_x(t, x)) = \alpha \); it shows that any solution of \( \dot{x} = \alpha \), i.e., any straight line of slope \( \alpha \), will be optimal relative to its endpoints for this Lagrangian. In particular, if one has in mind a particular straight line, a suitable choice of \( \alpha \) will confirm its optimality.

It is worth noting that the family of solutions produced here does not include the value function for the problem of minimizing the integral of \( L \) over all arcs satisfying \( x(0) = 0 \), which we computed in Example C.2(iii).

F.3. Example. For the Lagrangian \( L(t, x, v) = \frac{1}{2} v^2 - \frac{1}{2} x^2 \), we have

\[
H(t, x, p) = \frac{1}{2} p^2 + \frac{1}{2} x^2, \quad H_p(t, x, p) = p.
\]

Thus the Hamilton-Jacobi equation has the form

\[
W_t(t, x) + \frac{1}{2} x^2 + \frac{1}{2} (W_x(t, x))^2 = 0.
\]

If we guess \( W(t, x) = u(t) + w(x) \) and choose \(-\frac{1}{2} \alpha^2 \) for the separation constant, we arrive at the pair of equations

\[
\dot{u}(t) = -\frac{1}{2} \alpha^2, \quad \text{so } u(t) = -\frac{1}{2} \alpha^2 t + \beta_0, \quad \text{and}
\]

\[
w'(x)^2 = \alpha^2 - x^2, \quad \text{so } w(x) = \frac{1}{2} \alpha^2 \arcsin \left( \frac{x}{\alpha} \right) + \frac{1}{2} x \sqrt{\alpha^2 - x^2} + \beta_1.
\]

Combining these solutions leads to the complete integral

\[
W(t, x; \alpha, \beta) = -\frac{1}{2} \alpha^2 t + \frac{1}{2} \alpha^2 \arcsin \left( \frac{x}{\alpha} \right) + \frac{1}{2} x \sqrt{\alpha^2 - x^2} + \beta.
\]

This is a function completely different in appearance from the value functions computed in Examples D.3 and D.4, and yet it is easy to check that it does satisfy the Hamilton-Jacobi Equation in the region where \( t \in \mathbb{R}, -\alpha < x < \alpha \). Moreover, using it in Theorem C.1 will verify the optimality of every arc \( x \) in this region satisfying \( \dot{x} = \phi(t, x) \), where

\[
\phi(t, x) = H_p(t, x, W_x(t, x)) = W_x(t, x) = \sqrt{\alpha^2 - x^2}.
\]

The arcs satisfying \( \dot{x} = \phi(t, x) \) have the form \( x(t) = \alpha \sin(t + K) \) for an arbitrary constant \( K \), so every curve of this form is optimal relative to its endpoints provided the inequality \( |x(t)| < \alpha \) is respected. The latter condition requires the values of \( t + K \) in the sine function to stay in one of the open intervals between successive integer multiples of \( \pi \), as anyone acquainted with the implications of Jacobi’s necessary conditions in this problem might have predicted.

Envelopes. In Example F.2, each solution \( W(t, x; \alpha, \beta) = \alpha x - \frac{1}{2} \alpha^2 t + \beta \) corresponds to a plane

\[
z = \alpha x - \frac{1}{2} \alpha^2 t + \beta
\]

(26)
in three-dimensional \((t, x, z)\)-space. These planes have an envelope, which can be found by eliminating the parameter \(\alpha\) between the two equations (26) and

\[
0 = \frac{\partial z}{\partial \alpha} = x - \alpha t.
\]  

(27)

The latter equation gives \(\alpha = x/t\), so (26) gives the envelope equation

\[
z = \left(\frac{x}{t}\right) x - \frac{1}{2} \left(\frac{x}{t}\right)^2 t + \beta = \frac{x^2}{2t} + \beta.
\]

It follows that the Hamilton-Jacobi equation in Example F.2 has the so-called “singular integral” \(W(t, x) = x^2/(2t) + \beta, \beta \in \mathbb{R}\), which we recognize as the value function from Example C.3(iii).

In Example F.3, each solution \(W(t, x; \alpha, \beta) = -\frac{1}{2} \alpha^2 t + \frac{1}{2} \alpha^2 \arcsin \left(\frac{x}{\alpha}\right) + \frac{1}{2} x \sqrt{\alpha^2 - x^2} + \beta\)
corresponds to the three-dimensional figure

\[
z = -\frac{1}{2} \alpha^2 t + \frac{1}{2} \alpha^2 \arcsin \left(\frac{x}{\alpha}\right) + \frac{1}{2} x \sqrt{\alpha^2 - x^2} + \beta.
\]

(28)
in \((t, x, z)\)-space. The envelope of this family of figures can be found by eliminating the parameter \(\alpha\) between the two equations (28) and

\[
0 = \frac{\partial z}{\partial \alpha} = -\alpha t + \alpha \arcsin \left(\frac{x}{\alpha}\right) - \frac{x\alpha}{2\sqrt{\alpha^2 - x^2}} + \frac{x\alpha}{2\sqrt{\alpha^2 - x^2}}
\]

\[
\quad = \alpha \left[ \arcsin \left(\frac{x}{\alpha}\right) - t \right].
\]

(29)
The latter equation gives \(\alpha = x/\sin t\), so (28) gives the envelope equation

\[
z = -\frac{1}{2} \left(\frac{x}{\sin t}\right)^2 t + \frac{1}{2} \left(\frac{x}{\sin t}\right)^2 t + \frac{1}{2} x^2 \cot t + \beta
\]

\[
\quad = \frac{1}{2} x^2 \cot t + \beta.
\]

It follows that a singular integral for the Hamilton-Jacobi equation in Example F.3 is \(W(t, x) = \frac{1}{2} x^2 \cot t + \beta, \beta \in \mathbb{R}\), which we recognize as \((\frac{1}{2} \times)\) the value function from Example D.3. (The objective functional here is \(\frac{1}{2} \times\) the one in Example D.3.)

**Translation Invariance.** The unknown function \(W\) appears in the Hamilton-Jacobi equation (11) only through its partial derivatives. Thus if \(W\) is a solution of the equation, so is \(W + \beta\) for any real constant \(\beta\). This is an obvious form of a translation invariance principle, and the solutions of (11) in the previous paragraphs illustrate it.

Consider next the case where the Hamiltonian \(H\) has no explicit dependence on \(t\), and some function \(W\) satisfies (11) on the set \(\Omega\). In this case we may take any constant \(\tau\) and check that the new function \(\tilde{W}(t, x) = W(t - \tau, x)\) formed by translation in \(t\) is also a solution of (11) on the set \((\tau, 0) + \Omega\). The value functions in Examples D.3 and D.4 (for which the Lagrangian is identical) illustrate this fact: the first is \(V(t, x) = x^2 \cot t\), valid in \(\Omega = (0, \pi) \times \mathbb{R}\), while the second, \(-x^2 \tan t\), is precisely the translate \(V(t + \pi/2, x)\), valid on the set \((-\pi/2, 0) + \Omega\).
Finally, suppose the Hamiltonian $H$ is independent of some component of the vector $x$—say the $j$-th component, with corresponding standard basis vector $\hat{e}_j$. Then for any solution $W$ of (11) in $\Omega$, the translated function $\widetilde{W}(t, x) = W(t, x - c\hat{e}_i)$ satisfies (11) in the translated set $(0, c\hat{e}_i) + \Omega$. In the one-dimensional case of Example C.3, where $H(t, x, p) = \frac{1}{2}p^2$ is independent of both $t$ and $x$, this result together with those of the previous paragraphs implies that the function

$$W(t, x) = \frac{(x - \xi)^2}{2(t - \tau)} + \beta,$$

obtained from the value function in Example C.3(iii) by three translations, obeys the Hamilton-Jacobi equation in the region where $t > \tau$ for any choice of the three constants $\xi$, $\tau$, and $\beta$.

**Convexity.** When the given integrand $L(t, x, v)$ is jointly convex in $(x, v)$ on all of $\mathbb{R}^n \times \mathbb{R}^n$, every extremal arc $\hat{x}$ is globally optimal relative to its endpoints. This can be confirmed directly through the subgradient inequality and the vector form of the Euler-Lagrange equation in Theorem A.1(b), or else indirectly, by means of the following verification function:

$$W(t, x) = p(t) [x - \hat{x}(t)] + \int_a^t \hat{L}(s) \, ds,$$

where $p(t) = \hat{L}_v(t)$. Notice that $W$ is linear in $x$.

To see that (30) provides a suitable verification function, note first that $W_x(t, x) = p(t) = \hat{L}_v(t)$, so condition (iii) of Theorem C.1 is fulfilled. To demonstrate the other two conditions, we write (DEL) in the vector form introduced in Theorem A.1(b):

$$[\dot{p}(t) \ p(t)] = \nabla_{x, v} L(t, \hat{x}(t), \dot{x}(t)).$$

We may now compute

$$W_t(t, x) + H(t, x, W_x(t, x))$$

$$= \dot{p}(t) [x - \hat{x}(t)] - \dot{p}(t) \hat{x}(t) + \hat{L}(t) + \sup_v \{p(t)v - L(t, x, v)\}$$

$$= \sup_v \left\{ p(t) \left[ v - \dot{x}(t) \right] + \dot{p}(t) [x - \hat{x}(t)] + \hat{L}(t) - L(t, x, v) \right\}$$

$$= \sup_v \left\{ \hat{L}(t) + (\dot{p}(t), p(t)) \cdot [(x, v) - (\hat{x}(t), \dot{x}(t))] - L(t, x, v) \right\}.$$

Now the subgradient inequality for $L$ (see the vector form of (DEL) above) implies that the quantity in braces on the right-hand side is never positive, so condition (i) of Theorem C.1 holds. Moreover, setting $x = \hat{x}(t)$ above and choosing $v = \dot{x}(t)$ makes the right-hand side vanish: this confirms condition C.1(ii). Consequently $W$ is indeed a suitable verification function for $\hat{x}$, and it establishes that $\hat{x}$ is a global minimum.
G. Fields of Extremals

In this section we restrict our attention to the case \( n = 1 \).

G.1. Definition. Let \( \Omega \subseteq \mathbb{R} \times \mathbb{R} \) be open and simply connected (no holes). Let \( L \in C^2(\Omega \times \mathbb{R}) \). To say that \( \Omega \) is covered by a field of extremals with slope function \( \phi \) means that \( \phi(t, x) : \Omega \to \mathbb{R} \) is a \( C^1 \) function with the property that every integral curve \( x \) for the differential equation \( \dot{x} = \phi(t, x) \) in \( \Omega \) is an extremal for \( L \). We use the notation \( F \) for the collection of these integral curves, and sometimes refer to \( F \) as “the field of extremals.”

We have already met many fields of extremals in the examples above: they came from solving the Hamilton-Jacobi Equation (11) and then defining the slope function as \( \phi(t, x) = H_p(t, x, W_x(t, x)) \). The point of this section is that the process can be reversed. One can start with a field of extremals, and use it to construct a solution of the Hamilton-Jacobi equation that will verify that each extremal in the field is optimal. To produce a field of extremals “by hand”, one looks for an open set \( \Omega \) with two properties:

(i) for each \( (\tau, \xi) \) in \( \Omega \), exactly one extremal \( x \) in \( F \) obeys \( x(\tau) = \xi \); and

(ii) defining \( \phi(\tau, \xi) = \dot{x}(\tau) \) for this unique arc \( x \) produces a function \( \phi \) of class \( C^1 \) in \( \Omega \).

G.2. Example. For \( L(t, x, v) = \sqrt{1 + v^2} \), the extremals are straight lines. There are many ways to construct fields of straight lines, including the two below.

(i) \( \Omega_1 = \mathbb{R}^2 \), \( F_1 = \{ x(t) := t + b : b \in \mathbb{R} \} \), \( \phi_1(\tau, \xi) = 1 \).

(ii) \( \Omega_2 = \{(t, x) : t > 0, x \in \mathbb{R} \} \), \( F_2 = \{ x(t) := mt : m \in \mathbb{R} \} \), \( \phi_2(\tau, \xi) = \xi/\tau \).

The extremal arc \( \tilde{x}(t) = 2t \) is embedded in \( F_2 \) but not in \( F_1 \); the extremal \( \tilde{x}(t) = 2t+1 \) in neither.

G.3. Notes. Any arc \( x \) in the field \( F \) obeys \( \dot{x}(t) = \phi(t, x(t)) \) for all \( t \). Thus

(a) \( \ddot{x}(t) = \dot{\phi}(t, x(t)) + \dot{\phi}_x(t, x(t)) \dot{x}(t) = [\dot{\phi} + \phi \dot{x}]_{t, x(t)} \); also,

(b) \( \frac{d}{dt} L_v(t, x(t), \dot{x}(t)) = L_x(t, x(t), \dot{x}(t)) \), i.e., \( L_{vt} + L_{vx} \dot{x} + L_{vv} \ddot{x} = L_x \).

Since \( F \) contains an extremal through every point in \( \Omega \), we may call upon (a) to produce the identity

\[-L_x(t, x) + L_{vt}(t, x, \phi) + \phi L_{vx}(t, x, \phi) + (\phi_t + \phi \phi_x)L_{vv}(t, x, \phi) = 0 \quad \forall (t, x) \in \Omega.\]

(Here the function \( \phi \) is to be evaluated at \( (t, x) \) wherever it appears.)

Hilbert’s Invariant Integral. Consider the vector-valued function \( f(t, x) : \Omega \to \mathbb{R}^2 \) defined by

\[ f(t, x) := [L(t, x, \phi(t, x)) - \phi(t, x)L_v(t, x, \phi(t, x))]\mathbf{i} + [L_v(t, x, \phi(t, x))]\mathbf{j} \]

\[ =: f_1(t, x)\mathbf{i} + f_2(t, x)\mathbf{j}. \] (31)

It turns out that integrals of this function along curves in \( \Omega \) depend only on the endpoints, and not on the path between them. Recall the following result from vector calculus:
G.4. Theorem. For any $C^1$ function $f: \Omega \to \mathbb{R}^2$ ($f = f_1 \mathbf{i} + f_2 \mathbf{j}$), the following assertions are equivalent:

(a) The line integral $\int_\gamma f \cdot d\mathbf{r}$ is the same for all curves $\gamma$ having the same endpoints;

(b) $\frac{\partial f_1}{\partial x} = \frac{\partial f_2}{\partial t}$ in $\Omega$;

(c) $f = \nabla W$ in $\Omega$, for some “potential” $W(t,x): \Omega \to \mathbb{R}$.

Proof. See Adams, Calculus of Several Variables, 2nd ed., Theorem 7.2.2 and page 278.

Now the function $f$ introduced in (31) satisfies the equivalent conditions of Theorem G.4. To see this, just check condition (b):

$$\frac{\partial f_2}{\partial t} - \frac{\partial f_1}{\partial x} = [L_{vt} + \phi_t L_{vv}] - [L_x + \phi_x L_v - \phi_x L_v - \phi_{xx} L_{vv} - \phi L_{vx}]$$

$$= 0,$$ by the identity in G.3 above.

Of course, any arc $x$ in $\Omega$ joining $(a,A)$ to $(b,B)$ gives rise to a curve $\gamma$ via $\gamma(t) = ti + x(t)j$, and along this curve the line integral in condition (a) of the theorem equals

$$U[x] := \int_\gamma f \cdot d\mathbf{r} = \int_a^b [f_1(t,x(t)) + f_2(t,x(t))\dot{x}(t)] \, dt$$

$$= \int_a^b [L(t,x(t),\phi(t,x(t))) - (\phi(t,x(t)) - \dot{x}(t))L_v(t,x(t),\phi(t,x(t)))] \, dt.$$ 

It follows that any two arcs in $\Omega$ with the same endpoints $(a,A)$ and $(b,B)$ give the same value to the integral $U[x]$. This important property gives the functional $U$ its name: it is called Hilbert’s Invariant Integral. What is more, if $x$ is an extremal from the family $\mathcal{F}$, then $\dot{x}(t) = \phi(t,x(t))$, so

$$U[x] = \int_a^b L(t,x(t),\dot{x}(t)) \, dt = \Lambda[x]$$

(32)

coincides with the objective value of $x$ relative to $L$.

The following result makes reference to the Weierstrass Excess Function

$$\mathcal{E}(t,x,v,w) := L(t,x,w) - L(t,x,v) - L_v(t,x,v)(w - v).$$

(33)

Recall that if the function $v \mapsto L(t,x,v)$ is convex on $\mathbb{R}$ for some fixed $(t,x)$ (as we are assuming throughout these notes), then one has $\mathcal{E}(t,x,v,w) \geq 0$ for all $v$ and all $w$ by the subgradient inequality.

G.5. Theorem (Sufficiency). Let $\Omega$ be a simply connected open subset of $\mathbb{R} \times \mathbb{R}$. Let $L \in C^2(\Omega \times \mathbb{R})$. Let $\mathcal{F}$ be a field of extremals covering $\Omega$, with slope function $\phi$, and suppose an extremal $\hat{x}$ in $\mathcal{F}$ joins the points $(a,A)$ and $(b,B)$. Then for any arc $x$ in $\Omega$ with the same endpoints as $\hat{x}$, one has

$$\Lambda[x] - \Lambda[\hat{x}] = \int_a^b \mathcal{E}(t,x(t),\phi(t,x(t)),\dot{x}(t)) \, dt.$$
In particular, $\hat{x}$ is optimal relative to its endpoints.

**Proof.** Pick any arc $x$ as described in the theorem statement, and calculate

\[
\Lambda[x] - \Lambda[\hat{x}] = (\Lambda[x] - U[x]) - (\Lambda[\hat{x}] - U[\hat{x}]) \quad \text{(since } U[x] = U[\hat{x}])
\]

\[
= \Lambda[x] - U[x] \quad \text{(by (32))}
\]

\[
= \int_a^b [L(t, x, \dot{x}) - L(t, x, \phi(t, x)) + (\phi(t, x) - \dot{x})L_v(t, x, \phi(t, x))] \, dt
\]

\[
= \int_a^b \mathcal{E}(t, x(t), \phi(t, x(t)), \dot{x}(t)) \, dt.
\]

Now our standing hypothesis that $L$ is strictly convex in its $v$-component implies that the function $\mathcal{E}$ is everywhere nonnegative-valued. Thus the equation above leads directly to the minimality of $\hat{x}$.

/

**G.6. Remarks.** (i) The convexity condition on $L$ mentioned in the proof is not required for any of the developments in this section. One could drop this requirement and instead impose directly the hypothesis that

\[
\mathcal{E}(t, x, \phi(t, x), w) \geq 0 \quad \forall (t, x) \in \Omega, \forall w \in \mathbb{R}.
\]

(ii) The “potential function” $W$ of Theorem G.4 above works as a verification function in the Hamilton-Jacobi equation. The beauty of the theory of fields, however, is that one never actually has to evaluate the function $W$. Perhaps this is to be expected: in our study of the Hamilton-Jacobi theory, the typical approach was first to use path integrals along extremal curves to generate a suitable candidate function $W$, but then to use this function in a sufficiency theorem that only required looking at its partial derivatives. Somehow the theory of fields manages to go straight from the extremal trajectories to the gradient of $W$—and the desired conclusions—without an intermediate step of integration. Of course, the proof of Theorem G.4 also provides a mechanism for constructing $W$: one simply fixes some starting point $(a, A)$ in $\Omega$ and calculates $W(T, X)$ at any point $(T, X)$ in $\Omega$ by choosing a convenient curve $\gamma$ from $(a, A)$ to $(T, X)$, and evaluating $W(T, X) = \int_\gamma f \cdot ds$.

**G.7. Example.** Consider the integrand $L(t, x, v) = v^2/t$. Provided $t > 0$, this function is strictly convex in $v$, so any extremal will be a $C^2$ solution of

\[
\frac{d}{dt} \left( \frac{2\dot{x}}{t} \right) = 0, \text{ i.e., } \frac{2\dot{x}}{t} = 4A, \text{ i.e., } x(t) = At^2 + B, \ A, B \in \mathbb{R}.
\]

The unique extremal joining $(1,1)$ to $(3,9)$, for example, is $\hat{x}(t) = t^2$. It can be embedded in the field $\mathcal{F}_1 = \{ x(t) := t^2 + B : B \in \mathbb{R} \}$, for which the slope function is $\phi_1(t, x) = 2t$. (To calculate $\phi_1$, note that $x(\tau) = \xi$ means $B = \xi - \tau^2$, so $x(t) = t^2 - \tau^2 + \xi$: thus $\phi_1(\tau, \xi) = \dot{x}(\tau) = 2\tau$.) The slope function and the family of extremals both make sense in the whole plane $\mathbb{R} \times \mathbb{R}$, but the conditions of Theorem G.5 include the requirement that $L$ be convex with respect to $v$. Thus these conditions hold (only) in the domain $\Omega_1 = \{(t, x) : t > 0, x \in \mathbb{R} \}$, and the arc $\hat{x}$ (and indeed every other arc in $\mathcal{F}_1$ with graph in $\Omega$) is globally optimal relative to its endpoints.
Another field containing $\hat{x}$ is $F_2 = \{ x(t) := At^2 + A - 1 : A \in \mathbb{R} \}$. Here the unique extremal satisfying $x(\tau) = \xi$ must have

$$\xi = A\tau^2 + A - 1, \text{ i.e., } A = \frac{\xi + 1}{\tau^2 + 1},$$

so the corresponding slope function $\phi_2$ must obey $\phi_2(\tau, \xi) = \dot{x}(\tau) = 2A\tau = \frac{2(\xi + 1)\tau}{\tau^2 + 1}$.

Again, this slope function and the family of extremals make sense in a wider context, but the convexity condition in Theorem G.5 requires that we restrict attention to the region $\Omega_2 = \Omega_1$: we get the same conclusion as before, namely, that any extremal arc in $F_2$ with graph in $\Omega_2$ is optimal relative to its endpoints.

H. Future Considerations

The next iteration of this document will expand on the following points.

**Transversality.** Consider the surface $S$ in $\mathbb{R} \times \mathbb{R}^n$ defined by the equation $\psi(t, x) = 0$, and assume that the $n + 1$-dimensional row vector $(\psi_t, \psi_x)$ is nonzero at every point of $S$. If a function $W(t, x)$ is constant on the surface $S$, then every point of $S$ provides a solution for the optimization problem, “Minimize $W(t, x)$ subject to $\psi(t, x) = 0$.” Thus for every fixed point $(t, x)$ in $S$, there is a pair of Lagrange multipliers $\lambda_0 \in \{0, 1\}$, $\lambda \in \mathbb{R}$, not both zero, such that

$$\lambda_0(W_t(t, x), W_x(t, x)) + \lambda(\psi_t(t, x), \psi_x(t, x)) = 0.$$

The choice $\lambda_0 = 0$ is impossible, because then $\lambda \neq 0$ and then our regularity assumption on $\psi$ is contradicted. Hence we may take $\lambda_0 = 1$ and deduce that

$$W_t = -\lambda \psi_t, \quad W_x = -\lambda \psi_x.$$

Multiplying the first of these equations by $\psi_x$ and the second by $\psi_t$ allows us to equate their right-hand sides and conclude that

$$W_t(t, x)\psi_x(t, x) = \psi_t(t, x)W_x(t, x) \quad \forall (t, x) \in S. \quad (34)$$

In reasonable situations it is also true that condition (34) implies that $W$ is constant on $S$.

Now consider the extremal arcs in the Verification Theorem C.2, say, under the hypothesis that conditions (i)–(ii) hold because the function $W$ actually solves the Hamilton-Jacobi Equation (11). Then we have $W_x(t, x(t)) = L_v(t, x(t), \dot{x}(t))$ from condition (iii) and line (2), while (11) gives

$$W_t(t, x) = -H(t, x(t), W_x(t, x(t)))$$

$$= -H(t, x(t), L_v(t)).$$

Thus the verification function $W$ will be constant on the surface $S$ if and only if its associated extremal arcs satisfy

$$[L(t, x(t), \dot{x}(t)) - L_v(t, x(t), \dot{x}(t))\dot{x}(t)] \psi_x(t, x(t)) = L_v(t, x(t), \dot{x}(t))\psi_t(t, x(t)) \quad (35)$$

at the instant when they cross the target surface. This is precisely the transversality condition we always use for point-to-curve problems, and a coherent account of variable-endpoint problems should consider it.
Regularity. All the theory in these notes deals with “arcs”, i.e., piecewise smooth functions. However, the proof of Thm. C.2 goes through for any functions $x$ and $\hat{x}$ for which the integral converges as an improper Riemann integral. This may be useful in some nearly-singular problems.

Canonical Transformations. These fit in here too. See Goldstein, *Classical Mechanics*. 