HAMILTON'S EQUATIONS AND WEIERSTRASS' SIDE CONDITION

Héctor J. Sussmann

November 15, 2000

CONTENTS

1. THE TWO FORMS OH HAMILTON'S SYSTEM OF EQUATIONS

We will now discuss an interesting story of "missed opportunities," by showing how Hamilton in the 1830s, Weierstrass in the 1870s, and Carathéodory[1] in the 1930s, had all the knowledge that was needed to make a dramatic discovery and get close to what we now call the "maximum principle" of optimal control, but failed to make it for a fairly trivial reason. Specifically, it was merely the fact that they chose one particular formalism rather than another that on the surface looked nearly identical, that led them astray and prevented them from making the discovery.

We will argue that

[1]This handout just deals with Hamilton and Weierstrass. The discussion of Carathéodory's attempt to rewrite the Weierstrass condition in Hamiltonian form will be given next week.

> *The evolution of the calculus of variations reached a critical fork in the road when Hamilton's equations were introduced. The way the equations were actually written was only one of two possible formulations, and may not have been the best one. If Hamilton's equations had been written in a slightly different way—using a Hamiltonian that is a function of all three sets of variables (positions, velocities and momenta) rather than a function of the positions and momenta only—then some important discoveries that were not made until the 1950s would almost certainly have been made much sooner. Weierstrass and Carathéodory were misled by the fact that only the "wrong" form of the equations was available to them.*

Before we get to Hamilton, Weierstrass, and Carathédory, it will be cobvenient to analyze the second-order necessary condition due to Legendre.

1.1. The Legendre condition. Recall that the Euler-Lagrange system is given by

$$\boxed{\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = \frac{\partial L}{\partial x}} \tag{1.1}$$

or

$$\boxed{\frac{d}{dt}\left[\frac{\partial L}{\partial u}\Big(\xi_*(t),\dot{\xi}_*(t),t\Big)\right] = \frac{\partial L}{\partial x}\Big(\xi_*(t),\dot{\xi}_*(t),t\Big)\,,} \tag{1.2}$$

or

$$\boxed{\frac{d}{dt}\left[\frac{\partial L}{\partial u^i}\Big(\xi_*(t),\dot{\xi}_*(t),t\Big)\right] = \frac{\partial L}{\partial x^i}\Big(\xi_*(t),\dot{\xi}_*(t),t\Big)\,,\qquad i=1,\ldots,n\,.} \tag{1.3}$$

This necssary condition for an optimum can be derived, as we shall see later, by looking at the "first variation" δI of the cost functional I, and requiring that $\delta I = 0$. So in fact (1.1) is a necessary condition for the *stationarity* of I.

The next natural step was to look at the second variation of I, and this was done in 1786 by *Adrien-Marie Legendre* (1752-1833), who found an additional necessary condition for a minimum. His condition, derived for the scalar case, is

$$\boxed{\frac{\partial^2 L}{\partial \dot{x}^2}\Big(\xi_*(t),\dot{\xi}_*(t),t\Big) \ge 0 \quad \Big(\text{i.e., } \frac{\partial^2 L}{\partial u^2}\Big(\xi_*(t),\dot{\xi}_*(t),t\Big) \ge 0\Big)\,.} \tag{1.4}$$

With an appropriate reinterpretation, Legendre's condition (1.4) is also necessary in the vector case: all we have to do is read (1.4) as asserting that *the Hessian matrix* $\left\{\frac{\partial^2 L}{\partial u^i \partial u^j}\Big(\xi_*(t),\dot{\xi}_*(t),t\Big)\right\}_{1\le i,j\le n}$ *has to be nonnegative definite*, that is, that

$$\boxed{\sum_{i,j=1}^n \alpha^i\alpha^j\frac{\partial^2 L}{\partial u^i \partial u^j}\Big(\xi_*(t),\dot{\xi}_*(t),t\Big) \ge 0 \text{ for all } \alpha=(\alpha^1,\ldots,\alpha^n)\in\mathbb{R}^n.} \tag{1.5}$$

1.2. Hamilton's missed opportunity. The work of *William Rowan Hamilton* (1805-1865), published in 1834 and 1835, produced a far-reaching reformulation of the Euler-Lagrange equations. But we will see that this reformulation could have been even more powerful if the new equations had been written in a "slightly different" form. To explain this, we will have to write and analyze in detail both versions, the one that Hamilton used, and the one that he could have used but did not, so that we can then compare them and show that what looks like a small difference has in fact enormous implications.

In a sense, the issue at stake will seem rather trivial, just a matter of rewriting the Euler-Lagrange system in a different formalism. However, sometimes formalisms can make a tremendous difference. To understand what happened and what could have happened but did not, let us first see if we can make sense of the two necessary conditions for a minimum that have been presented so far, namely, the Euler-Lagrange equation (1.1) and the Legendre condition (1.4), and look for a single simple property that would imply both conditions.

The Legendre condition is clearly the second-order necessary condition for a minimum of a function, namely, the function $\mathbb{R}^n \ni u \mapsto L(\xi_*(t), u, t)$. But (1.1) does not look at all like the first-order condition for a minimum of that same function. It is natural to ask whether there might be a way to relate the two conditions. Is it possible that both could be expressed as necessary conditions for a minimum of one and the same function? The answer is "yes," and understanding how this is done leads straight to optimal control theory, the maximum principle, and far-reaching generalizations of the classical theory. But before we get there let us tell the story of how Hamilton almost got there himself but missed, and Weierstrass got even closer but missed as well.

Let us look at another way of writing (1.1). Suppose a curve $t \mapsto \xi_*(t)$ is a solution of (1.1). Define a function $(x, u, p, t) \mapsto H(x, u, p, t)$ of three vector variables x, u, p in $\mathbb{R}^n$, and of $t \in \mathbb{R}$, by letting

$$\boxed{H(x, u, p, t) = \langle p, u\rangle - L(x, u, t)\,.} \tag{1.6}$$

Then define the "momentum" p by

$$\pi(t) = \frac{\partial L}{\partial u}\Big(\xi_*(t), \dot{\xi}_*(t), t\Big)\,. \tag{1.7}$$

It is then clear that

$$\frac{\partial H}{\partial p} = u\,,$$

so along our curve ξ_*:

$$\frac{d\xi_*}{dt}(t) = \frac{\partial H}{\partial p}\Big(\xi_*(t), \dot{\xi}_*(t), \pi(t), t\Big)\,. \tag{1.8}$$

Also,

$$\frac{\partial H}{\partial x} = -\frac{\partial L}{\partial x}\,.$$

Therefore (1.2), with $\pi(t)$ defined by (1.7), says that

$$\frac{d\pi}{dt}(t) = -\frac{\partial H}{\partial x}\Big(\xi_*(t), \dot{\xi}_*(t), \pi(t), t\Big)\,. \tag{1.9}$$

Finally,

$$\frac{\partial H}{\partial u} = p - \frac{\partial L}{\partial u}\,, \tag{1.10}$$

so (1.7) says:

$$\frac{\partial H}{\partial u}\Big(\xi_*(t), \dot{\xi}_*(t), \pi(t), t\Big) = 0\,. \tag{1.11}$$

So we have shown that

The system of equations (1.8), (1.9), (1.11), usually written more concisely as

$$\boxed{\frac{dx}{dt} = \frac{\partial H}{\partial p}\,, \qquad \frac{dp}{dt} = -\frac{\partial H}{\partial x}\,, \qquad \frac{\partial H}{\partial u} = 0\,,} \tag{1.12}$$

is *exactly equivalent* to (1.1), provided that H is defined as in (1.6).

More precisely,

THEOREM 1.2.1. *Assume n is a positive integer, Ω is an open subset of $\mathbb{R}^n$, a, b are real numbers such that $a < b$, L is a real-valued function on $\Omega \times \mathbb{R}^n \times [a, b]$, and $\bar{x}$, $\hat{x}$ are given points in Ω. Assume that L is a function of class C^1. Let $\xi_* : [a, b] \mapsto \Omega$ be a curve of class C^1. Then the Euler-Lagrange equation (1.2) holds for all $t \in [a, b]$ if and only if the "control Hamilton equations" (1.8), (1.9), (1.11) are satisfied for some choice of the function $t \mapsto \pi(t)$. Moreover, in that case the momentum function π is necessarily given by (1.7).* ♢

We will call the function H the **control Hamiltonian,** and refer to (1.12) as the *control Hamiltonian form* of the Euler-Lagrange equations. In our view, Formula (1.6) is the definition that Hamilton should have given for the Hamiltonian, and Equations (1.12) are *"Hamilton's equations as he should have written them."*

What Hamilton actually wrote was (in our notation, not his)

$$\frac{dx}{dt} = \frac{\partial \mathcal{H}}{\partial p}\,, \qquad \frac{dp}{dt} = -\frac{\partial \mathcal{H}}{\partial x}\,, \tag{1.13}$$

where $(x, p, t) \mapsto \mathcal{H}(x, p, t)$ is a function of x, p and t alone, defined by the formula

$$\mathcal{H}(x, p, t) = \langle p, \dot{x}\rangle - L(x, \dot{x}, t)\,, \tag{1.14}$$

which resembles (1.6) but is not at all the same. The difference is that in Hamilton's definition $\dot{x}$ is supposed to be treated *not* as an independent variable, but as a function of x, p, t, defined implicitly by the equation

$$p = \frac{\partial L}{\partial \dot{x}}(x, \dot{x}, t)\,. \tag{1.15}$$

It is easy to prove the following:

If the map $(x, \dot{x}, t) \mapsto (x, p, t)$ defined by (1.15) can be inverted, *i.e., if we can "solve (1.15) for $\dot{x}$ as a function of x, p, t," then (1.13) is equivalent to (1.12).*

More precisely,

THEOREM 1.2.2. *Assume n is a positive integer, Ω is an open subset of $\mathbb{R}^n$, a, b are real numbers such that $a < b$, L is a real-valued function on $\Omega \times \mathbb{R}^n \times [a,b]$, and $\bar{x}$, $\hat{x}$ are given points in Ω. Assume that L is a function of class C^1. Let $\xi_* : [a,b] \mapsto \Omega$ be a curve of class C^1. Let π be the function given by (1.7). Assume that for some $\delta > 0$ there exists a map*

$$U_\delta \ni (x,p,t) \mapsto v(x,p,t)\,,$$

of class C^1, defined on the set

$$U_\delta = \{(x,p,t) : a \le t \le b\,,\ \|x - \xi_*(t)\| < \delta\,,\ \|p - \pi(t)\| < \delta\}\,,$$

such that

$$p = \frac{\partial L}{\partial u}(x,u,t) \Longleftrightarrow u = v(x,p,t) \quad \text{whenever} \quad (x,p,t) \in U_\delta\,,\ u \in \mathbb{R}^n\,. \tag{1.16}$$

Define

$$\mathcal{H}(x,p,t) = \langle p, v(x,p,t)\rangle - L(x, v(x,p,t), t)\,. \tag{1.17}$$

Then the Euler-Lagrange equation (1.2) holds for all $t \in [a,b]$ if and only if the "Hamilton equations"

$$\frac{dx}{dt} = \frac{\partial \mathcal{H}}{\partial p}\,, \qquad \frac{dp}{dt} = -\frac{\partial \mathcal{H}}{\partial x}\,, \tag{1.18}$$

are satisfied along (ξ_*, π), that is, if and only if

$$\begin{aligned} \dot{\xi}_*(t) &= \frac{\partial \mathcal{H}}{\partial p}(\xi_*(t), \pi(t), t)\,,\\ \dot{\pi}(t) &= -\frac{\partial \mathcal{H}}{\partial x}(\xi_*(t), \pi(t), t)\,, \end{aligned}$$

for all $t \in [a,b]$. $\diamondsuit$

PROOF. If $(x,p,t) \in U_\delta$, then it is clear that

$$\mathcal{H}(x,p,t) = H(x, v(x,p,t), p, t)\,.$$

So the chain rule tells us that

$$\frac{\partial \mathcal{H}}{\partial x} = \frac{\partial H}{\partial x} + \frac{\partial H}{\partial u} \cdot \frac{\partial v}{\partial x}\,, \tag{1.19}$$

that is,

$$\frac{\partial \mathcal{H}}{\partial x}(x,p,t) = \frac{\partial H}{\partial x}(x, v(x,p,t), p, t) + \frac{\partial H}{\partial u}(x, v(x,p,t), p, t) \cdot \frac{\partial v}{\partial x}(x,p,t)\,. \tag{1.20}$$

Now, (1.10) implies

$$\frac{\partial H}{\partial u}(x, v(x,p,t), p, t) = p - \frac{\partial L}{\partial u}(x, v(x,p,t), p, t) = p - p = 0\,. \tag{1.21}$$

Therefore

$$\frac{\partial \mathcal{H}}{\partial x}(x,p,t) = \frac{\partial H}{\partial x}(x, v(x,p,t), p, t)\,, \tag{1.22}$$

and, similarly,

$$\frac{\partial \mathcal{H}}{\partial p}(x,p,t) = \frac{\partial H}{\partial p}(x, v(x,p,t), p, t)\,. \tag{1.23}$$

If ξ_* is a solution of Euler-Lagrange and π is defined by (1.7), then (1.7) and (1.16) imply $\dot{\xi}_*(t) = v(\xi_*(t), \pi(t), t)$, and Theorem 1.2.1 tells us that

$$\begin{aligned}\dot{\xi}_*(t) &= \frac{\partial H}{\partial p}(\xi_*(t), \dot{\xi}_*(t), \pi(t), t)\,,\\ \dot{\pi}(t) &= -\frac{\partial H}{\partial x}(\xi_*(t), \dot{\xi}_*(t), \pi(t), t)\,,\end{aligned}$$

so

$$\begin{aligned}\dot{\xi}_*(t) &= \frac{\partial H}{\partial p}(\xi_*(t), \dot{\xi}_*(t), \pi(t), t)\\ &= \frac{\partial H}{\partial p}(\xi_*(t), v(\xi_*(t), \pi(t), t), \pi(t), t)\\ &= \frac{\partial \mathcal{H}}{\partial p}(\xi_*(t), \pi(t), t)\,,\\ \dot{\pi}(t) &= -\frac{\partial H}{\partial x}(\xi_*(t), \dot{\xi}_*(t), \pi(t), t)\\ &= -\frac{\partial H}{\partial x}(\xi_*(t), v(\xi_*(t), \pi(t), t), \pi(t), t)\\ &= -\frac{\partial \mathcal{H}}{\partial x}(\xi_*(t), \pi(t), t)\,,\end{aligned}$$

showing that (1.18) is satisfied. The converse is also easily proved. □

1.3. The "control Hamiltonian" versus Hamilton's Hamiltonian. It should be clear from the above discussion that the Hamiltonian reformulation of the Euler-Lagrange equations in terms of the "control Hamiltonian" is at least as natural as the classical version (1.13)-(1.14)-(1.15), and perhaps even simpler. Moreover, the control formulation has at least one obvious advantage, namely,

> **(A1)** *the control version of the Hamilton equations is* exactly equivalent *to the Euler-Lagrange system under completely general conditions, whereas the classical version (1.13)-(1.14)-(1.15) only makes sense when the transformation (1.15) can be inverted, at least locally, to solve for $\dot{x}$ as a function of x, p, t.*

We now show that (A1) is not the only advantage of the control view over the classical one. To see this, we must take another look at Legendre's condition (1.4). Since $H(x, u, p, t)$ is equal to $-L(x, u, t)$ plus a linear function of u, (1.4) is completely equivalent to

$$\text{(1.24)}\qquad \frac{\partial^2 H}{\partial u^2}\Big(\xi_*(t), \dot{\xi}_*(t), \pi(t), t\Big) \le 0\,, \quad \text{i.e.,} \quad \frac{\partial^2 H}{\partial \dot{x}^2}\Big(\xi_*(t), \dot{\xi}_*(t), \pi(t), t\Big) \le 0\,.$$

Now let us write (1.24) alongside the third equation of (1.12):

$$\text{(1.25)}\qquad \boxed{\frac{\partial H}{\partial u} = 0 \quad \text{and} \quad \frac{\partial^2 H}{\partial u^2} \le 0}\,,$$

and stare at the result for a few seconds.

These equations unmistakably suggest something! They show that *the Euler-Lagrange and Legendre conditions have to do with critical points—with respect to the velocity variable u—of a function H. Precisely, Euler-Lagrange says that the gradient $\nabla_u H$ of this function has to vanish, and then Legendre adds to this the extra requirement that the second u-derivative of H—i.e., the Hessian matrix $\{\frac{\partial^2 H}{\partial u_i \partial u_j}\}_{1\leq i,j\leq n}$—must be nonpositive definite.*

The message should be clear: **what must be happening here is that H has a maximum as a function of u.**

We state this formally as a conjecture.

CONJECTURE M: Besides (1.12) (or the equivalent form (1.1)), an additional necessary condition for optimality of an arc ξ_* should be that the function $\mathbb{R}^n \ni u \mapsto H(\xi_*(t), u, \pi(t), t) \in \mathbb{R}$ have a maximum at $\dot{\xi}_*(t)$ for each t. That is, "the velocity that is actually used must maximize the Hamiltonian—for fixed time, position and momentum—among all possible values of the velocity vector."

We contend that

1. Conjecture M is an extremely natural consequence of rewriting Hamilton's equations "as Hamilton should have done it," in terms of we are calling the "control Hamiltonian."
2. It is reasonable to guess that, if Hamilton had actually written "Hamilton's equations" in this alternative way, then he himself, or some other 19th century mathematician, would have written (1.25) and been led by it to the conjecture.
3. On the other hand, it is only by using the Hamiltonian of (1.6), as opposed to Hamilton's own form of the Hamiltonian, that one can see that the Euler-Lagrange and Legendre condition really are first- and second-order conditions about critical points **of the same function.** This function cannot be L itself, because the first order conditions do *not* say that $\frac{\partial L}{\partial u} = 0$. Nor can it be Hamilton's Hamiltonian $\mathcal{H}$, which isn't even a function of u. *Only the use of the "control" Hamiltonian leads naturally to Conjecture M.*

It will turn out that

Conjecture M is true. Moreover, once its truth is known then vast generalizations are possible.

2. WEIERSTRASS MISSES ANOTHER OPPORTUNITY

The contributions of *Karl Weierstrass* (1815-1897) to the calculus of variations are contained in the notes of his three series of lectures on the topic, given in 1875, 1879 and 1882, and not formally published as books or journal articles.

2.1. The Weierstrass side condition. Weierstrass considered the problem of minimizing an integral $I(\xi)$ of the form

$$I(\xi) = \int_a^b L(\xi(s), \dot{\xi}(s))\, ds\,,$$

for Lagrangians L such that

(W) $L(x, \dot{x})$ is positively homogeneous with respect to the velocity $\dot{x}$ (that is, $L(x, \alpha\,\dot{x}) = \alpha L(x, \dot{x})$ for all $x, \dot{x}$ and all $\alpha \geq 0$) and does not depend on time. $\diamondsuit$

As will become clear from Remark 2.1.1, we have a good reason for using s rather than t as the "time" variable in the expression for I.

Remark 2.1.1. In a sense, one can always assume (W) "without loss of generality." Indeed, suppose we are interested in a completely arbitrary problem $\mathcal{P}(L, a, b, \bar{x}, \hat{x})$ of the standard form considered earlier, namely,

$$\left\{\begin{array}{ll} \text{given} & L,\ a,\ b,\ \bar{x},\ \hat{x}\,, \\ \text{minimize} & I = \int_a^b L(\xi(t), \dot{\xi}(t), t)dt\,, \\ \text{subject to} & \xi(a) = \bar{x}\ \text{ and }\ \xi(b) = \hat{x}\,. \end{array}\right. \tag{2.1}$$

Then we can always define a new function $\Lambda(x, t, u, \tau) = \tau L(x, u/\tau, t)$, and think of t as a new x variable, say x^0, and of τ as, $\frac{dx^0}{ds}$, were s is a new time variable, or "pseudotime," not to be confused with the true time variable t. Then, using x_{new} for (x, x^0)—so that $x_{new} = (x, t)$—we have

$$\begin{aligned} \int L(x, \dot{x}, t)dt &= \int L(x, \frac{dx}{dt}, t)dt \\ &= \int L\left(x, \left(\frac{dt}{ds}\right)^{-1}\frac{dx}{ds}, t\right)\frac{dt}{ds}\, ds \\ &= \int \Lambda\left(x_{new}, \frac{dx_{new}}{ds}\right) ds\,, \end{aligned}$$

so the minimization problem for Λ is equivalent to (2.1). Moreover, if $\alpha>0$, then

$$\Lambda(x, t, \alpha\, u, \alpha\, \tau) = \alpha\, \Lambda(x, t, u, \tau)\,,$$

i. e.,

$$\Lambda(x_{new}, \alpha\, \dot{x}_{new}) = \alpha\, \Lambda(x_{new}, \dot{x}_{new})\,,$$

where we are now using the dot to denote differentiation with respect to s rather than t. So the new Lagrangian Λ is positively homogeneous with respect to the new velocity variable $\dot{x}_{new}$ and does not depend on the new time s. $\diamondsuit$

So when Weierstrass imposed Condition (W) on his Lagrangians, he was doing so "without loss of generality." However, "without loss of generality" is a dangerous phrase, and does not at all entail "without loss of insight." We shall argue below that this restriction, in conjunction with the dominant view that Hamilton's equations had to be written in the form (1.13), may have served to conceal from Weierstrass the true meaning and the far-reaching implications of the new condition he discovered.

Weierstrass introduced what is now known as

THE WEIERSTRASS EXCESS FUNCTION,

$$\mathcal{E}(x,\bar{u},u) = L(x,u) - \frac{\partial L}{\partial u}(x,\bar{u})\cdot u\,. \tag{2.2}$$

This function depends on three sets of independent variables, namely, x, $\bar{u}$ and u. That is, *$\mathcal{E}$ is a function of position and two velocity vectors.*

He then proved his *side condition*:

> *(SC) For a curve $s \mapsto \xi_*(s)$ to be a solution of the minimization problem, the function $\mathcal{E}$ has to be ≥ 0 when evaluated for $x = \xi_*(s)$, $\bar{u} = \dot{\xi}_*(s)$, and a completely arbitrary u.*

Notice that Lagrangians with Property (W) satisfy the identity

$$L(x,u) = \frac{\partial L}{\partial u}(x,u)\cdot u\,. \tag{2.3}$$

Therefore Weierstrass could equally well have written his excess function as

$$\mathcal{E}(x,\bar{u},u) = L(x,u) - \frac{\partial L}{\partial u}(x,\bar{u})\cdot u - \Big(L(x,\bar{u}) - \frac{\partial L}{\partial u}(x,\bar{u})\cdot\bar{u}\Big)\,, \tag{2.4}$$

and he could even have generalized this formula to the time-dependent case, by writing

$$\mathcal{E}(x,\bar{u},u,t)=L(x,u,t)-\frac{\partial L}{\partial u}(x,\bar{u},t)\cdot u-\Big(L(x,\bar{u},t)-\frac{\partial L}{\partial u}(x,\bar{u},t)\cdot\bar{u}\Big). \tag{2.5}$$

Remark 2.1.2. Formulae (2.4) and (2.5) have a simple geometric interpetation. Assume, for simplicity, that L does not depend on t. For a given point x, let G_x be the graph of

$$y = L(x,u) \stackrel{\text{def}}{=} L_x(u)$$

as a function of u. Fix a value $\bar{u}$ of u, and let

$$\bar{y} = L(x,\bar{u})\,,$$

so the point

$$\bar{P} = (\bar{u},\bar{y})$$

lies on G_x. Let $\Gamma_{x,\bar{u}}$ be the tangent hyperplane to G_x at $\bar{P}$. Then the equation of $\Gamma_{x,\bar{u}}$ is

$$y = \bar{y} + \frac{\partial L}{\partial u}(x,\bar{u})\cdot(u-\bar{u})\,,$$

that is

$$y = L(x,\bar{u}) + \frac{\partial L}{\partial u}(x,\bar{u})\cdot(u-\bar{u})\,.$$

Given any u, the ordinates $y_{graph}(u)$, $y_{plane}(u)$ of the points of G_x, $\Gamma_{x,\bar{u}}$, that lie on the line $\{u\} \times \mathbb{R}$ are given by

$$y_{graph}(u) = L(x,u)\,, \quad y_{plane}(u) = L(x,\bar{u}) + \frac{\partial L}{\partial u}(x,\bar{u}) \cdot (u - \bar{u})\,.$$

Therefore

$$\mathcal{E}(x,\bar{u},u) = y_{graph}(u) - y_{plane}(u)\,.$$

In other words (cf. Figure 1):

> *The number $\mathcal{E}(x,\bar{u},u)$ tells us by how much the actual value of the function $y = L_x(u)$ at a particular u is larger than the value at u of the linearization (i.e., tangent approximation) of this function at the point $(\bar{u}, L(x,\bar{u}))$.*

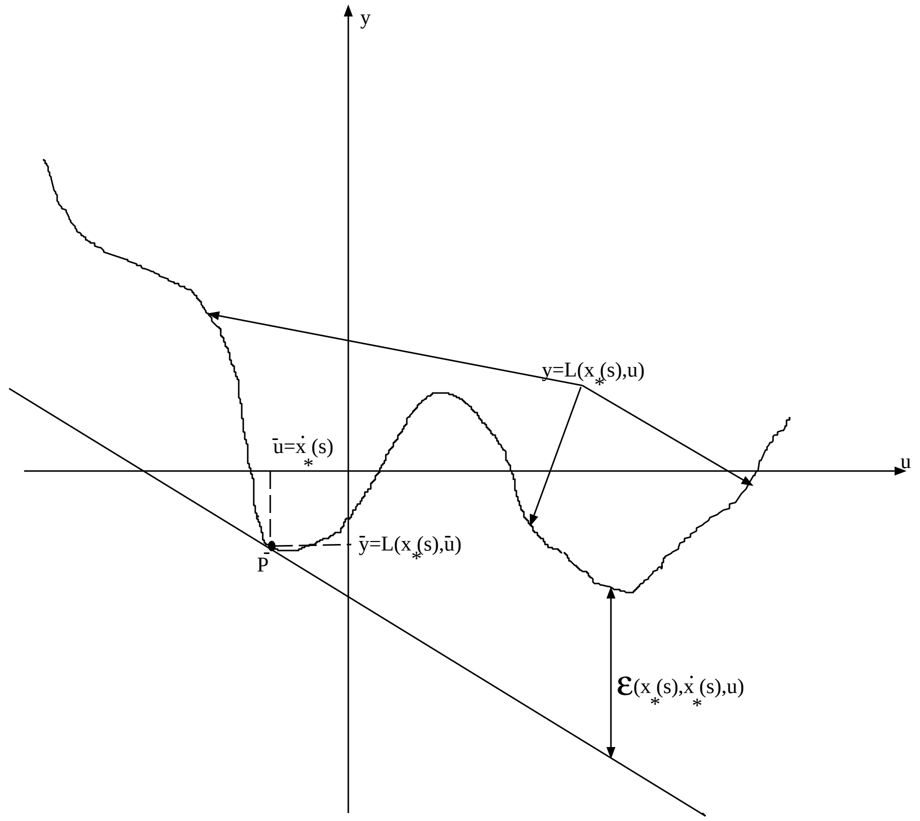

FIGURE 1. The Weierstrass side condition

Therefore the Weierstrass side condition says that

(SC) If a curve $t \mapsto \xi_*(t)$ is a solution of the minimization problem, then for every t, if we draw the tangent hyperplane to the graph of the function $u \mapsto L(\xi_*(t), u, t)$ at the point of this graph corresponding to $u = \dot{\xi}_*(t)$, then the graph must lie entirely above the hyperplane. ♢

2.2. A Hamiltonian rewriting of the Weierstrass condition. Using

$$\pi(t) = \frac{\partial L}{\partial u}(\xi_*(t), \dot{\xi}_*(t), t)\,,$$

as in (1.7), we see that

$$\begin{aligned}(2.6)\qquad & \mathcal{E}(\xi_*(t), \dot{\xi}_*(t), u, t) \\ & = \Big(L(\xi_*(t), u, t) - \langle \pi(t), u\rangle\Big) - \Big(L(\xi_*(t), \dot{\xi}_*(t), t) - \langle \pi(t), \dot{\xi}_*(t)\rangle\Big)\,,\end{aligned}$$

which the reader will immediately recognize as saying that

$$\mathcal{E}(\xi_*(t), \dot{\xi}_*(t), u, t) = H(\xi_*(t), \dot{\xi}_*(t), \pi(t), t) - H(\xi_*(t), u, \pi(t), t)\,, \tag{2.7}$$

where H is our "control Hamiltonian." So Weierstrass' condition, expressed in terms of the control Hamiltonian, simply says that

(MAX) Along an optimal curve $t \mapsto \xi_*(t)$, if we define $\pi(t)$ via (1.7), then for every t the value $u = \dot{\xi}_*(t)$ must maximize the (control) Hamiltonian $H(\xi_*(t), u, \pi(t), t)$ as a function of u.

Moreover, (MAX) can be considerably simplified. Indeed,

The requirement that $\pi(t)$ be defined via (1.7) is now redundant.

To see this, observe that, if $H(\xi_*(t), u, \pi(t), t)$, regarded as a function of u, has a maximum at $u = \dot{\xi}_*(t)$, then $\frac{\partial H}{\partial u}(\xi_*(t), \dot{\xi}_*(t), \pi(t), t)$ has to vanish, so $\pi(t)$ must be given by (1.7). Moreover, the vanishing of $\frac{\partial H}{\partial u}(\xi_*(t), \dot{\xi}_*(t), \pi(t), t)$ is also one of the conditions of (1.12). So we can just drop (1.7) altogether, and state (1.12) and (MAX) together:

(NCO) If a curve $t \mapsto \xi_*(t)$ is a solution of the minimization problem (2.1), then there has to exist a function $t \mapsto \pi(t)$ such that the following three conditions hold for all t:

$$\dot{\xi}_*(t) = \frac{\partial H}{\partial p}(\xi_*(t), \dot{\xi}_*(t), \pi(t), t)\,, \tag{2.8}$$

$$\dot{\pi}(t) = -\frac{\partial H}{\partial x}(\xi_*(t), \dot{\xi}_*(t), \pi(t), t)\,, \tag{2.9}$$

$$H(\xi_*(t), \dot{\xi}_*(t), \pi(t), t) = \max_u H(\xi_*(t), u, \pi(t), t)\,. \tag{2.10}$$

As a version of the necessary conditions for optimality, (NCO) encapsulates in one single statement the combined power of the Euler-Lagrange necessary conditions and the Weierstrass side condition as well, of course, as the Legendre condition, which obviously follows from (NCO). Notice the elegance and economy of language achieved by this unified statement: there is no need to bring in an extra entity called the "excess function." Nor does one need to include a formula specifying how $\pi(t)$ is defined, since (2.10) does this automatically. So the addition of the new Weierstrass condition to the three equations of (1.12) results in a new set of three, rather than four, conditions, a set much "simpler than the sum of its parts." Notice moreover that (2.10) is exactly Conjecture M. So we can surmise at this point that (NCO), as stated, could probably have been discovered soon after the work of Hamilton—since it is strongly suggested by (1.25)—and almost certainly by Weierstrass, if only Hamilton's equations had been written in the form (1.6), (1.12).

So we can now add two new items to our list of advantages of the "control formulation" of Hamilton's equations over the classical one:

(A2) *Using the control Hamiltonian, it would have been an obvious next step to write Legendre's condition in "Hamiltonian form," as in (1.25), and this would have led immediately to the formulation of Conjecture M, a proof of which would then have been found soon after.*

(A3) *In terms of the control Hamiltonian, the Weierstrass side condition has a much simpler statement, not requiring the introduction of an "excess function," and can be combined with the Hamilton equations into an elegant unified formulation (NCO) of the necessary conditions for optimality.*

2.3. What Weierstrass missed. This is by no means the end of our story. There is much more to the new formulation (NCO) than just elegance and simplicity. If you compare (NCO) with all the other necessary conditions that we had written earlier, a remarkable new fact becomes apparent. Quite amazingly, *the derivatives with respect to the u variable are completely gone.* All the earlier equations involved u-derivatives of L or of H, and even if we use the classical version (1.13) of Hamilton's equations—which involves no functions of u and therefore no u-derivatives—the fact remains that in order to get to (1.13) we first have to solve (1.15), which does involve a u-derivative.

Now, if our necessary conditions for optimality can be stated without any references to u-derivatives, it is reasonable to guess that the u-derivatives of L are not needed. Also, the minimization that occurs in (2.10) makes sense over any subset U of $\mathbb{R}^n$, so there is no longer any reason to insist that the range of values of u be the whole space. This leads us to

> *CONJECTURE M2: (NCO) should still be a necessary condition for optimality even for problems where the derivative $\dot{x}$ is restricted to belong to some subset U of $\mathbb{R}^n$, and $L(x, u, t)$ is not required to be differentiable or even continuous with respect to u.* ♢

Naturally, once u is restricted to a subset U of $\mathbb{R}^n$, and $L(x, u, t)$ is not required to be differentiable with respect to u, the maximization condition (2.10) no longer implies (1.7). So we see that ***(2.10) is more fundamental than (1.7)***.

Notice also that there has been a subtle change in the formulation of the necessary conditions for optimality. In the classical conditions of the calculus of variations (including Weierstrass' own version of the side condition), the covector $\pi(t)$—the "momentum"—is defined by (1.7), and the necessary conditions are equations happening along the trajectory ξ_*. These equations can be formulated somewhat more simply using $\pi(t)$, but $\pi(t)$ can also be completely eliminated, since one can always substitute for it its value given by (1.7). The new version of the necessary conditions for an optimum given by Conjecture M2 is of a different form: rather than state that certain equations hold along ξ_*, the conditions are now an *existential statement*, asserting the *existence* of a vector-valued function $t \mapsto \pi(t)$ with certain properties. The elimination of $\pi(t)$, by expressing $\pi(t)$ in terms of $\xi_*(t)$, $\dot{\xi}_*(t)$ and t, is only possible when $U = \mathbb{R}^n$ and L is differentiable with respect to u, in which case (1.7) follows from (2.10). Otherwise, $\pi(t)$ is a "new," "independent" object.

Since Weierstrass was working under assumptions that imply (1.7), he missed the chance to see that his condition was really about $\pi(t)$, not about $\frac{\partial L}{\partial u}$ evaluated along ξ_*, and that, stated this way, his condition was valid for much more general problems, for which $\pi(t)$ no longer equals $\frac{\partial L}{\partial u}(\xi_*(t), \dot{\xi}_*(t), t)$.

2.4. A toy problem. To test Conjecture M2, let us look at a toy problem where the answer is reasonably easy to guess directly, and see whether using (NCO) gives the same result.

EXAMPLE 2.4.1. Let T be a fixed positive number. Suppose that we want to find a real-valued (Lipschitz) function $t \mapsto x(t)$ on the interval $[0, T]$ that satisfies

$$x(0) = 1\,, \quad x(T) = 1\,, \text{ and } |\dot{x}(t)| \leq 1 \text{ for all } t \in [0, T]\,, \tag{2.11}$$

and minimizes the integral

$$I = \int_0^T x(t)^2 dt$$

among all the functions satisfying the constraints (2.11). This looks exactly like a calculus of variations problem of the classical type—with a Lagrangian L given by

$$L(x, \dot{x}, t) = x^2\,,$$

except that the derivative $\dot{x}$ is required to satisfy an "inequality constraint" $|\dot{x}| \leq 1$. If we apply the classical Euler-Lagrange equations formally, we get

$$\frac{\partial L}{\partial x} = 0\,,$$

since L does not depend on $\dot{x}$, so $\frac{\partial L}{\partial \dot{x}} = 0$, and then

$$\frac{\partial L}{\partial x} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = 0\,.$$

But

$$\frac{\partial L}{\partial x} = 2x\,.$$

So $x(t) \equiv 0$, contradicting the endpoint constraints $x(0) = 1$, $x(T) = 1$. So *there is no way at all to satisfy Euler-Lagrange together with the boundary conditions.*

On the other hand, we can apply (NCO) formally, making the sensible guess that in this case the maximization with respect to u should be made over the set $U = [-1, 1]$ of permissible values of u. Then (2.8) and (2.9) become the equations

$\dot{x} = u$, $\dot{p} = 2x$. This suggests, among other things, that p as a function of t should be of class C^1, since $\dot{p}$ has to be continuous. In addition, (2.10) says that $u = 1$ when $p > 0$, and $u = -1$ when $p < 0$. (When $p = 0$ all values of $u \in U$ are solutions.) So when $p > 0$ we have $\ddot{p} = 2$, and when $p < 0$ we have $\ddot{p} = -2$. This shows that $p(\cdot)$ is a convex function of t when $p(t) > 0$ and a concave function when $p(t) < 0$. Now, a continuous function on $[0, T]$ which is convex wherever it is > 0 and concave wherever it is < 0 can only be of one of the 13 types P, N, Z, PZ, PN, ZP, ZN, NP, NZ, PZP, PZN, NZP or NZN, where "Z" means "identically zero," "P means "> 0," "N" means "< 0," and, for example, a symbol such as PZN means "a positive piece followed by a zero piece and then by a negative piece." (This is easy to prove. For example, ZPZ is impossible because the P part would have to vanish at two points and be convex and > 0 in between, which clearly cannot happen. The reader should analyze all possibilities and verify that only our 13 cases can arise.) Clearly, P corresponds to $u = 1$, and N to $u = -1$. As for Z, it corresponds to $p(t) \equiv 0$, which yields $x(t) \equiv 0$ (since $\dot{p} = 2x$) and then $u(t) \equiv 0$ (since $\dot{x} = u$). So Z corresponds to $x = u = 0$. (Notice that (2.10) does not directly determine the value of u when $p = 0$, since in this case every value of u satisfies (2.10), but by suitably combining the three conditions of (NCO) we have still been able to find u for $p = 0$.) With this in mind, it is clear that the only way to satisfy the endpoint conditions $x(0) = 1$, $x(T) = 1$, is with a PN or an NP or an NZP curve. Of these three possibilities, PN is obviously not optimal, so we are left with NP and NZP, and then a direct analysis shows that the minimum is NP if $T \leq 2$ and NZP if $T > 2$. This result can of course be proved directly, and corresponds to the obvious intuition that to minimize $\int x^2$ one should try to move from the initial value $x = 1$ towards $x = 0$ as fast as possible, stay at $x = 0$ as long as as possible, and then move back towards $x = 1$, choosing the departure time from $x = 0$ so as to arrive to $x = 1$ exactly at time T.

So *(NCO) gives us the correct solution whereas the Euler-Lagrange equation is not satisfied by this solution.* ♢