Archive Rat. Mech. Anal. **125** (1994), 297–310

ON THE DERIVATION OF HAMILTON'S EQUATIONS

Patrizia Pucci & James Serrin

Dedicated to Walter Noll

§1. Introduction

In analytical dynamics the Lagrange system governing a holonomic mechanical system with N degrees of freedom takes the form

$$\big(\mathcal{L}_p(t,u,u')\big)' - \mathcal{L}_u(t,u,u') = 0.$$

Letting $\mathcal{H}$ be the Legendre transform of $\mathcal{L}$ and $v = \mathcal{L}_p(t,u,u')$ the conjugate momentum. The Lagrange system is customarily transformed into Hamiltonian form, namely

$$\begin{aligned} u' &= \ \ \mathcal{H}_v(t,u,v) \\ v' &= -\mathcal{H}_u(t,u,v). \end{aligned}$$

For elementary mechanics the Lagrangian $\mathcal{L}(t,u,p)$ takes the simple form

$$\mathcal{L}(t,u,p) = \tfrac{1}{2}|p|^2 - V(u)$$

with the corresponding Hamiltonian

$$\mathcal{H}(t,u,v) = \tfrac{1}{2}|v|^2 + V(u).$$

Almost equally easy is the mechanically natural situation when

$$\mathcal{L}(t,u,p) = \tfrac{1}{2}(A(t,u)p,p) - V(t,u)$$

and A is a non–singular real symmetric matrix, the corresponding Hamiltonian then being

$$\mathcal{H}(t,u,p) = \tfrac{1}{2}(A^{-1}(t,u)v,v) + V(t,u)$$

Beyond this case, standard derivations of the Hamilton equations presume that the Hessian matrix of the Lagrangian is non–singular, and carry out the derivation in relatively straightforward terms. At the same time, this procedure generally yields only a local result, or, from another point of view, a multiple valued Hamiltonian on the domain of the conjugate variable. Such behavior is possible, in fact, even for positive definite Hessians, provided the domain of p is non–convex.

This being the case, a first purpose of this paper is to set up conditions guaranteeing that the Hamiltonian will be a single valued function over the entire domain of the conjugate momenta – an objective which will be attained by means of *convexity conditions* on the Lagrangian.

Another goal, which is equally important and also of interest for applications, concerns the degree of smoothness which must be required of the Lagrangian in order for the Hamilton system to be valid. Thus our second intention is to obtain the Hamilton equations even when the Lagrangian is only of class C^1, so that the Hessian is not defined, and equally when the Hessian *is* defined but not everywhere non–singular (note that convexity in itself forces the Hessian matrix only to be non–negative definite). Obviously the standard derivations do not apply in either of these cases.

Specifically, we shall show that the Hamilton system is globally valid on the domain of the conjugate momenta, solely under the following natural assumptions,

(i) the function $\mathcal{L} = \mathcal{L}(t, u, p)$ is defined and of class C^1 in a domain $I \times \Omega$, where I is an open set of $\mathbb{R}^{1+N}$ and Ω is an open *convex* set in $\mathbb{R}^N$;

(ii) the function $\mathcal{L}(t, u, \cdot)$ is *strictly convex* for every fixed $(t, u) \in I$.

In view of the convexity assumptions in (i) and (ii) it is evident that $\mathcal{L}$ also satisfies the condition

$$\mathcal{L}(t, u, p) - \mathcal{L}(t, u, p_0) > \big(\mathcal{L}_p(t, u, p_0), p - p_0\big)$$

for all p, $p_0 \in \Omega$ with $p \neq p_0$; here $(\cdot, \cdot)$ denotes the inner product in $\mathbb{R}^N$. This relation, called *strict convexity in the sense of Weierstrass*, is actually all that is required for our analysis. That is, we can replace (i) and (ii) by

(i′) the function $\mathcal{L} = \mathcal{L}(t, u, p)$ is defined and of class C^1 in a domain $I \times \Omega$, where I is an open set of $\mathbb{R}^{1+N}$ and Ω an *open* set of $\mathbb{R}^N$;

(ii′) the function $\mathcal{L}(t, u, \cdot)$ is *strictly convex in the sense of Weierstrass* for every fixed $(t, u) \in I$.

Note that (i′), (ii′) are equivalent to (i), (ii) when Ω is convex; otherwise, if Ω is not convex, then (i′), (ii′) are in general significantly weaker assumptions.

A derivation of Hamilton's equations under these minimal conditions not only has an intrinsic interest, but also allows one to consider singular Lagrangians, for example of the form

$$\mathcal{L}(t, u, p) = \frac{1}{m}|p|^m - V(u), \qquad m > 1, \qquad m \neq 2,$$

(see e.g. [10]) as well as Lagrangians of the kind introduced into variational theory by Ekeland and Temam, Clarke and others.

In order to understand our purpose more clearly, it is worth commenting at some length on the derivations of Hamilton's equations which are presently available. First, if one makes the classical assumptions that the function $\mathcal{L}_p$ is itself of class C^1 *and its Hessian matrix* $\mathcal{L}_{pp} = \left(\partial^2\mathcal{L}/\partial p_i \partial p_j\right)$ *is non–singular,* then the derivation can be carried out by routine differentiation together with an application of the standard implicit function theorem.[1] It is evident that this procedure, unaided by other considerations, can produce only a local result. To show that the Hamiltonian is globally well–defined and single valued on the *complete domain* of the conjugate variable requires further argument and assumption, a fact seldom remarked in standard treatises. To underline what may occur, consider the following non–quadratic Lagrangian in $\mathbb{R}^2$

$$\mathcal{L}(p_1, p_2) = e^{p_2} \sin p_1.$$

Here it is easily checked that $|\det\, \mathcal{L}_{pp}| = e^{2p_2} > 0$, while the conjugate mapping

$$v_1 = e^{p_2} \cos p_1, \qquad v_2 = e^{p_2} \sin p_1$$

is obviously not globally one–to–one.

To deal with these questions one may simply *postulate*, as in Abraham & Marsden's treatise, that the conjugate mapping is a diffeomorphism. Here we avoid such a hypothesis and rely instead on the convexity asssumptions introduced above. In this regard it is worth noting that Arnold in his influential text [3] also introduces the hypothesis of convexity, but then leaves the whole issue at a heuristic level (see pages 65–66).

When $\mathcal{L}$ is simply of class C^1, or if $\mathcal{L}_{pp}$ is allowed to be singular, then we are aware of full proofs only when $\mathcal{L}$ can be written in separated form

$$\mathcal{L}(t, u, p) = G(p) - V(t, u), \qquad (t, u) \in I, \quad p \in \Omega,$$

with G and Ω convex, in which case a derivation can be given on the basis of work of Rockafellar or of Mawhin & Willem in the context of convex analysis.

[1] This is the almost universal procedure found in treatises on analytical dynamics which deal at all with general Lagrangians (see e.g., Amaldi & Levi–Civita, Lanczos, Gantmacher and Santilli, among others). Of course many books treat only the elementary cases noted at the beginning of the introduction, or fail to state any hypotheses whatever when they come to general Lagrangians. The latter is the situation even for the famous monograph of Whittaker ([13], see pages 263, 264 of the last edition), which in other respects is a reliable and scholarly work, containing as well historical comments on the Hamilton equations (page 264) which are well worth reading. We remark finally the common, though incorrect, view that the classical assumptions on $\mathcal{L}$ are *necessary* for the derivation of Hamilton's equations; for example, Santilli [12, pages 158 and 160] notes that when the Hessian is singular the implicit function theorem is inapplicable, but goes on to state, inaccurately, that in this case the Hamiltonian *cannot be defined with a conventional Legendre transform.* Moreover, on page 164 he adds that:

The case when $\mathcal{L}$ is only of class C^1 will not be considered. It generally implies the breakdown of the above indicated properties of the Legendre transform due to the [inapplicability of the Implicit Function Theorem]. Therefore, the minimal continuity property of the Lagrangian, which we shall assume for the validity of the direct Legendre transform, is $\mathcal{L} \in C^2$.

More specifically, Mawhin & Willem require in addition to (i) and (ii) that the function $G = G(p)$, when extended to all of $\mathbb{R}^N$ by setting $G(p) = \infty$ for $p \notin \Omega$, should satisfy the condition $G(p)/|p| \to \infty$ as $|p| \to \infty$. Rockafellar on the other hand requires that G be *essentially smooth*, namely that (i) and (ii) hold and $|G_p(p)| \to \infty$ as $p \to p_0 \in \partial\Omega$, where $\partial\Omega$ denotes the set $\overline{\Omega} \setminus \Omega$. It follows at once that Mawhin & Willem's derivation applies whenever Ω is convex and bounded, and Rockafellar's whenever $\Omega = \mathbb{R}^N$. On the other hand, it is easy to construct examples in which (i) and (ii) hold yet neither Mawhin & Willem's nor Rockafellar's condition is satisfied, and moreover in which $G \notin C^2$ or G_{pp} is not everywhere positive definite. For such cases, therefore, even when $\mathcal{L}$ is separable no direct proof of the Hamilton system seems to be available in the literature. Some examples illustrating these points are of interest.

(1) $G(p) = \frac{1}{m}|p|^m, \qquad m > 1,\ m \neq 2; \qquad \Omega = \mathbb{R}^N.$

Here, because of the point $p = 0$, the function $\mathcal{L}$ is either not of class C^2 $(m < 2)$ or $\mathcal{L}_{pp}$ is singular at $p = 0$ $(m > 2)$. On the other hand, both Mawhin & Willem's and Rockafellar's conditions apply. Note that in this case

$$\mathcal{H}(t,u,v) = \frac{m-1}{m}|v|^{m/(m-1)} + V(t,u), \qquad (t,u) \in I, \quad v \in \mathbb{R}^N.$$

(2) $G(p) = \sqrt{1 + |p|^2} - 1; \quad \Omega = \mathbb{R}^N.$

Here $\mathcal{L}$ is of class C^2 and $\mathcal{L}_{pp}$ is positive definite. Hence the classical proof and Rockafellar's proof both apply, but not Mawhin & Willem's. In this case, note that $\mathcal{H}(t,u,v) = 1 - \sqrt{1 - |v|^2} + V(t,u)$, $(t,u) \in I$, $|v| < 1$.

(3) $G(p) = |p|^2 + p_1^2/p_2; \quad \Omega = \{p \in \mathbb{R}^2 \ :\ p_2 > 0\}$, $N = 2$.

Here $\mathcal{L}$ is of class C^2 and $\mathcal{L}_{pp}$ is positive definite, while also Mawhin & Willem's condition is satisfied, but not Rockafellar's (because $|\mathcal{L}_p| \not\to \infty$ as $p \to 0$). This example is interesting also because the domain of the conjugate variable $v = G_p(p)$, namely

$$\{v \in \mathbb{R}^2 \ :\ v_1^2 + 4v_2 > 0\},$$

is *not* convex.

(4) $G(p) = (1 + |p_1|^m)^{1/m}/p_2, \quad m > 1,\ m \neq 2;$

with

$$\Omega = \{p \in \mathbb{R}^2 \ :\ |p_1|^m < 2(m-1),\ p_2 > 0\},\ N = 2.$$

Neither Mawhin & Willem's condition nor Rockafellar's is satisfied; moreover, although the function is strictly convex and of class C^1 in Ω, it fails to be twice differentiable on the ray $p_1 = 0$, $p_2 > 0$ if $m < 2$, while if $m > 2$ it is of class C^2 in Ω but $\det \mathcal{L}_{pp} = 0$ on the ray. On the other hand, conditions (i) and (ii) and a fortiori (i′) and (ii′) *are* satisfied.

Another line of inquiry in which Hamilton's equations are obtained under weak conditions on the Lagrangian is due to Clarke. In his study, the minimization of the action integral is basic, and the Hamilton equations arise only in the form of a differential inclusion. In consequence of the different goals intended in Clarke's work, we shall not pursue further relations between his conditions and ours.

As stated above, our purpose is to derive the Hamiltonian system under the conditions (i′) and (ii′) given at the beginning of the paper. Our principal results are the following Theorems 1–4 and the corollary applications of Theorems 5–7. Theorem 7 in particular is the famous energy identity of Hamilton, presented here in both its original version and in its less well known Lagrangian form, while Theorem 8 is its extension to the quasi–variational case. In stating these results we assume that conditions (i′) and (ii′) hold throughout.

For definiteness in what follows, we understand a *solution* of Lagrange's equation to be a function $u = u(t)$ which is of class C^1 from some interval $J \subset \mathbb{R}$ into $\mathbb{R}^N$ and is such that

(1) $(t, u(t), u'(t)) \in I \times \Omega$ for $t \in J$;

(2) the conjugate momentum $v(t) = \mathcal{L}_p(t, u(t), u'(t))$ is continuously differentiable in J;

(3) $v'(t) = \mathcal{L}_u(t, u(t), u'(t))$ for all $t \in J$.

These conditions are, moreover, the natural assumptions under which Hamilton's equations are well–defined.

§2. Main Results

Here we state our main conclusions, it being assumed that *conditions* (i′) *and* (ii′) *hold throughout*. Proofs will be given in Section 3.

Theorem 1. *The function $\mathcal{L}_p(t, u, \cdot) : \Omega \to \Omega_{t,u}$ for fixed $(t, u) \in I$ is a homeomorphism.*

In view of Theorem 1 we can define the inverse map

$$\mathcal{P}(t, u, v) = \big(\mathcal{L}_p(t, u, \cdot)\big)^{-1}(v)$$

from D into Ω, where

$$D = \{(t, u, v) \in I \times \mathbb{R}^N \ : \ (t, u) \in I, \ v \in \Omega_{t,u}\}.$$

Theorem 2. *The domain D is open and $\mathcal{P}$ is continuous in D.*

We may now introduce the Legendre transform $\mathcal{H}$ of $\mathcal{L}$ as follows. Let

$$H(t, u, p) = \big(\mathcal{L}_p(t, u, p), p\big) - \mathcal{L}(t, u, p). \tag{2.1}$$

Then we define $\mathcal{H} : D \to \mathbb{R}$ by

$$\mathcal{H}(t,u,v) = H(t,u,p), \tag{2.2}$$

where

$$p = \mathcal{P}(t,u,v), \qquad v = \mathcal{L}_p(t,u,p).$$

Clearly $\mathcal{H} \in C(D)$.

Theorem 3. *The function $\mathcal{H}$ is continuously differentiable with respect to v, and for $(t,u,v) \in D$ we have*

$$\mathcal{H}_v(t,u,v) = p. \tag{2.3}$$

Moreover for each fixed $(t,u) \in I$ the function $\mathcal{H}(t,u,\cdot)$ is strictly convex in the sense of Weierstrass, namely for every v, $w \in \Omega_{t,u}$ with $v \neq w$ there holds

$$\mathcal{H}(t,u,v) - \mathcal{H}(t,u,w) > \big(\mathcal{H}_v(t,u,w), v - w\big). \tag{2.4}$$

The domain $\Omega_{t,u}$ of the variable v need not be convex, even if Ω is convex, cf. Example 3. Moreover for this example it is easy to see that the Hamiltonian cannot be extended as a *strictly convex* function to all of $\mathbb{R}^N$, which is here the convex hull of $\Omega_{t,u}$ (though it *can* be so extended as a convex function).

The proofs of Theorems 1–3 are straightforward, though not entirely obvious; the arguments employed are a direct minimization procedure together with a difference quotient argument to calculate $\mathcal{H}_v$. It is worth noting also that the inverse function theorem is not required. Proofs using the Fenchel transform, as e.g. in Mawhin & Willem [8], are possible as well – they would seem to be somewhat less direct, however, and less easy to understand.

With the aid of Theorems 1 and 3, it is easy to derive Hamilton's equations for the special case when

$$\mathcal{L}(t,u,p) = G(p) - V(t,u)$$

(of course here it is assumed only that $\mathcal{L}$ satisfies conditions (i$'$) and (ii$'$) in the introduction, and in particular that G and V are only of class C^1). First we have by Theorem 1

$$v = G_p(p), \qquad p = G_p^{-1}(v) = \mathcal{P}(v),$$

and in turn one calculates without difficulty that

$$\mathcal{H}(t,u,v) = G^*(v) + V(t,u),$$

where G^* is the Legendre transform of G. Moreover, since p is constant when v is constant, we get

$$\mathcal{H}_t(t,u,v) = -\mathcal{L}_t(t,u,p)$$
$$\mathcal{H}_u(t,u,v) = -\mathcal{L}_u(t,u,p).$$

Thus, in view of Theorem 3, the function $\mathcal{H}$ is of class $C^1(D)$ and Hamilton's equations follow at once; see Theorem 6 below. A similar but slightly more delicate computation applies also to Lagrangians of the form $\mathcal{L}(t,u,p) = \Phi(t,u)G(p) - V(t,u)$, $\Phi(t,u) > 0$, yielding again that $\mathcal{H}$ is of class $C^1(D)$ and that $\mathcal{H}_t = -\mathcal{L}_t$, $\mathcal{H}_u = -\mathcal{L}_u$.

The last conclusions are also valid under the assumptions (i′) and (ii′) alone, *without assuming that* $\mathcal{L}$ *is separable*. In particular we have the following general result, which complements Theorem 3 by showing that in all cases $\mathcal{H}$ is necessarily of class $C^1(D)$.

Theorem 4. *The function* $\mathcal{H}$ *is continuously differentiable with respect to* t, u, *and*

$$\mathcal{H}_t(t,u,v) = -\mathcal{L}_t(t,u,p), \tag{2.5}$$

$$\mathcal{H}_u(t,u,v) = -\mathcal{L}_u(t,u,p). \tag{2.6}$$

Theorems 3 and 4 have a number of important corollaries.

Theorem 5 (Duality). *The Legendre transform of* $\mathcal{H}(t,u,\cdot)$ *is* $\mathcal{L}(t,u,\cdot)$ *for every fixed* $(t,u) \in I$.

The duality in Theorem 5 can be expressed elegantly in the form

$$\mathcal{L}(t,u,p) + \mathcal{H}(t,u,v) = (p,v),$$

where

$$v = \mathcal{L}_p(t,u,p), \qquad p = \mathcal{H}_v(t,u,v).$$

Similarly the result of Theorem 4 can be written dually as

$$\mathcal{L}_x(x,p) + \mathcal{H}_x(x,v) = 0,$$

where x denotes the variable $(t,u) \in I$.

In the following Theorems 6–8 the variable t is always understood to be in the domain of the solution in question.

Theorem 6. *Let $u = u(t)$ be a solution of the Lagrangian system*

$$\big(\mathcal{L}_p(t,u,u')\big)' - \mathcal{L}_u(t,u,u') = 0. \tag{2.7}$$

Then the pair

$$u = u(t), \qquad v = v(t) = \mathcal{L}_p(t,u(t),u'(t))$$

is a solution of the Hamiltonian system

$$\begin{aligned} u' &= \ \ \mathcal{H}_v(t,u,v) \\ v' &= -\mathcal{H}_u(t,u,v) \end{aligned} \tag{2.8}$$

and conversely.

Theorem 7. *Let $(u,v) = (u(t),v(t))$ be a solution of the Hamilton system* (2.8)*. Then*

$$\{\mathcal{H}(t,u(t),v(t))\}' = \mathcal{H}_t(t,u(t),v(t)). \tag{2.9}$$

Similarly if $u = u(t)$ is a solution of the Lagrange system (2.7)*, then $H(t,u(t),u'(t))$ is of class C^1 and*

$$\{H(t,u(t),u'(t))\}' = -\mathcal{L}_t(t,u(t),u'(t)). \tag{2.10}$$

In the second part of Theorem 7 neither the function $H(t,u,p)$, defined in (2.1), nor the function $u'(t)$ need be of class C^1. Even so, by the result of Theorem 7, the *composite function $H(t,u(t),u'(t))$ is* of class C^1.

Our results also imply that the quasi–variational system[2]

$$\big(\mathcal{L}_p(t,u,u')\big)' - \mathcal{L}_u(t,u,u') = Q(t,u,u') \tag{2.11}$$

can be written in the Hamiltonian form

$$\begin{aligned} u' &= \ \ \mathcal{H}_v(t,u,v) \\ v' &= -\mathcal{H}_u(t,u,v) + Q(t,u,u'); \end{aligned}$$

in the last line u' can of course be replaced by $\mathcal{P}(t,u,v)$. Hence, with the notation $(\cdot,\cdot)$ as inner product, we have the following

[2]See, e.g. Pars, equations (6.5.9) and (10.13.14), though in both these cases the intended functions Q are, unnecessarily, supposed to depend at most linearly on u'.

More recently Santilli, equation (I.18) in [12], has presented a general system equivalent to (1.11) for forces not derivable from a potential. In [10] Pucci & Serrin have initiated a stability analysis for systems governed by (2.11) when the perturbation term Q represents a non–linear damping.

Theorem 8. *Along any solution $u = u(t)$ of (2.11) there holds*

$$\{H(t, u(t), u'(t)\}' = \big(Q(t, u(t), u'(t)), u'(t)\big) - \mathcal{L}_t(t, u(t), u'(t)).$$

§3. Proofs.

For simplicity in the notation, we denote by x any pair $(t, u) \in I$.

Proof of Theorem 1. Fix $x \in I$. First we show that $\mathcal{L}_p(x, \cdot) : \Omega \to \Omega_x$ is 1–1. Otherwise there would be points $p_1 \neq p_2$ in Ω with $\mathcal{L}_p(x, p_1) = \mathcal{L}_p(x, p_2)$. Since $\mathcal{L}$ is strictly convex in the sense of Weierstrass we get

$$\mathcal{L}(x, p_2) - \mathcal{L}(x, p_1) > \big(\mathcal{L}_p(x, p_1), p_2 - p_1\big), \qquad \mathcal{L}(x, p_1) - \mathcal{L}(x, p_2) > \big(\mathcal{L}_p(x, p_2), p_1 - p_2\big).$$

Adding these yields an immediate contradiction. Hence $\mathcal{L}_p(x, \cdot)$ is 1–1.

We next show that $\big(\mathcal{L}_p(x, \cdot)\big)^{-1}$ is continuous in $\Omega_x \equiv \mathcal{L}_p(x, \Omega)$, or what comes to the same thing, that the map $\mathcal{L}_p(x, \cdot)$ is open. Suppose $v_0 = \mathcal{L}_p(x, p_0) \in \Omega_x$ and let $C = B(p_0, \rho)$ be an open ball centered at p_0 and with radius ρ, whose closure is in Ω. We shall show that there exists an open ball $B(v_0, r) \subset \mathbb{R}^N$ such that $B(v_0, r) \subset \mathcal{L}_p(x, C) \subset \Omega_x$.

By the strict convexity of $\mathcal{L}(x, \cdot)$ in C there exists a number $d > 0$ such that

$$\mathcal{L}(x, p) - \mathcal{L}(x, p_0) - (v_0, p - p_0) \geq d > 0 \qquad \text{for every } p \in \partial C.$$

Choose $r \in (0, d/\rho)$ and let $v \in B(v_0, r)$. The function

$$g(p) = \mathcal{L}(x, p) - \mathcal{L}(x, p_0) - (v, p - p_0)$$

is zero at $p = p_0$ and for $p \in \partial C$ we have

$$g(p) = \mathcal{L}(x, p) - \mathcal{L}(x, p_0) - (v_0, p - p_0) - (v - v_0, p - p_0) \geq d - r \cdot \rho > 0.$$

Hence g has a minimum at some point $p_v \in C$ and $0 = g_p(p_v) = \mathcal{L}_p(x, p_v) - v$, which proves the claim. It follows that $\mathcal{L}_p(x, \cdot)$ is open, which completes the proof.

Remark. If in place of condition (ii′) one assumes only that $\mathcal{L}$ is of class C^2 and $\mathcal{L}_{pp}$ is non–singular, or even positive definite, in $I \times \Omega$, then the result of Theorem 1 may not hold, since for non–convex domains Ω the mappings $\mathcal{L}_p(x, \cdot)$ need not be 1–1.

This comment also shows the value of the assumptions (i′) and (ii′) as minimal hypotheses in the derivation of the function $\mathcal{P}$ and its domain D.

Proof of Theorem 2. Consider the mapping $f : I \times \Omega \to D$ defined by

$$f(x,p) = \big(x, \mathcal{L}_p(x,p)\big).$$

By Theorem 1 this mapping is 1–1. Thus it is enough to prove that f is open.

Let $(x_0, v_0) \in D$. Hence $v_0 = \mathcal{L}_p(x_0, p_0)$ for some $p_0 \in \Omega$. Let $\gamma > 0$ be so small that the closure of

$$\hat{C} = \{\, (x,p) \ : \ |x - x_0| < \gamma,\ |p - p_0| < \gamma \,\}$$

is contained in $I \times \Omega$. Define

$$d = \inf\{\, \mathcal{L}(x,p) - \mathcal{L}(x,p_0) - \big(\mathcal{L}_p(x,p_0), p - p_0\big) \ : \ |x - x_0| < \gamma,\ |p - p_0| = \gamma \,\} > 0$$

and put

$$\hat{B} = \{\, (x,v) \in I \times \mathbb{R}^N \ : \ |x - x_0| < \gamma,\ |v - \mathcal{L}_p(x,p_0)| < d/\gamma \,\}.$$

This set is open, by the continuity of $\mathcal{L}_p(\cdot, p_0)$. Moreover it is contained in $f(\hat{C}) \subset D$, as is easily seen using the argument of Theorem 1. Hence the map f is open, and so in turn $\mathcal{P}$ is continuous, which completes the proof.

Proof of Theorem 3. Fix $(x,v) \in D$ and $e \in S^{N-1}$. For small, non–zero $h \in \mathbb{R}$ define

$$\sigma = \sigma(h) = \mathcal{P}(x, v + he) - \mathcal{P}(x,v);$$

clearly $\sigma \to 0$ as $h \to 0$. Since $p = \mathcal{P}(x,v)$ it follows also

$$v = \mathcal{L}_p(x,p), \qquad v + he = \mathcal{L}_p(x, p + \sigma). \tag{3.1}$$

Then from (2.1), (2.2) and (3.1) we get for sufficiently small h

$$\begin{aligned} \mathcal{H}(x, v + he) - \mathcal{H}(x,v) &= H(x, p + \sigma) - H(x,p) \\ &= (v + he, p + \sigma) - \mathcal{L}(x, p + \sigma) - (v,p) + \mathcal{L}(x,p) \\ &= (p,e)h + (v + he, \sigma) - \mathcal{L}(x, p + \sigma) + \mathcal{L}(x,p). \end{aligned} \tag{3.2}$$

The assertion (2.3) now results if we show that

$$\mathcal{I} = \mathcal{L}(x, p + \sigma) - \mathcal{L}(x,p) - (v + he, \sigma) = o(h) \qquad \text{as } h \to 0, \tag{3.3}$$

for then it follows from (3.2) that $\mathcal{H}$ is differentiable with respect to v and $\mathcal{H}_v(x,v) = p$, as required.

Now from (ii′) we have (since $\sigma \neq 0$)

$$\mathcal{L}(x, p + \sigma) - \mathcal{L}(x,p) > \big(\mathcal{L}_p(x,p), \sigma\big) = (v, \sigma),$$

so that $\mathcal{I} > -(\sigma, e)h$ and also

$$\mathcal{I} < \big(\mathcal{L}_p(x, p+\sigma), \sigma\big) - (v + he, \sigma) = 0$$

by (3.1). Thus $|\mathcal{I}| < |\sigma| \cdot |h| = o(h)$ and (3.3) is proved.

Finally, to prove (2.4), we have in the obvious notation

$$\begin{aligned}\mathcal{H}(x,v) - \mathcal{H}(x,w) &= (v,p) - \mathcal{L}(x,p) - (w,q) + \mathcal{L}(x,q)\\ &= (q, v-w) + \mathcal{L}(x,q) - \mathcal{L}(x,p) - (v, q-p)\\ &= \big(\mathcal{H}_v(x,w), v-w\big) + \mathcal{L}(x,q) - \mathcal{L}(x,p) - \big(\mathcal{L}_p(x,p), q-p\big)\end{aligned}$$

by (2.1), (2.2) and (2.3). Thus (2.4) follows from (ii′).

Proof of Theorem 4. First suppose that $\mathcal{L}_p$ is differentiable in $I \times \Omega$, and that $\mathcal{L}_{pp}$ is everywhere positive definite in $I \times \Omega$. The formulas (2.5) and (2.6) are then standard, as we have observed in the introduction. For completeness we recall the proof. By the identity

$$v = \mathcal{L}_p(x, \mathcal{P}(x,v)) \qquad \text{in } D$$

and the regularity of $\mathcal{L}_p$ it is easy to prove directly that $\mathcal{P}_x$ exists in D and

$$\mathcal{P}_x(x,v) = -\big(\mathcal{L}_{pp}(x,p)\big)^{-1}\mathcal{L}_{px}(x,p).$$

Therefore $\mathcal{H}$ is differentiable with respect to x and by the chain rule

$$\begin{aligned}\mathcal{H}_x(x,v) &= (\mathcal{P}_x(x,v), v) - \mathcal{L}_x(x, \mathcal{P}(x,v)) - \big(\mathcal{L}_p(x, \mathcal{P}(x,v)), \mathcal{P}_x(x,v)\big)\\ &= -\mathcal{L}_x(x, \mathcal{P}(x,v)),\end{aligned}$$

as required.

We now turn to the main case, when $\mathcal{L}_p$ is not smooth. Let $\hat{\mathcal{L}} = \hat{\mathcal{L}}(h,x,p)$ be defined in $E_h = (-1,1) \times I \times \Omega_h$ by

$$\hat{\mathcal{L}}(h,x,p) = \begin{cases} \displaystyle\int_{\mathbb{R}^N} \mathcal{L}(x,y)K_h(y-p)dy, & h \neq 0,\\ \mathcal{L}(x,p), & h = 0,\end{cases}$$

where

$$\Omega_h = \{\, p \in \Omega \;:\; B(p,h) \subset \Omega \,\}, \qquad K_h(s) = \frac{1}{h}K(|s|/h), \quad s \in \mathbb{R}^N,$$

and K is a symmetric mollification kernel with support in the interval $(-1,1)$.

Since $\mathcal{L} \in C^1(I \times \Omega)$ it follows that $\hat{\mathcal{L}}$, $\hat{\mathcal{L}}_x$, $\hat{\mathcal{L}}_p$ are continuous in E_h. Moreover an easy computation shows that $\hat{\mathcal{L}}(h, x, \cdot)$ is strictly convex in the sense of Weierstrass in the open set Ω_h.

Replacing x in Theorem 2 by (h, x), it is also clear that the map

$$\hat{\mathcal{P}}(h, x, v) = \big(\hat{\mathcal{L}}_p(h, x, \cdot)\big)^{-1}(v)$$

is continuous in the open set

$$\hat{D} = \{\, (h, x, v) \ : \ (h, x) \in (-1, 1) \times I \text{ and } v = \hat{\mathcal{L}}_p(h, x, p) \text{ for } p \in \Omega_h \,\}.$$

Consequently

$$\hat{\mathcal{H}}(h, x, v) = \hat{H}(h, x, p) = \big(\hat{\mathcal{L}}_p(h, x, p), p\big) - \hat{\mathcal{L}}(h, x, p),$$

$$v = \hat{\mathcal{L}}_p(h, x, p), \qquad p = \hat{\mathcal{P}}(h, x, v),$$

is also continuous in $\hat{D}$. Finally $\hat{\mathcal{L}}_p(h, \cdot, \cdot)$ is of class $C^1(I \times \Omega_h)$ for $h \neq 0$.

To complete the proof we need the following result, of interest in itself.

Lemma. *The matrix $\hat{\mathcal{L}}_{pp}(h, \cdot, \cdot)$ is positive definite in $I \times \Omega_h$ for $h \neq 0$.*

Proof of the Lemma. Let $\xi \in S^{N-1}$be fixed. It is easily seen that

$$\big(\hat{\mathcal{L}}_{pp}(h, x, p)\xi, \xi\big) = -\int_{|s|<h} \big(\mathcal{L}_p(x, p + s), \xi\big) \cdot \big(K_{h,s}(s), \xi\big)\, ds.$$

Let $(e_1, \ldots, e_{N-1}, \xi)$ be an orthonormal basis for $\mathbb{R}^N$, with corresponding coordinates $s = (z, \tau) = (z_1, \ldots, z_{N-1}, \tau)$. Then by Fubini's theorem, with $\zeta = \sum_1^{N-1} z_i e_i$,

$$\big(\hat{\mathcal{L}}_{pp}(h, x, p)\xi, \xi\big) = -\int_{|z|<h} \int_{|\tau|<r_z} \big(\mathcal{L}_p(x, p + \zeta + \tau\xi), \xi\big) \cdot \big(K_{h,s}(\zeta + \tau\xi), \xi\big)\, d\tau dz,$$

where r_z is defined in the obvious way. Next note from the strict convexity of $\mathcal{L}(x, \cdot)$ that $\big(\mathcal{L}_p(x, p{+}\zeta{+}\tau\xi), \xi\big)$ is a strictly increasing function of τ in $(-r_z, r_z)$, while $\big(K_{h,s}(\zeta{+}\tau\xi), \xi\big)$ is an odd function of τ in $(-r_z, r_z)$, positive for negative τ and negative for positive τ. Therefore the inner integral is negative and consequently

$$\big(\hat{\mathcal{L}}_{pp}(h, x, p)\xi, \xi\big) > 0,$$

as required.

It now follows, as shown above, that $\hat{\mathcal{H}}(h,\cdot,\cdot)$ is differentiable with respect to x in $\hat{D}$ when $h \neq 0$ and

$$\hat{\mathcal{H}}_x(h,x,v) = -\hat{\mathcal{L}}_x(h,x,p),$$

$$p = \hat{\mathcal{P}}(h,x,v), \qquad v = \hat{\mathcal{L}}_p(h,x,p).$$

We now complete the proof of the theorem. By the continuity of $\hat{\mathcal{H}}$, $\hat{\mathcal{L}}_x$ and $\hat{\mathcal{P}}$ in $\hat{D}$ we obtain

$$\hat{\mathcal{H}}(h,x,v) \;\to\; \mathcal{H}(x,v) \quad \text{and} \quad \hat{\mathcal{H}}_x(h,x,v) \;\to\; -\mathcal{L}_x(x,p) \quad \text{as } h \to 0,\;\; h \neq 0,$$

uniformly in compact sets of $\hat{D}$. Hence $\mathcal{H}$ is differentiable with respect to x and moreover

$$\mathcal{H}_x(x,v) = -\mathcal{L}_x(x,p), \qquad p = \mathcal{P}(x,v).$$

Remark. It would also be possible to prove Theorem 3 by the alternative method of mollification, as in Theorem 4. This would slightly shorten the development, but at the same time would obscure the direct and straightforward derivation given here. Moreover, when the Lagrangian has the separated form $\mathcal{L}(t,u,p) = \Phi(t,u)G(p) - V(t,u)$, the present approach is certainly the easiest (see the discussion after the statement of Theorem 3 in the Introduction).

Proof of Theorem 5. Let $x \in I$ be fixed and let $\mathcal{L}^*(x,p)$ denote the Legendre transform with respect to v of $\mathcal{H}(x,v)$, which exists since $\mathcal{H} \in C^1(D)$ and $\mathcal{H}(x,\cdot)$ is strictly convex in the sense of Weierstrass. Then

$$\mathcal{L}^*(x,p) = \big(\mathcal{H}_v(x,v), v\big) - \mathcal{H}(x,v),$$

where p and v are related by the formula

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

But Theorem 3, formula (2.3), yields the relation

$$\mathcal{H}_v(x,v) = \mathcal{P}(x,v) = \big(\mathcal{L}_p(x,\cdot)\big)^{-1}(v)$$

and so

$$v = \mathcal{L}_p(x,p).$$

Hence, using also (2.1) and (2.2),

$$\mathcal{L}^*(x,p) = (p,v) - H(x,p) = (p,v) - (v,p) + \mathcal{L}(x,p) = \mathcal{L}(x,p),$$

as required.

Proof of Theorem 6. A solution of (2.7) is a C^1 function u from some interval $J \subset \mathbb{R}$ into $\mathbb{R}^N$ such that $v(t) = \mathcal{L}_p(t, u(t), u'(t))$ is a C^1 function in J and $v'(t) = \mathcal{L}_u(t, u(t), u'(t))$ in J. Hence by (2.6) of Theorem 4 we have $v'(t) = -\mathcal{H}_u(t, u(t), v(t))$. Furthermore $u'(t) = \mathcal{P}(t, u(t), v(t)) = \mathcal{H}_v(t, u(t), v(t))$, by (2.3) of Theorem 3.

Proof of Theorem 7. Along a solution (u, v) of (2.8) we have by Theorems 3 and 4 that $\mathcal{H}(\cdot, u(\cdot), v(\cdot))$ is of class C^1 and so

$$\begin{aligned} \{\mathcal{H}(t, u(t), v(t))\}' &= \mathcal{H}_t(t, u(t), v(t)) + \big(\mathcal{H}_u(t, u(t), v(t)), u'(t)\big) \\ &\qquad + \big(\mathcal{H}_v(t, u(t), v(t)), v'(t)\big) \\ &= \mathcal{H}_t(t, u(t), u'(t)), \end{aligned} \tag{3.4}$$

proving (2.9). To obtain the second part of Theorem 7, observe that if u is a solution of the Lagrange system (2.7), and (u, v) is the corresponding solution of (2.8), then

$$H(t, u(t), u'(t)) = \mathcal{H}(t, u(t), v(t))$$

by (2.2). Hence $H(t, u(t), u'(t))$ is of class C^1, and (2.10) follows from (2.9) and (2.5).

Proof of Theorem 8. This is proved in the same way as Theorems 6 and 7. Note that the last line of (3.4) should be replaced by

$$\mathcal{H}_t(t, u(t), u'(t)) + \big(Q(t, u(t), u'(t)), u'(t)\big)$$

since $v' = -\mathcal{H}_u + Q$. The required result now follows as before from (2.2) and (2.5).

Concluding Remark. Under the minimal assumptions used in this paper, solutions of the initial value problem for either Lagrange's system or Hamilton's equations need not be unique. It is of course clear that uniqueness will hold when $\mathcal{L}$ is of class C^2 and the Hessian is positive definite. On the other hand, if we consider the Lagrangian of example (1) in the Introduction, this assumption fails at $p = 0$ when $m \neq 2$. At the same time, provided that $m \leq 2$ and $V \in C^2$ *the Hamiltonian*

$$\mathcal{H}(t, u, v) = \frac{m-1}{m}|v|^{m/(m-1)} + V(t, u)$$

is of class C^2. Hence for this example *the initial value problem for the Hamiltonian system – and thus also for the Lagrange system – necessarily has a unique solution.* Thus there are cases where $\mathcal{L}$ is *not* of class C^2 but for which uniqueness *does* hold. We intend to return to this situation in more detail in a later paper.

§4. Appendix on the general Young inequality.

The following result is well–known through the Fenchel transform. Here we give a direct proof based on the results of Theorems 1 and 3.

Theorem (Young's inequality). *Let $\mathcal{L} = \mathcal{L}(p)$ be of class C^1 and strictly convex in the sense of Weierstrass in an open set $\Omega \subset \mathbb{R}^N$. Also let $\mathcal{H} = \mathcal{H}(v)$ be the Legendre transform of $\mathcal{L}$, defined for $v \in \Omega^* = \mathcal{L}_p(\Omega)$. Then for all $p \in \Omega$ and $q \in \Omega^*$ we have*

$$(p, q) \leq \mathcal{L}(p) + \mathcal{H}(q) \tag{4.1}$$

with equality if and only if p and q are related by $q = \mathcal{L}_p(p)$.

Proof. Let $v = \mathcal{L}_p(p)$. Then by (2.1), (2.2) we have $\mathcal{H}(v) = (p, v) - \mathcal{L}(p)$, so (4.1) holds with equality.

If $q \neq v$, then

$$\mathcal{H}(q) + \mathcal{L}(p) = \mathcal{H}(v) + \mathcal{L}(p) + \mathcal{H}(q) - \mathcal{H}(v) > (p, v) + (\mathcal{H}_v(v), q - v),$$

since $\mathcal{H}$ is strictly convex in the sense of Weierstrass (see Theorem 3). Thus in turn, by (2.3),

$$\mathcal{H}(q) + \mathcal{L}(p) > (p, v) + (p, q - v) = (p, q),$$

completing the proof.

The choice

$$\mathcal{L}(p) = \frac{1}{m}|p|^m, \qquad \mathcal{H}(v) = \frac{m-1}{m}|v|^{m/(m-1)},$$

gives the original case envisaged by Young:

$$(p, v) \leq \frac{1}{m}|p|^m + \frac{1}{s}|v|^s, \qquad \text{where } \frac{1}{m} + \frac{1}{s} = 1$$

and equality holds if and only if $v = |p|^{m-2}p$. When $\mathcal{L}(p) = \sqrt{1 + |p|^2} - 1$ we get

$$(p, q) \leq \sqrt{1 + |p|^2} - \sqrt{1 - |q|^2}$$

for all $p \in \mathbb{R}^N$ and $q \in \mathbb{R}^N$ with $|q| < 1$. Equality holds if and only if p and q are parallel and $(1 + |p|^2)(1 - |q|^2) = 1$.

Acknowledgment. P. Pucci is a member of *Gruppo Nazionale di Analisi Funzionale e sue Applicazioni* of the *Consiglio Nazionale delle Ricerche*. This research has been partly supported by the Italian *Ministero della Università e della Ricerca Scientifica e Tecnologica.*

References

[1] R. Abraham and J. Marsden, *Foundations of Mechanics*, 2nd edition, Addison-Wesley, Redwood City, Cal., 1985.

[2] U. Amaldi and T. Levi-Civita, *Lezioni di Meccanica Razionale*, Bologna, 1922, 1927.

[3] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed., Springer-Verlag, New York, 1989.

[4] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Canadian Math. Soc. Ser. Monographs and Advanced Texts, Wiley & Sons Publ., New York, 1983.

[5] I. Ekeland and R. Temam, *Convex Analysis and Variational Principles*, North-Holland, Amsterdam, 1976.

[6] F.R. Gantmacher, *Lectures in Analytical Mechanics*, Mir Publishing Co., Moscow, 1970.

[7] C. Lanczos, *The Variational Principles of Mechanics*, University of Toronto Press, 1949.

[8] J. Mawhin and M. Willem, *Critical Point Theory and Hamiltonian Systems*, Appl. Math. Sci. Vol. 74, Springer-Verlag, 1989.

[9] L.A. Pars, *A Treatise on Analytical Dynamics*, John Wiley, New York, and Heineman, London, 1965.

[10] P. Pucci and J. Serrin, *Precise damping conditions for global asymptotic stability for nonlinear second order systems*, Acta Mathematica **170** (1993), 275–307.

[11] R.T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, 1970.

[12] R.M. Santilli, *Foundations of Theoretical Mechanics,* I, Springer-Verlag, New York, 1978.

[13] E.T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge Univ. Press, 1904, 1937.

Patrizia Pucci
Dipartimento di Matematica
Università di Perugia
Italy

James Serrin
Department of Mathematics
University of Minnesota
Minneapolis