SIAM Journal on Control and Optimization

Vol. 34, No. 4 (July), **this is a preprint from the final TEX file, prepared in September, 1995**

EQUIVALENT SUBGRADIENT VERSIONS OF HAMILTONIAN AND EULER-LAGRANGE EQUATIONS IN VARIATIONAL ANALYSIS *

R. TYRRELL ROCKAFELLAR †

Abstract. Much effort in recent years has gone into generalizing the classical Hamiltonian and Euler-Lagrange equations of the calculus of variations so as to encompass problems in optimal control and a greater variety of integrands and constraints. These generalizations, in which nonsmoothness abounds and gradients are systematically replaced by subgradients, have succeeded in furnishing necessary conditions for optimality which reduce to the classical ones in the classical setting, but important issues have remained unsettled, especially concerning the exact relationship of the subgradient versions of the Hamiltonian equations versus those of the Euler-Lagrange equations. Here it is shown that new, tighter subgradient versions of these equations are actually equivalent to each other. The theory of epi-convergence of convex functions provides the technical basis for this development.

Key words. Euler-Lagrange equations, Hamiltonian equations, variational analysis, nonsmooth analysis, subgradients, optimality.

AMS(MOS) subject classifications. 49B10, 49B34

1. Introduction.

In the classical theory of minimization problems involving an integral functional $\int_{t_0}^{t_1} L\big(t,x(t),\dot{x}(t)\big)dt$ with Lagrangian expression $L(t,x,v)$ on $[t_0,t_1]\times I\!R^n\times I\!R^n$, a key role in analyzing the optimality of an arc $x(\cdot):[t_0,t_1]\to I\!R^n$ is played by the Euler-Lagrange equation,

$$\dot{p}(t)=\nabla_x L\big(t,x(t),\dot{x}(t)\big)\quad\text{for}\quad p(t)=\nabla_v L\big(t,x(t),\dot{x}(t)\big). \tag{1.1}$$

When $L(t,x,v)$ is twice differentiable and the Hessian matrix in v is positive definite, the Legendre transform can be applied in the v argument to get a Hamiltonian $H(t,x,p)$ in terms of which the Euler-Lagrange equation can be expressed equivalently as the Hamiltonian system

$$\dot{x}(t)=\nabla_p H\big(t,x(t),p(t)\big),\qquad -\dot{p}(t)=\nabla_x H\big(t,x(t),p(t)\big). \tag{1.2}$$

The differentiability assumptions in this scheme have long posed difficulties, however.

Many problems of interest fail to meet all the criteria for utilizing the Legendre transform, and in such cases (1.2) may only be a consequence of (1.1), not equivalent to it. Then the arcs $x(\cdot)$ and $p(\cdot)$ can have "corners" where their derivatives are discontinuous. Tonelli's theory for the existence of optimal arcs demands an even broader setting: problems must be studied with $x(\cdot)$ merely assumed to be absolutely continuous, so that (1.1) and (1.2), to the degree that they are valid, have to be interpreted in an almost everywhere sense.

Questions of existence have also challenged the suitability of classical assumptions in other ways. Tonelli showed that the convexity of $L(t,x,v)$ in v is a crucial property.

---

*Received by the editors August, 1994; accepted for publication, February, 1995. This work was supported in part by the National Science Foundation under grant DMS–9200303 at the University of Washington, Seattle, and also by the U.S.-Israel Science Foundation under grant 90-00455..

†Department of Mathematics, University of Washington 354350, Seattle, WA 98115–4350.

If this is lacking, a convexification process can be introduced to achieve it as a justifiable sort of regularization (or relaxation) of a given problem, but convexification can disrupt differentiability. Thus, Lagrangians L need to be admitted for which certain derivatives may be absent. The theory of optimal control has pushed this direction of generalization much further through the recognition that a vast range of applications can be covered "neoclassically" in terms of Lagrangians that aren't even continuous everywhere and can take on the value ∞, as a device for representing constraints on $x(t)$ and $\dot{x}(t)$ through infinite penalization when they are violated.

As far as possible in the face of this far-reaching extension of the classical framework, one would nonetheless like to make sense of the Euler-Lagrange and Hamiltonian equations as necessary conditions for optimality. The Hamiltonian can always be defined by appealing to the Legendre-Fenchel transform of convex analysis [1] instead of the Legendre transform:

$$H(t,x,p) \;=\; \sup_{v\in\mathbb{R}^n}\big\{\langle p,v\rangle - L(t,x,v)\big\}, \tag{1.3}$$

where $\langle p,v\rangle$ denotes the inner product of two vectors p and v in $\mathbb{R}^n$. Provided that $L(t,x,v)$ as a function of v is convex and lower semicontinuous, one has

$$L(t,x,v) \;=\; \sup_{p\in\mathbb{R}^n}\big\{\langle p,v\rangle - H(t,x,p)\big\}, \tag{1.4}$$

so that a one-to-one correspondence is set up between Lagrangians and Hamiltonians without calling for their differentiability. In the possible absence of gradients of L and H, the idea is to try to rewrite (1.1) and (1.2) in terms of some kind of "subgradients."

This program was first carried out in the fully convex case, where $L(t,x,v)$ is convex as a function of (x,v) (rather than just v), which corresponds to $H(t,x,p)$ being concave in x and convex in p. Subgradients of convex analysis were used by Rockafellar [2], [3], [4], to establish an Euler-Lagrange condition

$$\big(\dot{p}(t),p(t)\big) \;\in\; \partial L\big(t,x(t),\dot{x}(t)\big) \quad \text{a.e. } t \tag{1.5}$$

and a Hamiltonian condition

$$\big(-\dot{p}(t),\dot{x}(t)\big) \;\in\; \partial H\big(t,x(t),p(t)\big) \quad \text{a.e. } t, \tag{1.6}$$

where in (1.5) the subgradients are those of $L(t,\cdot,\cdot)$ as a convex function, while in (1.6) they are those of $H(t,\cdot,\cdot)$ in the special sense employed for concave-convex functions. The equivalence of these Euler-Lagrange and Hamiltonian conditions was shown through the dualization rules for subgradient relations in convex analysis.

In a major advance, Clarke [5], [6], developed a robust concept of subgradient which could serve for nonconvex functions and be used in pushing the Euler-Lagrange and Hamiltonian conditions further. This concept has evolved considerably since its introduction, both in the pattern of definition and the role of the convex hull operation. The subgradients in question can now be described in several ways, but for purposes here it is easiest to start with proximal subgradients and then take limits.

Consider a function $f:\mathbb{R}^d\to\overline{\mathbb{R}}$ (where $\overline{\mathbb{R}}$ denotes the extended reals). A vector z is a *proximal subgradient* of f at $\bar{y}$ if $f(\bar{y})$ is finite and for some $\rho\geq 0$ and $\delta>0$ one has

$$f(y) \;\geq\; f(\bar{y}) + \langle z,\, y-\bar{y}\rangle - \tfrac{1}{2}\rho|y-\bar{y}|^2 \quad \text{when} \quad |y-\bar{y}|\leq\delta.$$

It is a *subgradient in the general sense*, expressed by $z \in \partial f(\bar{y})$, if there are sequences $y^\nu \to \bar{y}$ and $z^\nu \to z$ such that z^ν is a proximal subgradient of f at y^ν and $f(y^\nu) \to f(\bar{y})$. It is a *subgradient in the horizon sense*, expressed by $z \in \partial^\infty f(\bar{y})$, if this condition holds with the modification that, instead of $z^\nu \to z$, one has $\lambda^\nu z^\nu \to z$ for some sequence of scalars $\lambda^\nu \searrow 0$. (In these expressions and below, we use superscript ν as the generic index for sequences.)

When f is continuously differentiable, $\partial f(\bar{y})$ consists of just the gradient $\nabla f(\bar{y})$, while $\partial^\infty f(\bar{y})$ has just the zero vector. When f is convex, $\partial f(\bar{y})$ is a closed, convex set, the same as the subgradient set of convex analysis, which if nonempty has $\partial^\infty f(\bar{y})$ as its recession cone. In general, however, $\partial f(\bar{y})$ and $\partial^\infty f(\bar{y})$ aren't convex, although they are always closed. Seeking a subgradient set that always would be both closed and convex, Clarke, although his notation was different and his definition followed an alternate route, ended up with the set

$$\bar{\partial} f(\bar{y}) \;=\; \mathit{cl}\ \mathit{con}\left[\partial f(\bar{y}) + \partial^\infty f(\bar{y})\right],$$

where "*cl*" stands for closure and "*con*" for convex hull. He especially emphasized the case where f is Lipschitz continuous around $\bar{y}$; then $\partial f(\bar{y})$ is a nonempty compact set, whereas $\partial^\infty f(\bar{y}) = \{0\}$, so the formula simplifies to $\bar{\partial} f(\bar{y}) = \mathit{con}\, \partial f(\bar{y})$. See Loewen [7] for a recent exposition furnishing the details.

Nowadays the convexification in this definition is no longer seen as essential for most applications, thanks to improvements in subgradient calculus achieved by Mordukhovich, Ioffe and others. In the treatment of the class of problems under discussion here, which was Clarke's chief concern, it has a natural genesis in taking weak limits, however, and the question of the extent to which it is needed has been harder to answer.

With full recourse to such convexification, Clarke was able to demonstrate in some situations where $L(t,\cdot,\cdot)$ is locally Lipschitz continuous [8] the necessity of the Euler-Lagrange condition in the form

$$\big(\dot{p}(t), p(t)\big) \in \bar{\partial} L\big(t, x(t), \dot{x}(t)\big) \quad \text{a.e. } t, \tag{1.7}$$

where the subgradient set $\bar{\partial} L\big(t, x(t), \dot{x}(t)\big)$ refers to the function $L(t,\cdot,\cdot)$ at $\big(x(t), \dot{x}(t)\big)$ and because of the Lipschitz continuity equals $\mathit{con}\, \partial L\big(t, x(t), \dot{x}(t)\big)$. On the other hand, he established in some other situations [9] where $H(t,\cdot,\cdot)$ is locally Lipschitz continuous the necessity of the Hamiltonian condition in the form

$$\big(-\dot{p}(t), \dot{x}(t)\big) \in \bar{\partial} H\big(t, x(t), p(t)\big) \quad \text{a.e. } t, \tag{1.8}$$

where the subgradient set $\bar{\partial} H\big(t, x(t), \dot{x}(t)\big)$ is that of $H(t,\cdot,\cdot)$ at $\big(x(t), p(t)\big)$ and, again because of the Lipschitz continuity, is the same as $\mathit{con}\, \partial H\big(t, x(t), p(t)\big)$. (See the books [6], [10], for an overview of this development.)

Although Clarke's conditions (1.7) and (1.8) reduce to (1.1) and (1.2) in the classical case and to (1.5) and (1.6) in the convex case, and then are equivalent, neither necessarily implies the other in general, even when both $L(t,\cdot,\cdot)$ and $H(t,\cdot,\cdot)$ are locally Lipschitz continuous. Their precise relationship has therefore been a mystery.

Loewen and Rockafellar [11] showed, in building on Clarke's results, that for a major class of problems the Euler-Lagrange condition (1.7) and Hamiltonian condition (1.8) do at least have to hold simultaneously for some arc $p(\cdot)$ when $x(\cdot)$ is optimal.

Rockafellar proved in [12, Theorem 5.1] that when $H(t,\cdot,\cdot)$ is locally Lipschitz continuous the Hamiltonian condition implies

$$\big(\dot{p}(t),\dot{x}(t)\big)\in con\,\big\{(-w,v)\,\big|\,\big(w,p(t)\big)\in\partial L(t,x(t),v),\ p(t)\in\partial_v L(t,x(t),v)\big\},\quad (1.9)$$

which is a form of the Euler-Lagrange condition suggested by Mordukhovich [13], [14], [15]. For Hamiltonians arising from bounded differential inclusions, Ioffe [16] has established that this implication is an equivalence. Also identified in Rockafellar [12, Thm. 3.4] is a broadly applicable case, beyond the known classical and convex ones, where (1.7) and (1.8) are equivalent even with $\bar{\partial}L$ replaced by ∂L.

More recent work of Loewen and Rockafellar in [17] has raised the possibility of establishing the Euler-Lagrange condition in the form

$$\dot{p}(t)\in con\,\big\{w\,\big|\,\big(w,p(t)\big)\in\partial L\big(t,x(t),\dot{x}(t)\big)\big\}\quad\text{a.e. } t \tag{1.10}$$

with the companion property that

$$p(t)\in\partial_v L\big(t,x(t),\dot{x}(t)\big)\quad\text{a.e. } t. \tag{1.11}$$

They were able to do this in a case where L is the indicator of a possibly unbounded differential inclusion, which should allow extension to other Lagrangians L through consideration of epigraphical mappings. This case also covers, for instance, the one where L is the indicator of a Lipschitz continuous differential inclusion of the kind underlying Clarke's Hamiltonian results. Such an Euler-Lagrange condition has also been obtained now by Mordukhovich [18] for a class of nonconvex differential inclusions, and by Ioffe and Rockafellar [19] for certain finite functions L. In the special case where $L(t,x,v)$ is essentially strictly convex in v, which corresponds in the theory of the Legendre-Fenchel transform to $H(t,x,p)$ being smooth in p, a case used as a technical stepping stone in [17], (1.9) comes out as saying the same thing as (1.10) and (1.11). In general, though, the combination of (1.10) with (1.11) is distinctly sharper than the versions of Euler-Lagrange in (1.7) and (1.9), because the process of convexification is much more limited.

Here we sidestep the exploration of the full range of situations in which the Euler-Lagrange condition in form (1.10) might be necessary for the optimality of an arc $x(\cdot)$. Instead we focus on the relationship between (1.10) and a corresponding version of the Hamiltonian condition, namely

$$\dot{p}(t)\in con\,\big\{w\,\big|\,\big(-w,\dot{x}(t)\big)\in\partial H\big(t,x(t),p(t)\big)\big\}\quad\text{a.e. } t \tag{1.12}$$

along with

$$\dot{x}(t)\in\partial_p H\big(t,x(t),p(t)\big)\quad\text{a.e. } t. \tag{1.13}$$

This is sharper than the Hamiltonian condition (1.8) and has not previously been considered. We'll show it's in fact equivalent to (1.10) in the kinds of circumstances that are typically present in derivations of necessary conditions for the optimality of an arc $x(\cdot)$. Efforts aimed enlarging the range of cases in which the Euler-Lagrange condition holds in version (1.10) can thus count on the side benefit of improving Clarke's Hamiltonian condition in a hitherto unsuspected way.

The following theorem is our main result. Through [17] it brings to light, among other things, that (1.8) can be strengthened to (1.12) in Clarke's context [9].

In stating this theorem, we say that the Lagrangian L has the *epi-continuity property along* $x(\cdot)$ if, for almost every t, there is an open set $O(t)$ containing $x(t)$ such that

(a) $L(t,\cdot,\cdot)$ is lower semicontinuous on $O(t)\times I\!R^n$,

(b) for every point $(\bar{x},\bar{v})\in O(t)\times I\!R^n$ with $L(t,\bar{x},\bar{v})<\infty$, and every sequence $x^\nu\to\bar{x}$, there is a sequence $v^\nu\to\bar{v}$ with $L(t,x^\nu,v^\nu)\to L(t,\bar{x},\bar{v})$.

Clearly (a) and (b) are satisfied in particular when $L(t,\cdot,\cdot)$ is continuous on $O(t)\times I\!R^n$.

Theorem 1.1. *Let $L(t,x,v)$ be convex in v (possibly with the value ∞), and let $H(t,x,p)$ be defined by (1.3). Let $x(\cdot)$ be an arc along which L has the epi-continuity property. Suppose for almost every t that*

$$(w,0)\in\partial^\infty L\big(t,x(t),\dot{x}(t)\big) \quad\Longrightarrow\quad w=0, \tag{1.14}$$

this being true in particular if $L(t,\cdot,\cdot)$ is Lipschitz continuous around $\big(x(t),\dot{x}(t)\big)$. Then version (1.10) of the Euler-Lagrange condition is equivalent to version (1.12) of the Hamiltonian condition and automatically entails (1.11) and (1.13). The same holds when (1.14) is replaced by

$$(w,0)\in\partial^\infty H\big(t,x(t),p(t)\big) \quad\Longrightarrow\quad w=0, \tag{1.15}$$

this being true in particular if $H(t,\cdot,\cdot)$ is Lipschitz continuous around $\big(x(t),p(t)\big)$.

The epi-continuity property invoked in Theorem 1.1 concerns the continuity of the set-valued mapping that associates with each x the epigraph of the function $L(t,x,\cdot)$, as will become clearer in the next section. Assumption (1.14) concerns a kind of localized Lipschitz continuity of this mapping. Such properties of epigraphical mappings have long been implicit in most developments of the subject, in consequence for instance of Lipschitz assumptions placed on L or H, or on some underlying differential inclusion mapping, but their effects on subgradients haven't been explored directly. Here they emerge finally in the foreground. Also coming on stage for the first time in such a setting, through the technique we'll use to prove Theorem 1.1, will be a number of tools of convex analysis. These include Fenchel's duality theorem in convex optimization, Moreau's theory of proximal regularizations of convex functions, Wijsman's epi-continuity theorem for the Legendre-Fenchel transform, and Attouch's theorem on convergence of subgradients.

2. Dualization framework and epi-continuity.

For the task to be accomplished, the t argument doesn't matter; all questions revolve around properties that hold for a fixed t. We therefore suppress t. We consider an open subset O of $I\!R^m$ and take $L(x,v)$ to be an expression defined for $(x,v)\in O\times I\!R^n$ such that, for each $x\in O$, $L(x,\cdot)$ is a convex, lsc (lower semicontinuous) function on $I\!R^n$ which is *proper*, i.e., although possibly extended-real-valued does not take on $-\infty$ and is not identically ∞. In the targeted applications to Euler-Lagrange and Hamiltonian conditions we'll have $m=n$, but for the sake of other potential uses of the results to be obtained, we allow the dimensions m and n to differ. We define $H(x,p)$ for $(x,p)\in O\times I\!R^n$ by

$$H(x,p) \;=\; \sup_{v\in\mathbb{R}^n}\big\{\langle p,v\rangle - L(x,v)\big\}. \tag{2.1}$$

The Legendre-Fenchel transformation, on which this formula is based, has the property that for each $x\in O$, $H(x,\cdot)$ is, like $L(x,\cdot)$, a proper, convex, lsc function on $I\!R^n$, moreover with

$$L(x,v) \;=\; \sup_{p\in\mathbb{R}^n}\big\{\langle p,v\rangle - H(x,p)\big\}. \tag{2.2}$$

This symmetric relationship between L and H will enable us to apply any result proved for either function to the other function as well. We can later interpret O as a neighborhood of some particular point of $I\!R^m$ that happens to be under scrutiny.

The lower semicontinuity of $L(x,v)$ in v, and of $H(x,p)$ in p, has already been incorporated into our framework, but nothing has been said yet about continuity properties relative to x. At the very least we'll need $L(x,v)$ to be lsc in $(x,v) \in O \times I\!R^n$, and similarly for $H(x,p)$ in (x,p), but we're going to go further, clarifying on the way the property that provides the simplest dualization scheme and best supports our subsequent analysis. We'll be working with the concept of epi-continuity in the dependence of the functions $L(x,\cdot)$ and $H(x,\cdot)$ on x.

Recall that the *epigraph* of a function $f : I\!R^n \to \overline{I\!R}$ is the set

$$epi\, f = \big\{ (y,\alpha) \in I\!R^n \times I\!R \,\big|\, \alpha \geq f(y) \big\}.$$

In general, a sequence of functions $f^\nu : I\!R^n \to \overline{I\!R}$ is said to *epi-converge* to a function $f : I\!R^n \to \overline{I\!R}$ if the corresponding epigraphs converge:

$$epi\, f = \limsup_{\nu\to\infty} epi\, f^\nu = \liminf_{\nu\to\infty} epi\, f^\nu$$

in the Painlevé-Kuratowski sense as subsets of $I\!R^n \times I\!R$. This is true if and only if

$$\begin{cases} \liminf_\nu f^\nu(v^\nu) \geq f(v) & \text{for every sequence } v^\nu \to v, \\ \limsup_\nu f^\nu(v^\nu) \leq f(v) & \text{for some sequence } v^\nu \to v. \end{cases} \tag{2.3}$$

Epi-convergence was introduced for proper, lsc, convex functions by Wijsman [20], who proved that the Legendre-Fenchel transformation was continuous with respect to it. For more background on this topic, see Wets [21], Salinetti and Wets [22].

PROPOSITION 2.1. *The following six properties are equivalent and imply in particular that L and H are lsc in both arguments jointly, as functions on $O \times I\!R^n$:*

(a) *The set $epi\, L(x,\cdot)$ in $I\!R^n \times I\!R$ depends continuously on $x \in O$.*
(b) *Whenever $x^\nu \to \bar{x}$ in O, the function $L(x^\nu,\cdot)$ epi-converges to $L(\bar{x},\cdot)$.*
(c) *For any $(\bar{x},\bar{v}) \in O \times I\!R^n$ and sequence $x^\nu \to \bar{x}$, one has*

$$\begin{cases} \liminf_\nu L(x^\nu,v^\nu) \geq L(\bar{x},\bar{v}) & \textit{for every sequence } v^\nu \to \bar{v}, \\ \limsup_\nu L(x^\nu,v^\nu) \leq L(\bar{x},\bar{v}) & \textit{for some sequence } v^\nu \to \bar{v}. \end{cases}$$

(d) *The set $epi\, H(x,\cdot)$ in $I\!R^n \times I\!R$ depends continuously on $x \in O$.*
(e) *Whenever $x^\nu \to \bar{x}$ in O, the function $H(x^\nu,\cdot)$ epi-converges to $H(\bar{x},\cdot)$.*
(f) *For any $(\bar{x},\bar{p}) \in O \times I\!R^n$ and sequence $x^\nu \to \bar{x}$, one has*

$$\begin{cases} \liminf_\nu H(x^\nu,p^\nu) \geq H(\bar{x},\bar{p}) & \textit{for every sequence } p^\nu \to \bar{p}, \\ \limsup_\nu H(x^\nu,p^\nu) \leq H(\bar{x},\bar{p}) & \textit{for some sequence } p^\nu \to \bar{p}. \end{cases}$$

Proof. Conditions (a) and (b) mean the same, by the definition of epi-convergence, and (c) characterizes this property in accordance with the facts just cited. This pattern holds for (d), (e) and (f) as well. But because $L(x^\nu,\cdot)$ and $H(x^\nu,\cdot)$ are proper convex functions conjugate to each other under the Legendre-Fenchel transformation, which preserves epi-convergence according to Wijsman's theorem, (b) is equivalent to (e). Hence, all the conditions are equivalent to each other. □

For short, we'll say that the *epi-continuity assumption* is satisfied when the six equivalent properties in Proposition 2.1 are present. Obviously this is true in particular when L is continuous on $O \times I\!R^n$ (through (c)), or when H is continuous on $O \times I\!R^n$ (through (f)). In typical applications the epi-continuity assumption merely means (through property (c)) that, in addition to taking $L(x, v)$ to be lsc in (x, v), rather than just in x, we suppose that whenever $(\bar{x}, \bar{v})$ is a point of $O \times I\!R^n$ where L is finite, and $\{x^\nu\}$ is a sequence in O converging to $\bar{x}$, there must be a sequence $\{v^\nu\}$ converging to $\bar{v}$ for which $L(x^\nu, v^\nu)$ converges to $L(\bar{x}, \bar{v})$.

Note that the epi-continuity condition used in the hypothesis of Theorem 1.1 merely requires for almost every t that this should hold relative to some neighborhood $O(t)$ of $x(t)$. Proposition 2.1 shows that the condition in question could be expressed in terms of the Hamiltonian just as well as the Lagrangian. It's actually symmetric between the two functions (as long as the Lagrangian is lsc and convex with respect to v).

The study of subgradients of L and H with respect to both of their arguments in $O \times I\!R^n$ requires working with the definition in Section 1 in terms of limits of proximal subgradients. But subgradients of L in the v argument, and of H in the p argument, enjoy the benefits of convexity. Convex analysis informs us that

$$p \in \partial_v L(x, v) \iff v \in \partial_p H(x, p) \iff L(x, v) + H(x, p) = \langle p, v \rangle, \tag{2.3}$$

cf. [1, Thm. 23.5], where from (2.1) and (2.2) we know that $L(x, v) + H(x, p) \geq \langle v, p \rangle$ for all choices of $(x, v, p) \in O \times I\!R^n \times I\!R^n$.

PROPOSITION 2.2. *Under the epi-continuity assumption,*

$$\begin{aligned} (w, p) \in \partial L(x, v) &\implies p \in \partial_v L(x, v), \\ (w, v) \in \partial H(x, p) &\implies v \in \partial_p H(x, p). \end{aligned} \tag{2.4}$$

Proof. Due to symmetry, it suffices to deal with the first of these implications. Suppose $(\bar{w}, \bar{p}) \in \partial L(\bar{x}, \bar{v})$. By definition there exist $(x^\nu, v^\nu) \to (\bar{x}, \bar{v})$ and $(w^\nu, p^\nu) \to (\bar{w}, \bar{p})$ such that (w^ν, p^ν) is a proximal subgradient of L at (x^ν, v^ν), and $L(x^\nu, v^\nu) \to L(\bar{x}, \bar{v})$. The proximal subgradient condition refers to the existence of $\rho^\nu \geq 0$ and $\delta^\nu > 0$ such that

$$L(x, v) \;\geq\; L(x^\nu, v^\nu) + \big\langle (w^\nu, p^\nu),\, (x, v) - (x^\nu, v^\nu) \big\rangle - \tfrac{1}{2}\rho^\nu \big(|x - x^\nu|^2 + |v - v^\nu|^2 \big)$$

when $\big|(x - x^\nu,\, v - v^\nu)\big| \leq \delta^\nu$. In taking $x = x^\nu$ we see that the convex function

$$f^\nu(v) \;:=\; L(x^\nu, v) - \big\langle p^\nu,\, v - v^\nu \big\rangle + \tfrac{1}{2}\rho^\nu |v - v^\nu|^2$$

must have a local minimum at v^ν. This implies that $0 \in \partial f^\nu(v^\nu) = \partial_v L(x^\nu, v^\nu) - p^\nu$, or in other words, $p^\nu \in \partial_v L(x^\nu, v^\nu)$, a subgradient condition which, because of the convexity of $L(x, v)$ in v, can be written as the inequality

$$L(x^\nu, v) \;\geq\; L(x^\nu, v^\nu) + \big\langle p^\nu,\, v - v^\nu \big\rangle \quad \text{for all} \quad v \in I\!R^n.$$

Consider now an arbitrary $v \in I\!R^n$ for which $L(\bar{x}, v) < \infty$. Our epi-continuity assumption ensures the existence of a sequence $\hat{v}^\nu \to v$ with $L(x^\nu, \hat{v}^\nu) \to L(x^\nu, v)$. For each index ν we have

$$L(x^\nu, \hat{v}^\nu) \;\geq\; L(x^\nu, v^\nu) + \big\langle p^\nu,\, \hat{v}^\nu - v^\nu \big\rangle.$$

In passing to the limit as $\nu \to \infty$ and using the fact that $L(x^\nu, v^\nu) \to L(\bar{x}, v)$ in particular, we obtain

$$L(\bar{x}, v) \ \geq \ L(\bar{x}, \bar{v}) + \langle \bar{p},\, v - \bar{v} \rangle.$$

We have shown this inequality to hold for any v with $L(\bar{x}, v)$ finite, but it holds trivially when $L(\bar{x}, v) = \infty$. Hence it holds for all $v \in \mathbb{R}^n$, confirming that $\bar{p} \in \partial_v L(\bar{x}, \bar{v})$. □

Proposition 2.2 suggests approaching the subgradients of L and H in general by looking at the set

$$M \ := \ \big\{ (x, v, p) \in O \times \mathbb{R}^n \times \mathbb{R}^n \,\big|\, \text{ properties (2.3) hold } \big\} \tag{2.5}$$

and the set-valued mappings

$$\begin{aligned} S_L &: (x, v, p) \mapsto \big\{ w \,\big|\, (w, p) \in \partial L(x, v) \big\}, \\ S_H &: (x, v, p) \mapsto \big\{ w \,\big|\, (w, v) \in \partial H(x, p) \big\}. \end{aligned} \tag{2.6}$$

The graph of S_L, consisting by definition of all (x, v, p, w) such that $w \in S_L(x, v, p)$, is the same then as the graph of ∂L, except for a permutation of arguments; likewise for the graph of S_H in comparison with the graph of ∂H.

Proposition 2.3. *Under the epi-continuity assumption, M is closed in $O \times \mathbb{R}^n \times \mathbb{R}^n$ and the functions $(x, v, p) \mapsto L(x, v)$ and $(x, v, p) \mapsto H(x, p)$ are finite and continuous on M. Moreover, the effective domains of the set-valued mappings S_L and S_H on $O \times \mathbb{R}^n \times \mathbb{R}^n$ (the effective domains being the sets on which the mappings are nonempty-valued) lie in M, and the graphs of these mappings are closed as subsets of $O \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$.*

Proof. We can view M as the graph of the mapping G that associates with each $x \in O$ the set of all $(v, p) \in \mathbb{R}^n \times \mathbb{R}^n$ such that p is a subgradient of the convex function $L(x, \cdot)$ at v. According to Attouch's theorem on subgradient convergence (see [23]), the epi-convergence of $L(x^\nu, \cdot)$ to $L(x, \cdot)$ implies the set convergence of $G(x^\nu)$ to $G(x)$. In particular this entails the closedness of the graph of G in $O \times \mathbb{R}^n \times \mathbb{R}^n$.

As functions of (x, v, p), both $L(x, v)$ and $H(x, p)$ are lower semicontinuous by Proposition 2.1 and never take on $-\infty$. But on M they are related by $H(x, p) = \langle v, p \rangle - L(x, v)$ and $L(x, v) = \langle v, p \rangle - H(x, p)$, so they can't take on ∞ either and must be upper semicontinuous as well. Hence they are finite and continuous on M.

The assertion about the effective domains of S_L and S_H just restates Proposition 2.2. Verifying the closedness of the graphs of S_L and S_H comes down to verifying the closedness of the graphs of ∂L and ∂H. The graph of ∂L consists by definition of the closure, in a special way relative to $O \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$, of the set of all (x, v, w, p) such that (w, p) is a proximal subgradient of L at (x, v). The closure consists of all limits of sequences $(x^\nu, v^\nu, w^\nu, p^\nu)$ that not only converge themselves but have the additional property that the values $L(x^\nu, v^\nu)$ converge. But whenever (w^ν, p^ν) is a proximal subgradient of L at (x^ν, v^ν) we have in particular that p^ν is a proximal subgradient of the convex function $L(x^\nu, \cdot)$ at v^ν. This implies $p^\nu \in \partial_v L(x^\nu, v^\nu)$, hence $(x^\nu, v^\nu, p^\nu) \in M$. The convergence of the values $L(x^\nu, v^\nu)$ is then automatic because L is continuous as a function on M. Thus, the special feature of the closure process falls away, and the graph of ∂L is seen to be a closed set in the ordinary sense relative to $O \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$. For ∂H the argument is parallel. □

Our strategy for proving Theorem 1.1 is now perhaps becoming clear. Under the epi-continuity assumption, we need only come up with additional conditions on a point $(\bar{x}, \bar{v}, \bar{p}) \in M$ which guarantee that the sets $S_L(\bar{x}, \bar{v}, \bar{p})$ and $-S_H(\bar{x}, \bar{v}, \bar{p})$ have the same convex hull. The fact that these sets don't necessarily agree in advance of taking convex hulls is evident from simple examples. For instance, if $L(x,v) = c(x) + l(v)$ for a finite, continuous function c on $I\!R^m$ and a proper, lsc, convex function l on $I\!R^n$, we have $H(x,p) = -c(x) + h(p)$ with h the proper, lsc, convex function on $I\!R^n$ conjugate to l. Then M is the product of $I\!R^m$ with the graph of ∂l, and for $(\bar{x},\bar{v},\bar{p}) \in M$ we see that

$$S_L(\bar{x},\bar{v},\bar{p}) = \partial c(\bar{x}), \qquad S_H(\bar{x},\bar{v},\bar{p}) = -\partial(-c)(\bar{x}).$$

While $\partial c(\bar{x})$ is the set of subgradients of c at $\bar{x}$ as defined in the manner explained, from limits of proximal subgradients introduced "from below," $-\partial(-c)(\bar{x})$ has the analogous interpretation with proximal subgradients introduced instead "from above." These two sets are known often to differ for a nonsmooth function c, although they have the same convex hull when c is Lipschitz continuous on a neighborhood of $\bar{x}$.

The key to further progress will be the following *regularized functions* associated with L and H:

$$\begin{aligned} R_L(x,u) &:= \inf_{v\in \boldsymbol{R}^n} \Big\{ L(x,v) + \tfrac{1}{2}|v-u|^2 \Big\}, \\ R_H(x,u) &:= \inf_{p\in \boldsymbol{R}^n} \Big\{ H(x,p) + \tfrac{1}{2}|p-u|^2 \Big\}, \end{aligned} \tag{2.7}$$

where $|\cdot|$ is the Euclidean norm. Also important will be the corresponding *proximal mappings*:

$$\begin{aligned} P_L(x,u) &:= \operatorname*{argmin}_{v\in \boldsymbol{R}^n} \Big\{ L(x,v) + \tfrac{1}{2}|v-u|^2 \Big\}, \\ P_H(x,u) &:= \operatorname*{argmin}_{p\in \boldsymbol{R}^n} \Big\{ H(x,p) + \tfrac{1}{2}|p-u|^2 \Big\}. \end{aligned} \tag{2.8}$$

In dealing with these regularized functions and proximal mappings for a fixed u, we draw heavily on the theory of Moreau [24]; for reference see also Rockafellar [1, Thm. 31.5]. For any fixed $x \in O$, the functions $u \mapsto R_L(x,u)$ and $u \mapsto R_H(x,u)$ are finite, convex and continuously differentiable on $I\!R^n$. They satisfy the identity

$$R_L(x,u) + R_H(x,u) = \tfrac{1}{2}|u|^2. \tag{2.9}$$

The mappings $u \mapsto P_L(x,u)$ and $u \mapsto P_H(x,u)$ are single-valued from $I\!R^n$ into $I\!R^n$ and nonexpansive—globally Lipschitz continuous with constant 1—and they are related by

$$P_L(x,u) + P_H(x,u) = u, \tag{2.10}$$

$$P_L(x,u) = \nabla_u R_H(x,u), \qquad P_H(x,u) = \nabla_u R_L(x,u). \tag{2.11}$$

The proximal mappings P_L and P_H are especially of interest in providing a convenient parameterization of M in terms of (x,u).

PROPOSITION 2.4. *Under the epi-continuity assumption, $R_L(x,u)$ and $R_H(x,u)$ are not just continuous in u, but with respect to $(x,u) \in O \times I\!R^n$. This holds also for $P_L(x,u)$ and $P_H(x,u)$. The one-to-one correspondence between points $(x,v,p) \in M$ and points $(x,u) \in O \times I\!R^n$ that is set up by the relations*

$$u = v + p, \qquad (v,p) = \big(P_L(x,u), P_H(x,u)\big),$$

is then a homeomorphism.

Proof. From the theory of epi-convergence [21], [23], the convex functions $L(x^\nu,\cdot)$ epi-converge to $L(x,\cdot)$ if and only if the regularized functions $R_L(x^\nu,\cdot)$ converge pointwise on $I\!R^n$ to $R_L(x,\cdot)$. These regularized functions being not only convex but finite, their pointwise convergence implies uniform convergence on all bounded subsets of $I\!R^n$, cf. [1, Thm. 10.8], and then their gradient mappings $\nabla R_L(x^\nu, \cdot)$ converge in such a manner to $\nabla R_H(x, \cdot)$ as well [1, Thm. 24.5]. The assertions about R_L and P_L, and similarly those about R_H and P_H, thus follow from the meaning of the epi-continuity assumption as designating the six conditions in Proposition 2.1. (The indicated correspondence is one-to-one because the elements $v = P_L(x,u)$ and $p = P_H(x,u)$ satisfy (2.10) and are characterized by the relations $0 \in \partial_v L(x,v) + (v-u)$ and $0 \in \partial_p H(x,p) + (v-p)$.) $\square$

The gradients of $R_L(x,u)$ and $R_H(x,u)$ with respect to u are pinpointed by (2.11), but we can also determine, or at least estimate, subgradients with respect to (x,u).

Proposition 2.5. *Under the epi-continuity assumption, consider any $(\bar{x},\bar{u}) \in O \times I\!R^n$ and let $\bar{v} = P_L(\bar{x},\bar{u})$ and $\bar{p} = P_H(\bar{x},\bar{u})$. Then*

$$\begin{aligned}
\partial R_L(\bar{x},\bar{u}) &\subset \big\{(w,p) \,\big|\, p = \bar{p},\ (w,\bar{p}) \in \partial L(\bar{x},\bar{v})\big\},\\
\partial^\infty R_L(\bar{x},\bar{u}) &\subset \big\{(w,p) \,\big|\, p = 0,\ (w,0) \in \partial^\infty L(\bar{x},\bar{v})\big\},\\
\partial R_H(\bar{x},\bar{u}) &\subset \big\{(w,v) \,\big|\, v = \bar{v},\ (w,\bar{v}) \in \partial H(\bar{x},\bar{p})\big\},\\
\partial^\infty R_H(\bar{x},\bar{u}) &\subset \big\{(w,v) \,\big|\, v = 0,\ (w,0) \in \partial^\infty H(\bar{x},\bar{p})\big\}.
\end{aligned}$$

Proof. Let $f(x,v,u) = L(x,v) + \frac{1}{2}|v-u|^2$, so that $R_L(x,u) = min_v\, f(x,v,u)$. For $(\bar{x},\bar{u})$ the minimum is attained uniquely at $\bar{v}$. A general calculus rule in [25, Thm. 3.1] gives us

$$\begin{aligned}
\partial R_L(\bar{x},\bar{u}) &\subset \big\{(w,p) \,\big|\, (w,0,p) \in \partial f(\bar{x},\bar{v},\bar{u})\big\},\\
\partial^\infty R_L(\bar{x},\bar{u}) &\subset \big\{(w,p) \,\big|\, (w,0,p) \in \partial^\infty f(\bar{x},\bar{v},\bar{u})\big\},
\end{aligned}$$

but also in terms of the smooth function $f_0(x,v,u) := \frac{1}{2}|v-u|^2$ we have (cf. [7, Lem. 5A.1]) that

$$\begin{aligned}
\partial f(\bar{x},\bar{v},\bar{u}) &= \big[\partial L(\bar{x},\bar{v}) \times \{0\}\big] + \nabla f_0(\bar{x},\bar{v},\bar{u}),\\
\partial^\infty f(\bar{x},\bar{v},\bar{u}) &= \big[\partial^\infty L(\bar{x},\bar{v}) \times \{0\}\big],
\end{aligned}$$

where $\nabla f_0(\bar{x},\bar{v},\bar{u}) = (0,\ \bar{v}-\bar{u},\ \bar{u}-\bar{v})$. The combination of these two sets of relations yields the inclusions claimed in the proposition for $\partial R_L(\bar{x},\bar{u})$ and $\partial^\infty R_L(\bar{x},\bar{u})$. The ones for $\partial R_H(\bar{x},\bar{u})$ and $\partial^\infty R_H(\bar{x},\bar{u})$ follow by symmetry. $\square$

Lipschitz continuity of the regularized functions with respect to x will be critical to us at a certain stage. This property is the subject of the next proposition.

Proposition 2.6. *Under the epi-continuity assumption, the following four properties are equivalent to each other at a point $(\bar{x},\bar{u}) \in O \times I\!R^n$:*

(a) *R_L is Lipschitz continuous around $(\bar{x},\bar{u})$.*
(b) *$R_L(x,\bar{u})$ is Lipschitz continuous in x around $\bar{x}$.*
(c) *R_H is Lipschitz continuous around $(\bar{x},\bar{u})$.*
(d) *$R_H(x,\bar{u})$ is Lipschitz continuous in x around $\bar{x}$.*

These properties are present in particular when the unique vectors $\bar{v}$ and $\bar{p}$ satisfying $\bar{u} = \bar{v} + \bar{p}$ and $(\bar{x},\bar{v},\bar{p}) \in M$ are such that

$$(w,0) \in \partial^\infty L(\bar{x},\bar{v}) \quad\Longrightarrow\quad w = 0, \tag{2.12}$$

or such that

$$(w,0) \in \partial^\infty H(\bar{x},\bar{p}) \quad\Longrightarrow\quad w = 0. \tag{2.13}$$

As a special case, the first of these two conditions is implied by L being Lipschitz continuous around $(\bar{x}, \bar{v})$, whereas the second is implied by H being Lipschitz continuous around $(\bar{x}, \bar{p})$.

Proof. The equivalence is apparent from the identity in (2.9) and the fact that $R_L(x, u)$ and $R_H(x, u)$ are finite, convex functions of u, hence locally Lipschitz continuous in u. (As a matter of fact, they are globally Lipschitz continuous in u with constant 1.)

By a result of Rockafellar [26], the function R_L, because of its lower semicontinuity on $O \times I\!R^n$ (Proposition 2.1), is Lipschitz continuous around $(\bar{x}, \bar{u})$ if and only if the set $\partial^\infty R_L(\bar{x}, \bar{u})$ contains only $(0, 0)$. This is true under (2.12) by virtue of the second inclusion in Proposition 2.5. In the same way, (2.13) suffices for Lipschitz continuity of R_H.

If L is Lipschitz continuous around $(\bar{x}, \bar{v})$, so that $\partial^\infty L(\bar{x}, \bar{v}) = \{(0, 0)\}$ by the result cited, we have (2.12) trivially. Likewise, if H is Lipschitz continuous around $(\bar{x}, \bar{p})$, so that $\partial^\infty H(\bar{x}, \bar{p}) = \{(0, 0)\}$, we have (2.13) trivially. □

For the record, conditions (2.12) and (2.13) aren't equivalent to each other, and they therefore can't actually be equivalent to the Lipschitz continuity property in Proposition 2.6, but merely sufficient for it. This is seen through the example of $L(x, v) = x^{4/3} v^2$ on $I\!R^1 \times I\!R^1$, which is convex in v and continuously differentiable in (x, v). Since $\nabla L(0, 0) = (0, 0)$, the point $(\bar{x}, \bar{v}, \bar{p}) = (0, 0, 0)$ belongs to M. We have $\partial^\infty L(0, 0) = \{(0, 0)\}$ because L is Lipschitz continuous around $(0, 0)$; thus (2.12) is satisfied. But (2.13) isn't satisfied, which is seen as follows. Our choice of L corresponds to

$$H(x, p) = \begin{cases} p^2/4x^{4/3} & \text{when } x \neq 0, \\ 0 & \text{when } x = 0 \text{ and } p = 0, \\ \infty & \text{when } x = 0 \text{ and } p \neq 0. \end{cases}$$

Away from $x = 0$, H is twice continuously differentiable, so its gradients $\nabla H(x, p) = \big(-p^2/3x^{7/3},\, p/2x^{4/3}\big)$ are proximal subgradients. We aim at constructing a nonzero vector $(\bar{w}, 0)$ in $\partial^\infty H(0, 0)$ in accordance with the definition of that set in terms of limits of proximal subgradients. Consider any sequence $t^\nu \searrow 0$ and let $x^\nu = (t^\nu)^6$, $p^\nu = (t^\nu)^5$. Then $(x^\nu, p^\nu) \to (\bar{x}, \bar{p}) = (0, 0)$ and $H(x^\nu, p^\nu) = (t^\nu)^2/4 \to 0$. Let $(w^\nu, v^\nu) = \nabla H(x^\nu, p^\nu) = \big(-(t^\nu)^{-4},\, (t^\nu)^{-3}/2\big)$. Then for $\lambda^\nu = (t^\nu)^4$ we have $\lambda^\nu \searrow 0$ and $\lambda^\nu (w^\nu, v^\nu) = (-1, t^\nu/2) \to (-1, 0)$. This limit vector belongs to $\partial^\infty H(0, 0)$ and demonstrates that (2.13) fails.

3. The main arguments.

With this foundation in place, we can turn to the subgradient arguments that lead to the equivalence relation in Theorem 1.1.

LEMMA 3.1. *Under the epi-continuity assumption, suppose that $(\bar{w}, \bar{p})$ is a proximal subgradient to L at a point $(\bar{x}, \bar{v}) \in O \times I\!R^n$; in other words, $L(\bar{x}, \bar{v})$ is finite and there exist $\rho > 0$ and $\delta > 0$ such that*

$$\begin{aligned} L(x, v) \;&\geq L(\bar{x}, \bar{v}) + \langle \bar{w},\, x - \bar{x} \rangle + \langle \bar{p},\, v - \bar{v} \rangle - \tfrac{1}{2}\rho |x - \bar{x}|^2 - \tfrac{1}{2}\rho |v - \bar{v}|^2 \\ &\qquad \textit{when } x \in O,\ |x - \bar{x}| \leq \delta,\ |v - \bar{v}| \leq \delta. \end{aligned} \tag{3.1}$$

Then there exists $\varepsilon \in (0, \delta)$ such that

$$\begin{aligned} L(x, v) \;&\geq L(\bar{x}, \bar{v}) + \langle \bar{w},\, x - \bar{x} \rangle + \langle \bar{p},\, v - \bar{v} \rangle - \tfrac{1}{2}\rho |x - \bar{x}|^2 - \tfrac{1}{2}\rho |v - \bar{v}|^2 \\ &\qquad \textit{for all } v \in I\!R^n \textit{ when } x \in O,\ |x - \bar{x}| \leq \varepsilon. \end{aligned} \tag{3.2}$$

Proof. We can write (3.1) equivalently as

$$\rho^{-1}L(\bar{x},\bar{v}) + \langle \rho^{-1}\bar{w},\, x-\bar{x}\rangle - \tfrac{1}{2}|x-\bar{x}|^2 \;\le\; f(x) \quad \text{when} \quad x\in O,\ |x-\bar{x}|\le\delta, \tag{3.3}$$

where

$$\begin{aligned} f(x) := &\inf_{|v-\bar{v}|\le\delta} \big\{\rho^{-1}L(x,v) + \tfrac{1}{2}|v-\bar{v}|^2 - \langle \rho^{-1}\bar{p},\, v-\bar{v}\rangle\big\} \\ = &\inf_{|v-\bar{v}|\le\delta} \big\{\rho^{-1}L(x,v) + \tfrac{1}{2}\big|v-(\bar{v}+\rho^{-1}\bar{p})\big|^2\big\} - \tfrac{1}{2}|\rho^{-1}\bar{p}|^2. \end{aligned}$$

The question is whether (3.3) will continue to hold when the constraint $|v-\bar{v}|\le\delta$ is dropped in the definition of f, at least if δ is replaced by some smaller value in (3.3).

We can answer this by applying the facts about proximal regularization to the function

$$\widehat{L}(x,v) := \begin{cases} \rho^{-1}L(x,v) & \text{when } |v-\bar{v}|\le\delta, \\ \infty & \text{when } |v-\bar{v}|>\delta, \end{cases}$$

in terms of which we have $f(x) = R_{\widehat{L}}(x,\bar{u})$ for $\bar{u} = \bar{v}+\rho^{-1}\bar{p}$. Here $\widehat{L}(x,\cdot)$ is the sum of two convex functions, namely $\rho^{-1}L(x,\cdot)$ and the indicator of $\big\{v \,\big|\, |v-\bar{v}|\le\delta\big\}$. For $x=\bar{x}$, the effective domain of the first function meets the interior of the effective domain of the second, and this is enough to guarantee through the convergence theorem of McLinden and Bergstrom [27] that whenever $x^\nu\to\bar{x}$ and $L(x^\nu,\cdot)$ epi-converges to $L(\bar{x},\cdot)$, the sum $\widehat{L}(x^\nu,\cdot)$ epi-converges to $\widehat{L}(\bar{x},\cdot)$. It follows then from our epi-continuity assumption that, for x in some open neighborhood O' of $\bar{x}$ within O, $\widehat{L}(x,\cdot)$ depends epi-continuously on x.

Through Proposition 2.4 we conclude that the associated proximal mapping $P_{\widehat{L}}$, which gives the unique minimizing v in the formula for f, is continuous on $O'\times I\!R^n$. Moreover, the minimum defining $f(\bar{x})$ is attained at $\bar{v}$, because of the proximal subgradient inequality; thus, $P_{\widehat{L}}(\bar{x},\bar{u}) = \bar{v}$. There must, then, exist an $\varepsilon\in(0,\delta)$ such that when $x\in O'$ and $|x-\bar{x}|\le\varepsilon$ we have $\big|P_{\widehat{L}}(x,\bar{u})-\bar{v}\big|<\delta$. For such x the constraint $|v-\bar{v}|\le\delta$ in the formula for $f(x)$ is inactive. Since the function of v being minimized in this formula is convex, the constraint can in this case be suppressed without affecting the minimum value that is attained. □

LEMMA 3.2. *Under the epi-continuity assumption, suppose $(\bar{w},\bar{p})$ is a proximal subgradient of L at a point $(\bar{x},\bar{v})\in O\times I\!R^n$, and let $\bar{u}=\bar{v}+\bar{p}$. If R_H is Lipschitz continuous around $(\bar{x},\bar{u})$, then $(-\bar{w},\bar{v})\in\bar{\partial}R_H(\bar{x},\bar{u}) = con\,\partial R_H(\bar{x},\bar{u})$.*

Proof. We start by noting that in particular $(\bar{w},\bar{p})\in\partial L(\bar{x},\bar{v})$, hence $\bar{p}\in\partial_v L(\bar{x},\bar{v})$ by Proposition 2.2. Thus $(\bar{x},\bar{v},\bar{p})\in M$. We have

$$\bar{v} = \nabla_u R_H(\bar{x},\bar{u}) \tag{3.4}$$

by (2.11); this will be needed later.

Through Lemma 3.1 the proximal subgradient condition can be identified with the existence of $\rho>0$ and $\varepsilon>0$ such that (3.2) holds. Replacing the second occurrence of ρ in (3.2) by $\rho+1$, which certainly maintains the inequality, we get

$$F(\bar{x}) + \langle\bar{v}, x-\bar{x}\rangle - \tfrac{1}{2}\rho|x-\bar{x}|^2 \le F(x) \quad \text{when} \quad |x-\bar{x}|\le\varepsilon \tag{3.5}$$

for the function

$$F(x) \;:=\; \inf_{v\in \boldsymbol{R}^n} \big\{L(x,v) - \langle\bar{p},\, v-\bar{v}\rangle + \tfrac{1}{2}(\rho+1)|v-\bar{v}|^2\big\},$$

which has $F(\bar{x}) = L(\bar{x}, \bar{v})$. As a consequence of (3.5) we certainly have $\bar{w} \in \partial F(\bar{x})$.

To carry the analysis of $\bar{w}$ further, we'll make use of Fenchel's duality theorem [1, Thm. 31.1] to represent F in a different way. For any fixed x the definition of $F(x)$ can be interpreted as saying that

$$F(x) = \inf_{v \in \boldsymbol{R}^n} \{\varphi(v) - \psi(v)\}$$

for the convex function $\varphi(v) := L(x, v) + \frac{1}{2}|v - \bar{v}|^2$ and the concave function $\psi(v) := \langle \bar{p},\, v - \bar{v}\rangle - \frac{1}{2}\rho|v - \bar{v}|^2$. Because ψ is finite everywhere, the effective domains of these functions overlap in the manner required by the duality theorem in question, and we are able to conclude from it that

$$-F(x) = \inf_{p \in \boldsymbol{R}^n} \{\varphi^*(p) - \psi^*(p)\}$$

for the functions φ^* and ψ^* conjugate to φ and ψ. From the definition of the convex conjugate φ^* we calculate that

$$\begin{aligned}
\varphi^*(p) &= \sup_{v \in \boldsymbol{R}^n} \{\langle p, v\rangle - L(x, v) - \tfrac{1}{2}|v - \bar{v}|^2\} \\
&= -\inf_{v \in \boldsymbol{R}^n} \{L(x, v) + \tfrac{1}{2}|v|^2 - \langle \bar{v} + p, v\rangle + \tfrac{1}{2}|\bar{v} + p|^2 - \tfrac{1}{2}|\bar{v} + p|^2 + \tfrac{1}{2}|\bar{v}|^2\} \\
&= \tfrac{1}{2}|\bar{v} + p|^2 - \tfrac{1}{2}|\bar{v}|^2 - R_L(x, \bar{v} + p) = R_H(x, \bar{v} + p) - \tfrac{1}{2}|\bar{v}|^2,
\end{aligned}$$

where the final steps use definition (2.7) and the identity (2.9). The definition of the concave conjugate ψ^* yields

$$\psi^*(p) = \inf_{v \in \boldsymbol{R}^n} \{\langle p, v\rangle - \langle \bar{p},\, v - \bar{v}\rangle + \tfrac{1}{2}\rho|v - \bar{v}|^2\} \;=\; \langle p, \bar{v}\rangle - \tfrac{1}{2}\rho^{-1}|p - \bar{p}|^2.$$

Out of these calculations we get

$$-F(x) = \inf_{p \in \boldsymbol{R}^n} \{R_H(x, \bar{v} + p) - \tfrac{1}{2}|\bar{v}|^2 - \langle p, \bar{v}\rangle + \tfrac{1}{2}\rho^{-1}|p - \bar{p}|^2\},$$

which can be transformed to

$$-\rho F(x) = \inf_{p \in \boldsymbol{R}^n} \{\rho R_H(x, \bar{v} + p) + \tfrac{1}{2}\big|p - (\rho\bar{v} + \bar{p})\big|^2\} - \tfrac{1}{2}\rho(\rho + 1)|\bar{v}|^2. \tag{3.6}$$

In terms of $\widetilde{H}(x, p) := \rho R_H(x, \bar{v} + p)$ this has the interpretation that

$$\rho F(x) = -R_{\widetilde{H}}(x, \tilde{u}) + \tfrac{1}{2}\rho(\rho + 1)|\bar{v}|^2 \quad \text{for} \quad \tilde{u} = \rho\bar{v} + \bar{p}. \tag{3.7}$$

Proposition 2.4 assures us that $\widetilde{H}(x, p)$ is finite and continuous in $(x, p) \in O \times \mathbb{R}^n$. It's convex in p besides. We can therefore apply our proximal regularization results to this function in place of H. By assumption, $\widetilde{H}$ is Lipschitz continuous around $(\bar{x}, \bar{p})$. We have $\rho\bar{v} \in \partial_p \widetilde{H}(\bar{x}, \bar{p})$ by (3.4), hence also $\bar{p} = P_{\widetilde{H}}(\bar{x}, \tilde{u})$ and $\rho\bar{v} = \nabla_u R_{\widetilde{H}}(\bar{x}, \tilde{u})$.

By means of Proposition 2.6 we see that $R_{\widetilde{H}}$ is Lipschitz continuous around $(\bar{x}, \tilde{u})$. The fact that $\bar{w} \in \partial F(\bar{x})$ gives us in (3.7) that $\rho\bar{w} \in \partial_x(-R_{\widetilde{H}})(\bar{x}, \tilde{u})$. But by the

Lipschitz continuity of $R_{\widetilde{H}}(\cdot,\tilde{u})$ at $\bar{x}$ the Clarke subgradient relation $\bar{\partial}_x(-R_{\widetilde{H}})(\bar{x},\tilde{u}) = -\bar{\partial}_x R_{\widetilde{H}}(\bar{x},\tilde{u})$ holds, i.e., *con* $\partial_x(-R_{\widetilde{H}})(\bar{x},\tilde{u}) = -$ *con* $\partial_x R_{\widetilde{H}}(\bar{x},\tilde{u})$, cf. [5], [7]. Therefore

$$-\rho\bar{w} \in con\ \partial_x R_{\widetilde{H}}(\bar{x},\tilde{u}). \tag{3.8}$$

Next we analyze the set $\partial_x R_{\widetilde{H}}(\bar{x},\tilde{u})$. Because of the Lipschitz continuity of $R_{\widetilde{H}}$ around $(\bar{x},\tilde{u})$, the rule holds that

$$\partial_x R_{\widetilde{H}}(\bar{x},\tilde{u}) \subset \big\{ w \,\big|\, \exists\, v \text{ with } (w,v) \in \partial R_{\widetilde{H}}(\bar{x},\tilde{u}) \big\},$$

see [7, Lem. 5A.3]. Now Proposition 2.5, as applied with $\widetilde{H}$ in place of H (using the fact that $\tilde{u} = \rho\bar{v} + \bar{p}$ with $\rho\bar{v} \in \partial_p R_{\widetilde{H}}(\bar{x},\tilde{u})$), says that

$$\partial R_{\widetilde{H}}(\bar{x},\tilde{u}) \ \subset\ \big\{ (w,v) \,\big|\, v = \rho\bar{v},\ (w,\rho\bar{v}) \in \partial\widetilde{H}(\bar{x},\bar{p}) \big\},$$

where from the choice of $\widetilde{H}$ we have $\partial\widetilde{H}(\bar{x},\bar{p}) = \rho\partial R_H(\bar{x},\bar{u})$. In combination with (3.8), this gives us $(-\bar{w},\bar{v}) \in con\ \partial R_H(\bar{x},\bar{u})$, as claimed. □

THEOREM 3.3. *Under the epi-continuity assumption, consider the mappings S_L and S_H of (2.6) at a point $(\bar{x},\bar{v},\bar{p})$ such that either $S_L(\bar{x},\bar{v},\bar{p})$ or $S_H(\bar{x},\bar{v},\bar{p})$ is nonempty, or merely $(\bar{x},\bar{v},\bar{p}) \in M$. Suppose for $\bar{u} = \bar{v} + \bar{p}$ that $R_L(\cdot,\bar{u})$ is Lipschitz continuous around $\bar{x}$, or equivalently, that $R_H(\cdot,\bar{u})$ is Lipschitz continuous around $\bar{x}$. Then both $S_L(\bar{x},\bar{v},\bar{p})$ and $S_H(\bar{x},\bar{v},\bar{p})$ are nonempty and compact, and*

$$con\, S_L(\bar{x},\bar{v},\bar{p}) = -\, con\, S_H(\bar{x},\bar{v},\bar{p}).$$

Proof. In all cases we have $(\bar{x},\bar{v},\bar{p}) \in M$, since otherwise both $S_L(\bar{x},\bar{v},\bar{p})$ and $S_H(\bar{x},\bar{v},\bar{p})$ are empty by Proposition 2.3. The equivalence of the Lipschitz continuity assumptions is shown by Proposition 2.6, which also reveals they imply that R_L and R_H are Lipschitz continuous around $(\bar{x},\bar{u})$.

It will be demonstrated there is a compact set W such that whenever $\bar{w} \in S_L(\bar{x},\bar{v},\bar{p})$ we have $\bar{w} \in W$ and $-\bar{w} \in con\, S_H(\bar{x},\bar{v},\bar{p})$. The full conclusion of the theorem will follow then by symmetry.

By the definition of subgradients in general, the relation $\bar{w} \in S_L(\bar{x},\bar{v},\bar{p})$ implies the existence of proximal subgradients (w^ν,p^ν) to L at points $(x^\nu,v^\nu) \to (\bar{x},\bar{v})$ such that $(w^\nu,p^\nu) \to (\bar{w},\bar{p})$. The points $u^\nu = v^\nu + p^\nu$ then converge to $\bar{u}$, so that eventually (x^ν,u^ν) lies in the neighborhood of $(\bar{x},\bar{u})$ in which R_H is Lipschitz continuous. Once this is true, we can apply Lemma 3.2 at (x^ν,u^ν) to ascertain that $(-w^\nu,v^\nu) \in \bar{\partial}R_H(x^\nu,u^\nu)$. In the limit this yields $(-\bar{w},\bar{v}) \in \bar{\partial}R_H(\bar{x},\bar{u})$. In particular $\bar{w}$ belongs to the image W of $\bar{\partial}R_H(\bar{x},\bar{u})$ under the projection $(w,v) \mapsto -w$. This image set W is compact because $\bar{\partial}R_H(\bar{x},\bar{u})$ is compact (in consequence of the Lipschitz continuity of R_H). Since $\bar{\partial}R_H(x^\nu,u^\nu) = con\,\partial R_H(x^\nu,u^\nu)$, the inclusion for $\partial R_H(\bar{x},\bar{u})$ in Proposition 2.5 gives us

$$(-\bar{w},\bar{v}) \in con\,\big\{ (-w,v) \,\big|\, v = \bar{v},\ (-w,\bar{v}) \in \partial H(\bar{x},\bar{p}) \big\}.$$

This implies that $-\bar{w} \in con\, S_H(\bar{x},\bar{v},\bar{p})$ as required. □

The proof of Theorem 3.3 discloses an additional property of the mappings S_L and S_H which is worth noting. Recall that a set-valued mapping is *locally bounded* at a given point if some neighborhood of that point has bounded image under the mapping.

PROPOSITION 3.4. *Under the hypothesis of Theorem 3.3, the mappings S_L and S_H are locally bounded at $(\bar{x},\bar{v},\bar{p})$. The same is true also of the mappings $(x,v,p)\mapsto con\, S_L(x,v,p)$ and $(x,v,p)\mapsto con\, S_H(x,v,p)$, which in this case must, like S_L and S_H, have closed graph relative to some neighborhood of $(\bar{x},\bar{v},\bar{p})$.*

Thus, whenever $(x^\nu,v^\nu,p^\nu)\to(\bar{x},\bar{v},\bar{p})$ and $w^\nu\in con\, S_L(x^\nu,v^\nu,p^\nu)$, the sequence $\{w^\nu\}$ must be bounded, and all of its cluster points must belong to $con\, S_L(\bar{x},\bar{v},\bar{p})$; likewise with S_L replaced by S_H.

Proof. This comes from the observation in the proof of Theorem 3.3 that when $(x,\ v{+}p)$ belongs to the neighborhood of $(\bar{x},\ \bar{v}{+}\bar{p})$ on which R_H is Lipschitz continuous, we have $S_L(x,v,p)\subset\big\{w\,\big|\,(-w,v)\in\bar{\partial}R_L(x,v)\big\}$. The mapping $\bar{\partial}R_L$ is known to be locally bounded on such a neighborhood. The local boundedness of S_L along with the closedness of its graph (Proposition 2.3) ensures the same properties of the mapping $con\, S_L$. The case of S_H follows by symmetry. □

In conclusion we summarize how the results we have obtained fit together to produce the main result stated in Section 1.

Proof of Theorem 1.1. Let $L_t=L(t,\cdot,\cdot)$ and $H_t=H(t,\cdot,\cdot)$. The assumption that L has the epi-continuity property along the arc $x(\cdot)$ puts us for almost every t in the picture of L_t and H_t satisfying the epi-continuity assumption of Section 2 relative to some open set $O(t)$ containing $x(t)$. Then by Proposition 2.2, if either $\big(w,p(t)\big)\in\partial L_t\big(x(t),\dot{x}(t)\big)$ or $\big(-w,\dot{x}(t)\big)\in\partial H_t\big(x(t),p(t)\big)$ we have $p(t)\in\partial_v L_t\big(x(t),\dot{x}(t)\big)$ and $\dot{x}(t)\in\partial_p H_t\big(x(t),p(t)\big)$. Thus, the Euler-Lagrange condition (1.10) and the Hamiltonian condition (1.12) automatically give (1.11) and (1.13). The equivalence of (1.10) and (1.12) follows from Theorem 3.3 whenever the regularized function R_{L_t} happens to be Lipschitz continuous around $\big(x(t),\dot{x}(t)\big)$ for almost every t, or equivalently, the function R_{H_t} is Lipschitz continuous around $\big(x(t),p(t)\big)$ for almost every t. Proposition 2.6 shows that such cases occur under the additional assumptions furnished in Theorem 1.1. □

REFERENCES

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

[2] R. T. Rockafellar, "Conjugate convex functions in optimal control and the calculus of variations," J. Math. Analysis Appl. 32 (1970), 174–222.

[3] R. T. Rockafellar, "Existence and duality theorems for convex problems of Bolza," Trans. Amer. Math. Soc. 159 (1971), 1–40.

[4] R. T. Rockafellar, "Generalized Hamiltonian equations for convex problems of Lagrange," Pacific J. Math. 33 (1970), 411–428.

[5] F. H. Clarke, "Generalized gradients and applications," Trans. Amer. Math. Soc. 205 (1975), 247–262.

[6] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, 1983; reprinted as Vol. 5 of *Classics in Applied Mathematics*, SIAM (Philadelphia), 1990.

[7] P. D. Loewen, *Optimal Control via Nonsmooth Analysis*, CRM Proceedings & Lecture Notes 2, Amer. Math. Soc. (Providence), 1993.

[8] F. H. Clarke, "The Euler-Lagrange differential inclusion," J. Diff. Eqns. 19 (1975), 80–90.

[9] F. H. Clarke, "Extremal arcs and extended Hamiltonian systems," Trans. Amer. Math. Soc. 64 (1977), 349–367.

[10] F. H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*, CBMS-NSF Regional Conference Series in Appl. Math. 57, SIAM Publications, 1989.

[11] P. D. Loewen and R. T. Rockafellar, "The adjoint arc in nonsmooth optimization," Trans. Amer. Math. Soc. 325 (1991), 39–72.

[12] R. T. Rockafellar, "Dualization of subgradient conditions for optimality," Nonlinear Analysis Th. Meth. Appl. 20 (1993), 627–646.

[13] B. S. Mordukhovich, "Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems," Soviet Math. Dok. 22 (1980), 526–530.

[14] B. S. Mordukhovich, "Optimization and approximation of differential inclusions," Cybernetics 24 (1988), 781–788.

[15] B. S. Mordukhovich, "On variational analysis of differential inclusions," in *Optimization and Nonlinear Analysis* (A. D. Ioffe et al., eds.), Longman, 1992, 199–213.

[16] A. D. Ioffe, "Nonsmooth subdifferentials: their calculus and applications," in *Proceedings of the 1st World Congress of Nonlinear Analysist*, de Grueter (Berlin), 1995.

[17] P. D. Loewen and R. T. Rockafellar, "Optimal control of unbounded differential inclusions," SIAM J. Control Opt. 32 (1994), 442–470.

[18] B. S. Mordukhovich, "Discrete approximations and refined Euler-Lagrange conditions for nonconvex differential inclusions," SIAM J. Control Opt., to appear.

[19] A. D. Ioffe and R. T. Rockafellar, "The Euler and Weierstrass conditions for nonsmooth variational problems," to appear.

[20] R. J. Wijsman, "Convergence of sequences of convex sets, cones, and function, II," Trans. Amer. Math. Soc. 123 (1966), 32–45.

[21] R. J-B Wets, "Convergence of convex functions, variational inequalities, and convex optimization problems," in *Variational Inequalities and Complementarity Problems* (R. W. Cottle et al., eds.), Wiley, 1980, 376–403.

[22] G. Salinetti and R. J-B Wets, "On the convergence of convex sets in finite dimensions," SIAM Review 21 (1979), 18–33.

[23] H. Attouch, *Variational Convergence of Functions and Operators*, Pitman, 1984.

[24] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," Bull. Soc. Math. France 93 (1965), 273–299.

[25] R. T. Rockafellar, "Extensions of subgradient calculus with applications to optimization," Nonlinear Analysis Th. Meth. App. 7 (1985), 665–698.

[26] R. T. Rockafellar, "Clarke's tangent cones and the boundaries of closed sets in $\mathbb{R}^n$," Nonlinear Analysis Th. Meth. Appl. 3 (1979), 145–154.

[27] L. McLinden and R. C. Bergstrom, "Preservation of convergence of convex sets and functions," Trans. Amer. Math. Soc. 268 (1981), 127–142.