DUALIZATION OF SUBGRADIENT CONDITIONS FOR OPTIMALITY

R. T. Rockafellar*

Abstract. A basic relationship is derived between generalized subgradients of a given function, possibly nonsmooth and nonconvex, and those of a second function obtained from it by partial conjugation. Applications are made to the study of multiplier rules in finite-dimensional optimization and to the theory of the Euler-Lagrange condition and Hamiltonian condition in nonsmooth optimal control.

Keywords. Subgradients, nonsmooth analysis, Lagrange multiplier rules, Euler-Lagrange condition, Hamiltonian condition, optimality.

June, 1991

* This work was supported in part by the National Science Foundation under grant DMS–8819586 at the University of Washington, Seattle.

1. Introduction

Optimality conditions can usually be approached from more than one angle, and questions often arise as to which of the forms that may be derived is the stronger or more general. This issue has been troublesome especially in connection with techniques of nonsmooth analysis, whose unprecedented power to handle situations beyond those previously manageable has sometimes outstripped a full understanding of the consequences. In optimal control, for instance, robust generalizations of both the Euler-Lagrange equation and the Hamiltonian equation in the classical calculus of variations have been derived in the absence of smoothness, but in contrast to the smooth case the two appear to be different, neither subsuming the other. A similar puzzle has come up in the study of broad "multiplier rules" in the context of perturbational models in finite-dimensional optimization. In both areas the mystery revolves around what happens to the subgradients of a function when the function is dualized in part by applying the Legendre-Fenchel transform in certain variables where convexity is present, but not in others.

Consider a proper function $f : I\!R^n \times I\!R^m \to \overline{I\!R}$ such that $f(x, w)$ is lower semicontinuous (lsc) in (x, w) and convex in w. The Legendre-Fenchel transform can be taken in the w component to get a function

$$g(x, y) := \sup_{w \in \boldsymbol{R}^m} \{ \langle y, w \rangle - f(x, w) \}, \tag{1.1}$$

which in turn will yield

$$f(x, w) = \sup_{y \in \boldsymbol{R}^m} \{ \langle y, w \rangle - g(x, y) \}. \tag{1.2}$$

What relationships exist between the subgradients of f, in one general sense or another, and those of g?

Connections are already known in cases where f is smooth or fully convex, and these provide guidelines on what might be expected. In this paper, after some elaboration of the machinery of nonsmooth analysis, we establish relationships for nonsmooth, nonconvex functions f under appropriate assumptions. Moving on to applications, we develop multiplier rules of great scope which can be written in either of two forms, one involving a generalized Lagrangian function. Finally, we verify that in a broad situation in optimal control and the calculus of variations the Hamiltonian condition on optimal arcs implies a version of the Euler-Lagrange condition, and that the two are often equivalent.

In the smooth case the following fact is well known for functions f that are of class $\mathcal{C}^2$ and such that the Hessian matrix $\nabla^2_{ww} f(x, w)$ is always nonsingular. Usually it is established for such functions by viewing the calculation of $g(x,y)$ in (1.1) in terms of solving the equation $y - \nabla_w f(x,w) = 0$ for w as an expression in (x,y) and then substituting this into $\langle y,w\rangle - f(x,w)$ (the original Legendre transform); the implicit mapping theorem, when applied to the equation, yields the derivatives in question. However, instead of relying on such a traditional argument, we give a proof that bypasses the implicit mapping theorem and requires only first-derivative assumptions. In this way we extend a result in Rockafellar [1, Thm. 26.6] for smooth, purely convex functions conjugate to each other—where no x component is present.

Theorem 1.1. *Suppose f is finite and continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^m$ with $f(x,w)$ strictly convex in w. If g is finite everywhere, then g likewise is continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^m$ with $g(x,y)$ strictly convex in y, and one has*

$$(z,y) = \nabla f(x,w) \quad \Longleftrightarrow \quad (-z,w) = \nabla g(x,y), \tag{1.3}$$

in which case $g(x,y) = \langle y,w\rangle - f(x,w)$.

Proof. For each x the finite convex functions $f(x,\cdot)$ and $g(x,\cdot)$ are conjugate to each other. The strict convexity of f corresponds to the differentiability of g, whereas the differentiability of f corresponds to the strict convexity of g, cf. [1, Thm. 26.3]. The finiteness and strict convexity of both functions insures that the maximum in (1.1) or (1.2) is always attained at a unique point. The mapping $M_x : w \mapsto \nabla_w f(x,w)$ is thus one-to-one from $\mathbb{R}^n$ onto $\mathbb{R}^n$ with continuous inverse, this inverse being the mapping $y \mapsto \nabla_y g(x,y)$, cf. [1, Thm. 26.6]. We shall subsequently require the continuity of $\nabla_y g(x,y)$ not only with respect to y for fixed x, as just asserted, but with respect to (x,y). To establish this, we only have to verify that $g(x,y)$ is continuous in x for fixed y, since the desired continuity property then follows from [1, Thms. 24.5 and 25.1]. For a fixed vector $\bar{y}$ let

$$\varphi(x) := -g(x,\bar{y}) = \inf_w \psi_x(w) \;\text{ for }\; \psi_x(w) := f(x,w) - \langle x,\bar{y}\rangle.$$

We wish to show that φ is continuous at a point $\bar{x}$. From our assumptions, $\psi_x(w)$ converges to $\psi_{\bar{x}}(w)$ as $x \to \bar{x}$. This implies by [1, Theorem 10.8] that as $x \to \bar{x}$ the function ψ_x converges uniformly to $\psi_{\bar{x}}$ on all bounded sets. Let $\bar{w}$ be the unique point minimizing $\psi_{\bar{x}}$, and for any $\varepsilon > 0$ let $W := \{w \mid \psi_{\bar{x}}(w) \le \varphi(\bar{x}) + 2\varepsilon\}$, a bounded set by [1, Cor. 8.7.1] because the set $\{w \mid \psi_{\bar{x}}(w) \le \varphi(\bar{x})\}$ is the singleton $\{\bar{w}\}$. We can therefore find a neighborhood X of $\bar{x}$ such that $|\psi_x(w) - \psi_{\bar{x}}(w)| < \varepsilon$ when $x \in X$ and $w \in W$. Then for all $x \in X$ the set $\{w \mid \psi_{\bar{x}}(w) \le \varphi(\bar{x}) + \varepsilon\}$ must be included within W and contain $\bar{w}$, so

$$\varphi(\bar{x}) + \varepsilon \ge \varphi(x) = \inf_{w\in W} \psi_x(w) \ge \inf_{w\in W} \{\, \psi_{\bar{x}}(w) - \varepsilon \,\} = \varphi(\bar{x}) - \varepsilon.$$

Thus, $|\varphi(x) - \varphi(\bar{x})| \le \varepsilon$ when $x \in X$, and the continuity of $\nabla_y g(x,y)$ in (x,y) is assured.

Another way of writing (1.1), as we now know, is

$$g(x,y) = \max_{w \in \boldsymbol{R}^n} h(x,y,w) \text{ with } h(x,y,w) := \langle y,w \rangle - f(x,w), \tag{1.4}$$

where the maximum is attained uniquely at the point $w = \nabla_y g(x,y)$. Fix any $\widehat{x}$ and $\widehat{y}$ and let $\widehat{w} = \nabla_y g(\widehat{x},\widehat{y})$. The claim to be verified is that the joint gradient $\nabla g(\widehat{x},\widehat{y}) = \nabla_{x,y} g(\widehat{x},\widehat{y})$ exists and equals $\big(-\nabla_x f(\widehat{x},\widehat{w}), \nabla_w f(\widehat{x},\widehat{w})\big)$, which is the vector $\nabla_{x,y} h(\widehat{x},\widehat{y},\widehat{w})$. This vector depends continuously on the choice of $(\widehat{x},\widehat{y})$ through the continuous differentiability of h and the continuity of the mapping $\nabla_y g$, so we will be able to conclude that g is not only differentiable but continuously differentiable.

Let $w^\nu = \nabla_y g(\widehat{x} + t^\nu \xi^\nu, \widehat{y} + t^\nu \eta^\nu)$ for arbitrary sequences $\xi^\nu \to \xi$, $\eta^\nu \to \eta$ and $t^\nu \searrow 0$. Then $w^\nu \to \widehat{w}$. We have $g(\widehat{x},\widehat{y}) = h(\widehat{x},\widehat{y},\widehat{w}) \ge h(\widehat{x},\widehat{y},w^\nu)$, whereas $g(\widehat{x} + t^\nu \xi^\nu, \widehat{y} + t^\nu \eta^\nu) = h(\widehat{x}+t^\nu \xi^\nu, \widehat{y}+t^\nu \eta^\nu, w^n u) \ge h(\widehat{x}+t^\nu \xi^\nu, \widehat{y}+t^\nu \eta^\nu, w)$. These inequalities furnish the estimates

$$\begin{aligned}
&\liminf_{\nu\to\infty} \Big[g(\widehat{x} + t^\nu \xi^\nu, \widehat{y} + t^\nu \eta^\nu) - g(\widehat{x},\widehat{y}) \Big] \Big/ t^\nu \\
&\qquad \ge \liminf_{\nu\to\infty} \Big[h(\widehat{x} + t^\nu \xi^\nu, \widehat{y} + t^\nu \eta^\nu, \widehat{w}) - h(\widehat{x},\widehat{y},\widehat{w}) \Big] \Big/ t^\nu = \Big\langle \nabla_{x,y} h(\widehat{x},\widehat{y},\widehat{w}), (\xi,\eta) \Big\rangle, \\
&\limsup_{\nu\to\infty} \Big[g(\widehat{x} + t^\nu \xi^\nu, \widehat{y} + t^\nu \eta^\nu) - g(\widehat{x},\widehat{y}) \Big] \Big/ t^\nu \\
&\qquad \le \limsup_{\nu\to\infty} \Big[h(\widehat{x} + t^\nu \xi^\nu, \widehat{y} + t^\nu \eta^\nu, w^\nu) - h(\widehat{x},\widehat{y},w^\nu) \Big] \Big/ t^\nu = \Big\langle \nabla_{x,y} h(\widehat{x},\widehat{y},\widehat{w}), (\xi,\eta) \Big\rangle,
\end{aligned}$$

where the last inequality uses the fact that $w^\nu \to \widehat{w}$ along with the mean value theorem:

$$\Big[h(\widehat{x} + t^\nu \xi^\nu, \widehat{y} + t^\nu \eta^\nu, w^\nu) - h(\widehat{x},\widehat{y},w^\nu) \Big] / t^\nu = \Big\langle \nabla_{x,y} h(\widehat{x} + \tau^\nu \xi^\nu, \widehat{y} + \tau^\nu \eta^\nu, w^\nu), (\xi,\eta) \Big\rangle$$

for some $\tau^\nu \in (0, t^\nu)$. The estimates show that $\big[g(\widehat{x}+t^\nu \xi^\nu, \widehat{y}+t^\nu \eta^\nu) - g(\widehat{x},\widehat{y}) \big] / t^\nu$ converges to $\big\langle \nabla_{x,y} h(\widehat{x},\widehat{y},\widehat{w}), (\xi,\eta) \big\rangle$, as required. □

The fully convex case, where $f(x,w)$ is not just convex in w for fixed x but convex in (x,w), is characterized by $g(x,y)$ being concave in x as well as convex in y [1, Thm. 33.1]. The subgradient sets $\partial_y g(x,y)$ and $\partial_x(-g)(x,y)$ both make sense then in the framework of convex analysis, and it is appropriate to compare them with subgradient sets $\partial f(x,w)$.

Theorem 1.2. *If $f(x,w)$ is convex in (x,w), one has*

$$(z,y) \in \partial f(x,w) \quad \Longleftrightarrow \quad z \in \partial_x(-g)(x,y), \quad w \in \partial_y g(x,y), \tag{1.5}$$

in which case $g(x,y) = \langle y,w \rangle - f(x,w)$.

Proof. This is part of [1, Thm. 37.5]. □

Our goal is to extend Theorems 1.1 and 1.2 to possibly nonsmooth or nonconvex functions. For this purpose several concepts of subgradient are available, and it is crucial to employ the right one. Subgradients in sense of Clarke [2][3] will not give the desired result in all cases, due to the operation of convexification which is built into their definition—unless the situation being treated is one of the many where this convexification happens to be superfluous. Rather, we must focus on the set of basic subgradients generated before the convexification, which has played a considerable role in the development of rules of subdifferential calculus for Clarke subgradients, e.g. in Rockafellar [4][5], but has been emphasized by Mordukhovich also as a direct vehicle for the statement of optimality conditions, cf. [6][7].

2. Subgradient Background

Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be lower semicontinuous. A vector $z \in \mathbb{R}^n$ is called a *proximal subgradient* of h at $\bar{x}$ if $h(\bar{x})$ is finite and for some $r \geq 0$ and $\delta > 0$ one has

$$h(x) \geq h(\bar{x}) + \langle z, x - \bar{x} \rangle - \tfrac{1}{2} r |x - \bar{x}|^2 \text{ when } |x - \bar{x}| \leq \delta. \tag{2.1}$$

On the other hand, z is a *basic subgradient* if there are sequences $x^\nu \to \bar{x}$ and $z^\nu \to z$ such that z^ν is a proximal subgradient of h at x^ν and $h(x^\nu) \to h(\bar{x})$. Finally, z is a *horizon subgradient* if the latter holds except 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 this paper the notation will be used that

$$\begin{aligned} \partial h(\bar{x}) &= \text{ set of basic subgradients (no convexification)} \\ \partial^\infty h(\bar{x}) &= \text{ set of horizon subgradients (no convexification)} \\ \overline{\partial} h(\bar{x}) &= \text{ set of Clarke subgradients } = \operatorname{cl}\operatorname{con}\big[\partial h(\bar{x}) + \partial^\infty h(\bar{x})\big]. \end{aligned} \tag{2.2}$$

Here "con" stands for convex hull and "cl" for closure. (For more on this mode of definition, see Rockafellar [4].) It is known that h is Lipschitz continuous around $\bar{x}$ if and only if it is lower semicontinuous relative to a neighborhood of $\bar{x}$ and $\partial^\infty h(\bar{x}) = \{0\}$, and then $\overline{\partial} h(\bar{x}) = \operatorname{con} \partial h(\bar{x})$. When h is $\mathcal{C}^1$ around $\bar{x}$, the subgradient sets $\partial h(\bar{x})$ and $\overline{\partial} h(\bar{x})$ reduce to $\{\nabla h(\bar{x})\}$. When h is convex, they both coincide with the subgradient set in convex analysis. The details behind these assertions can be found in Clarke [3].

In the case of the indicator δ_C of a set $C \subset \mathbb{R}^n$ and a point $\bar{x} \in C$, one has

$$N(\bar{x}|C) = \partial \delta_C(\bar{x}), \qquad \overline{N}(\bar{x}|C) = \overline{\partial} \delta_C(\bar{x}),$$

where $N(\bar{x}|C)$ is the cone of *basic normals* to C at $\bar{x}$ and $\overline{N}(\bar{x}|C)$ is the cone of *Clarke normals* to C at $\bar{x}$, with $\overline{N}(\bar{x}|C) = \operatorname{cl}\operatorname{con} N(\bar{x}|C)$. These cones provide an alternative, more geometric, description of subgradients of a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ relative to the epigraph set $\operatorname{epi} h$ in $\mathbb{R}^{n+1}$, cf. [1],[3][4]:

$$\begin{aligned} \partial h(\bar{x}) &= \Big\{ z \,\Big|\, (z,-1) \in N\big(\bar{x}, h(\bar{x}) \,\big|\, \operatorname{epi} h\big) \Big\}, \\ \partial^\infty h(\bar{x}) &= \Big\{ z \,\Big|\, (z,0) \in N\big(\bar{x}, h(\bar{x}) \,\big|\, \operatorname{epi} h\big) \Big\}, \\ \overline{\partial} h(\bar{x}) &= \Big\{ z \,\Big|\, (z,-1) \in \overline{N}\big(\bar{x}, h(\bar{x}) \,\big|\, \operatorname{epi} h\big) \Big\}. \end{aligned} \tag{2.3}$$

When C is convex, $N(\bar{x}|C)$ and $\overline{N}(\bar{x}|C)$ are the usual normal cone of convex analysis.

Sometimes it is desirable to speak of subgradients of functions that are not lower semicontinuous. Such a situation has already been encountered with subgradients of convex analysis. The convex function $x \mapsto -g(x,y)$ in Theorem 1.2 need not be lsc despite the assumption that $f(x,w)$ is lsc in (x,w), cf. [1, Sec. 33], and an attempted restriction to cases where it did happen to be lsc would turn out to be quite inconvenient. Likewise, the expression $-g(x,y)$ in our general framework cannot be guaranteed to be lsc with respect to x, not to mention with respect to (x,y), without substantial pain, although it may well have such a property in particular applications. To get around this, we make a simple extension of the definitions: when $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is not necessarily lsc, we take

$$\partial h(\bar{x}) = \begin{cases} \partial \bar{h}(\bar{x}) & \text{if } h(\bar{x}) = \bar{h}(\bar{x}), \\ \emptyset & \text{if } h(\bar{x}) > \bar{h}(\bar{x}), \end{cases} \quad \text{where } \bar{h}(\bar{x}) := \liminf_{x \to \bar{x}} h(x), \tag{2.4}$$

and similarly for $\partial^\infty h(\bar{x})$ and $\overline{\partial} h(\bar{x})$. When h is convex we retain under this convention the property that the subgradient set of convex analysis coincides with the one defined through proximal limits, because (2.4) is known to hold in convex analysis, cf. [1, Thm. 23.5]. Likewise in the case of a set C that might not be closed we take

$$N(\bar{x}|C) = N(\bar{x}|\operatorname{cl} C) \text{ for any } \bar{x} \in C. \tag{2.5}$$

In understanding the circumstances in which $\partial h(\bar{x})$ and $\overline{\partial} h(\bar{x})$ may turn out to be the same, a notion of regularity introduced by Clarke in [2] has turned out to be important. A lower semicontinuous function h is *subdifferentially regular* at a point $\bar{x}$ where it is finite if $\operatorname{epi} h$ is tangentially regular at $\big(\bar{x}, h(\bar{x})\big)$; a closed set C is *tangentially regular* at $\bar{x}$ if it is closed relative to a neighborhood of $\bar{x}$ and the Clarke tangent cone and the ordinary tangent cone (Bouligand contingent cone) to C at $\bar{x}$ coincide. (The tangent cones in question are discussed in Clarke [3] but will not be dealt with directly here, so we do not review their meaning.) We extend the definitions to the case where h is not lsc, or C is not closed, in the manner just adopted. Thus, we say that h is subdifferentially regular at

$\bar{x}$ if $h(\bar{x}) = \overline{h}(\bar{x})$ and $\overline{h}$ is subdifferentially regular at $\bar{x}$, and that C is tangentially regular at a point $\bar{x} \in C$ if $\operatorname{cl} C$ is tangentially regular at $\bar{x}$.

Our main use of this kind of regularity will come through the facts stated next.

Proposition 2.1. *Consider a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $\bar{x}$ where $h(\bar{x})$ is finite. If a vector $z \in \mathbb{R}^n$ satisfies*

$$h(x) \geq h(\bar{x}) + \langle z, x - \bar{x} \rangle + o(x - \bar{x}), \tag{2.6}$$

then $z \in \partial h(\bar{x}) \subset \overline{\partial} h(\bar{x})$. Under the assumption that $\overline{\partial} h(\bar{x}) \neq \emptyset$, the converse property, namely that every $z \in \overline{\partial} h(\bar{x})$ satisfies this inequality, holds if and only if h is subdifferentially regular at $\bar{x}$. In that case $\partial h(\bar{x}) = \overline{\partial} h(\bar{x})$.

Proof. Since (2.6) implies the lower semicontinuity of h at $\bar{x}$, we may as well suppose, on the basis of our conventions, that h is lower semicontinuous everywhere. The claims all follow then from [8, Cor. 1 of Prop. 1], except that the cited result only shows that a vector z satisfying (2.6) belongs to $\overline{\partial} h(\bar{x})$. That the vector actually belongs to the basic subgradient set $\partial h(\bar{x})$ must be argued separately. The argument has been furnished by Poliquin [9, Thm. 2.1]. □

A special chain rule in subgradient calculus will be critical to some of our main results. This rule is close to one of Ward and Borwein [10, Prop. 6.7], but it does not directly require lower semicontinuity of the functions involved (a flexibility that will be needed for reasons already explained). Furthermore, in contrast to the cited result it provides a formula for the horizon subgradients of the composite function and establishes the subdifferential regularity of that function.

In the statement we denote by $\nabla F(x)$ the Jacobian matrix at x for a mapping F and by $\nabla F(x)^*$ its transpose.

Proposition 2.2. *Suppose $h(x) = h_0\big(G(x)\big)$ for a $\mathcal{C}^1$ mapping $G : \mathbb{R}^n \to \mathbb{R}^m$ and a proper, convex function $h_0 : \mathbb{R}^m \to \overline{\mathbb{R}}$, and let*

$$C_0 = \operatorname{dom} h_0, \qquad C = \operatorname{dom} h = \left\{\, x \,\middle|\, G(x) \in C_0 \,\right\}.$$

Let $\bar{x}$ be a point with $h(\bar{x}) < \infty$ such that the only vector v in the normal cone $N\big(G(\bar{x})\big|C_0\big)$ having $\nabla G(\bar{x})^* v = 0$ is $v = 0$. Then there is a neighborhood X of $\bar{x}$ such that for all $x \in X$ with $h(x)$ finite, h is subdifferentially regular at x and

$$\begin{aligned} \partial h(x) &= \left\{\, \nabla G(x)^* v \,\middle|\, v \in \partial h_0\big(G(x)\big) \,\right\}, \\ \partial^\infty h(x) &= \left\{\, \nabla G(x)^* v \,\middle|\, v \in N\big(G(x)\big|C_0\big) \,\right\} = N(x|C). \end{aligned} \tag{2.7}$$

Moreover the neighborhood X can be chosen such that the closures (lsc hulls) $\overline{h}$ of h and $\overline{h}_0$ of h_0 are related by $\overline{h}(x) = \overline{h}_0\big(G(x)\big)$ for all $x \in X$.

Before proceeding with the proof we take special note of the case of Proposition 2.2 in which h_0 is just an indicator function.

Corollary 2.3. *Suppose $C = \{\, x \,|\, G(x) \in C_0 \,\}$ for a $\mathcal{C}^1$ mapping $G : I\!R^n \to I\!R^m$ and a nonempty, convex set $C_0 \subset I\!R^m$. Let $\bar{x}$ be a point of C such that the only vector v in the normal cone $N\big(G(\bar{x})\big|C_0\big)$ having $\nabla G(\bar{x})^* v = 0$ is $v = 0$. Then there is a neighborhood X of $\bar{x}$ such that, for all $x \in X \cap C$, C is tangentially regular at x with*

$$N(x|C) = \Big\{\, \nabla G(x)^* v \,\Big|\, v \in N\big(G(x)\big|C_0\big) \,\Big\} \tag{2.8}$$

and $X \cap \operatorname{cl} C = \{\, x \in X \,|\, G(x) \in \operatorname{cl} C_0 \,\}$.

Proof of Proposition 2.2. The constraint qualification on $\bar{x}$ means the convex set C_0 cannot be separated from the affine set M that is the graph of the affine mapping $x \mapsto G(\bar{x}) + \nabla G(\bar{x})(x - \bar{x})$. Equivalently, the point $G(\bar{x}) - \nabla G(\bar{x})\bar{x}$ lies in the interior of the convex set $C_0 - \nabla G(\bar{x}) I\!R^n$. This equivalent version of the constraint qualification was shown by Ward and Borwein [10, Prop. 6.7] to yield

$$\overline{\partial} h(\bar{x}) = \Big\{\, \nabla G(\bar{x})^* v \,\Big|\, v \in \partial h_0\big(G(\bar{x})\big) \,\Big\} \tag{2.9}$$

when h_0 is lsc and therefore h is lsc. (Here $\partial h_0\big(G(\bar{x})\big) = \overline{\partial} h_0\big(G(\bar{x})\big)$ because h_0 is convex.) Any vector $z = \nabla G(\bar{x})^* v$ of the kind on the right side of (2.9) gives the inequality

$$\begin{aligned} h(x) = h_0\big(G(x)\big) &\geq h_0\big(G(\bar{x})\big) + \big\langle v,\, G(x) - G(\bar{x}) \big\rangle \\ &= h(\bar{x}) + \big\langle v, \nabla G(\bar{x})(x - \bar{x}) + o(x - \bar{x}) \big\rangle \\ &= h(\bar{x}) + \big\langle z,\, x - \bar{x} \big\rangle + o(x - \bar{x}), \end{aligned}$$

which comes out as (2.6). Thus, when h_0 is lsc and $\overline{\partial} h(\bar{x}) \neq \emptyset$, formula (2.9) implies through Proposition 2.1 that h is subdifferentially regular at $\bar{x}$. Then $\partial h(\bar{x}) = \overline{\partial} h(\bar{x})$ as well, so that the first formula in (2.7) holds. If h_0 is not lsc, these facts apply at least to $\overline{h}_0$ and the corresponding composite function $\overline{h}_0 \circ G$ (because the convex sets $\operatorname{dom} h_0$ and $\operatorname{dom} \overline{h}_0$ have the same closure and therefore the same normal cone at $\bar{x}$, cf. [1, Cor. 7.4.1]), so that if $\overline{h} = \overline{h}_0 \circ G$ on a neighborhood of $\bar{x}$ as claimed, they will apply to h_0 and h themselves under our convention (2.4).

To justify the full statement of Proposition 2.2, we must verify that the condition $\overline{\partial} h(\bar{x}) \neq \emptyset$ can be obviated in establishing subdifferential regularity at $\bar{x}$, and that the second formula in (2.7) then holds as well as the first. Further, we must prove not only that $\overline{h} = \overline{h}_0 \circ G$ on a neighborhood of $\bar{x}$ but that the conclusions obtained at $\bar{x}$ must hold equally well for all points $x \in C$ in some neighborhood of $\bar{x}$.

Our strategy for this is to argue first that Proposition 2.2 is actually equivalent to Corollary 2.3, the case where h_0 and h are the indicators δ_{C_0} and δ_C. That will get around the issue over $\overline{\partial} h(\bar{x})$ having to be assumed nonempty, because such is always true for an indicator function (the set $\overline{\partial}\delta_C(\bar{x})$ being the Clarke normal cone $\overline{N}(\bar{x}|C)$). It will also eliminate the need for working with the second formula in (2.7), since just the specialization of the first formula in (2.7) is present in Corollary 2.3. Let $\widetilde{C}$ and $\widetilde{C}_0$ be the epigraphs of h and h_0, and consider the mapping $\widetilde{G} : I\!R^n \times I\!R \to I\!R^m \times I\!R$ defined by $\widetilde{G}(x, \alpha) = (G(x), \alpha)$. Clearly $\widetilde{G}$ is $\mathcal{C}^1$, the set $\widetilde{C}_0$ is nonempty and convex, and we have $\widetilde{C} = \{ (x, \alpha) \,|\, \widetilde{G}(x, \alpha) \in \widetilde{C}_0 \}$. In fact Proposition 2.2 can be identified with the content of Corollary 2.3 as obtained in the case of this choice of elements, cf. (2.3).

We can focus henceforth on proving Corollary 2.3, or rather, the assertions in Corollary 2.3 not covered through specialization of the facts already obtained. Specifically, we aim at proving the claim about $\operatorname{cl} C$ and the claim that when the constraint qualification holds at $\bar{x}$ it also holds at all points $x \in C$ in some neighborhood of $\bar{x}$.

We shall demonstrate in several steps that the setting can be reduced to the case where $\operatorname{int} C_0 \neq \emptyset$. The setting is coordinate-free, in the sense that the assumptions and claims persist under any affine change of coordinates in $I\!R^m$, or on the other hand, under any smooth, local change of coordinates in $I\!R^n$. Taking advantage of an affine change of coordinates in $I\!R^m$, we can reduce the argument to the case of $C_0 = C_1 \times \{0\}$ in $I\!R^{m_1} \times I\!R^{m_2}$, where $m_1 = \dim C_0$ and $m_2 = m - m_1$; then C_1 has nonempty interior in $I\!R^{m_1}$. Correspondingly we have a component representation $G(x) = \big(G_1(x), G_2(x)\big)$ with $G_1(\bar{x}) \in C_1$ and $G_2(\bar{x}) = 0$. The constraint qualification on $\bar{x}$ translates to the condition that the only vector pair $(v_1, v_2) \in I\!R^{m_1} \times I\!R^{m_2}$ with $v_1 \in N_{C_1}\big(G_1(\bar{x})\big)$ and $\nabla G_1(\bar{x})^* v_1 + \nabla G_2(\bar{x})^* v_2 = 0$ is $(0, 0)$.

In particular this condition implies that $\nabla G_2(\bar{x})$ has rank m_2. A local, smooth change of coordinates around $\bar{x}$ can therefore reduce the equation $G_2(x) = 0$ to a linear system. There is no loss of generality then in writing each $x \in I\!R^n$ as $(x_1, x_2) \in I\!R^{n-m_2} \times I\!R^{m_2}$ and supposing the equation $G_2(x) = 0$ to mean that the x_2 component of x vanishes. In that case the vectors of the form $\nabla G_2(\bar{x})^* v_2$ can be identified with the vectors in $I\!R^n$ whose initial $n - m_2$ coordinates vanish. Thus, without loss of generality we can focus on the case of

$$C = \{ (x_1, x_2) \,|\, x_2 = 0,\ G_1(x_1, 0) \in C_1 \},$$

where the only vector v_1 for which the initial $n-m_2$ coordinates of $\nabla G_1(\bar{x}_1,0)^* v_1$ vanish is $v_1=0$. In the notation $\overline{G}_1(x_1)=G_1(x_1,0)$ this constraint qualification says that the only vector v_1 with $\nabla\overline{G}_1(\bar{x}_1)^* v_1=0$ is $v_1=0$. The x_2 components of vectors x no longer have a significant role, and everything hinges on the mapping $\overline{G}_1 : I\!R^{n-m_2}\to I\!R^{m_1}$ and the convex set $C_1\subset I\!R^{m_1}$ with nonempty interior. Our claims need only be verified for $\overline{G}_1$ and C_1. We may as well revert to the original claims and notation—but with the added assumption that $\operatorname{int} C_0\neq\emptyset$.

The constraint qualification on $\bar{x}$ was seen earlier to mean that the point $G(\bar{x})-\nabla G(\bar{x})\bar{x}$ lies in the interior of the convex set $C_0-\nabla G(\bar{x})I\!R^n$. This interior can now be identified as $\operatorname{int} C_0-G(\bar{x})I\!R^n$ [1, Cor. 6.6.2]. The constraint qualification is thus: there exists ξ with $G(\bar{x})+\nabla G(\bar{x})\xi\in\operatorname{int} C_0$. Then there is an open neighborhood X of $\bar{x}$ such that

$$\text{for all } x\in X \text{ there exists } \xi \text{ with } G(x)+\nabla G(x)\xi\in\operatorname{int} C_0. \tag{2.10}$$

In consequence, the constraint qualification holds at all $x\in C$ belonging to the neighborhood X. The existence of such a neighborhood was one of our needs.

The remainder of the proof is concerned with showing that because (2.10) holds for all $x\in X$ we have $X\cap\operatorname{cl} C=X\cap G^{-1}(\operatorname{cl} C_0)$. Since $G^{-1}(C_0)\subset\operatorname{cl} C\subset G^{-1}(\operatorname{cl} C_0)$, it will suffice in this to demonstrate that for any $x\in X$ with $G(x)\in\operatorname{cl} C_0$ but $G(x)\notin\operatorname{int} C_0$ one has $x\in\operatorname{cl} C$. For a point x as hypothesized, it is clear that $\nabla G(x)\xi\neq 0$, inasmuch as $G(x)+\nabla G(x)\xi\in\operatorname{int} C_0$. For scalars $t^\nu\searrow 0$ the vectors $\big[G(x+t^\nu\xi)-G(x)\big]/t^\nu$ converge to $\nabla G(x)\xi$, so that the point $u^\nu:=G(x)+\big[G(x+t^\nu\xi)-G(x)\big]/t^\nu$ lies in $\operatorname{int} C_0$ for all ν sufficiently large. Then $G(x+t^\nu\xi)$ belongs to $\operatorname{int} C_0$ as well, because $G(x+t^\nu\xi)=(1-t^\nu)G(x)+t^\nu u^\nu$ with $G(x)\in\operatorname{cl} C_0$, cf. [1, Thm. 6.1]. From the fact that $x+t^\nu\xi\in G^{-1}(\operatorname{int} C_0)\subset C$ we conclude that $x\in\operatorname{cl} C$, as we wanted. □

3. Dualization Theorems

Our challenge is to determine in the framework of this theory what can be said about subgradients of the partially conjugated functions f and g in (1.1) and (1.2). Along with the relationship between subgradients, we are interested in the corresponding values of f and g. For this we record at the outset the fact from convex analysis that

$$g(x,y)=\langle y,w\rangle-f(x,y) \quad\Longleftrightarrow\quad y\in\partial_w f(x,w) \quad\Longleftrightarrow\quad w\in\partial_y g(x,y). \tag{3.1}$$

First we prove a general inclusion which goes in one direction and helps to identify situations where conditions stated in terms of Clarke subgradients of g imply conditions in terms of basic subgradients of f.

Theorem 3.1. *For any pair of functions f and g satisfying the partial conjugacy relations (1.1) and (1.2), with f proper and lsc on $\mathbb{R}^n \times \mathbb{R}^m$,*

$$(-z,w) \in -\partial(-g)(x,y) \quad\Longrightarrow\quad (z,y) \in \partial f(x,w), \quad y \in \partial_w f(x,w). \tag{3.2}$$

When g is Lipschitz continuous on a neighborhood of (x,y), one further has

$$\begin{aligned} &\overline{\partial} g(x,y) \subset \operatorname{con} S(x,y), \ \text{ where} \\ &\qquad S(x,y) = \big\{ (-z,w) \,\big|\, (z,y) \in \partial f(x,w),\ y \in \partial_w f(x,w) \big\}, \end{aligned} \tag{3.3}$$

so that if $S(x,y)$ happens to be convex, one has

$$(-z,w) \in \overline{\partial} g(x,y) \quad\Longrightarrow\quad (z,y) \in \partial f(x,w), \quad y \in \partial_w f(x,w). \tag{3.4}$$

The proof of this theorem will utilize an elementary fact which we state as a lemma, because it is not widely known but has independent interest.

Lemma 3.2. *Let φ and ψ be functions from $\mathbb{R}^n \times \mathbb{R}^m$ to $\overline{\mathbb{R}}$ such that $\varphi(x,\cdot)$ and $\psi(x,\cdot)$ are convex functions conjugate to each other for each x. Then the saddle point condition*

$$\psi(x,\bar{y}) \le \psi(\bar{x},\bar{y}) \le \psi(\bar{x},y) \ \text{ for all } x \text{ and } y$$

is equivalent to the condition

$$\varphi(x,w) \ge \varphi(\bar{x},0) + \langle \bar{y}, w\rangle \ \text{ for all } x \text{ and } w, \tag{3.5}$$

and these two conditions imply $\psi(\bar{x},\bar{y}) = -\varphi(\bar{x},0)$.

Proof. The saddle point condition can be written as $\sup \psi(\cdot,\bar{y}) = \inf \psi(\bar{x},\cdot)$, where by virtue of the conjugacy between $\psi(x,\cdot)$ and $\varphi(x,\cdot)$ we have

$$\begin{aligned} \inf \psi(\bar{x},\cdot) &= -\sup\nolimits_y \big\{ \langle y, 0\rangle - \psi(\bar{x},y) \big\} = -\varphi(\bar{x},0), \\ \sup \psi(\cdot,\bar{y}) &= \sup\nolimits_x \Big[\sup\nolimits_w \big\{ \langle \bar{y}, w\rangle - \varphi(x,w) \big\} \Big]. \end{aligned}$$

The saddle point condition is therefore equivalent to having $-\varphi(\bar{x},0) = \psi(\bar{x},\bar{y})$ and

$$(\bar{x},0) \in \operatorname*{argmin}_{x,\,w} \big\{ \varphi(x,w) - \langle \bar{y}, w\rangle \big\}. \tag{3.6}$$

Conditions (3.6) and (3.5) mean the same thing. □

Proof of Theorem 3.1. First consider the situation where $(\bar{z}, -\bar{w})$ is a proximal subgradient of $-g$ at $(\bar{x},\bar{y})$. We shall verify that $(\bar{z},\bar{y})$ is a proximal subgradient of f at $(\bar{x},\bar{w})$, and that $g(\bar{x},\bar{y}) = \langle \bar{y},\bar{w}\rangle - f(\bar{x},\bar{w})$. By assumption there exist $r>0$ and neighborhoods X of $\bar{x}$ and Y of $\bar{y}$ such that

$$-g(x,y) \geq -g(\bar{x},\bar{y}) + \langle \bar{z},\, x-\bar{x}\rangle - \langle \bar{w},\, y-\bar{y}\rangle - \tfrac{1}{2}r|x-\bar{x}|^2 - \tfrac{1}{2}r|y-\bar{y}|^2$$

when $x\in X$ and $y\in Y$. In particular this gives

$$g(x,\bar{y}) \leq g(\bar{x},\bar{y}) - \langle \bar{z},\, x-\bar{x}\rangle + \tfrac{1}{2}r|x-\bar{x}|^2 \text{ when } x\in X, \tag{3.7}$$

$$g(\bar{x},y) \leq g(\bar{x},\bar{y}) + \langle \bar{w},\, y-\bar{y}\rangle + \tfrac{1}{2}r|y-\bar{y}|^2 \text{ when } y\in Y. \tag{3.8}$$

The function $g(\bar{x},\cdot)$ is not only finite at $\bar{y}$ and convex but, by (3.8), less than ∞ on a neighborhood of $\bar{y}$. Thus, $\partial_y g(\bar{x},\bar{y})$ is nonempty [1, Thm. 23.4] and included in the set of subgradients at $\bar{y}$ of the function on the right side of (3.8), which is the singleton set $\{\bar{w}\}$. Hence $\partial_y g(\bar{x},\bar{y}) = \{\bar{w}\}$ (and $g(\bar{x},\cdot)$ is actually differentiable at $\bar{y}$ with $\bar{w} = \nabla_y g(\bar{x},\bar{y})$). It follows that

$$g(\bar{x},y) \geq g(\bar{x},\bar{y}) + \langle \bar{w},\, y-\bar{y}\rangle \text{ for all } y\in \mathbb{R}^m, \tag{3.9}$$

and also through (3.1) that $g(\bar{x},\bar{y}) = \langle \bar{y},\bar{w}\rangle - f(\bar{x},\bar{w})$. Define $\psi : \mathbb{R}^n\times\mathbb{R}^m \to \overline{\mathbb{R}}$ by

$$\psi(x,y) := \begin{cases} g(x,y) + \langle \bar{z},\, x-\bar{x}\rangle - \langle \bar{w},\, y-\bar{y}\rangle - \frac{1}{2}r|x-\bar{x}|^2 & \text{if } x\in X,\\ -\infty & \text{if } x\notin X. \end{cases}$$

We have from (3.7) and (3.9) the saddle point condition: $\psi(x,\bar{y}) \leq \psi(\bar{x},\bar{y}) \leq \psi(\bar{x},y)$ for all x and y. Here for each x such that $\psi(x,\cdot)\not\equiv -\infty$ the function $\psi(x,\cdot)$ is lsc, proper and convex. Denote by $\varphi(x,\cdot)$ the convex function conjugate to $\psi(x,\cdot)$. For $x\notin X$ we have $\varphi(x,\cdot)\equiv\infty$, whereas for $x\in X$ we have

$$\begin{aligned}\varphi(x,w) &= \sup_y \left\{ \langle y,w\rangle - \psi(x,y) \right\}\\ &= -\langle \bar{z},\, x-\bar{x}\rangle + \tfrac{1}{2}r|x-\bar{x}|^2 - \langle \bar{y},\bar{w}\rangle + \sup_y \left\{ \langle y,\, \bar{w}+w\rangle - g(x,y) \right\}\\ &= -\langle \bar{z},\, x-\bar{x}\rangle + \tfrac{1}{2}r|x-\bar{x}|^2 - \langle \bar{y},\bar{w}\rangle + f(x,\bar{w}+w).\end{aligned}$$

We are able then with the aid of Lemma 3.2 to translate the saddle point condition on ψ at $(\bar{x},\bar{y})$ into the condition that

$$\begin{aligned}(\bar{x},0) &\in \operatorname*{argmin}_{x\in\mathbf{R}^n,\, w\in\mathbf{R}^m} \varphi(x,w)\\ &= \operatorname*{argmin}_{x\in X,\, w\in\mathbf{R}^m} \left\{ f(x,\bar{w}+w) - \langle \bar{z},\, x-\bar{x}\rangle + \tfrac{1}{2}r|x-\bar{x}|^2 - \langle \bar{y},w\rangle \right\},\end{aligned}$$

or equivalently

$$(\bar{x},\bar{w}) \in \operatorname*{argmin}_{x\in X,\, w\in\mathbf{R}^m} \left\{ f(x,w) - \langle \bar{z},\, x-\bar{x}\rangle + \tfrac{1}{2}r|x-\bar{x}|^2 - \langle \bar{y},\, w-\bar{w}\rangle \right\}.$$

This condition means that

$$f(x,w) \geq f(\bar{x},\bar{w}) + \langle \bar{z},\, x-\bar{x}\rangle + \langle \bar{y},\, w-\bar{w}\rangle - \tfrac{1}{2}r|x-\bar{x}|^2 \ \text{ for all } x \in X,\, w \in I\!R^m.$$

In particular we see that $(\bar{z},\bar{y})$ is a proximal subgradient of f at $(\bar{x},\bar{w})$.

Consider now the situation where merely $(\bar{z},-\bar{w}) \in \partial(-g)(\bar{x},\bar{y})$. By definition, $-g$ must be finite and lower semicontinuous at $(\bar{x},\bar{y})$, and there must there exist sequences of points $z^\nu \to \bar{z}$, $w^\nu \to \bar{w}$, $x^\nu \to \bar{x}$ and $y^\nu \to \bar{y}$ with $g(x^\nu,y^\nu) \to g(\bar{x},\bar{y})$, such that $(z^\nu,-w^\nu)$ is a proximal subgradient of $(-g)$ at (x^ν,y^ν). Then, by the argument presented so far, (z^ν,y^ν) is a proximal subgradient of f at (x^ν,w^ν) and

$$g(x^\nu,y^\nu) = \langle y^\nu,w^\nu\rangle - f(x^\nu,w^\nu).$$

Taking the limit in this equation and utilizing the assumption that f is lsc, we obtain

$$g(\bar{x},\bar{y}) \leq \langle \bar{y},\bar{w}\rangle - f(\bar{x},\bar{w}).$$

The opposite inequality always holds by conjugacy (cf. (1.2)), so $g(\bar{x},\bar{y}) = \langle \bar{y},\bar{w}\rangle - f(\bar{x},\bar{w})$ and consequently $\bar{y} \in \partial_w f(\bar{x},\bar{w})$ by (3.1). Further, we have $f(x^\nu,w^\nu) \to f(\bar{x},\bar{w})$ along with $(x^\nu,w^\nu) \to (\bar{x},\bar{w})$. Inasmuch as (z^ν,y^ν) is a proximal subgradient of f at (x^ν,w^ν) and $(z^\nu,y^\nu) \to (\bar{z},\bar{y})$, it follows that $(\bar{x},\bar{y})$ is a basic subgradient of f at $(\bar{x},\bar{y})$, i.e., $(\bar{z},\bar{y}) \in \partial f(\bar{x},\bar{w})$.

This argument proves (3.2). If g is Lipschitz continuous on a neighborhood of (x,y) we have $\overline{\partial} g(x,y) = -\overline{\partial}(-g)(x,y)$ (through the characterization of $\overline{\partial} g(x,y)$ as the convex hull of the set of limits of gradients at nearby points of differentiability, cf. Clarke [3]). But $\overline{\partial}(-g)(x,y) = \operatorname{con}\partial(-g)(x,y)$ in the case of Lipschitz continuity. Therefore $\overline{\partial} g(x,y) = \operatorname{con}\big[-\partial(-g)(x,y)\big]$, so (3.3) is a consequence of (3.2). It is obvious that (3.3) yields (3.4) when $S(x,y)$ is convex. □

Although Theorem 3.1 supports an implication in one direction, we are able to demonstrate a symmetric relationship between subgradients of f and g when a kind of composite representation is available.

Theorem 3.3. *Suppose $f(x,w) = f_0\big(F(x),w\big)$ for a $\mathcal{C}^1$ mapping $F: I\!R^n \to I\!R^d$ and a proper, lsc, convex function $f_0 : I\!R^d \times I\!R^m \to \overline{I\!R}$, so that $g(x,y) = g_0\big(F(x),y\big)$ for the concave-convex function $g_0 : I\!R^d \times I\!R^m$ obtained by taking $g_0(u,\cdot)$ to be conjugate to $f_0(u,\cdot)$ for each $u \in I\!R^d$. Define the convex set D by*

$$D := \big\{\, u \in I\!R^d \,\big|\, f_0(u,\cdot) \not\equiv \infty \,\big\} = \big\{\, u \in I\!R^d \,\big|\, g_0(u,\cdot) \not\equiv -\infty \,\big\}.$$

Consider a point $\bar{x}$ with $F(\bar{x}) \in D$ such that the following constraint qualification is satisfied: the only vector v in the normal cone $N\big(F(\bar{x})\big|D\big)$ with $\nabla F(\bar{x})^* v = 0$ is $v = 0$. Then there is a neighborhood X of $\bar{x}$ such that when $x \in X$ one has

$$(z, y) \in \partial f(x, w) \quad \Longleftrightarrow \quad z \in \partial_x(-g)(x, y), \quad w \in \partial_y g(x, y), \tag{3.10}$$

and for all the points $x \in X$, w, y and z in this relation f is subdifferentially regular at (x, w) and $-g(\cdot, y)$ is subdifferentially regular at x (the convex function $g(x, \cdot)$ also being subdifferentially regular at y), so that

$$\partial f(x, w) = \overline{\partial} f(x, w), \qquad \partial_x(-g)(x, y) = \overline{\partial}_x(-g)(x, y), \qquad \partial_y g(x, y) = \overline{\partial}_y g(x, y). \tag{3.11}$$

Proof. Regarding f as the composite $f_0 \circ G$ with $G(x, w) := \big(F(x), w\big)$, we apply the chain rule in Proposition 2.2. We have $\nabla G(x, w) = \operatorname{diag}\big[\nabla F(x), I\big]$. The constraint qualification in Proposition 2.2, as invoked in the present context at a point $(\bar{x}, \bar{w})$, requires that there be no vector pair $(z, y) \in N\big(F(\bar{x}), \bar{w} \big| \operatorname{dom} f_0\big)$ with $(0, 0) = \big(\nabla F(\bar{x})^* z, Iy\big)$, except $(z, y) = (0, 0)$. The condition thus only concerns pairs (z, y) with $y = 0$. But if $(z, 0)$ is normal to the convex set $\operatorname{dom} f_0$ at $\big(F(\bar{x}), \bar{w}\big)$, the vector z is normal to the projected convex set D at $F(\bar{x})$. The hypothesis of the theorem excludes the existence of a vector $z \neq 0$ with this property. Therefore, the constraint qualification in Proposition 2.2 is satisfied at $(\bar{x}, \bar{w})$ regardless of the location of $\bar{w}$, as long as $f(\bar{x}, \bar{w}) < \infty$. We are able to conclude then that for some neighborhood X_1 of $\bar{x}$ the function f is subdifferentially regular at all points (x, w) such that $x \in X_1$ and $f(x, w)$ is finite, with

$$\partial f(x, w) = \Big\{ \big(\nabla F(x)^* v, y\big) \,\Big|\, (v, y) \in \partial f_0\big(F(x), w\big) \Big\}. \tag{3.12}$$

We know from Theorem 1.2 that

$$(v, y) \in \partial f_0\big(F(x), w\big) \quad \Longleftrightarrow \quad v \in \partial_u(-g_0)\big(F(x), y\big),\ w \in \partial_y g_0\big(F(x), y\big), \tag{3.13}$$

in which case $g_0\big(F(x), y\big) = \langle y, w \rangle - f_0\big(F(x), w\big)$.

Because g_0 is concave-convex, there is a convex set Y such that for each $u \in D$ the proper convex function $g_0(u, \cdot)$ has its effective domain between Y and $\operatorname{cl} Y$ and has $\partial g_0(u, y) = \emptyset$ when $y \notin Y$ [1, Thm. 34.3]. For each $y \in Y$ the convex function $-g_0(\cdot, y)$ is proper with D as its effective domain. We can view $-g(\cdot, y)$ as the composite of this function with the mapping F and again apply the chain rule in Proposition 2.2 to determine the subgradients. The constraint qualification being satisfied at $\bar{x}$, we get for all x in some neighborhood of $\bar{x}$ relative to D the formula

$$\partial_x(-g)(x, y) = \Big\{ \nabla F(x)^* v \,\Big|\, v \in \partial_u(-g_0)\big(F(x), y\big) \Big\}.$$

We also deduce that $(-g)(\cdot, y)$ is subdifferentially regular at such points x.

It only remains to combine this formula with (3.12) and (3.13). The equivalence in (3.10) is then apparent. □

Theorem 3.4. *Suppose f can be represented in the form*

$$f(x,w) = f_0\big(F(x),w\big) + \big\langle G(x),w\big\rangle \tag{3.14}$$

for $\mathcal{C}^1$ mappings $F : I\!R^n \to I\!R^d$ and $G : I\!R^n \to I\!R^m$ and a proper, lsc, convex function $f_0 : I\!R^d \times I\!R^m \to \overline{I\!R}$, so that $g(x,y) = g_0\big(F(x),\, y - G(x)\big)$ for the concave-convex function g_0 obtained by taking $g_0(u,\cdot)$ to be the conjugate to $f_0(u,\cdot)$ for each u. Let $(\bar{x},\bar{y})$ be such that $\bar{x}$ satisfies the constraint qualification in Theorem 3.3 and g is Lipschitz continuous around $(\bar{x},\bar{y})$, the latter being true in particular when g_0 is finite on a neighborhood of $\big(F(\bar{x}), \bar{y} - G(\bar{x})\big)$. Then for all (x,y) in a neighborhood of $(\bar{x},\bar{y})$ one has

$$(z,y) \in \partial f(x,w) \quad \iff \quad (-z,w) \in \overline{\partial} g(x,y), \tag{3.15}$$

with these conditions implying that $g(x,y) = \langle y,w\rangle - f(x,w)$.

Proof. The criterion for Lipschitz continuity of g in terms of g_0 stems from the fact that a concave-convex function is locally Lipschitz continuous on any open set where it is finite [1, Thm. 35.1]. We begin the proof of the main assertion of the theorem with the case where $G(x) \equiv 0$, so f and g are as in Theorem 3.3. Then in a neighborhood of $(\bar{x},\bar{y})$ we have (3.10) and (3.11). Since for fixed (x,y) the set of (z,y) satisfying

$$z \in \overline{\partial}_x(-g)(x,y), \qquad w \in \overline{\partial}_y g(x,y),$$

is convex, we also have the inclusion in (3.4) of Theorem 3.1. The issue then is simply one of showing that

$$\overline{\partial} g(x,y) \;\supset\; \big[-\partial_x(-g)(x,y)\big] \times \partial_y g(x,y)$$

when g is Lipschitz continuous around (x,y) and g as the specified representation. We may as well concentrate on the pair $(\bar{x},\bar{y})$.

Fix any $\bar{z} \in \partial_x(-g)(\bar{x},\bar{y})$ and $\bar{w} \in \partial_y g(\bar{x},\bar{y})$. We have

$$\begin{aligned} &\bar{w} \in \partial_y g_0\big(F(\bar{x}),\bar{y}\big),\\ &\bar{z} = \nabla F(\bar{x})^* \bar{v} \;\text{ for some } \bar{v} \in \partial_u(-g_0)\big(F(\bar{x}),\bar{y}\big), \end{aligned} \tag{3.16}$$

the latter stemming from the chain rule in Proposition 2.2 as already explained in the proof of Theorem 3.3. The support function of the compact, convex set $\overline{\partial} g(\bar{x},\bar{y})$ is the Clarke derivative

$$g^\circ(\bar{x},\bar{y};\xi,\eta) := \limsup_{\substack{(x,y)\to(\bar{x},\bar{y})\\ t\searrow 0}} \big[g(x+t\xi, y+t\eta) - g(\bar{x},\bar{y})\big]/t,$$

cf. [3]. To check that $(-\bar z,\bar w)\in\overline{\partial} g(\bar x,\bar y)$, it suffices therefore to show for arbitrary vectors ξ and η that for some choice of sequences $x^\nu\to\bar x$, $y^\nu\to\bar y$, $t^\nu\searrow 0$, one has

$$\limsup_{\nu\to\infty}\big[g(x^\nu+t^\nu\xi,y^\nu+t^\nu\eta)-g(x^\nu,y^\nu)\big]/t^\nu \;\geq\; -\langle\bar z,\xi\rangle+\langle\bar w,\eta\rangle. \tag{3.17}$$

This is accomplished by choosing any sequence $t^\nu\searrow 0$ and taking $x^\nu=\bar x-t^\nu\xi$ and $y^\nu\equiv\bar y$. We then have

$$\begin{aligned}\big[g(x^\nu+t^\nu\xi,y^\nu+t^\nu\eta)&-g(x^\nu,y^\nu)\big]/t^\nu\\ &=\big[g(\bar x,\bar y+t^\nu\eta)-g(\bar x-t^\nu\xi,\bar y)\big]/t^\nu\\ &=\Big[g_0\big(F(\bar x),\bar y+t^\nu\eta\big)-g_0\big(F(\bar x-t^\nu\xi),\bar y\big)\Big]\Big/t^\nu.\end{aligned} \tag{3.18}$$

The subgradient relations in (3.16) in terms of the concave-convex function g_0 give us

$$\begin{aligned}&g_0\big(F(\bar x),\bar y+t^\nu\eta\big)\geq g_0\big(F(\bar x),\bar y\big)+t^\nu\langle\bar w,\eta\rangle,\\ &g_0\big(F(\bar x-t^\nu\xi),\bar y\big)\leq g_0\big(F(\bar x),\bar y\big)-t^\nu\big\langle\bar v,\,F(\bar x-t^\nu\xi)-F(\bar x)\big\rangle.\end{aligned}$$

Utilizing these inequalities in (3.18) we obtain

$$\begin{aligned}\limsup_{\nu\to\infty}\big[g(x^\nu+t^\nu\xi,y^\nu&+t^\nu\eta)-g(x^\nu,y^\nu)\big]/t^\nu\\ &\geq\langle\bar w,\eta\rangle+\lim_{\nu\to\infty}\Big\langle\bar v,\big[F(\bar x-t^\nu\xi)-F(\bar x)\big]/t^\nu\Big\rangle\\ &=\langle\bar w,\eta\rangle+\big\langle\bar v,-\nabla F(\bar x)\xi\big\rangle=\langle\bar w,\eta\rangle-\big\langle\bar z,\xi\big\rangle,\end{aligned}$$

so (3.17) is true.

Finally we look at the case where $G(x)\not\equiv 0$. The facts already developed apply to the functions $f_1(x,w):=f_0\big(F(x),w\big)$ and $g_1(x,y)=g_0\big(F(x),y\big)$: for all (x,y) in some neighborhood of $(\bar x,\bar y)$ we have

$$(z,s)\in\partial f_1(x,w)\quad\Longleftrightarrow\quad(-z,w)\in\overline{\partial} g_1(x,s), \tag{3.19}$$

with these conditions implying that $g_1(x,s)=\langle s,w\rangle-f(x,w)$. Since $f(x,y)=f_1(x,y)+\big\langle G(x),w\big\rangle$, the subgradients of f are given by $\partial f(x,w)=\partial f_1(x,w)+\big\langle\nabla G(x)^*w,G(x)\big\rangle$, so that

$$(z,y)\in\partial f(x,w)\quad\Longleftrightarrow\quad\big(z-\nabla G(x)^*w,\,y-G(x)\big)\in\partial f_1(x,w). \tag{3.20}$$

On the other hand, we have $g(x,y)=g_1\big(M(x,y)\big)$ for the smooth mapping $M:(x,y)\mapsto\big(x,y-G(x)\big)$. This mapping is one-to-one from $\mathbb{R}^n\times\mathbb{R}^m$ onto itself, and its inverse is smooth as well; the Jacobian is

$$\nabla M(x,y)=\begin{bmatrix}I&0\\-\nabla G(x)&I\end{bmatrix},\ \text{ with }\ \nabla M(x,y)^{-1}=\begin{bmatrix}I&0\\\nabla G(x)&I\end{bmatrix}.$$

Thus, M merely specifies a global change of coordinates, so that

$$\overline{\partial} g(x,y) = \nabla M(x,y)^* \overline{\partial} g_1\big(M(x,y)\big),$$

or in other words,

$$(-z,w) \in \overline{\partial} g(x,y) \quad \Longleftrightarrow \quad (-z + \nabla G(x)^* w, w) \in \overline{\partial} g_1\big(x, y - G(x)\big). \tag{3.21}$$

In applying the equivalences in (3.20) and (3.21) to the one in (3.19), we obtain the targeted relation (3.15). □

Corollary 3.5. *If f has the representation in Theorem 3.3, and (x,y) is such that the function g_0 corresponding to that representation is finite on a neighborhood of $\big(F(x), y\big)$, then*

$$(-z,w) \in \overline{\partial} g(x,y) \quad \Longleftrightarrow \quad z \in \partial_x(-g)(x,y),\ w \in \partial_y g(x,y). \tag{3.22}$$

Proof. This is the case of Theorem 3.4 where $G \equiv 0$. The combination of (3.10) with (3.15) gives the claimed equivalence (3.22). □

As a special case we obtain a characterization of the Clarke subgradients of a finite, concave-convex function, identifying them with the subgradients introduced in convex analysis.

Corollary 3.6. *Let $C \subset I\!\!R^n$ and $D \subset I\!\!R^m$ be convex sets, and let $h : C \times D \to I\!\!R$ be a finite function such that $h(x,y)$ is concave in x and convex in y. Then at any point $(x,y) \in \operatorname{int} C \times \operatorname{int} D$ one has*

$$\overline{\partial} h(x,y) \;=\; \big[-\partial_x(-h)(x,y)\big] \times \partial_y h(x,y).$$

Proof. On $\operatorname{int} C \times \operatorname{int} D$ the function h is Lipschitz continuous [1, Thm. 35.1]. Fix any point $(\bar{x},\bar{y}) \in \operatorname{int} C \times \operatorname{int} D$ and take X and Y to be compact neighborhoods of $\bar{x}$ and $\bar{y}$. Define $g : I\!\!R^n \times I\!\!R^m \to \overline{I\!\!R}$ by

$$g(x,y) = \begin{cases} h(x,y) & \text{if } x \in X \text{ and } y \in Y, \\ \infty & \text{if } x \in X \text{ but } y \notin Y, \\ -\infty & \text{if } x \notin X. \end{cases}$$

Then $g(x,y)$ is convex and lsc in y as well as convex and lsc in x, therefore a closed convex-concave function on $I\!\!R^n \times I\!\!R^m$ in the sense of convex analysis, so that g corresponds by way of (1.1) and (1.2) to some proper, lsc, convex function f, cf. [1, Thm. 33.3]. We can apply the theorem with $F = I$, $f_0 = f$, $g_0 = g$. The combination of (3.15) with the equivalence in Theorem 1.2 gives the result. □

4. Application to Multiplier Rules.

The proper, lsc function $f : I\!R^n \times I\!R^m \to \overline{I\!R}$ can be regarded as describing a parameterized family of optimization problems. Specifically, for any fixed vector $\bar{w} \in I\!R^m$ we may consider the problem

$$(\mathcal{P}) \qquad \text{minimize } f(x, \bar{w}) \text{ over all } x \in I\!R^n.$$

Under the assumption that $f(x, w)$ is convex in w, we can define the *Lagrangian function* for problem $(\mathcal{P})$ by

$$l(x, y) = \langle \bar{w}, y \rangle - g(x, y), \tag{4.1}$$

where g corresponds to f as before through (1.1) and (1.2). This fits with the way Lagrangians are introduced in the general perturbational theory of convex optimization problems and also with numerous examples in nonconvex optimization. For instance, if

$$f(x, w) = \begin{cases} F_0(x) & \text{if } F_1(x) \geq w_1,\ F_2(x) = w_2, \\ \infty & \text{otherwise}, \end{cases} \tag{4.2}$$

for $w = (w_1, w_2) \in I\!R^{m_1} \times I\!R^{m_2}$, a closed set $C \subset I\!R^n$, a continuous function $F_0 : I\!R^n \to I\!R$, and continuous mappings $F_1 : I\!R^n \to I\!R^{m_1}$ and $F_2 : I\!R^n \to I\!R^{m_2}$, problem $(\mathcal{P})$ takes the form

$$\text{minimize } F_0(x) \text{ subject to } F_1(x) \geq \bar{w}_1,\ F_2(x) = \bar{w}_2, \tag{4.3}$$

and the Lagrangian function in (1.4) comes out to be the ordinary one:

$$l(x, y) = \begin{cases} F_0(x) + \langle y_1, \bar{w}_1 - F_1(x) \rangle + \langle y_2, \bar{w}_2 - F_2(x) \rangle & \text{if } y_1 \geq 0, \\ -\infty & \text{if } y_1 \not\geq 0. \end{cases}$$

Other examples involve penalty representations of constraints and cover a vast array of applications.

In various cases, optimality conditions can be developed for $(\mathcal{P})$ in terms of subgradient conditions on the Lagrangian l and its "multiplier" argument y. The Kuhn-Tucker conditions for the nonlinear programming problem just described, with smooth F_0, F_1 and F_2, are the prototype. Clarke [11] was the first to establish such conditions without smoothness or convexity assumptions. Lagrangian conditions for nonconvex problems beyond the format of (4.3), but with a partial degree of smoothness, have been furnished recently in Rockafellar [12]. Multiplier rules for fully general *convex* optimization problems $(\mathcal{P})$, where f is convex and the Lagrangian l is therefore convex-concave, have been known for much longer [1].

Questions about the status of optimality conditions in terms of l arise because the sharpest techniques in nonsmooth analysis apply directly to f instead of l. The next theorem gives the pattern.

Theorem 4.1. *Let $\bar{x}$ be a locally optimal solution to $(\mathcal{P})$ (the value $f(\bar{x}, \bar{w})$ being finite), and suppose $\bar{x}$ satisfies the following constraint qualification: there is no vector $y \neq 0$ such that $(0, y) \in \partial^\infty f(\bar{x}, \bar{w})$. Then there is a vector $\bar{y}$ such that $(0, \bar{y}) \in \partial f(\bar{x}, \bar{w})$.*

Proof. Let X be a compact neighborhood of $\bar{x}$ relative to which $\bar{x}$ minimizes the objective function in $(\mathcal{P})$. For any $r > 0$ define

$$\widehat{f}(x, w) = \begin{cases} f(x, w) + \frac{1}{2} r|x - \bar{x}|^2 & \text{if } x \in X, \\ \infty & \text{if } x \notin X, \end{cases}$$

and consider alongside of $(\mathcal{P})$ the modified problem $(\widehat{\mathcal{P}})$ in which $\widehat{f}(x, \bar{w})$ is minimized over all x. Clearly, $\bar{x}$ is the unique optimal solution to $(\widehat{\mathcal{P}})$. Define

$$p(w) = \inf_x \widehat{f}(x, w).$$

The function $p : I\!R^m \to \overline{I\!R}$ is lower semicontinuous, and $p(\bar{w})$ is the optimal value in $(\widehat{\mathcal{P}})$. According to Rockafellar [5, Thm. 3.1], one has

$$\partial p(\bar{w}) \subset \{\, y \,|\, (0, y) \in \partial \widehat{f}(\bar{x}, \bar{w}) \,\}, \qquad \partial^\infty p(\bar{w}) \subset \{\, y \,|\, (0, y) \in \partial^\infty \widehat{f}(\bar{x}, b) \,\},$$

where $\partial \widehat{f}(\bar{x}, b) = \partial f(\bar{x}, \bar{w})$ and $\partial^\infty \widehat{f}(\bar{x}, \bar{w}) = \partial^\infty f(\bar{x}, \bar{w})$. The constraint qualification on $\bar{x}$ gives us $\partial^\infty p(\bar{w}) = \{0\}$, and from this it follows that p is Lipschitz continuous in a neighborhood of $\bar{w}$, and $\partial p(\bar{w})$ is a nonempty, compact set. In particular, there is at least one vector $\bar{y} \in \partial p(\bar{w})$, and this vector fulfills the requirement of $(0, \bar{y}) \in \partial f(\bar{x}, \bar{w})$. □

Theorem 4.2. *Suppose the function f underlying the parameterization in $(\mathcal{P})$ can be represented in the form specified in Theorem 3.3, and that the locally optimal solution $\bar{x}$ satisfies the constraint qualification in that theorem as well as the one in Theorem 4.1. Then the necessary condition asserted by Theorem 4.1 can be written equivalently as a Lagrange multiplier rule in the form:*

$$\textit{there exists } \bar{y} \textit{ such that } \; 0 \in \partial_x l(\bar{x}, \bar{y}), \;\; 0 \in \partial_y(-l)(\bar{x}, \bar{y}).$$

Proof. This is immediate from Theorem 3.3. □

In order to apply Theorems 4.1 and 4.2 to a given problem of minimizing an extended-real-valued function φ over $I\!R^n$, the first step is to introduce a parameterization involving elements of an auxiliary space $I\!R^m$ so that $\varphi(x) = f(x, \bar{w})$ for some choice of a proper, lsc function $f : I\!R^n \times I\!R^m$ and a vector $\bar{w} \in I\!R^m$. In this there is considerable latitude, and a representation of f as in Theorem 3.3 may well be attainable. The surprisingly wide coverage of such representations is explored in Poliquin and Rockafellar [13]. In particular applications, the chain rule in Proposition 2.2 can be used to bring out the individual details in the problem's structure as manifested in elaborating the multiplier rule in Theorem 4.2.

Example 4.3. *In the standard nonlinear programming problem in (4.3), the representing function f in (4.2) does have the composite structure demanded in Theorem 4.2. Namely, $f(x,w) = f_0\big(F(x),w\big)$ for*

$$F(x) := \big(F_0(x), F_1(x), F_2(x)\big) \in I\!R \times I\!R^{m_1} \times I\!R^{n_2},$$
$$f_0(u_0,u_1,u_2,w_1,w_2) := \begin{cases} u_0 & \text{if } u_1 \geq w_1,\ u_2 = w_2, \\ \infty & \text{otherwise.} \end{cases}$$

The constraint qualification in Theorem 4.1 corresponds to the Mangasarian-Fromovitz condition, and the multiplier rule in Theorem 4.2 reduces to the Kuhn-Tucker conditions.

Example 4.4. *Consider the problem*

$$(\mathcal{P}_0) \qquad \text{minimize } \varphi_0\big(F_0(x)\big) + \varphi_1\big(c - F_1(x)\big) \text{ over all } x \in I\!R^n,$$

where the mappings $F_0 : I\!R^n \to I\!R^{m_0}$ and $F_1 : I\!R^n \to I\!R^{m_1}$ are smooth and the functions $\varphi_0 : I\!R^{m_0} \to \overline{I\!R}$ and $\varphi_1 : I\!R^{m_1} \to \overline{I\!R}$ are proper, lsc and convex; the implicit constraints in problem $(\mathcal{P}_0)$ are

$$F_0(x) \in D_0 \text{ and } F_1(x) \in D_1, \text{ where } D_0 = \operatorname{dom}\varphi_0 \text{ and } D_1 = \operatorname{dom}\varphi_1.$$

View this as problem $(\mathcal{P})$ in the case of $f(x,w) = f_0\big(F_0(x), F_1(x), w\big)$ with $f_0(u_0,u_1,w) = \varphi_0(u_0) + \varphi_1(w - u_1)$ and $\bar{w} = c$, so that the corresponding Lagrangian is

$$l(x,y) = \varphi_0\big(F_0(x)\big) + \big\langle y, c - F_1(x)\big\rangle - \varphi_1^*(y)$$

(under the convention $\infty - \infty = \infty$ if needed). Suppose $\bar{x}$ is a locally optimal solution to $(\mathcal{P}_0)$ such that the following constraint qualification is satisfied: the only choice of $v_0 \in N\big(F_0(\bar{x})\big|D_0\big)$ and $v_1 \in N\big(c - F_1(\bar{x})\big|D_1\big)$ with $\nabla F_0(\bar{x})^* v_0 - \nabla F_1(\bar{x})^* v_1 = 0$ is $v_0 = 0$ and $v_1 = 0$. Then the multiplier rule in Theorem 4.2 holds; in other words, there is a vector $\bar{y}$ such that

$$c \in \partial\varphi_1^*(\bar{y}) \text{ and } 0 \in \partial\Phi(\bar{x}), \text{ where } \Phi(x) = \varphi_0\big(F_0(x)\big) - \big\langle \bar{y}, F_1(x)\big\rangle.$$

In this problem model the function φ_1 expresses penalty conditions on the deviation of $F_1(x)$ from c. The penalties may be finite or infinite, and for some directions of deviation they may be 0.

5. Application to Optimal Control

A fundamental problem model in optimal control and modern forms of the calculus of variations is the generalized Bolza problem in which a functional

$$J(x) := \int_{t_0}^{t_1} L\big(t, x(t), \dot{x}(t)\big)dt + l\big(x(t_0), x(t_1)\big) \tag{5.1}$$

is minimized over the space of all absolutely continuous trajectories $x : [t_0, t_1] \to I\!R^n$. Here $\dot{x}(t) = dx/dt$, which exists for almost every t. The functions $L(t,\cdot,\cdot)$ and $l(\cdot,\cdot)$ are proper and lsc on $I\!R^n \times I\!R^n$, and L is $\mathcal{L} \times \mathcal{B} \times \mathcal{B}$-measurable on $[t_0, t_1] \times I\!R^n \times I\!R^n$, where $\mathcal{L}$ refers to Lebesgue sets and $\mathcal{B}$ to Borel sets.

The extended-real-valuedness of L and l allows a wide variety of problems to be reduced to the generalized Bolza form. These include problems with constraints and even ones with explicit control variables; see Clarke [3] for a full discussion. When $L(t,\cdot,\cdot)$ is for each t the indicator function for a closed subset of $I\!R^n \times I\!R^n$, which may be interpreted as the graph of a set valued mapping $x \mapsto S(t,x)$, the minimization of $J(x)$ amounts to the minimization of $l\big(x(t_0), x(t_1)\big)$ subject to the differential inclusion

$$\dot{x}(t) \in S\big(t, x(t)\big) \text{ for a.e. } t. \tag{5.2}$$

Our interest here is in necessary conditions for optimality of a trajectory x that can be derived for this problem, specifically ones involving the integrand L. The *Euler-Lagrange condition* in the classical case where $L(t,\cdot,\cdot)$ is smooth refers to the existence of an absolutely continuous adjoint trajectory p such that

$$\big(\dot{p}(t), p(t)\big) = \nabla L\big(t, x(t), \dot{x}(t)\big) \text{ for a.e. } t. \tag{5.3}$$

For the nonsmooth case where $L(t,\cdot,\cdot)$ is convex and possibly not finite-valued, the Euler-Lagrange condition has been cast instead in the form

$$\big(\dot{p}(t), p(t)\big) \in \partial L\big(t, x(t), \dot{x}(t)\big) \text{ for a.e. } t, \tag{5.4}$$

cf. Rockafellar [14]. (In these statements the convention is followed that gradient and subgradient symbols apply only to the x and $\dot{x}$ arguments, never to the t argument.) Certain cases where $L(t,\cdot,\cdot)$ is not necessarily smooth or convex have been treated by Clarke [15][3] in the form

$$\big(\dot{p}(t), p(t)\big) \in \overline{\partial} L\big(t, x(t), \dot{x}(t)\big) \text{ for a.e. } t, \tag{5.5}$$

which is equivalent to (5.4) when $L(t,\cdot,\cdot)$ is subdifferentially regular (as for instance when $L(t,\cdot,\cdot)$ is convex) and reduces to (5.3) in the presence of smoothness. Mordukhovich, on the other hand, has investigated in [6][16][17] versions of the Euler-Lagrange condition for nonsmooth, nonconvex problems in which there is less convexification than in the definition of $\overline{\partial}L\big(t,x(t),\dot{x}(t)\big)$. These efforts have mainly kept to the context of a differential inclusion (5.2), where L is an indicator, but a generalized formulation of Mordukhovich's kind of condition, for the sake of comparison with (5.5), would be

$$\begin{gathered}\big(\dot{x}(t),\dot{p}(t)\big)\in \operatorname{con} R\big(t,x(t),p(t)\big),\ \text{ where}\\ R(t,x,p)=\big\{\,(v,z)\,\big|\,(z,p)\in\partial L(t,x,v),\ p\in\partial_v L(t,x,v)\,\big\}.\end{gathered}\tag{5.6}$$

This version of the Euler-Lagrange condition is stronger than the one in (5.5) in some situations, so it would be good to know when it can be established for a category of integrands L other than indicators.

An alternative mode of development of necessary conditions has centered on the associated *Hamiltonian function* H, which is defined by

$$H(t,x,p)=\sup\nolimits_v\big\{\,\langle p,v\rangle-L(t,x,v)\,\big\}.\tag{5.7}$$

When $L(t,x,v)$ is convex in v, as virtually dictated by consideration of the existence of a solution to the problem of minimizing the functional J in (5.1), one has the reciprocal formula

$$L(t,x,v)=\sup\nolimits_p\big\{\,\langle p,v\rangle-H(t,x,p)\,\big\}.\tag{5.8}$$

The associated *Hamiltonian condition* takes the form

$$\dot{x}(t)=\nabla_p H\big(t,x(t),p(t)\big),\quad \dot{p}(t)=-\nabla_x H\big(t,x(t),p(t)\big)\ \text{ for a.e. } t\tag{5.9}$$

in the classical, smooth case and the form

$$\dot{x}(t)\in\partial_p H\big(t,x(t),p(t)\big),\quad \dot{p}(t)\in\partial_x(-H)\big(t,x(t),p(t)\big)\ \text{ for a.e. } t\tag{5.10}$$

in the convex case [18]. In working with problems where $H(t,\cdot,\cdot)$ is finite and Lipschitz continuous, Clarke [19][3][20] has studied the version

$$\big(-\dot{p}(t),\dot{x}(t)\big)\in\overline{\partial}H\big(t,x(t),p(t)\big)\ \text{ for a.e. } t,\tag{5.11}$$

In contrast to the smooth and convex cases, where the Hamiltonian condition is readily seen to be equivalent to the corresponding Euler-Lagrange condition, the connection between (5.11) and (5.5) has been unclear. In recent work of Loewen and Rockafellar [21] the existence of an adjoint trajectory p satisfying, for the optimal trajectory x, *both* the Hamiltonian condition in version (5.11) and the Euler-Lagrange condition in version (5.5) has been established in some rather general circumstances. This still leaves the relationship between the two conditions unsettled, however, along with the possible sharpening of the Euler-Lagrange condition toward (5.6) and beyond.

The theorems in Section 3 go a long way toward dispelling the fog over the relationship, although not every question is answered. Obviously, when $L(t, x, v)$ is convex in v as well as lower semicontinuous in (x, v) the functions $L(t, \cdot, \cdot)$ and $H(t, \cdot, \cdot)$ form a pair f, g, of the very kind we have been studying. We are not concerned in this paper with the precise assumptions under which the Euler-Lagrange condition or the Hamiltonian condition can be established as necessary, but rather with the extent to which one may imply the other, so we limit our conclusions here to the following.

Theorem 5.1. *Suppose the Hamiltonian $H(t, x, p)$ is locally Lipschitz continuous in x. Then the Hamiltonian condition in the Clarke version (5.11) implies the Euler-Lagrange condition in the Mordukhovich version (5.6), so that in cases where the set $R\big(t, x(t), p(t)\big)$ in (5.6) happens to be convex the Hamiltonian condition (5.11) actually guarantees that*

$$\big(\dot{p}(t), p(t)\big) \in \partial L\big(t, x(t), \dot{x}(t)\big) \quad \text{and} \quad p(t) \in \partial_v L\big(t, x(t), \dot{x}(t)\big) \quad \text{for a.e. } t. \tag{5.12}$$

When $L(t, \cdot, \cdot)$ can be represented in the form in Theorem 3.4 (namely $L\big(t, x(t), \dot{x}(t)\big) = L_0\big(t, F\big(t, x(t)\big), \dot{x}(t)\big) + \big\langle G\big(t, x(t)\big), \dot{x}(t)\big\rangle$ with the constraint qualification in that result satisfied at $x(t)$), then the Hamiltonian condition (5.11) is equivalent to the Euler-Lagrange condition in the sharpest version (5.12). If the representation is of the more special kind in Theorem 3.3 (i.e., one has $G \equiv 0$), the Hamiltonian condition can also be written equivalently in the product form

$$\dot{x}(t) \in \partial_p H\big(t, x(t), p(t)\big) \ \text{and} \ \dot{p}(t) \in \partial_x(-H)\big(t, x(t), p(t)\big) \ \text{for a.e. } t. \tag{5.13}$$

Proof. For fixed t, when $H(t, x, p)$ is Lipschitz continuous in x it is Lipschitz continuous jointly in x and p by virtue of its convexity in p, cf. [1, Thm. 10.6]. The first assertion in the theorem is clear then from Theorem 3.1. The second follows from Theorem 3.4. The equivalent expression in (5.13) follows from Corollary 3.5. □

Example 5.2. *Suppose*

$$L(x, v) = \min_u \left\{ \varphi_0\big(F_0(x)\big) + \psi(u) \,\middle|\, F_1(x) + Bu = v \right\},$$

as befits an underlying control model with dynamics $\dot{x}(t) = F_1\big(x(t)\big) + Bu(t)$ *and cost expression* $\varphi_0\big(F_0(x(t))\big) + \psi\big(u(t)\big)$. *Assume the mappings* F_0 *and* F_1 *are smooth and the function* φ_0 *is finite and convex on* $\mathbb{R}^{m_0}$, *whereas the function* ψ *is lsc, proper and convex on* $\mathbb{R}^m$ *with* ψ^* *finite. Then also*

$$L(x,v) = \varphi_0\big(F_0(x)\big) + \varphi_1\big(v - F_1(x)\big) \text{ for } \varphi_1(v) := \min_u \big\{ \psi(u) \,\big|\, Bu = v \big\},$$

the function φ_1 *being lsc, proper and convex on* $\mathbb{R}^n$, *so that* $L(x,v) = L_0\big(F_0(x), F_1(x), v\big)$ *for* $L_0(\xi_0, \xi_1, v) = \varphi_0(\xi_0) + \varphi_1(v - \xi_1)$. *Here the set of* (ξ_0, ξ_1) *such that the convex function* $L_0(\xi_0, \xi_1, \cdot)$ *is proper is all of* $\mathbb{R}^{m_0} \times \mathbb{R}^n$, *so the constraint qualification demanded of this representation is sure to be satisfied. The Hamiltonian function is*

$$H(x,p) = -\varphi_0\big(F_0(x)\big) + \big\langle p, F_1(x)\big\rangle + \psi^*\big(B^* p\big).$$

This function is locally Lipschitz continuous, so by Theorem 5.1 the Hamiltonian condition (5.11) is equivalent to the Euler-Lagrangian condition (5.12).

Example 5.3. *On* $\mathbb{R} \times \mathbb{R}$ *let*

$$L(x,v) = \begin{cases} \frac{1}{2}x^2 + xv & \text{if } -1 \leq v \leq 1, \\ \infty & \text{otherwise,} \end{cases}$$

so that the Hamiltonian is the locally Lipschitz function

$$H(x,p) = -\tfrac{1}{2}x^2 + |p - x|.$$

The function L *has the kind of representation specified in Theorem 5.1 (with* $L_0(x,v) = \frac{1}{2}x^2$ *if* $-1 \leq v \leq 1$, $L_0(x,v) = \infty$ *otherwise,* $F(x) = x = G(x)$, *the constraint qualification being satisfied). The Hamiltonian condition (5.11) is therefore equivalent to the Euler-Lagrangian condition (5.12), despite the fact that* $-H(x,p)$ *is not subdifferentially regular in* x. *Moreover,*

$$-\partial(-H)(x,p) \subset \overline{\partial} H(x,p) \subset \big[-\overline{\partial}_x(-H)(x,p)\big] \times \partial_p H(x,p), \text{ strictly when } x = p.$$

References

1. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton NJ, 1970.
2. F. H. Clarke, "Generalized gradients and applications," Transactions Amer. Math. Soc. **205** (1975), 247–262.
3. F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
4. R. T. Rockafellar, "Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization," Math. Operat. Res. **6** (1981), 427–437.
5. R. T. Rockafellar, "Extensions of subgradient calculus with applications to optimization," Nonlin. Analysis Th. Meth. Appl. **9** (1985), 665–698.
6. B. S. Mordukhovich, "The maximum principle in the problem of time-optimal control with nonsmooth constraints," J. Appl. Math. Mech. **40** (1976), 960–969.
7. B. S. Mordukhovich, *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1988 (in Russian; English translation to appear).
8. R. T. Rockafellar, "Directionally Lipschitzian functions and subdifferential calculus," Proc. London Math. Soc. **39** (1979), 331–355.
9. R. A. Poliquin, "Subgradient monotonicity and convex functions," Nonlinear Analysis Th. Meth. Appl. **14** (1990), 305–317.
10. D. E. Ward and J. M. Borwein, "Nonsmooth calculus in finite dimensions," SIAM J. Control Opt. **25** (1987), 1312–1340.
11. F. H. Clarke, "A new approach to Lagrange multipliers," Math. Oper. Res. **1** (1976), 165–174.
12. R. T. Rockafellar, "Perturbation of generalized Kuhn-Tucker points in finite-dimensional optimization," in *Nonsmooth Optimization and Related Topics*, F.H. Clarke et al. (eds.), Plenum Press, 1989, 393–402.
13. R. A. Poliquin and R. T. Rockafellar, "A calculus of epi-derivatives with applications to optimization," Canadian J. Math., submitted 1991.
14. R. T. Rockafellar, "Existence and duality theorems for convex problems of Bolza," Transactions Amer. Math. Soc. **159** (1971), 1–40.
15. F. H. Clarke, "The Euler-Lagrange differential inclusion," J. Differ. Eq. **19** (1975), 80–90.
16. B. S. Mordukhovich, "Optimization and approximation of differential inclusions," Cybernetics **24** (1988), 781–788.

17. B. S. Mordukhovich, “On variational analysis of differential inclusions,” Dept. of Math., Wayne State Univ., 1991.

18. R. T. Rockafellar, “Generalized Hamiltonian equations for convex problems of Lagrange,” Pacific J. Math. **33** (1970), 411–428.

19. F. H. Clarke, “Extremal arcs and extended Hamiltonian systems,” Transactions Amer. Math. Soc. **64** (1977), 349–367.

20. F. H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*, No. 57 in the CBMS-NSF Regional Conference Series in Applied Mathematics, S.I.A.M. Publications, Philadelphia, 1989.

21. P. D. Loewen and R. T. Rockafellar, “The adjoint arc in nonsmooth optimization,” Transactions Amer. Math. Soc. **325** (1991), 39–72.