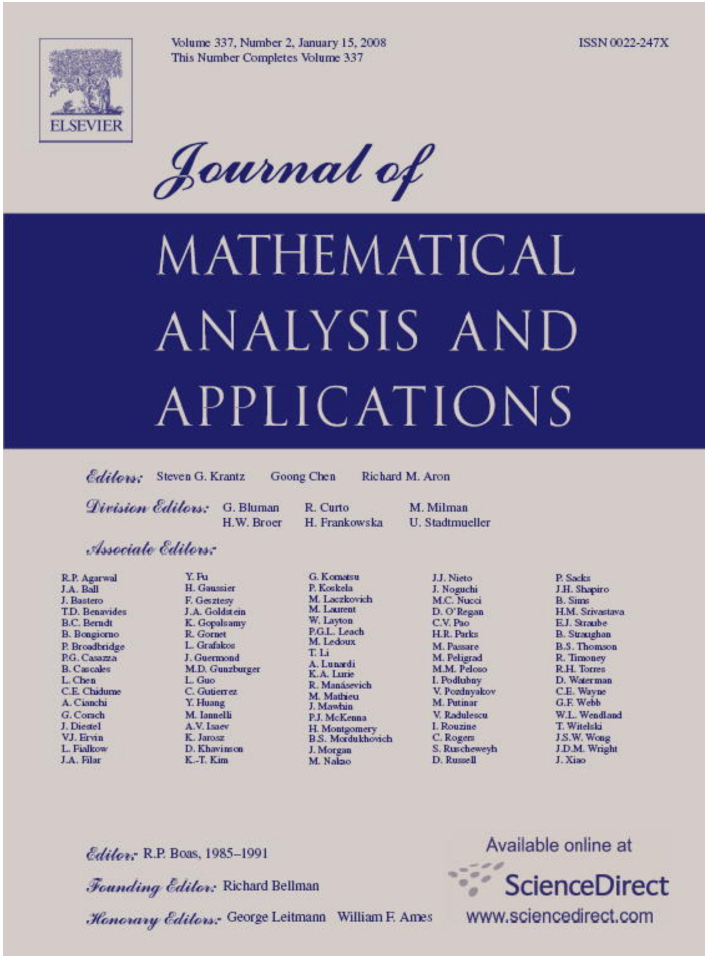
Volume 337, Number 2, January 15, 2008
This Number Completes Volume 337
ISSN 0022-247X
ELSEVIER
Journal of
MATHEMATICAL
ANALYSIS AND
APPLICATIONS
Editors: Steven G. Krantz Goong Chen Richard M. Aron
Division Editors: G. Bluman R. Curto M. Milman H.W. Broer H. Frankowska U. Stadtmueller
Associate Editors:
Editor: R.P. Boas, 1985–1991
Founding Editor: Richard Bellman
Honorary Editors: George Leitmann William F. Ames
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
J. Math. Anal. Appl. 337 (2008) 1064–1071
Journal of MATHEMATICAL ANALYSIS AND APPLICATIONS
www.elsevier.com/locate/jmaa

A multiplier rule on a metric space

Michael McAsey *, Libin Mou

Department of Mathematics, Bradley University, Peoria, IL 61625, USA

Received 2 April 2007
Available online 24 April 2007
Submitted by Richard M. Aron

Abstract

A multiplier rule is proved for constrained minimization problems defined on a metric spaces. The proof requires a generalization of the values of a derivative in the classical case that the metric space is a normed space.

Keywords: Multiplier rule; Derivative; Derivate

1. Introduction

Lagrange multiplier rules form a powerful tool for deriving necessary conditions for solutions of constrained minimization problems. Most of the existing multiplier rules are established for optimization problems defined on Banach spaces, on which appropriate notions of differentiability are well-defined; see [6,7] for example. One of the exceptions is an abstract multiplier rule for a mathematical programing problem using the notion of cone differential in [4, Theorem 13.1, Section II.13]. In this paper we prove a multiplier rule for a general constrained optimization problem formulated on a metric space. The motivation for this is a proof of the maximum principle for an optimal control problem with quite general constraints [10]. A key idea in the proof of the multiplier rule is to apply the Ekeland variational principle to a penalized objective function with no constraint. This is a typical approach in optimization problems; see [3] and [8] for example.

In Section 2, we define the notion of a set of sequential strict "derivates" for a map on a metric space and prove the multiplier rule (Theorem 6) for minimizers of a general constrained optimization problem. The notion of sequential strict derivate is a generalization of the value of a classical derivative on a Banach space as well as sequential and directional derivatives. It enables us to derive desired inequalities satisfied by the multipliers for the optimization problem whose existence is guaranteed by the subdifferential of a distance function. This form of the multiplier theorem appears to be new even when the metric space is a closed subset of a Banach space. In this latter case, there are many related interesting works on nondifferentiable multiplier rules; see [2,5,9,11,13–17] and [18].

* Corresponding author.
E-mail addresses: mcasey@bradley.edu (M. McAsey), mou@bradley.edu (L. Mou).

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2007.04.027

2. A general multiplier rule

To formulate the general optimization problem on a metric space let $(\mathcal{W}, d)$ be a complete metric space and $(Z, \|\cdot\|)$ be a Banach space. Let $(J(\cdot), S(\cdot)) : \mathcal{W} \to \mathbb{R} \times Z$ be continuous maps and $Q \subset Z$ a subset. Consider

Problem.

$$\text{minimize } J(w), \quad w \in \mathcal{W} \text{ with } S(w) \in Q. \tag{1}$$

If $w_0 \in \mathcal{W}$, $S(w_0) \in Q$ and $J(w_0) \leqslant J(w)$ for all $w \in \mathcal{W}$ with $S(w) \in Q$, then we say that w_0 is a *minimum point of* $J(\cdot)$ *on* $\mathcal{W}$ *subject to* $S(\cdot) \in Q$.

To establish a multiplier rule for the Problem above, appropriate notions of derivatives have to be defined for maps from a metric space to a Banach space.

In the rest of this section, $(\mathcal{W}, d)$ is a complete metric space and X, Y, and Z are Banach spaces with norms all denoted by $\|\cdot\|$. We start by recalling the notions of directional derivatives and strict differentiability of maps between Banach spaces.

Definition 1. (a) The *directional derivative* $g'(x_0; v)$ of a map $g : X \to Y$ at $x_0 \in X$ in the direction of $v \in X$ is

$$g'(x_0; v) = \lim_{t \to 0+} \frac{1}{t}\big[g(x_0 + tv) - g(x_0)\big].$$

(b) We say that g is *strictly differentiable* at x_0 if there exists a function $\omega(\cdot; x_0) : \mathbb{R}^+ = [0, \infty) \to \mathbb{R}^+$ with $\omega(\delta; x_0) \to 0$ as $\delta \downarrow 0$ and a linear and bounded operator, denoted by $Dg(x_0)$, from X to Y such that

$$\big\|g(x) - g(y) - \big\langle Dg(x_0), x - y\big\rangle\big\| \leqslant \omega_g(\delta; x_0)\|x - y\|$$

for all $x, y \in x_0 + \delta\mathbb{B}$, where $\mathbb{B}$ is the unit ball in X centered at 0. Sometimes the function $\omega_g(\delta; x_0)$ will be written simply as $\omega(\delta)$.

The notation $\langle\cdot,\cdot\rangle$ will often be used to denote an operator acting on a vector, so $\langle Dg(x_0), x - y\rangle$ replaces the function notation $Dg(x_0)(x - y)$. Note that the more usual definition of strict differentiability is that the difference quotient $\|g(x) - g(y) - \langle Dg(x_0), x - y\rangle\|/\|x - y\| \to 0$ as $x \to x_0$ and $y \to x_0$ and so does not involve the function ω explicitly; see [1] or [12] for example. But the two definitions are equivalent and notion of ω will be useful for us later.

We next wish to define a derivative of a map $S : \mathcal{W} \to X$ from a metric space to a Banach space. However since $\mathcal{W}$ does not have a linear structure, the derivative, a linear operator from $\mathcal{W}$ to X, does not make sense, so we instead define three related objects that generalize the familiar objects from Definition 2. The first notion will generalize the directional derivative where "direction" is replaced by a convergent sequence $w^i \to w$. We will write $w^i \to w$ in $\mathcal{W}$ for "$w^i \in \mathcal{W}$ and $w^i \to w$ as $i \to \infty$." (The sequence indices i, j are always assumed to go to ∞.) For the other two objects we will settle for defining what would be the *value* of a derivative in the classical case of a map between two Banach spaces. We call the resulting object in X a *derivate* of the map S. The special case in which $\mathcal{W}$ is a Banach space will be discussed in Proposition 1.

Definition 2. (a) The *sequential derivate* of S along $w^i \to w$ in $\mathcal{W}$ is

$$DS(w_0; w^i) = \lim_{i \to \infty} \frac{S(w^i) - S(w)}{d(w^i, w)},$$

if it exists. The notation is chosen to be reminiscent of the directional derivative in a Banach space. We have resisted the urge to use the more complete but cumbersome $DS(w_0; \{w^i\}_{i=1}^{\infty})$.

(b) For a given $\delta \geqslant 0$, we say that $x \in X$ is a *sequential* δ*-derivate* of S at w if there exists a sequence $d^i \downarrow 0$ and $w^i \in \mathcal{W}$ such that $d(w, w^i) \leqslant d^i$ and

$$\limsup_{i \to \infty} \left\| \frac{S(w^i) - S(w)}{d^i} - x \right\| \leqslant \delta. \tag{2}$$

The set of all δ-derivates of S at w is denoted by $D^\delta S(w)$.

(c) We say that $x \in X$ is a *sequential strict derivate* of S at w_0 if there exists a function $\delta : \mathcal{W} \to \mathbb{R}^+$ such that $\delta(w) \to 0$ as $d(w, w_0) \to 0$ and for all $w \in \mathcal{W}$,

$$x \in D^{\delta(w)} S(w).$$

The set of all sequential strict derivates x is denoted by $D_s S(w_0)$.

The sequential δ-derivate and sequential strict derivate have similarities with the notions of (the values of) sequential, directional, and strict derivatives on Banach spaces (Proposition 1 below) and possess a desirable property for proving the multiplier rule. The use of the sequence $\{d^i\}$ in the denominator rather than the sequence of distances $\{d(w^i, w)\}$ gives the needed flexibility in determining the set $D_s S(w_0)$. This flexibility is useful in applications, and is used in a proof of the maximum principle in optimal control [10].

Here are three elementary examples showing the kinds of sets that occur as $D^\delta S(w)$ and $D_s S(w_0)$—at least in the scalar case. Note that it follows from the definition that zero is always in $D_s S(w_0)$.

Example 1. (a) Let $f \in C^1(\mathbb{R})$. Then for any $w, w_0 \in \mathbb{R}$,

$$D^\delta f(w) = \left[-\left(\left|f'(w)\right| + \delta\right), \left|f'(w)\right| + \delta\right] \quad \text{and} \quad D_s f(w_0) = \left[-\left|f'(w_0)\right|, \left|f'(w_0)\right|\right].$$

(b) (Modified absolute value function)

$$f(w) = \begin{cases} \alpha w & w \geqslant 0, \\ \beta w & w < 0, \end{cases} \quad \text{with } \beta < 0 < \alpha.$$

By example (a), for $w > 0$, $D^\delta f(w) = [-\alpha - \delta, \alpha + \delta]$ and similarly for $w < 0$, $D^\delta f(w) = [-|\beta| - \delta, |\beta| + \delta]$. Then $D^0 f(0) = [0, \max\{|\beta|, \alpha\}]$ while $D_s f(0) = [0, \min\{|\beta|, \alpha\}]$.

(c) Let $f(x) = x \sin(1/x)$. Choosing w^k to be the positive relative minimizers of $\sin(1/x)$, $w^k = \frac{2}{(2k+1)\pi}$ so that $w^k \downarrow 0$ as $k \to 0$, and $f'(w^k) = 0$. So from example (a) $D^\delta f(w^k) = [-\delta, \delta]$. Therefore $D_s f(0) = \{0\}$.

When $\mathcal{W}$ is a Banach space and $w^i = w + d^i v$ with $d^i \downarrow 0$ and $\|v\| \neq 0$, then the sequential derivate is related to the classical directional derivative by

$$DS(w; w + d^i v) = S'\big(w; v/\|v\|\big). \tag{3}$$

The relationship between sequential strict derivates on a metric space and strict derivatives on a Banach space is given in the next proposition. It shows that for a strictly differentiable function, the set of all strict derivates, $D_s S(w_0)$, is essentially a set of values of the strict derivative.

Proposition 1. *Let S be a map between two Banach spaces $\mathcal{W}$ and X.* (a) *If S is strictly differentiable* (*in the Banach space sense*) *at w_0, then for any $v \in \mathcal{W}$ with $\|v\| \leqslant 1$,*

$$\big\langle DS(w_0), v\big\rangle \in D_s S(w_0). \tag{4}$$

(b) *If $\mathcal{W}$ is finite dimensional, then*

$$D_s S(w_0) = \big\{\big\langle DS(w_0), v\big\rangle \colon v \in \mathcal{W} \text{ with } \|v\| \leqslant 1\big\}.$$

(c) *A necessary and sufficient condition that $x \in D^\delta S(w)$ is that there exists a sequence $d^i \downarrow 0$ and a sequence $v^i \in \mathcal{W}$ with $||v^i|| \leqslant 1$ such that*

$$\limsup_{i \to \infty} \left\| \frac{S(w + d^i v^i) - S(w)}{d^i} - x \right\| \leqslant \delta. \tag{5}$$

Proof. (a) Fix $v \in \mathcal{W}$, $\|v\| \leqslant 1$. To show $\langle DS(w_0), v\rangle \in D_s S(w_0)$ let $\delta : \mathcal{W} \to \mathbb{R}^+$ be defined by $\delta(w) = \omega(\|w - w_0\|)$, where ω is the function from the definition of strict derivative. Then δ satisfies the conditions of the function in the definition of (metric space) sequential strict derivate. We need to show next that for any $w \in \mathcal{W}$, we have $\langle DS(w_0), v\rangle \in D^{\delta(w)} S(w)$. For each $w \in \mathcal{W}$, let $w^i = w + d^i v$, $\|v\| \leqslant 1$, with any sequence $d^i \downarrow 0$. Then $\|w^i - w\| = d^i \|v\| \leqslant d^i$. It follows from the definition of sequential strict derivate, as $i \to \infty$,

$$\left\| \frac{S(w^i) - S(w)}{d^i} - \langle DS(w_0), v\rangle \right\| = \left\| \frac{S(w + d^i v) - S(w)}{d^i} - \langle DS(w_0), v\rangle \right\| \leqslant \omega\big(\|w - w_0\| + d^i\big)\|v\| \leqslant \omega\big(\|w - w_0\| + d^i\big).$$

Taking the lim sup of this expression we get

$$\limsup_{i\to\infty} \left\| \frac{S(w^i) - S(w)}{d^i} - \langle DS(w_0), v\rangle \right\| \leqslant \limsup_{i\to\infty} \omega\big(\|w - w_0\| + d^i\big) = \omega\big(\|w - w_0\|\big) = \delta(w).$$

We have shown that $\langle DS(w_0), v\rangle \in D^{\delta(w)} S(w)$ with $\delta(w) = \omega(\|w - w_0\|) \to 0$. That is, $\langle DS(w_0), v\rangle \in D_s S(w_0)$.

(b) Since this fact is not used below, we leave the proof to interested readers.

(c) If (5) holds, then clearly $x \in D^\delta S(w)$. Conversely, if $x \in D^\delta S(w)$, then (2) holds for some sequence $w^i \to w$. Let $d^i = \|w^i - w\|$ and $v^i = (w^i - w)/d^i \in \mathcal{W}$, then $\|v^i\| = 1$, $d^i \to 0$ and (5) holds. □

We show some properties of δ-derivates in the next two propositions. The first one states an implication of the Ekeland variational principle. The second is a version of the chain rule for δ-derivates and directional derivatives.

Proposition 2. *Let $(\mathcal{W}, d)$ be a complete metric space, $F : \mathcal{W} \to (-\infty, \infty]$ proper, lower-semicontinuous, bounded from below, and $w_0 \in \mathcal{W}$ with $F(w_0) < \infty$. Then for every $\lambda > 0$, there exists $\overline{w} \in \mathcal{W}$ such that*

$$F(\overline{w}) + \lambda d(\overline{w}, w_0) \leqslant F(w_0),$$
$$F(\overline{w}) < F(w) + \lambda d(w, \overline{w}) \quad \text{for all } w \in \mathcal{W} \setminus \{\overline{w}\}. \tag{6}$$

In addition, every $x \in D^\delta F(\overline{w})$ satisfies $x \geqslant -\lambda - \delta$.

Proof. The display (6) is the well-known Ekeland variational principle; it is proved in many references; see [3] or [8] for example. For the last sentence, if $x \in D^\delta F(\overline{w})$, then there exists a sequence $w^i \in \mathcal{W}$ and $d^i \downarrow 0$ such that $d(w^i, \overline{w}) \leqslant d^i$ and (2) holds with $w = \overline{w}$. From (5) and (6) we have

$$\delta \geqslant \limsup_{i\to\infty} \frac{F(w^i) - F(\overline{w})}{d^i} - x \geqslant -\lambda \frac{d(w^i, \overline{w})}{d^i} - x \geqslant -\lambda - x.$$

This implies that $x \geqslant -\lambda - \delta$. □

Proposition 3. (a) *Let $S : \mathcal{W} \to X$. Let $g : X \to Y$ be Lipschitz near $S(w)$ with rank K. If $v \in D^\delta S(w)$ for some $\delta \geqslant 0$ and $g'(S(w); v)$ exists, then*

$$g'\big(S(w); v\big) \in D^{\delta K}(g \circ S)(w). \tag{7}$$

In particular, if $v = DS(w; w^i)$ where $w^i \to w$ in $\mathcal{W}$ and $g'(S(w); v)$ exist, then $D(g \circ S)(w; w^i)$ exists and

$$D(g \circ S)(w; w^i) = g'\big(S(w); v\big). \tag{8}$$

(b) *If $\mathcal{W}$ in* (a) *is a Banach space and $g'(S(w); S'(w; u))$ exists for some $u \in \mathcal{W}$, then*

$$(g \circ S)'(w; u) = g'\big(S(w); S'(w; u)\big). \tag{9}$$

Proof. (a) Since $v \in D^\delta S(w)$, there exists $d^i \downarrow 0$ and $w^i \in \mathcal{W}$ with $d(w, w_i) \leqslant d^i$ such that

$$\limsup_{i\to\infty} \left\| \frac{S(w^i) - S(w)}{d^i} - v \right\| \leqslant \delta.$$

Because g is Lipschitz with rank K near $S(w)$ and $g'(S(w); v)$ exists, we get

$$\begin{aligned}\limsup_{i\to\infty} \left\| \frac{g(S(w^i)) - g(S(w))}{d^i} - g'\big(S(w); v\big) \right\| &\leqslant \limsup_{i\to\infty} \left\| \frac{g(S(w^i)) - g(S(w) + d^i v)}{d^i} \right\| \\ &\quad + \limsup_{i\to\infty} \left\| \frac{g(S(w) + d^i v) - g(S(w))}{d^i} - g'\big(S(w); v\big) \right\| \\ &\leqslant \limsup_{i\to\infty} K \left\| \frac{S(w^i) - S(w)}{d^i} - v \right\| + 0 \leqslant K\delta.\end{aligned}$$

Equality (8) follows basically by repeating the preceding calculation proving (7). Indeed let $d^i = d(w^i, w)$, then

$$\lim_{i\to\infty}\left\|\frac{S(w^i)-S(w)}{d^i}-v\right\| = 0 \quad \text{for } v = DS(w; w^i).$$

Repeat the rest of the calculation replacing the "limit supremum" with "limit" and end with both limits equal to zero.

To prove (b), note that since $\mathcal{W}$ is a Banach space if $\|u\| = 1$, then equality (9) follows from (8) and the relation between directional derivatives given in (3) with $v = S'(w, u)$. Since (9) is homogeneous in u, it holds for all u. □

Let $Q \subset Z$ be a given subset, as in the Problem stated in Section 1. The distance function $d_Q(\cdot)$ of Q plays a crucial role in the multiplier rule, where

$$d_Q(z) = \inf_{z' \in Q} \|z - z'\|.$$

Suppose that Q is closed and convex, then $d_Q(z)$ is Lipschitz with rank 1 and d_Q is convex. Therefore, the directional derivative $d'_Q(x_0; v)$ exists for $(x_0, v) \in Z \times Z$ because $\frac{1}{t}[d_Q(x_0 + tv) - d_Q(x_0)]$ is increasing in $t \in (0, \infty)$. Recall that the subdifferential (in the sense of convex analysis) of $d_Q(z)$ is defined as the set

$$\partial d_Q(z) = \bigl\{\zeta \in Z^* \mid d_Q(\eta) - d_Q(z) \geqslant \langle \zeta, \eta - z\rangle,\ \forall \eta \in Z\bigr\}. \tag{10}$$

The following lemma collects some fundamental properties of $d_Q(\cdot)$. For a proof, see [8, Proposition 3.11], for example.

Lemma 4. *Suppose that $Q \subset Z$ is closed and convex. Then*

(1) *$d_Q(z)$ is convex and Lipschitz with rank* 1.
(2) *$d'_Q(z; \xi)$ is positively homogeneous and subadditive in ξ.*
(3) *$\partial d_Q(z)$ is nonempty, convex and weak*-compact and*

$$\partial d_Q(z) = \bigl\{\zeta \in Z^* \mid \langle \zeta, \xi\rangle \leqslant d'_Q(z; \xi) \text{ for all } \xi \in Z\bigr\}.$$

(4) $d'_Q(z; \xi) = \max\{\langle \zeta, \xi\rangle \mid \zeta \in \partial d_Q(z)\}$.
(5) *The map $z \to \partial d_Q(z)$ is pseudo-continuous on Z in the sense that if $z_\alpha \in Z$ and $\zeta_\alpha \in \partial d_Q(z_\alpha)$ are two nets with $z_\alpha \xrightarrow{s} z$, $\zeta_\alpha \xrightarrow{*} \zeta$, then $\zeta \in \partial d_Q(z)$.* (*The nets converge strongly, i.e., in norm, and weak-* respectively.*)
(6) *For any $z \notin Q$, $|\zeta|_{Z^*} = 1$ for all $\zeta \in \partial d_Q(z)$.*
(7) *Finally if Z has a strictly convex dual Z^*, then $\partial d_Q(z)$ consists of one point for each $z \notin Q$.*

We need another lemma for the proof of Theorem 6 to guarantee that the multipliers are nontrivial. Recall that a subset Q of a Banach space Z is said to be *finite codimensional* in Z if there exists a point z_0 in the convex closure of Q such that the closed subspace spanned by $Q - z_0 \equiv \{q - z_0 \mid q \in Q\}$ is a finite-codimensional subspace of Z and the convex closure of $Q - z_0$ has a nonempty interior in this subspace. See [8, pp. 142, 135], for the proof of the next lemma.

Lemma 5. *Let $Q \subset Z$ be finite codimensional in Z and $\{f_k\} \subset Z^*$ satisfy*

$$\|f_k\|_{Z^*} \geqslant \delta > 0, \quad f_k \xrightarrow{*} f \in Z^*,$$
$$\langle f_k, z\rangle \geqslant -\varepsilon_k, \quad \forall z \in Q,\ k \geqslant 1,$$

where $\varepsilon_k \to 0$. Then $f \neq 0$.

Finally we state and prove the multiplier rule. Although the domain of the objective is a metric space, the conclusions have a classical flavor: at least one multiplier is not zero, the "derivative of the Lagrangian" is nonnegative, and lastly the multiplier determines a support linear functional for the translate $Q - S(w_0)$.

Theorem 6 *(Multiplier rule). Suppose that w_0 is a minimum point of $J(\cdot)$ subject to $S(\cdot) \in Q$. Suppose that Z has strictly convex dual Z^* and $Q \subset Z$ is closed, convex and finite codimensional. Then there exists $(\psi^0, \psi) \in \mathbb{R}^+ \times Z^*$ such that*

$$\begin{cases} |\psi^0|^2 + \|\psi\|^2_{Z^*} > 0, & (11.1)\\ \psi^0 z^0 + \langle \psi, z\rangle \geqslant 0 \quad \text{for all } (z^0, z) \in D_s(J, S)(w_0), & (11.2)\\ \langle \psi, \eta - S(w_0)\rangle \leqslant 0 \quad \text{for all } \eta \in Q. & (11.3)\end{cases}$$

Proof. The first step in the proof is to use the Ekeland variational principle to produce minimizers of approximating functionals J_ε. These will yield approximations to the required multipliers (ψ^0, ψ) and prove (11.3).

We may assume that $J(w_0) = 0$. Fix an $\varepsilon > 0$ and consider the penalized functional

$$J_\varepsilon(w) = \sqrt{\max\{J(w) + \varepsilon, 0\}^2 + d_Q\big(S(w)\big)^2}$$

for $w \in \mathcal{W}$. Note that $J_\varepsilon(w)$ is the distance from $(J(w) + \varepsilon, S(w))$ to $Q_0 \times Q$, where $Q_0 = (-\infty, 0]$, that is,

$$J_\varepsilon(w) = d_{Q_0\times Q}\big(J(w) + \varepsilon, S(w)\big) = \sqrt{d_{Q_0}\big(J(w) + \varepsilon\big)^2 + d_Q\big(S(w)\big)^2}, \tag{12}$$

where $d_{Q_0}(y^0) = \max\{y^0, 0\}$ for $y^0 \in \mathbb{R}$. By the Ekeland variational principle in Proposition 2 with $\lambda = \sqrt{\varepsilon}$, there exists $w_\varepsilon \in \mathcal{W}$ such that

$$\begin{cases} J_\varepsilon(w_\varepsilon) + \sqrt{\varepsilon}\, d(w_\varepsilon, w_0) \leqslant J_\varepsilon(w_0), & (13.1)\\ J_\varepsilon(w_\varepsilon) < J_\varepsilon(w) + \sqrt{\varepsilon}\, d(w, w_\varepsilon) \quad \text{for all } w \in \mathcal{W} \setminus \{w_\varepsilon\}. & (13.2)\end{cases}$$

Note that

$$\big(J(w_\varepsilon) + \varepsilon, S(w_\varepsilon)\big) \notin Q_0 \times Q. \tag{14}$$

Indeed, if $S(w_\varepsilon) \notin Q$, then (14) is automatic, and if $S(w_\varepsilon) \in Q$, then $J(w_\varepsilon) \geqslant J(w_0) \geqslant 0$ by minimality of w_0, which implies that $J(w_\varepsilon) + \varepsilon \notin Q_0$. Therefore, by Lemma 4(7) applied to $Q_0 \times Q$,

$$\partial d_{Q_0\times Q}\big(J(w_\varepsilon) + \varepsilon, S(w_\varepsilon)\big) = \big\{\big(\psi^0_\varepsilon, \psi_\varepsilon\big)\big\} \tag{15}$$

is a singleton with $|\psi^0_\varepsilon|^2 + \|\psi_\varepsilon\|^2_{Z^*} = 1$. We claim that $(\psi^0_\varepsilon, \psi_\varepsilon)$ has the following representation:

$$\big(\psi^0_\varepsilon, \psi_\varepsilon\big) = \big(\alpha^0_\varepsilon \lambda^0_\varepsilon, \alpha_\varepsilon \lambda_\varepsilon\big) \tag{16}$$

where $(\alpha^0_\varepsilon, \alpha_\varepsilon) \in [0, 1] \times [0, 1]$ and $(\lambda^0_\varepsilon, \lambda_\varepsilon)$ are defined as

$$\begin{aligned} \big(\alpha^0_\varepsilon, \alpha_\varepsilon\big) &= \left(\frac{d_{Q_0}(J(w_\varepsilon) + \varepsilon)}{J_\varepsilon(w_\varepsilon)}, \frac{d_Q(S(w_\varepsilon))}{J_\varepsilon(w_\varepsilon)}\right),\\ \big(\lambda^0_\varepsilon, \lambda_\varepsilon\big) &\in \partial d_{Q_0}\big(J(w_\varepsilon) + \varepsilon\big) \times \partial d_Q\big(S(w_\varepsilon)\big). \end{aligned} \tag{17}$$

Note that if $\alpha^0_\varepsilon = 0$ or $\alpha_\varepsilon = 0$, then λ^0_ε or λ_ε may not be unique. However, $(\alpha^0_\varepsilon \lambda^0_\varepsilon, \alpha_\varepsilon \lambda_\varepsilon)$ is always uniquely defined. To prove (16), denote $(y^0_\varepsilon, y_\varepsilon) = (J(w_\varepsilon) + \varepsilon, S(w_\varepsilon))$. Then by (15) and Lemma 4(4),

$$\psi^0_\varepsilon z^0 + \langle \psi_\varepsilon, z\rangle = d'_{Q_0\times Q}\big(\big(y^0_\varepsilon, y_\varepsilon\big); \big(z^0, z\big)\big), \tag{18}$$

for all $(z^0, z) \in \mathbb{R} \times Z$. As shown in (12), $d_{Q_0\times Q}$ is the composition of (d_{Q_0}, d_Q) with the norm $g(r_1, r_2) = [r_1^2 + r_2^2]^{1/2}$ in $\mathbb{R}^2$. Since g Lipschitz with rank 1, by the chain rule in (9), we get the following directional derivative of the distance function:

$$d'_{Q_0\times Q}\big(\big(y^0_\varepsilon, y_\varepsilon\big); \big(z^0, z\big)\big) = g'\big(\big(d_{Q_0}\big(y^0_\varepsilon\big), d_Q(y_\varepsilon)\big); \big(d'_{Q_0}\big(y^0_\varepsilon; z^0\big), d'_Q(y; z)\big)\big). \tag{19}$$

To rewrite (19), first assume that $d_{Q_0}(y^0_\varepsilon) \neq 0, d_Q(y_\varepsilon) \neq 0$, then the directional derivative in (19) exists, and by Lemma 4(4), $d'_{Q_0}(y^0_\varepsilon; z^0) = \langle \lambda^0_\varepsilon, z^0\rangle$ and $d'_Q(y; z) = \langle \lambda_\varepsilon, z\rangle$. It follows from (18) and (19) that

$$\begin{aligned}\psi_\varepsilon^0 z^0 + \langle \psi_\varepsilon, z\rangle &= \frac{d_{Q_0}(y_\varepsilon^0) d'_{Q_0}(y_\varepsilon^0; z^0) + d_Q(y_\varepsilon) d'_Q(y; z)}{\sqrt{d_{Q_0}(y_\varepsilon^0)^2 + d_Q(y_\varepsilon)^2}} \\ &= \frac{\langle d_{Q_0}(y_\varepsilon^0)\lambda_\varepsilon^0, z^0\rangle + \langle d_Q(y)\lambda_\varepsilon, z\rangle}{\sqrt{d_{Q_0}(y_\varepsilon^0)^2 + d_Q(y_\varepsilon)^2}} \\ &= \big\langle \alpha_\varepsilon^0 \lambda_\varepsilon^0, z^0\big\rangle + \langle \alpha_\varepsilon \lambda_\varepsilon, z\rangle. \end{aligned} \tag{20}$$

Note that (20) continues to hold if only one of $d_{Q_0}(y_\varepsilon^0)$ and $d_Q(y_\varepsilon)$ is zero, and (14) implies that this is the case. Since (z^0, z) are arbitrary, (16) follows from (20).

Now consider a sequence $\varepsilon_k \downarrow 0$ as $k \to \infty$. The preceding discussion leads to sequences (with subindex ε_k changed to k) $w_k, \alpha_k^0, \alpha_k, \lambda_k^0, \lambda_k$, and $(\psi_k^0, \psi_k) \in \mathbb{R} \times Z^*$ satisfying (16) and (17) with $\varepsilon = \varepsilon_k$. By passing to a subsequence, we may assume that as $k \to \infty$,

$$\begin{aligned} &\big(\alpha_k^0, \alpha_k\big) \to (\alpha^0, \alpha) \in [0,1] \times [0,1], \\ &\big(\lambda_k^0, \lambda_k\big) \to \big(\lambda_0^0, \lambda_0\big) \quad \text{(weakly in } \mathbb{R} \times Z^*\text{)}, \quad \text{and} \\ &\text{either} \quad (1)\ \lambda_k = 0 \text{ for all } k \quad \text{or} \quad (2)\ \lambda_k \in \partial d_Q\big(S(w_k)\big) \quad \text{for all } k. \end{aligned}$$

Clearly $(\psi_k^0, \psi_k) \to (\alpha^0 \lambda_0^0, \alpha\lambda_0) \equiv (\psi^0, \psi)$. In the case (1), $\psi = \alpha\lambda_0 = 0$, which clearly satisfies (11.3). In the case (2), $\lambda_0 \in \partial d_Q(S(w_0))$ by the pseudo-continuity of the subdifferential in Lemma 4(5). In this case $\psi = \alpha\lambda_0$ also satisfies (11.3) by (10) for $\eta \in Q$.

Next we prove (11.2) by using (13.2). Let $(z^0, z) \in D_s(J, S)(w_0)$. By definition of D_s there exists $\delta_k = \delta(w_k) \downarrow 0$ so that

$$(z^0, z) \in D^{\delta_k}\big(J(\cdot), S(\cdot)\big)(w_k). \tag{21}$$

Again using the fact that J_{ε_k} is the composite function of $(J(\cdot) + \varepsilon_k\ , S(\cdot))$ and the distance function $d_{Q_0 \times Q}(\cdot,\cdot)$, which is Lipschitz with rank 1, (19), (21) and Proposition 3(a) imply that

$$\psi_k^0 z^0 + \langle \psi_k, z\rangle = d'_{Q_0\times Q}\big(\big(J(w_k) + \varepsilon_k, S(w_k)\big); (z^0, z)\big) \in D^{\delta_k} J_{\varepsilon_k}(w_k).$$

Now it follows from Proposition 2 that

$$\psi_k^0 z^0 + \langle \psi_k, z\rangle \geqslant -\sqrt{\varepsilon_k} - \delta_k. \tag{22}$$

Taking a limit as $k \to \infty$ in (22), we obtain (11.2) for all $(z^0, z) \in D_s(J, S)(w_0)$.

Finally we use Lemma 5 with $f_k = (\psi_k^0, \psi_k)$ and $f = (\psi^0, \psi)$ to show that (11.1) holds. Since $\|((\psi_k^0, \psi_k))\|^2 = |\psi_k^0|^2 + \|\psi_k\|_{Z^*}^2 = 1$, the condition that the norm of f_k be bounded away from zero is satisfied. Note that for each $\eta \in Q$, as $k \to \infty$, by (11.3)

$$\big\langle \psi_k, -\eta + S(w_0)\big\rangle \to \big\langle \psi, -\eta + S(w_0)\big\rangle \geqslant 0.$$

Adding this to (22) we get

$$\psi_k^0 z^0 + \big\langle \psi_k, z - \eta + S(w_0)\big\rangle \geqslant -\sqrt{\varepsilon_k} - \delta_k + \big\langle \psi_k, -\eta + S(w_0)\big\rangle,$$

for all $\eta \in Q$. So we can use $\min\{0, -\sqrt{\varepsilon_k} - \delta_k + \langle \psi_k, -\eta + S(w_0)\rangle\}$ for $-\varepsilon_k$ in the lemma. Because Q, and hence $z - Q - S(u_0)$, is finite codimensional, Lemma 5 implies that $(\psi^0, \psi) \neq 0$. □

Note that the proof shows that the element $\psi \in Z^*$ can be realized as a scalar multiple of an element in the subdifferential $\partial d_Q(S(w_0))$ where the scalar is in [0, 1].

References

[1] V.M. Alekseev, V.M. Tikhomirov, S.V. Fomin, Optimal Control, Plenum, New York, 1987.
[2] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[3] I. Ekeland, Nonconvex minimization problems, Bull. Amer. Math. Soc. (N.S.) 1 (3) (1979) 443–474.
[4] W.H. Fleming, R.W. Rishel, Deterministic and Stochastic Optimal Control, Springer-Verlag, Berlin, 1975.
[5] H. Halkin, Implicit functions and optimization problems without continuous differentiability of the data, SIAM J. Control 12 (1974) 229–236.

[6] M.R. Hestenes, Calculus of Variations and Optimal Control Theory, Wiley, New York, 1966.
[7] D. Luenberger, Optimization by Vector Space Methods, Wiley, New York, 1969.
[8] X. Li, J.M. Yong, Optimal Control Theory for Infinite Dimensional Systems, Birkhäuser, Boston, 1995.
[9] D.H. Martin, R.J. Gardner, G.G. Watkins, Indicating cones and the intersection principle for tangential approximants in abstract multiplier rules, J. Optim. Theory Appl. 33 (4) (1981) 515–537.
[10] M. McAsey, L. Mou, A proof of a general maximum principle for optimal controls via a multiplier rule on metric space, J. Math. Anal. Appl., in press.
[11] B.S. Mordukhovich, On necessary conditions for an extremum in nonsmooth optimization, Dokl. Akad. Nauk SSSR 283 (4) (1985) 816–822.
[12] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation I, Basic Theory, Springer, New York, 2006.
[13] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation II, Applications, Springer, New York, 2006.
[14] B.H. Pourciau, Modern multiplier rules, Amer. Math. Monthly 87 (6) (1980) 433–452.
[15] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[16] J.S. Treiman, Lagrange multipliers for nonconvex generalized gradients with equality, inequality, and set constraints, SIAM J. Control Optim. 37 (5) (1999) 1313–1329.
[17] X. Wang, V. Jeyakumar, A sharp Lagrange multiplier rule for nonsmooth mathematical programming problems involving equality constraints, SIAM J. Optim. 10 (4) (2000) 1136–1148.
[18] J.J. Ye, Nondifferentiable multiplier rules for optimization and bilevel optimization problems, SIAM J. Optim. 15 (1) (2004) 252–274.