Mediterr. j. math. 2 (2005), 437–450
1660-5446/040437-14, DOI 10.1007/s00009-005-0056-4

Mediterranean Journal of Mathematics

Proximal Calculus on Riemannian Manifolds

Daniel Azagra and Juan Ferrera

Abstract. We introduce a proximal subdifferential and develop a calculus for nonsmooth functions defined on any Riemannian manifold M. We give some applications of this theory, concerning, for instance, a Borwein-Preiss type variational principle on a Riemannian manifold M, as well as differentiability and geometrical properties of the distance function to a closed subset C of M.

Mathematics Subject Classification (2000). 49J52, 58E30, 58C30, 47H10.

Keywords. Proximal subdifferential, Riemannian manifold, variational principle, mean value theorem.

1. Introduction

The proximal subdifferential of lower semicontinuous real-valued functions is a very powerful tool which has been extensively studied and used in problems of optimization, control theory, differential inclusions, Lyapunov Theory, stabilization, and Hamilton-Jacobi equations (see [6] and references therein).

In this paper we will introduce a notion of proximal subdifferential for functions defined on a Riemannian manifold M (either finite or infinite dimensional) and we will develop the rudiments of a calculus for nonsmooth functions defined on M. Next we will prove an important result concerning inf-convolutions of lower semicontinuous functions with squared distance functions on M, from which a number of interesting consequences are deduced. For instance, we show a Borwein-Preiss type variational principle for functions defined on M, and we study some differentiability and geometrical properties of the distance function to a closed subset C of M.

This paper should be compared with [5], where a theory of viscosity subdifferentials for functions defined on Riemaniann manifolds is established and applied to

The first-named author was supported by a Marie Curie Intra-European Fellowship of the European Community, Human Resources and Mobility Programme under contract number MEIF CT2003-500927. The second-named author was supported by BFM2003-06420.

show existence and uniqueness of viscosity solutions to Hamilton-Jacobi equations on such manifolds.

In a sequel to the present paper we will elaborate on the applications of this proximal calculus on Riemannian manifolds, establishing a decrease principle from which one can deduce new theorems about functions having fixed points even after perturbing them with Lipschitzian functions (see [3]).

2. The main tools on proximal subdifferential

Let us recall the definition of the proximal subdifferential for functions defined on a Hilbert space X. A vector $\zeta \in X$ is called a *proximal subgradient* of a lower semicontinuous function f at $x \in dom f := \{y \in X : f(y) < +\infty\}$ provided there exist positive numbers σ and η such that

$$f(y) \geq f(x) + \langle \zeta, y - x \rangle - \sigma \|y - x\|^2 \text{ for all } y \in B(x, \eta).$$

The set of all such ζ is denoted $\partial_P f(x)$, and is referred to as the *proximal subdifferential*, or P-subdiferential. A comprehensive study of this subdifferential and its numerous applications can be found in [6].

Before giving the definition of proximal subdifferential for a function defined on a Riemannian manifold, we must establish a few preliminary results.

The following result is proved in [4, Corollary 2.4].

Proposition 2.1. *Let X be a real Hilbert space, and $f : X \longrightarrow (-\infty, +\infty]$ be a proper, lower semicontinuous function. Then,*

$$\partial_p f(x) = \{\varphi'(x) : \varphi \in C^2(X, \mathbb{R}),\ f - \varphi \text{ attains a local minimum at } x\}.$$

In particular this implies that $\partial_P f(x) \subseteq D^- f(x)$, where $D^- f(x)$ is the viscosity subdifferential of f at x.

Lemma 2.2. *Let X_1 and X_2 be two real Hilbert spaces, $\Phi : X_2 \to X_1$ a C^2 diffeomorphism, $f : X_1 \to (-\infty, +\infty]$ a lower semicontinuous function. Then $v \in \partial_P f(x_1)$ if and only if $D\Phi(x_2)^*(v) \in \partial(f \circ \Phi)(x_2)$, where $\Phi(x_2) = x_1$.*

Proof. This is a trivial consequence of Proposition 2.1, bearing in mind that compositions with diffeomorphisms preserve local minima. □

Corollary 2.3. *Let M be a Riemannian manifold, $p \in M$, (φ_i, U_i) $i = 1, 2$, two charts with $p \in U_1 \cap U_2$, and $\varphi_i(p) = x_i$. Then $\partial_P(f \circ \varphi_1^{-1})(x_1) \neq \emptyset$ if and only if $\partial_P(f \circ \varphi_2^{-1})(x_2) \neq \emptyset$. Moreover, $D(\varphi_1 \circ \varphi_2^{-1})(x_2)^*(\partial_P(f \circ \varphi_1^{-1})(x_1)) = \partial_P(f \circ \varphi_2^{-1})(x_2)$.*

Now we can extend the notion of P-subdifferential to functions defined on a Riemannian manifold.

Notation. In the sequel, M will stand for a Riemannian manifold defined on a real Hilbert space X (either finite dimensional or infinite dimensional). As usual, for a point $p \in M$, TM_p will denote the tangent space of M at p, and $\exp_p : TM_p \to M$

will stand for the exponential function at p. Recall that $\exp_p$ maps straight lines of the tangent space TM_p passing through $0_p \in TM_p$ into geodesics of M passing through p.

We will also use the parallel transport of vectors along geodesics. Recall that, for a given curve $\gamma : I \to M$, numbers $t_0, t_1 \in I$, and a vector $V_0 \in TM_{\gamma(t_0)}$, there exists a unique parallel vector field $V(t)$ along $\gamma(t)$ such that $V(t_0) = V_0$. Moreover, the mapping defined by $V_0 \mapsto V(t_1)$ is a linear isometry between the tangent spaces $TM_{\gamma(t_0)}$ and $TM_{\gamma(t_1)}$, for each $t_1 \in I$. In the case when γ is a minimizing geodesic and $\gamma(t_0) = x$, $\gamma(t_1) = y$, we will denote this mapping by L_{xy}, and we call it the parallel transport from TM_x to TM_y along the curve γ (see [9] for general reference on these topics).

The parallel transport allows us to measure the length of the "difference" between vectors (or forms) which are in different tangent spaces (or in duals of tangent spaces, that is, fibers of the cotangent bundle), and do so in a natural way. Indeed, let γ be a minimizing geodesic connecting two points $x, y \in M$, say $\gamma(t_0) = x, \gamma(t_1) = y$. Take vectors $v \in TM_x$, $w \in TM_y$. Then we can define the distance between v and w as the number

$$\|v - L_{yx}(w)\|_x = \|w - L_{xy}(v)\|_y$$

(this equality holds because L_{xy} is a linear isometry between the two tangent spaces, with inverse L_{yx}). Since the spaces T^*M_x and TM_x are isometrically identified by the formula $v = \langle v, \cdot \rangle$, we can obviously use the same method to measure distances between forms $\zeta \in T^*M_x$ and $\eta \in T^*M_y$ lying on different fibers of the cotangent bundle.

Definition 2.4. *Let M be a Riemannian manifold, $p \in M$, $f : M \to (-\infty, +\infty]$ a lower semicontinuous function. We define the proximal subdifferential of f at p, denoted by $\partial_P f(p) \subset TM_p$, as $\partial_P (f \circ \exp_p)(0)$ (understood that $\partial_P f(p) = \emptyset$ for all $p \notin domf$).*

The following result is an immediate consequence of Lemma 2.2.

Proposition 2.5. *Let M be a Riemannian manifold, $p \in M$, (φ, U) a chart, with $p \in U$, and $f : M \to (-\infty, +\infty]$ a lower semicontinuous function. Then*

$$\partial_P f(p) = D\varphi(p)^* [\partial_P (f \circ \varphi^{-1})(\varphi(p))].$$

As a consequence of the definition of $\partial_P (f \circ \exp_p)(0)$ we get the following result.

Corollary 2.6. *Let M be a Riemannian manifold, $p \in M$, $f : M \to (-\infty, +\infty]$ a lower semicontinuous function. Then $\zeta \in \partial_P f(p)$ if and only if there is a $\sigma > 0$ such that*

$$f(q) \geq f(p) + \langle \zeta, \exp_p^{-1}(q) \rangle - \sigma d(p, q)^2$$

for every q in a neighborhood of p.

We can also define the proximal superdifferential of a function f from a Hilbert space X into $[-\infty, +\infty)$ as follows. A vector $\zeta \in X$ is called a *proximal supergradient* of an upper semicontinuous function f at $x \in dom f$ if there are positive numbers σ and η such that

$$f(y) \leq f(x) + \langle \zeta, y - x \rangle + \sigma \|y - x\|^2 \text{ for all } y \in B(x, \eta)$$

and we denote the set of all such ζ by $\partial^P f(x)$, which we call P-subdifferential of f at x.

Now, let M be a Riemannian manifold, $p \in M$, $f : M \to [-\infty, +\infty)$ an upper semicontinuous function. We define the proximal superdifferential of f at p, denoted by $\partial^P f(p) \subset TM_p$, as $\partial^P(f \circ \exp_p)(0)$. As before, we have that $\zeta \in \partial^P f(p)$ if and only if there is a $\sigma > 0$ such that

$$f(q) \leq f(p) + \langle \zeta, \exp_p^{-1}(q) \rangle + \sigma d(p, q)^2$$

for every q in a neighborhood of p. It is also clear that $\partial^P f(p) = -\partial_P(-f)(p)$.

Before going into a study of the properties and applications of this proximal subdifferential, let us recall Ekeland's approximate version of the Hopf-Rinow theorem for infinite dimensional Riemannian manifolds (see [8]). In some of our proofs we will use Ekeland's theorem for the cases where the complete manifold M is infinite dimensional (so we cannot ensure the existence of a geodesic joining any two given points in the same connected component of M).

Theorem 2.7 (Ekeland). *If M is an infinite dimensional Riemannian manifold which is complete and connected then, for any given point p, the set*

$$\{q \in M : q \text{ can be joined to } p \text{ by a unique minimizing geodesic}\}$$

is dense in M.

3. Properties and applications of the proximal subdifferential

Most of the following properties are easily translated from the corresponding ones for $M = X$ a Hilbert space through charts (see [6]). Recall that a real-valued function f defined on a Riemannian manifold is said to be convex provided its composition $f \circ \alpha$ with any geodesic arc $\alpha : I \to M$ is convex as a function from $I \subset \mathbb{R}$ into $\mathbb{R}$.

Proposition 3.1. *Let M be a Riemannian manifold, $p \in M$, $f, g : M \to (-\infty, +\infty]$ lower semicontinuous functions. We have:*

(i) *if f is C^2, then $\partial_P f(p) = \{df(p)\}$;*
(ii) *if f is convex, then $\zeta \in \partial_P f(p)$ if and only if $f(q) \geq f(p) + \langle \zeta, v \rangle$ for every $q \in M$ and $v \in \exp_p^{-1}(q)$;*
(iii) *if f has a local minimum at p, then $0 \in \partial_P f(p)$;*
(iv) *every local minimum of a convex function f is global;*
(v) *if f is convex and $0 \in \partial_P f(p)$, then p is a global minimum of f;*
(vi) *$\partial_P f(p) + \partial_P g(p) \subseteq \partial_P (f + g)(p)$, with equality if f or g is C^2;*

(vii) $\partial_P(cf)(p) = c\partial_P f(p)$, *for* $c > 0$;
(viii) *if* f *is* K*-Lipschitz, then* $\partial_P f(p) \subset \overline{B}(0, K)$;
(ix) $\partial_P f(p)$ *is a convex subset of* TM_p;
(x) *if* $\zeta \in \partial_P f(p)$ *and* f *is differentiable at* p *then* $\zeta = df(p)$.

Proof. All the properties but perhaps (ii), (viii) and (x) are easily shown to be true. Property (viii) follows from the fact that $\exp_p^{-1}(\cdot)$ is almost 1-Lipschitz when restricted to balls of center 0_p and small radius.

Let us prove (ii). Let $q \in M$. Let $\gamma(t) = \exp_p(tv)$, $t \in [0, 1]$, which is a minimal geodesic joining p and q. The function $f \circ \gamma$ is convex and satisfies

$$\begin{aligned} f(\gamma(t)) &\geq f(\gamma(0)) + \langle \zeta, tv \rangle - \sigma d(\gamma(t)), \gamma(0))^2 \\ &= f(\gamma(0)) + \langle \zeta, t\gamma'(0)) \rangle - \sigma t^2 \end{aligned}$$

for some $\sigma > 0$ and $t > 0$ small. Hence $\langle \zeta, \gamma'(0)) \rangle \in \partial_P(f \circ \gamma)(0)$, and consequently (bearing in mind that $f \circ \gamma$ is convex on a Hilbert space) $f(\gamma(t)) \geq f(\gamma(0)) + \langle \zeta, t\gamma'(0)) \rangle$, which implies $f(q) \geq f(p) + \langle \zeta, v \rangle$.

To see (x), note that Proposition 2.1 implies that $\zeta \in D^- f(p)$, that is, ζ is a viscosity subdifferential of f at p in the sense of [5]. Then, since f is differentiable, we have that $\zeta \in D^- f(p) = D^+ f(p) = \{df(p)\}$, so we conclude that $\zeta = df(p)$. □

The following important result is also local and follows from [6, Theorem 1.3.1].

Theorem 3.2 (Density Theorem). *Let* M *be a Riemannian manifold,* $p \in M$, $f : M \to (-\infty, +\infty]$ *a lower semicontinuous function,* $\varepsilon > 0$. *Then there exists a point* q *such that* $d(p, q) < \varepsilon$, $f(p) - \varepsilon \leq f(q) \leq f(p)$, *and* $\partial_P f(q) \neq \emptyset$.

Now we arrive to one of the main results of this paper. We are going to extend the definition and main properties of the Moreau-Yosida regularization (see [1] for instance) to the category of functions defined on Riemannian manifolds of arbitrary dimension (finite or infinite).

Theorem 3.3. *Let* M *be a connected, complete Riemannian manifold, and let* $f : M \to \mathbb{R}$ *be a continuous function, bounded from below by a constant* c. *Then we have that for every* $\alpha > 0$ *the function*

$$f_\alpha(x) = \inf_{y \in M} \{f(y) + \alpha d(x, y)^2\}$$

is bounded from below by c, *it is Lipschitz on bounded sets and satisfies*

$$\lim_{\alpha \to +\infty} f_\alpha(x) = f(x).$$

Moreover, for every $x_0 \in M$ *with* $\partial_P f_\alpha(x_0) \neq \emptyset$, *there is a* $y_0 \in M$ *such that:*

a) *every minimizing sequence* $(y_n)_n$ *in the definition of* $f_\alpha(x_0)$ *converges to* y_0, *and consequently the infimum is a strong minimum;*
b) *there is a minimizing geodesic* γ *joining* x_0 *and* y_0;
c) f_α *is differentiable at* x_0;
d) $L_{x_0 y_0}[df_\alpha(x_0)] \in \partial_P f(y_0)$.

Finally, if we assume that M is finite dimensional (or, more generally, if we assume that M can be infinite dimensional but still has the property that every two points of M are connected by a minimizing geodesic) then the same remains true of every lower semicontinuous *function $f : M \to (-\infty, +\infty]$ which is bounded from below by c.*

Proof. It is clear that $f_\alpha(x) \geq \inf_{y\in M}\{c + \alpha d(x,y)^2\} = c$, and it is easily seen that $\lim_{\alpha\to+\infty} f_\alpha(x) = f(x)$. Let $A \subset M$ be a bounded set. We have that $f_\alpha(x) \leq f(z_0) + \alpha d(x, z_0)^2$ for a fixed z_0, hence there is a positive m such that $f_\alpha(x) \leq m$ provided $x \in A$. Let us consider $x, y \in A$ and $\varepsilon > 0$, choose $z = z_y \in M$ such that $f_\alpha(y) + \varepsilon \geq f(z) + \alpha d(y,z)^2$. We have that $d(y,z) \leq [\frac{1}{\alpha}(f_\alpha(y) + \varepsilon - c)]^{\frac{1}{2}} \leq [\frac{1}{\alpha}(m - c + \varepsilon)]^{\frac{1}{2}} := R$. Consequently,

$$\begin{aligned} f_\alpha(x) - f_\alpha(y) &\leq f_\alpha(x) - f(z) - \alpha d(y,z)^2 + \varepsilon \\ &\leq f(z) + \alpha d(x,z)^2 - f(z) - \alpha d(y,z)^2 + \varepsilon \\ &= \alpha(d(x,z) + d(y,z))(d(x,z) - d(y,z)) + \varepsilon \\ &\leq \alpha(2R + diam A)d(x,y) + \varepsilon. \end{aligned}$$

By letting ε go to 0, and changing x by y, we get that f_α is Lipschitz on A.

For the second part, fix $x_0 \in M$, $\zeta \in \partial_P f_\alpha(x_0)$, and a sequence $(y_n)_n$ such that $f(y_n) + \alpha d(x_0, y_n)^2$ converges to the infimum defining $f_\alpha(x_0)$. First of all let us observe that we can always assume that for each n there is a minimizing geodesic $\gamma_n : [0,1] \to M$ joining the point y_n to x_0. Indeed, for each couple of points y_n, x_0 we can apply Ekeland's Theorem 2.7 and continuity of f to find a point y_n' and a unique minimizing geodesic γ_n joining y_n' to x_0 in such a way that

$$d(y_n, y_n') \leq \frac{1}{n} \quad \text{and} \quad f(y_n') + \alpha d(x_0, y_n') \leq f(y_n) + \alpha d(x_0, y_n) + \frac{1}{n}.$$

Let us also notice that, if M is finite dimensional then we can directly apply the classical Hopf-Rinow theorem to find geodesics γ_n and we can dispense with the continuity assumption, thus requiring only that f be lower semicontinuous.

Since the sequence $(y_n)_n$ realizes the infimum defining $f_\alpha(x_0)$, so the sequence $(y_n')_n$ does. Then we can apply the argument which follows below to the sequence $(y_n')_n$ in order to find a point y_0 with the required properties. Finally the original sequence $(y_n)_n$ must also converge to y_0 because $d(y_n, y_n') \to 0$ as $n \to +\infty$. So, to save notation, we assume $y_n = y_n'$ for each n.

Because $\zeta \in \partial_P f_\alpha(x_0)$, there is $\sigma > 0$ such that, if y is in a neighborhood of x_0, we have

$$\langle \zeta, \exp_{x_0}^{-1}(y)\rangle \leq f_\alpha(y) - f_\alpha(x_0) + \sigma d(x_0, y)^2. \qquad (*)$$

Now, define $\varepsilon_n \geq 0$ by $f_\alpha(x_0) + \varepsilon_n^2 = f(y_n) + \alpha d(y_n, x_0)^2$. We have $\lim_n \varepsilon_n = 0$. From $(*)$, it follows that

$$\begin{aligned} \langle \zeta, \exp_{x_0}^{-1}(y)\rangle &\leq f(y_n) + \alpha d(y_n, y)^2 - [f(y_n) + \alpha d(y_n, x_0)^2 - \varepsilon_n^2] + \sigma d(x_0, y)^2 \\ &= \alpha d(y_n, y)^2 - \alpha d(y_n, x_0)^2 + \sigma d(x_0, y)^2 + \varepsilon_n^2, \end{aligned}$$

because $f_\alpha(y) \leq f(y_n) + \alpha d(y_n, y)^2$. Particularizing for $y = \exp_{x_0}(\varepsilon_n v)$, with $v \in TM_{x_0}$, $\|v\| = 1$, we have

$$\varepsilon_n \langle \zeta, v\rangle \leq (\sigma+1)\varepsilon_n^2 + \alpha[d(y_n, y)^2 - d(y_n, x_0)^2]. \tag{$**$}$$

Now, let us choose t_n close enough to 1 in order to ensure that the function $d(\cdot, \hat{x}_n)$ is differentiable at x_0, where $\hat{x}_n = \gamma_n(t_n)$. Let us denote the length of $\gamma_{n|[0,t_n]}$ by l_n. By using Taylor's Theorem, we have that:

$$\begin{aligned}
d(y_n, y)^2 - d(y_n, x_0)^2 &\leq (d(y, \hat{x}_n) + l_n)^2 - l(\gamma_n)^2 \\
&\leq (d(y, \hat{x}_n) + l_n)^2 - (d(x_0, \hat{x}_n) + l_n)^2 = \varphi(y) - \varphi(x_0) \\
&= \varphi'(x_0)(\varepsilon_n v) + \varphi''(\exp_{x_0}^{-1}(\lambda \varepsilon_n v))(\varepsilon_n v) \\
&= \varphi'(x_0)(\varepsilon_n v) + \varphi''(x)(\varepsilon_n v),
\end{aligned}$$

where $x = \exp_{x_0}^{-1}(\lambda \varepsilon_n v)$,

$$\varphi(y) = (d(y, \hat{x}_n) + l_n)^2,\ \varphi'(x_0) = 2(d(x_0, \hat{x}_n) + l_n)\frac{\partial d(x_0, \hat{x}_n)}{\partial x},$$

$$\varphi''(x) = 2[\frac{\partial d(x, \hat{x}_n)}{\partial x}]^2 + 2(d(x, \hat{x}_n) + l_n)\frac{\partial^2 d(x, \hat{x}_n)}{\partial x^2},$$

and $0 < \lambda < 1$. Hence,

$$\begin{aligned}
d(y_n, y)^2 - d(y_n, x_0)^2 \ &\leq\ 2\varepsilon_n(d(x_0, \hat{x}_n) + l_n)\frac{\partial d(x_0, \hat{x}_n)}{\partial x}(v) \\
&\quad +2\varepsilon_n^2\left[\frac{\partial d(x, \hat{x}_n)}{\partial x}(v)\right]^2 + 2(d(x, \hat{x}_n) + l_n)\frac{\partial^2 d(x, \hat{x}_n)}{\partial x^2}(\varepsilon_n v) \\
&\leq 2\varepsilon_n(d(x_0, \hat{x}_n) + l_n)\frac{\partial d(x_0, \hat{x}_n)}{\partial x}(v) + 2(d(x, \hat{x}_n) + l_n)\frac{\partial^2 d(x, \hat{x}_n)}{\partial x^2}(\varepsilon_n v) + 2\varepsilon_n^2 \\
&\leq 2\varepsilon_n(d(x_0, \hat{x}_n) + l_n)\frac{\partial d(x_0, \hat{x}_n)}{\partial x}(v) + \left[\frac{2(d(x, \hat{x}_n) + l_n)}{d(x, \hat{x}_n)} + 2\right]\ \varepsilon_n^2,
\end{aligned}$$

since $\|\frac{\partial d(x,\hat{x}_n)}{\partial x}\| = 1$ and $\|\frac{\partial^2 d(x,\hat{x}_n)}{\partial x^2}\| = \frac{1}{d(x,\hat{x}_n)}$. On the other hand, firstly we may assume that the sequence $(y_n)_n$ is bounded, hence so it is $(l_n)_n$; and secondly that $(\hat{x}_n)_n$ is uniformly away from x_0, hence $\frac{2(d(x,\hat{x}_n)+l_n)}{d(x,\hat{x}_n)} + 2$ is bounded by a constant K. Therefore,

$$d(y_n, y)^2 - d(y_n, x_0)^2 \leq 2\varepsilon_n(d(x_0, \hat{x}_n) + l_n)\frac{\partial d(x_0, \hat{x}_n)}{\partial x}(v) + K\varepsilon_n^2,$$

and from $(**)$ we get

$$\varepsilon_n\langle \zeta - 2\alpha(d(x_0, \hat{x}_n) + l_n)\frac{\partial d(x_0, \hat{x}_n)}{\partial x}, v\rangle \leq (K + \sigma + 1)\varepsilon_n^2.$$

This implies that

$$\lim_n \|\zeta - 2\alpha(d(x_0, \hat{x}_n) + l_n)\frac{\partial d(x_0, \hat{x}_n)}{\partial x}\|_{x_0} = 0.$$

From $\|\frac{\partial d(x_0,\hat{x}_n)}{\partial x}\|_{x_0} = 1$, it follows that

$$\lim_n d(x_0,\hat{x}_n) + l_n = \frac{1}{2\alpha}\|\zeta\|_{x_0} \qquad \text{and} \qquad \lim_n \frac{\partial d(x_0,\hat{x}_n)}{\partial x} = \frac{\zeta}{\|\zeta\|}.$$

Since γ_n are minimizing geodesics, we can deduce that

$$y_n = \exp_{x_0}[-(d(x_0,\hat{x}_n) + l_n)\frac{\partial d(x_0,\hat{x}_n)}{\partial x}].$$

Finally, the expected y_0 is necessarily $y_0 = \exp_{x_0}(-\frac{1}{2\alpha}\zeta)$. This proves (a).

The announced geodesic in part (b) is $\gamma : [0,1] \to M$ defined by $\gamma(t) = \exp_{x_0}(-\frac{1}{2\alpha}t\zeta)$, which is minimizing because

$$d(x_0,y_0) = \lim_n d(x_0,y_n) = \lim_n (d(x_0,\hat{x}_n) + l_n) = \frac{1}{2\alpha}\|\zeta\|_{x_0}.$$

In order to show (c), we observe that, for y near x_0, we have

$$f_\alpha(x_0) - f_\alpha(y) \geq f(y_0) + \alpha d(y_0,x_0)^2 - f(y_0) - \alpha d(y_0,y)^2 = \alpha[d(y_0,x_0)^2 - d(y_0,y)^2],$$

hence, using Taylor's Theorem again,

$$\begin{aligned} f_\alpha(y) &\leq f_\alpha(x_0) + \alpha[d(y_0,y)^2 - d(y_0,x_0)^2] = f_\alpha(x_0) + \alpha(\psi(y) - \psi(x_0)) \\ &= f_\alpha(x_0) + \alpha\psi'(x_0)(\exp_{x_0}^{-1}(y)) + \alpha\psi''(x_0)(\exp_{x_0}^{-1}(y)) + o(\|\exp_{x_0}^{-1}(y)\|^2) \\ &\leq f_\alpha(x_0) + \alpha\psi'(x_0)(\exp_{x_0}^{-1}(y)) + Cd(x_0,y)^2, \end{aligned}$$

where $\psi(y) = (d(\hat{y}_0,y) + d(\hat{y}_0,y_0))^2$ is C^2 at x_0, for some $\hat{y}_0$ lying on γ. This implies that $\alpha\psi'(x_0) \in \partial^P f_\alpha(x_0)$ and therefore f_α is differentiable at x_0.

Part (d) is trivial if $x_0 = y_0$. Otherwise the function $f + \alpha d(x_0,\cdot)^2 = f + \alpha[d(x_0,\hat{x}_0) + d(\hat{x}_0,\cdot)]^2$ attains its minimum at y_0 and therefore

$$0 \in \partial_P(f + \alpha d(x_0,\cdot)^2)(y_0) = \partial_P f(y_0) + 2\alpha d(x_0,y_0)\frac{\partial d(\hat{x}_0,y_0)}{\partial y}$$

since $[d(x_0,\hat{x}_0) + d(\hat{x}_0,\cdot)]^2$ is C^2 at y_0, provided that $\hat{x}_0 \in \gamma$ and is close enough to y_0. Then, according to the antisymmetry property of the partial derivatives of the distance function (see [5, Lemma 6.5]), we have

$$\begin{aligned} L_{x_0y_0}[df_\alpha(x_0)] &= L_{x_0y_0}\left[2\alpha d(x_0,y_0)\frac{\partial d(x_0,\hat{y}_0)}{\partial x}\right] \\ &= -2\alpha d(x_0,y_0)\frac{\partial d(\hat{x}_0,y_0)}{\partial y} \in \partial_P f(y_0). \end{aligned}$$

Let us observe that, as a consequence of part (c), the minimizing geodesic joining x_0 and y_0 is unique. $\square$

Now, we deduce a Borwein-Preiss variational principle for continuous functions defined on any complete Riemannian manifold M. Let us recall that, when M is infinite dimensional, generally a bounded continuous function $f : M \to \mathbb{R}$ does not attain any minima. In fact, as shown recently in [2], the set of smooth functions with no critical points is dense in the space of continuous functions on M. Therefore, in optimization problems one has to resort to perturbed minimization results, such as Ekeland's variational principle (which is applicable to any complete Riemannian manifold). Apart from Ekeland's result we have at least two other options.

If one wants to perturb a given function with a small smooth function which has a small derivative everywhere, in such a way that the sum of the two functions does attain a minimum, then one can use a Deville-Godefroy-Zizler smooth variational principle (originally proved for Banach spaces). An extension of the DGZ smooth variational principle is established in [5] for functions defined on those Riemannian manifolds which are *uniformly bumpable*.

If we wish to perturb the original function with small multiples of squares of distance functions (so that, among other interesting properties, we get local smoothness of the perturbing function near the approximate minimizing point) we can use the following Borwein-Preiss type variational principle.

Theorem 3.4. *Let M be a complete, connected Riemannian manifold, $f : M \to \mathbb{R}$ a continuous function which is bounded from below, and $\varepsilon > 0$. Let $x_0 \in M$ be such that $f(x_0) < \inf f + \varepsilon$. Then, for every $\lambda > 0$ there exist $z \in B(x_0, \lambda)$, $y \in B(z, \lambda)$ with $f(y) \leq f(x_0)$, and such that the function $\varphi(x) = f(x) + \frac{\varepsilon}{\lambda^2} d(x, z)^2$ attains a strong minimum at y.*

On the other hand, if M has the property that every two points of M are connected by a minimizing geodesic (such is the case, for instance, of any finite dimensional *manifold), then it is enough to assume that $f : M \to (-\infty, +\infty]$ is lower semicontinuous.*

Proof. Let us consider f_α as in Theorem 3.3, with $\alpha = \frac{\varepsilon}{\lambda^2}$. According to the Density Theorem 3.2, there is a z such that $d(x_0, z) \leq \lambda$, $\partial_P f_\alpha(z) \neq \emptyset$, and $f_\alpha(z) \leq f_\alpha(x_0) \leq f(x_0)$. Hence, by Theorem 3.3, φ attains a strong minimum at a point y_0.

Finally, $f(y_0) + \frac{\varepsilon}{\lambda^2} d(y_0, z)^2 = f_\alpha(z) \leq f(x_0) < \inf f + \varepsilon \leq f(y_0) + \varepsilon$, hence $d(y_0, z) < \lambda$. □

Next, as an application of Theorem 3.3 we establish three results concerning differentiability and geometrical properties of the distance function to a closed subset S of a Riemannian manifold M. These properties are probably known by the specialists, though we do not know of any suitable reference.

Theorem 3.5. *Let S be a nonempty closed subset of a complete connected finite dimensional Riemannian manifold M (or else, an infinite dimensional manifold M with the property that every two points of M are connected by a minimizing geodesic), and $x \in M - S$. If $\partial_P d_S(x) \neq \emptyset$, then d_S is differentiable at x. Moreover, there is an $s_0 \in S$ such that:*

(a) *every minimizing sequence of $d_S(x)$ converges to s_0;*
(b) *$d_S(x) = d(x, s_0)$ and $d(x, s) > d_S(x)$ for every $s \in S$, $s \neq s_0$;*
(c) *there is a unique minimizing geodesic joining x and s_0.*

Proof. Let assume that $\xi \in \partial_P d_S(x)$, this implies that there is a $\sigma > 0$ such that

$$d_S(y) - d_S(x) \geq \langle \xi, \exp_x^{-1}(y) \rangle - \sigma d(x,y)^2$$

if y is near x. We have that

$$\begin{aligned} d_S^2(y) - d_S^2(x) &= 2d_S(x)(d_S(y) - d_S(x)) + (d_S(y) - d_S(x))^2 \\ &\geq 2d_S(x)(d_S(y) - d_S(x)) \geq 2d_S(x)[\langle \xi, \exp_x^{-1}(y) \rangle - \sigma d(x,y)^2], \end{aligned}$$

which implies that $2d_S(x)\xi \in \partial_P d_S^2(x)$. On the other hand $d_S^2(y) = \inf_{z \in M}\{I_S(z) + d(z,y)^2\} = f_\alpha(y)$ with $\alpha = 1$, $f = I_S$, where I_S is the indicator function of S, that is $I_S(z) = 0$ if $z \in S$, and $I_S(z) = +\infty$ otherwise. Therefore, properties (a), (b) and (c), which are equivalent for d_S and d_S^2, follow from Theorem 3.3, as well as the fact that d_S^2 is differentiable at x. Hence d_S is differentiable at x because $d_S(x) > 0$. □

By combining this with the Density Theorem 3.2 we get, under the same assumptions on M and S, the following.

Corollary 3.6. *There is a dense subset of points $x \in M - S$ such that $d_S(x) = d(x, s_x)$ for a unique $s_x \in S$ and d_S is differentiable at x.*

Corollary 3.7. *Let x, x_0 be two different points of a complete connected finite dimensional Riemannian manifold M (or, more generally, of a Riemannian manifold with the property that every two points can be connected by a minimizing geodesic). The following statements are equivalent:*

(i) *the function $d(\cdot, x_0)$ is subdifferentiable (in the proximal sense) at x;*
(ii) *the function $d(\cdot, x_0)$ is Fréchet differentiable at x;*
(iii) *there is a unique minimizing geodesic joining x and x_0.*

Proof. (i) $\Longrightarrow$ (ii) and the existence of a unique minimizing geodesic follow from Theorems 3.3 and 3.5.

Let us assume that there is a unique minimizing geodesic joining x and x_0. We may prolong the geodesic up to a point $\hat{x}$ satisfying $d(\hat{x}, x_0) = d(\hat{x}, x) + d(x, x_0)$. In order to prove that $d(\cdot, x_0)$ is subdifferentiable, it is enough to see that $\varphi(y) = d(\hat{x}, x_0) - d(y, x_0)$ is superdifferentiable at x, which is a consequence of the following inequalities:

$$\begin{aligned} \varphi(y) - \varphi(x) &= d(\hat{x}, x_0) - d(y, x_0) - d(\hat{x}, x) \leq d(y, \hat{x}) - d(\hat{x}, x) \\ &\leq \langle \frac{\partial d(x, \hat{x})}{\partial x}, \exp_x^{-1}(y) \rangle + \sigma d(x,y)^2. \end{aligned}$$

□

Finally, we turn to another topic of the theory of proximal subdifferentials, namely mean value theorems. We begin with a few local results which will be used in the proof of a proximal mean value inequality. The following result can be deduced from [6, Theorem 1.8.3].

Theorem 3.8 (Fuzzy rule for the sum). *Let $f_1, f_2 : M \to (-\infty, +\infty]$ be two lower semicontinuous functions such that at least one of them is Lipschitz near x_0. If $\zeta \in \partial_P(f_1 + f_2)(x_0)$ then, for every $\varepsilon > 0$, there exist x_1, x_2 and $\zeta_1 \in \partial_P f_1(x_1)$, $\zeta_2 \in \partial_P f_2(x_2)$ such that*

(a) *$d(x_i, x_0) < \varepsilon$ and $|f_i(x_i) - f_i(x_0)| < \varepsilon$ for $i = 1, 2$;*
(b) *$\|\zeta - (L_{x_1 x_0}(\zeta)_1 + L_{x_2 x_0}(\zeta_2))\|_{x_0} < \varepsilon$.*

The following theorem is also local, a consequence of the Fuzzy Chain Rule known for functions defined on Hilbert spaces (see [6, Theorem 1.9.1, pp. 59]).

Theorem 3.9 (Fuzzy chain rule). *Let $g : N \to \mathbb{R}$ be lower semicontinuous, $F : M \to N$ be locally Lipschitz, and assume that g is Lipschitz near $F(x_0)$. Then, for every $\zeta \in \partial_P(g \circ F)(x_0)$ and $\varepsilon > 0$, there are $\tilde{x}$, $\tilde{y}$ and $\eta \in \partial_P g(\tilde{y})$ such that $d(\tilde{x}, x_0) < \varepsilon$, $d(\tilde{y}, F(x_0)) < \varepsilon$, $d(F(\tilde{x}), F(x_0)) < \varepsilon$, and*

$$L_{x\tilde{x}}\zeta \in \partial_P[\langle L_{\tilde{y}F(x_0)}(\eta), \exp^{-1}_{F(x_0)} \circ F(\cdot)\rangle](\tilde{x}) + \varepsilon B_{TM_{\tilde{x}}}.$$

The following result, which is local as well, relates the proximal subdifferential $\partial_P f(x)$ to the viscosity subdifferential $D^- f(x)$ of a function f defined on a Riemannian manifold M (see [5] for the definition of $D^- f(x)$ in the manifold setting).

Proposition 3.10. *Let $\xi_0 \in D^- f(x_0)$, $\epsilon > 0$. Then there exist $x \in B(x_0, \epsilon)$ and $\zeta \in \partial_P f(x)$ such that $|f(x) - f(x_0)| < \epsilon$ and $\|\xi_0 - L_{xx_0}(\zeta)\|_{x_0} < \epsilon$.*

Proof. This follows from [6, Proposition 3.4.5, pp. 138]. □

The above result is a *fuzzy* converse of the following obvious inclusion:

$$\partial_P f(x) \subseteq D^- f(x).$$

We conclude with the announced mean value inequality for the proximal subdifferential.

Theorem 3.11 (Proximal Mean Value Theorem). *Let $x, y \in M$, $\gamma : [0, T] \to M$ be a path joining x and y. Let f be a Lipschitz function around $\gamma[0, T]$. Then, for every $\varepsilon > 0$, there exist t_0, $z \in M$ and $\zeta \in \partial_P f(z)$ with $d(z, \gamma(t_0)) < \varepsilon$, and so that*

$$\frac{1}{T}\left(f(y) - f(x)\right) \leq \langle \zeta, L_{\gamma(t_0), z}(\gamma'(t_0))\rangle + \varepsilon.$$

Proof. Let us consider the function $\varphi : [0, T] \to \mathbb{R}$ defined as

$$\varphi(t) = f(\gamma(t)) - G(t),$$

where

$$G(t) = \frac{t}{T} f(y) + \frac{T - t}{T} f(x).$$

The function φ is continuous, and $\varphi(0) = \varphi(T) = 0$. Since the interval $[0, T]$ is compact, there exists $t_0 \in [0, T]$ such that $\varphi(t_0) \leq \varphi(t)$ for all $t \in [0, T]$. We will consider two cases.

Case 1. Assume that $t_0 \in (0,T)$. Since φ attains a local minimum at t_0, we know that $0 \in \partial_P \varphi(t_0)$. Since the function $G(t)$ is of class C^2, according to the *easy sum rule* Proposition 3.1(vi), we have that

$$\frac{1}{T}(f(y)-f(x)) = 0 + G'(t_0) \in \partial_P(f\circ\gamma)(t_0).$$

Now, by the fuzzy chain rule Theorem 3.9, there exist $\tilde{t}$, $\tilde{z}$, $\zeta \in \partial_P f(\tilde{z})$ such that $|\tilde{t}-t_0| < \varepsilon$, $d(\tilde{z},\gamma(t_0)) < \varepsilon$, $d(\gamma(\tilde{t}),\gamma(t_0)) < \varepsilon$, and

$$\begin{aligned}
\frac{1}{T}\left(f(y)-f(x)\right) &\in \partial_P \left(\langle L_{\tilde{z}\gamma(t_0)}(\zeta), \exp^{-1}_{\gamma(t_0)} \circ\gamma(\cdot)\rangle \right)(\tilde{t}) + [-\varepsilon,\varepsilon] \\
&= \frac{d}{dt}\left(\langle L_{\tilde{z}\gamma(t_0)}(\zeta), \exp^{-1}_{\gamma(t_0)} \circ\gamma(\cdot)\rangle \right)_{|_{t=\tilde{t}}} + [-\varepsilon,\varepsilon] \\
&= \langle L_{\tilde{z}\gamma(t_0)}(\zeta), \gamma'(t_0)\rangle + [-\varepsilon,\varepsilon] = \langle \zeta, L_{\gamma(t_0)\tilde{z}}\left(\gamma'(t_0)\right)\rangle + [-\varepsilon,\varepsilon].
\end{aligned}$$

In particular, we obtain that

$$\frac{1}{T}\left(f(y)-f(x)\right) \leq \langle \zeta, L_{\gamma(t_0),z}(\gamma'(t_0))\rangle + \varepsilon.$$

Case 2. Now let us suppose that $t_0 = 0$ or $t_0 = T$. Since $\varphi(0) = \varphi(T) = 0$, this means that $\varphi(t) \geq \varphi(0) = \varphi(T)$ for all $t \in [0,T]$. We may assume that φ attains no local minima in $(0,T)$ (otherwise the argument of Case 1 applies and we are done). Then there must exist $t_0' \in (0,T)$ such that φ is increasing on $(0,t_0')$ and φ is decreasing on (t_0',T). This implies that $\zeta \geq 0$ for every $\zeta \in \partial_P f(t)$ with $t \in (0,t_0')$, and $\eta \leq 0$ for every $\eta \in \partial_P f(t')$ with $t' \in (t_0',T)$. Indeed, assume for instance that $t \in (0,t_0')$ and take $\zeta \in \partial_P f(t)$. Then we have that $f(s) \geq f(t) + \zeta(s-t) - \sigma(s-t)^2$ for some $\sigma \geq 0$ and s in a neighborhood of t. By taking s close enough to t with $s < t$, we get

$$\zeta \geq \frac{f(t)-f(s)}{t-s} - \sigma(t-s),$$

and hence

$$\zeta \geq \liminf_{s\to t^-}\left[\frac{f(t)-f(s)}{t-s} - \sigma(t-s)\right] \geq 0.$$

Now, by the Density Theorem 3.2 there exist t_1, η_1 such that $t_1 \in (0,t_0')$ and $\eta_1 \in \partial_P\varphi(t_1)$. According to the preceding discussion, we have $\eta_1 \geq 0$. Since $G(t)$ is of class C^2, by the *easy sum rule* Proposition 3.1(vi), we have that

$$\eta_1 + \frac{1}{T}(f(y)-f(x)) = \eta_1 + G'(t_1) \in \partial_P(f\circ\gamma)(t_1).$$

Finally, by the fuzzy chain rule Theorem 3.9, there exist $\tilde{t}$, $\tilde{z}$, $\zeta \in \partial_P f(\tilde{z})$ such that $|\tilde{t} - t_1| < \varepsilon$, $d(\tilde{z}, \gamma(t_1)) < \varepsilon$, $d(\gamma(\tilde{t}), \gamma(t_1)) < \varepsilon$, and

$$\begin{aligned}
\eta_1 + \frac{1}{T}\left(f(y) - f(x)\right) &\in \partial_P \left(\langle L_{\tilde{z}\gamma(t_1)}(\zeta), \exp^{-1}_{\gamma(t_1)} \circ \gamma(\cdot)\rangle\right)(\tilde{t}) + [-\varepsilon, \varepsilon] \\
&= \frac{d}{dt}\left(\langle L_{\tilde{z}\gamma(t_1)}(\zeta), \exp^{-1}_{\gamma(t_1)} \circ \gamma(\cdot)\rangle\right)_{|_{t=\tilde{t}}} + [-\varepsilon, \varepsilon] \\
&= \langle L_{\tilde{z}\gamma(t_1)}(\zeta), \gamma'(t_1)\rangle + [-\varepsilon, \varepsilon] \\
&= \langle \zeta, L_{\gamma(t_1)\tilde{z}}\left(\gamma'(t_1)\right)\rangle + [-\varepsilon, \varepsilon].
\end{aligned}$$

In particular, we get

$$\frac{1}{T}\left(f(y) - f(x)\right) \leq \eta_1 + \frac{1}{T}\left(f(y) - f(x)\right) \leq \langle \zeta, L_{\gamma(t_1),z}(\gamma'(t_1))\rangle + \varepsilon. \qquad \square$$

References

[1] H. Attouch and R.J-B. Wets, *A convergence theory for saddle functions*. Trans. Amer. Math. Soc. **280** (1983), 1-41.

[2] D. Azagra and M. Cepedello, *Uniform approximation of continuous mappings by smooth mappings with no critical points on Hilbert manifolds*. Duke Math. J. **124** (2004), 47-66.

[3] D. Azagra and J. Ferrera, *Applications of proximal calculus to fixed point theory on Riemannian manifolds*. To appear on Nonlinear Anal.

[4] D. Azagra, J. Ferrera and F. López-Mesas, *Approximate Rolle's theorems for the proximal subgradient and the generalized gradient*. J. Math. Anal. Appl. **283** (2003), 180-191.

[5] D. Azagra, J. Ferrera and F. López-Mesas, *Nonsmooth analysis and Hamilton-Jacobi equations on Riemannian manifolds*. J. Funct. Anal. **220** (2005), 304-361.

[6] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern and P.R. Wolenski, *Nonsmooth Analysis and Control Theory*. Grad. Texts in Math. **178**, Springer, 1998.

[7] I.Ekeland, *Nonconvex minimization problems*. Bull. Amer. Math. Soc. (New series) **1** (1979), 443-474.

[8] I.Ekeland, *The Hopf-Rinow theorem in infinite dimension*. J. Differential Geom. **13** (1978), 287-301.

[9] W. Klingenberg, *Riemannian Geometry*, de Gruyter Stud. Math., de Gruyter & Co., Berlin-New York, 1982.

[10] C. Mantegazza and A.C. Mennucci, *Hamilton-Jacobi equations and distance functions on Riemannian manifolds*. Appl. Math. Optim. **47** (2003), 1-25.

Daniel Azagra and Juan Ferrera
Departamento de Análisis Matemático
Facultad de Matemáticas
Universidad Complutense
28040 Madrid
Spain
e-mail: azagra@mat.ucm.es
ferrera@mat.ucm.es

Received: February 3, 2005
Revised: September 28, 2005
Accepted: October 28, 2005