PACIFIC JOURNAL OF MATHEMATICS
Vol. 170 No. 2 1995

DIRECTIONAL DIFFERENTIABILITY OF THE METRIC PROJECTION IN HILBERT SPACE

DOMINIKUS NOLL

The differentiability properties of the metric projection P_C on a closed convex set C in Hilbert space are characterized in terms of the smoothness type of the boundary of C. Our approach is based on using variational type second derivatives as a sufficiently flexible tool to describe the boundary structure of the set C with regard to the differentiability of P_C. We extend results by R.B. Holmes and S. Fitzpatrick and R.R. Phelps.

1. Introduction.

Let H be a real Hilbert space, C a closed convex subset of H. Given an element x of H, consider the problem of finding the best approximation of x by elements of C. Let the unique best approximating element of C be denoted by P_Cx, or simply Px, that is

$$(1.1) \qquad ||x - Px|| = \inf\{||x-y|| : y \in C\}.$$

The operator P so defined is called the *nearest point mapping* or the *metric projection* onto C. One may ask how the solution Px of (1.1) behaves under slight perturbations of the data x.

Such sensitivity analysis of the best approximation problem is of course closely tied to the differentiability properties of the metric projection, and the latter therefore have been looked at by many authors. We mention in particular the work of E.H. Zarantonello [**29**, **30**], R.B. Holmes [**17**], S. Fitzpatrick and R. Phelps [**14**]. For other references see F. Mignot [**19**], A. Haraux [**15**], where applications to variational inequalities are considered, J. Sokolowski [**27**, **28**], K. Malanowski [**18**], or A. Shapiro [**25**, **26**] for sensitivity analysis, T. Abatzoglu [**1**] for relations to approximation theory, R.R. Phelps [**22**] for an application in nonlinear optimization, and [**7**, **8**], [**21**] for further information.

It is well-known that, at least in a separable Hilbert space, the operator P, being non-expansive, is Gâteaux differentiable almost everywhere. See N. Aronszajn [**2**], and also [**15**], [**19**], [**11**] for this extension of Rademacher's

Theorem to infinite dimensions. However, one could not hope to obtain a more refined analysis of the differentiability properties of P by using these techniques. In [**16**], therefore, J.B. Hiriart-Urruty posed the problem of characterizing the differentiability points of P as well as providing techniques which allow one to calculate the derivatives in a more or less explicit form.

It seems clear that the differentiability of P should be somehow tied to the smoothness of the boundary of the set C. Highlighting this observation, R.B. Holmes [**17**] has shown that if C has boundary of class $C^k (k \geq 2)$, then P is of class C^{k-1} in $H \backslash C$, and S. Fitzpatrick and R. Phelps [**14**] have shown that the converse is also true under an additional (in fact a necessary and sufficient) qualification hypothesis (see Section 6 for this). The situation becomes more complicated, however, when the smoothness type assumption is not satisfied throughout the boundary of the set C. For instance, if Px is a point of second order smoothness of C, (meaning that the gauge μ_C is twice differentiable at Px), is it true that P is differentiable at $x \in H \backslash C$? We shall present an answer to this and related problems concerning the first order differentiability of P.

It turns out that the key to understanding the differentiability of the metric projection is to consider *variational type* second order concepts such as *second order Mosco derivatives* or *second order Attouch-Wets derivatives.* In fact we will show here that P is Gâteaux (Fréchet) differentiable at $x \in H \backslash C$ if and only if the boundary of C is second order Mosco (Attouch-Wets) smooth at the point Px.

This gives new insight even in finite dimensions, for in this case, second order Mosco derivates coincide with *second order epi derivates* in the sense of R.T. Rockafellar [**23, 24**].

2. Differentiability.

It was shown by E. Zarantonello [**29**] that the metric projection P onto a closed convex set C with nonempty interior has a directional derivative at every boundary point $x \in \partial C$ in the sense that

$$P(x+th) = x + t\, d_+P(x)h + o(t), \tag{2.1}$$

$h \in H$, $t \to 0^+$. Here the operator $d_+P(x)$ turns out to be the orthogonal projection $P_{S(x)}$ onto the support cone of C at x:

$$S(x) = \overline{\bigcup_{\lambda \geq 0} \lambda(C-x)}.$$

In particular, $d_+P(x)$ can only be linear in the rather special case where $S(x)$ is a linear subspace. This observation suggests that one should in general consider directional type derivatives of the operator P.

We will say that P is *directionally Gâteaux differentiable* at $x \in H$, if the limit

$$\lim_{t\to 0^+} \frac{P(x+th)-P(x)}{t} = d_+P(x)h \tag{2.2}$$

exists in norm for every $h \in H$, or equivalently, if

$$P(x+th) = P(x) + t\, d_+P(x)h + \mathrm{o}(t), \tag{2.3}$$

for $h \in H$, $t \to 0^+$, where $\mathrm{o}(t)/t \to 0$ in norm as $t \to 0^+$. Notice that our approach is more general as for instance in [**14**], where the authors consider Gâteaux derivatives in the more restricted sense that the operator $d_+P(x)$ has to be linear. We preserve their notation $dP(x)$ for this particular case. If the limit (2.2) is uniform over $||h|| \le 1$, then P is said to have a *directional Fréchet derivative* at x, noted $P'_+(x)$. Again we use the notation $P'(x)$ (consistent with [**14**]) to indicate when $P'_+(x)$ is linear. Let us collect some basic information about d_+P. To start with, recall the fact that P is the Fréchet derivative of the continuous convex function f on H defined by

$$f(x) = \frac{1}{2}||x||^2 - \frac{1}{2}||x-Px||^2, \tag{2.4}$$

(see [**14**] or [**29**]). This gives rise to the following observation.

Proposition 2.1. *Let $x \in H \backslash C$, and suppose $d_+P(x)$ exists. Then f given by (2.4) is twice differentiable at x in the sense that it has a second order Taylor expansion at x of the form*

$$f(x+th) = f(x) + t\langle \nabla f(x), h\rangle + t^2 q(h) + o(t^2), \tag{2.5}$$

$h \in H$, $t \to 0^+$, $\nabla^F f(x) = Px$, for a positive and positively 2-homogeneous continuous convex function $q : H \to \mathbb{R}$. Moreover, we have

$$d_+P(x)h = \nabla^F q(h),\ \ h \in H. \tag{2.6}$$

Proof. As $P = \nabla f$, the convergence of $\frac{1}{t}\langle P(x+th) - Px, h\rangle$ implies the existence of a second order Taylor expansion of $t \to f(x+th)$ at $t=0$, that is

$$f(x+th) = f(x) + t\langle \nabla f(x), h\rangle + \frac{t^2}{2}\langle d_+P(x)h, h\rangle + \mathrm{o}(t^2),$$

$t \to 0^+$. (See for instance [**10**, §2]). In other terms, (2.5) is satisfied with q given by $q(h) = \frac{1}{2}\langle d_+P(x)h, h\rangle$, which is continuous, convex and positively 2-homogeneous. We have to show that $d_+P(x)h = \nabla^F q(h)$ for every h.

Consider the second order difference quotient of f at x, that is

$$\triangle_{f,x,y,t}(h) = \frac{f(x+th) - f(x) - t\langle y, h\rangle}{t^2}, \tag{2.7}$$

$t \neq 0$, $h \in H$, $(y = \nabla f(x) = Px)$. Then (2.5) means $\triangle_{f,x,y,t} \to q$ pointwise as $t \to 0^+$. Now observe that, for $t > 0$,

$$\frac{1}{t}(P(x+th) - Px) \in \partial \triangle_{f,x,y,t}(h),$$

so for any $k \in H$, the subgradient inequality gives

$$\frac{1}{t}\langle P(x+th) - Px, k\rangle \leq \triangle_{f,x,y,t}(h+k) - \triangle_{f,x,y,t}(h).$$

Passing to the limit $t \to 0^+$, we obtain

$$\langle d_+P(x)h, k\rangle \leq q(h+k) - q(h),$$

which shows $d_+P(x)h \in \partial q(h)$. But $d_+P(x)$ is nonexpansive, hence continuous, and so ∂q has a selection which is everywhere continuous. This means that q is everywhere Fréchet differentiable. □

The following result is essentially known, although the proofs in [**14**] and [**30**] seem to rely on the linearity of the derivative. We therefore include a proof of our own.

Proposition 2.2. *Let $x \in H \backslash C$, and suppose $d_+P(x)$ (rsp. $P'_+(x)$) exists. Then $d_+P(y)$ (rsp. $P'_+(y)$) exists for every point y on the ray from Px through x, that is, for y of the form*

$$y = \lambda x + (1-\lambda)Px, \ \lambda > 0.$$

Proof. Let $h \in H$ be fixed. Write $v = Px$, $v_t = P(x+th)$. Let $y = \lambda x + (1-\lambda)Px$. We choose k_t such that

$$v_t + \lambda(x + th - v_t) = y + tk_t,$$

which means that $P(y + tk_t) = P(x+th) = v_t$. We find

$$k_t = (1-\lambda)\frac{v_t - v}{t} + \lambda h \ \to \ (1-\lambda)d_+P(x)h + \lambda h =: k.$$

Now

$$\begin{aligned} \frac{1}{t}(P(y+tk) - Py) &= \frac{1}{t}(P(y+tk_t) - Py) + \frac{1}{t}(P(y+tk) - P(y+tk_t)) \\ &= \frac{1}{t}(P(x+th) - Py) + \mathrm{o}(1) \to d_+P(x)h. \end{aligned}$$

It remains to show that every $k \in H$ may be written in the form $k = (1 - \lambda)d_+P(x)h + \lambda h$ for some $h \in H$, i.e., that the operator $A = (1-\lambda)d_+P(x) + \lambda id$ has full range. Now observe that $d_+P(x)$, being a subdifferential by Proposition 2.1, is maximally monotone. For $0 < \lambda \leq 1$, we may therefore write

$$h = \left(id + \frac{1-\lambda}{\lambda} d_+P(x)\right)^{-1} \left(\frac{k}{\lambda}\right)$$

for the solution h of $Ah = k$. As for $\lambda > 1$, notice that $id - d_+P(x)$ as well is maximally monotone, being a subdifferential, too. So here we write

$$h = (id + (\lambda - 1)(id - d_+P(x)))^{-1}(k)$$

for the solution h of $Ah = k$. This completes our argument. Notice that a chain rule similar to the one obtained in [**14**] may readily be stated using the (nonlinear) operator A. □

The relation addressed in Proposition 2.1 seems to hint that we should even consider weak directional derivatives of the operator P, these being equivalent to f (given by (2.4)) having a second order Taylor expansion (2.5). However, this notion seems to have a drawback: It lacks the property derived for $d_+P(x)$ in Proposition 2.2. Namely, we have the following

Proposition 2.3. *Let $x \in H \backslash C$. Suppose P is weakly directionally Gâteaux differentiable at every point y on the ray from Px through x. That is, the limit* (2.2) *exists in the weak topology for any such y. Then P is already directionally Gâteaux differentiable in norm, that is, $d_+P(y)$ exists as a limit in norm.*

Remark. Suppose that for a point $x \in H \backslash C$, the limit (2.2) exists in the weak sense, but $d_+P(x)$ fails to exist. Then we must be able to find a point y (in fact there are many such points) on the ray from Px through x where (2.2) even fails to converge weakly. This is clearly hard to imagine, and in fact, we do not have an explicit example of such behaviour, so it may very well be true that weak convergence in (2.2) always implies norm convergence.

Proof. This is a special case of a result by H. Attouch [**3**, Théorème 1.2], stating that for a sequence f_n, f of proper lower semi-continuous convex functions satisfying $0 \in \partial f_n(0)$, weak convergence $(id + \lambda\partial f_n)^{-1} \rightharpoonup (id + \lambda\partial f)^{-1}$ for every $\lambda > 0$ entails norm convergence $(id + \lambda\partial f_n)^{-1} \to (id + \lambda\partial f)^{-1}$. Notice that the result applies with f_n the second order difference quotient $\triangle_{\sigma_C,x,Px,t_n}$ of the support function σ_C of C, and f the function q from Proposition 2.1 satisfying $d_+P(x) = \nabla q$. Here we have

$$(id + \lambda\partial\triangle_{\sigma_C,x,Px,t_n})^{-1}(h) = \frac{1}{\lambda}\frac{P(\lambda x + (1-\lambda)Px + t_n h) - Px}{t_n}.$$

□

3. Variational Second Derivatives.

As we have seen in Section 2, second order differentiability of a convex function may be defined as pointwise convergence of the second order difference quotient (2.7) to a fully defined continuous convex limit function q. Variational type second derivatives are now introduced in the same way by replacing pointwise convergence with any kind of variational type convergence, such as epi, Mosco or Attouch-Wets convergence, and moreover, by allowing for the corresponding limit function q to take on the value $+\infty$. To begin with, let us consider Mosco convergence of sequences of convex functions.

A sequence (f_n) of convex, proper lower semi-continuous functions is said to *Mosco converge* to a limit function f, noted $f_n \stackrel{m}{\to} f$, if the following conditions are satisfied:

(α) For any $x \in H$ there exist $x_n \to x$ (norm) such that $f_n(x_n) \to f(x)$;

(β) For any $x \in H$, $n_k \uparrow \infty$ and $x_k \rightharpoonup x$ (weakly), we have

$$\liminf_{k\to\infty} f_{n_k}(x_k) \geq f(x).$$

See Attouch [**3**, **4**] for basic information on this notion of convergence. **Definition 3.1** A continuous convex function f on H is said to be *second order Mosco differentiable* at $x \in H$ with respect to $y \in \partial f(x)$ if the second order difference quotient (2.7) Mosco converges to some limit function q, that is, $\triangle_{f,x,y,t} \stackrel{m}{\to} q$ as $t \to 0^+$. The function q is called the *second order Mosco derivative* of f at x with respect to $y \in \partial f(x)$.

Notice that $\text{dom}(q)$ might be a proper subcone of H; even $\text{dom}(q) = \{0\}$ is admitted. Unless $\text{dom}(q)$ is dense, we therefore may not infer from the existence of a second Mosco derivative that the first derivative exists. That explains why we have to specify the subgradient $y \in \partial f(x)$ in Definition 3.1.

This notion seems quite queer at first sight, especially since we allow for such limit functions as $\text{dom}(q) = \{0\}$. However, as we will see, one often gets situations where this *generalized second derivative* is actually the classical second derivative. This is made precise by the following result, proved in [**9**, Prop. 6.1].

Proposition 3.1. *Let f be twice Mosco differentiable at $x \in H$ with respect to $y \in \partial f(x)$, with second Mosco derivative q. That is $\triangle_{f,x,y,t} \xrightarrow{m} q(t \to 0^+)$. For q to be a second derivative in the usual sense (i.e., in the sense of formula* (2.5)*), it is necessary and sufficient that f be Lipschitz smooth at x. In this case,* $\mathrm{dom}(q) = H$*, and $y = \nabla^F f(x)$.*

Recall here from Fabian [**13**] that $f : H \to \mathbb{R}$ is *Lipschitz smooth* at x if there exist $C > 0$ and $\delta > 0$ such that

$$f(x+h) - f(x) - \langle \nabla f(x), h \rangle \le C||h||^2 \tag{3.1}$$

is satisfied for all $||h|| \le \delta$. Notice that (3.1) may be expressed equivalently by saying that $\triangle_{f,x,y,t}$ is uniformly bounded on $||h|| \le \delta$ for $0 < |t| \le 1$, say, (cf. [**9**, Section 2]).

It will be convenient to have a test for whether a function is second order Mosco differentiable. This is not always easy to check, but we have a reasonable method when the function f is of class $C^{1,1}$.

Proposition 3.2. *Let f be a convex $C^{1,1}$ function on H. Then the following are equivalent:*
(1) *f is second order Mosco differentiable at x;*
(2) *∇f is norm Gâteaux differentiable at x.*
Moreover, (2) *implies* (1) *even without any assumptions on f.*

Proof. First assume (1), that is, $\triangle_t := \triangle_{f,x,y,t} \xrightarrow{m} q$ for some limit function q. Now by Attouch's characterization of Mosco convergence, (cf. [**3**]), this means that for any fixed sequence $t_n \to 0^+$, $h \in H$ and $v \in \partial q(h)$ there exist $h_n \to h$ (norm) and $v_n \in \partial \triangle_{t_n}(h_n)$ having $v_n \to v$ (norm). The latter means $\frac{1}{t_n}(\nabla f(x + t_n h_n) - \nabla f(x)) = v_n \to v$ in norm. Using the local Lipschitz assumption on ∇f, we find $\frac{1}{t_n}(\nabla f(x+t_n h) - \nabla f(x)) \to v$ in norm, which is precisely the meaning of statement (2).

Conversely, assume (2) is satisfied. This implies $\triangle_t \to q$ pointwise for some limit function q (see [**9**, §2]). We have to show that convergence is as well in the Mosco sense. Since condition (α) is clear, it remains to check condition (β). Fix $t_n \to 0^+$ and $h_n \rightharpoonup h$ (weakly). Then $v_n = \frac{1}{t_n}(\nabla f(x + t_n h) - \nabla f(x))$ is an element of $\partial \triangle_{t_n}(h)$. The subgradient inequality therefore gives

$$\langle v_n, h_n - h \rangle \le \triangle_{t_n}(h_n) - \triangle_{t_n}(h).$$

Since by assumption v_n converges in norm, and $h_n - h \rightharpoonup 0$ weakly, the left hand side tends to 0 here. Since $\triangle_{t_n} \to q$, we find $\liminf \triangle_{t_n}(h_n) \ge q(h)$ as desired. This proves statement (1). ☐

Remark. Notice that (1) no longer implies (2) if the local Lipschitz assumption on ∇f is dropped, as shown by Example 7.3. Nevertheless, Proposition 3.2 provides a useful test for second order Mosco smoothness which in particular applies to the outer parallel sets of a convex set (see Section 4). As shown by Example 7.4, a function f of class $C^{1,1}$ may be second order differentiable at some point x, without being second order Mosco differentiable there. However, by the result of Aronszajn [**2**], in a separable space H, for almost all points $x \in H$, second order differentiability of f at x already implies second order Mosco differentiability.

With these preparations, we may now state our first main result.

Theorem 3.3. *Let C be a closed convex set in H, P_C its metric projection, σ_C its support function, d_C its distance function. Then the following statements are equivalent:*

(1) *σ_C is twice Mosco differentiable at x with respect to $y \in \partial\sigma_C(x)$, $(y = P_C(x+y))$;*

(2) *P_C is directionally Gâteaux differentiable at $z = x + y \in H\backslash C$, $(y = P_C z, x = z - P_C z)$;*

(3) *The distance function d_C is twice Mosco differentiable at $z = x + y$;*

(4) *The distance function d_C is twice differentiable in the classical sense at every point $z_\lambda = \lambda z + (1-\lambda)y = \lambda z + (1-\lambda)Pz$, $\lambda > 0$.*

Proof. The proof uses a general result by H. Attouch [**3**, Théorème 1.2]. Let us first show (1) $\Leftrightarrow$ (4). Indeed, statement (1) is: $\triangle_{\sigma_C,x,y,t} \stackrel{m}{\to} q$ for some q. Now a general fact is that

$$(3.2) \qquad (\triangle_{f,u,v,t})^* = \triangle_{f^*,v,u,t},$$

with $*$ denoting Young-Fenchel conjugation. As Mosco convergence is invariant under conjugation, (1) translates into the equivalent statement

$$(1') \ \triangle_{i_C,y,x,t} \stackrel{m}{\to} q^* \ (t \to 0^+),$$

where $i_C = \sigma_C^*$ is the indicator function of C. Next we use the quoted result by Attouch saying that Mosco convergence is equivalent to pointwise convergence of all the infimal convolutions with multiples of $||\cdot||^2$, that is, $(1')$ is equivalent to

$$(1'') \ \triangle_{i_C,y,x,t} \,\square\, \frac{1}{2\lambda}||\cdot||^2 \to q^* \,\square\, \frac{1}{2\lambda}||\cdot||^2,$$

pointwise (as $t \to 0^+$) for every fixed $\lambda > 0$. Now recall another general fact saying that

$$(3.3) \qquad \triangle_{f,u,v,t} \,\square\, \frac{1}{2\lambda}||\cdot||^2 = \triangle_{f \square \frac{1}{2\lambda}||\cdot||^2, u+\lambda v, v, t}\,.$$

That means, we have the new equivalent statement

$$(1''')\ \triangle_{i_C \Box \frac{1}{2\lambda}\|\cdot\|^2, y+\lambda x, x, t} \to q^* \Box \frac{1}{2\lambda}\|\cdot\|^2$$

pointwise (as $t \to 0^+$) for every fixed $\lambda > 0$. Now we remark that

$$i_C \Box \frac{1}{2\lambda}\|\cdot\|^2 = \frac{1}{2\lambda} d_C^2, \tag{3.4}$$

which shows that $(1''')$ is statement (4), since the differentiability properties of d_C and d_C^2 are equally good on $H \setminus C$.

Let us next prove (4) $\Leftrightarrow$ (2). Indeed, the Fréchet derivative of $\frac{1}{2}d_C^2$ is $id - P_C$, hence statement (4) is equivalent to saying that P_C is weakly directionally Gâteaux differentiable at every point $y + \lambda x$, that is, every point z_λ on the ray from y through $z = x + y$. By Proposition 2.3, this is equivalent to the existence of $d_+P(z)$, as claimed in statement (2).

Finally, let us show (1) $\Leftrightarrow$ (3). Indeed, we first observe that (1) is equivalent to

$$(3')\ \triangle_{\sigma_C, x, y, t} + \frac{\lambda}{2}\|\cdot\|^2 \overset{m}{\to} q + \frac{\lambda}{2}\|\cdot\|^2$$

for every $\lambda > 0$. This may be checked either directly using conditions (α) and (β), or by using the family of pseudo metrics defining the Mosco topology on the cone of proper convex and lower semi-continuous functions as presented in Attouch [**3**, §2]. Using once again the invariance of Mosco convergence under conjugation in tandem with formula (3.2), we see that $(3')$ amounts to the equivalent statement $\triangle_{i_C, y, x, t} \Box \frac{\lambda}{2}\|\cdot\|^2 \to q^* \Box \frac{\lambda}{2}\|\cdot\|^2$, pointwise, which by formula (3.4) and Attouch's Theorem is equivalent to statement (3). □

It turns out to be a more difficult problem to relate the differentiability of P to the second order differentiability properties of the gauge functional μ_C of C. The following result, which uses a technique from [**14**], gives an answer in the case where the derivative $dP(x)$ is linear.

Corollary 3.4. *Let C be closed bounded convex with 0 in its interior. Let $x \in H \setminus C$. Then the following are equivalent:*

(1) *P_C is Gâteaux differentiable at x, i.e., $dP_C(x)$ exists;*

(2) *μ_C is twice Mosco differentiable at $P_C x$ with respect to*

$$\langle x - P_C x, P_C x\rangle^{-1}(x - P_C x) \in \partial\mu_C(P_C x),$$

its second Mosco derivative q being a generalized quadratic form.

Proof. We know from Proposition 2.1 that, once P_C is differentiable at x, it is so at every point on the ray from $P_C x$ through x. We can find a point $\bar{x}$ on this ray which satisfies

$$\langle \bar{x} - P_C\bar{x}, P_C\bar{x}\rangle = 1. \tag{3.5}$$

Let us call $\bar{x}$ the *ideal point* on this ray. Now Fitzpatrick and Phelps [**14**] have shown that differentiability of P_C at the ideal point $\bar{x}$ is equivalent to differentiability of P_{C^o} at $\bar{x}$, where in both cases, $dP_C(\bar{x})$ and $dP_{C^o}(\bar{x})$ are supposed to be linear operators. Now we apply Theorem 3.3 to see that the latter is equivalent to second order Mosco differentiability of σ_{C^o} at $\bar{x} - P_{C^o}\bar{x}$ with respect to $P_{C^o}\bar{x} \in \partial\sigma_{C^o}(\bar{x} - P_{C^o}\bar{x})$, the second derivative being a generalized quadratic form. But recall that $\sigma_{C^o} = \mu_C$, so we are almost done. Notice that the choice of the ideal point $\bar{x}$ is such that

$$P_{C^o}(\bar{x}) = Q(\bar{x}) = \bar{x} - P_C\bar{x} = \bar{x} - P_C x,$$

with Q as in [**14**], so μ_C is finally seen to be twice Mosco differentiable at $\bar{x} - P_{C^o}\bar{x} = P_C x$ with respect to $\bar{x} - P_C x \in \partial\mu_C(P_C x)$. This proves the result, for $\bar{x} = P_C x + \langle x - P_C x, P_C x\rangle^{-1}(x - P_C x)$, and so $\bar{x} - P_C x = \langle x - P_C x, P_C x\rangle^{-1}(x - P_C x)$. □

As we will see later, this result may be exploited to obtain results in the spirit of [**17**] or [**14**] by finding the right conditions on C at $P_C x$ which force the second Mosco derivative of μ_C at $P_C x$ above to be a classical second derivative.

4. Epigraphical Analysis.

The task of this section is to prove that the differentiability of the metric projection at a point $x \in H \setminus C$ is equivalent to the second order Mosco smoothness of the boundary of C at Px. Equivalently, if ∂C is represented as the graph of a convex function f around Px, then f must be twice Mosco differentiable at the corresponding point.

Theorem 4.1. *Let f be a continuous convex function on H, and let (z, α) be a point not contained in the epigraph epif of f. Let $(y, f(y)) = P_{epif}(z, \alpha)$. Then the following are equivalent:*

(1) *P_{epif} is directionally Gâteaux differentiable at (z, α);*

(2) *f is twice Mosco differentiable at y with respect to*

$$(f(y) - \alpha)^{-1}(z - y) \in \partial f(y).$$

The proof will be obtained by a series of auxiliary results. To begin with, let us consider $f : H \to \mathbb{R}$ continuous convex with $f(0) = -1, 0 \in \partial f(0)$, so that $P_{epi f}(0,\mu) = (0,-1) = (0, f(0))$ for any $\mu < -1$. Now let $C = \{(x,\alpha) \in H \times \mathbb{R} \; ; \; ||x|| \leq 1, \alpha \leq 1, \alpha \geq f(x)\}$. Then C is a bounded closed convex set having (0,0) in its interior. Moreover, $P_{epif} = P_C$ in a neighbourhood of (0,-2). Under these conditions, we have the following:

Lemma 4.2. *If P_C is directionally Gâteaux differentiable at $x = (0,-2)$, then so is P_{C^0}.*

Proof. Notice that the result would follow directly from [**FP**, Proposition 4.1] if we knew that $d_+P_C(x)$ was linear. Indeed, as $\langle (0,-2) - (0,-1), (0,-1) \rangle = 1$, $x = (0,-2)$ is just the ideal point on the ray from $P_C x$ through x to which the quoted reference applies. However, as $d_+P(x)$ need not be linear, we have to check the proof given in [**14**] in detail.

Let T be the operator defined by

$$Th = h - \langle h - d_+P(x)h, Px \rangle x, \tag{4.1}$$

which in [**14**, Prop. 4.1] arises as the derivative $dA(x)$ of $Ay = \langle y - Py, y \rangle^{-1} y$ at x. We do not claim here that T is invertible, as it is in [**14**]. However, it suffices to know that T is surjective in order to carry out the second half of the argument in [**14**] following formula (4) there. So the argument rests on showing that T is surjective. While this is easy to see in the linear case, here we only know that T is positively homogeneous, i.e., $T(\lambda h) = \lambda T h$ for $\lambda \geq 0$. We have to show that, setting $\phi(h) = \langle h - d_+P(x)h, Px \rangle$, the equation

$$h - \phi(h)x = k \tag{4.2}$$

is solvable for any fixed k. Naturally, any solution must lie in the two dimensional subspace $L = \operatorname{lin}\{x, k\}$, and clearly T maps L into itself. So the problem of solving (4.2) is reduced to the case of two dimensions.

Let $r > 0$ be arbitrary, and define $\Phi : [0,1] \times B_r(0) \to L$ by setting

$$\Phi(t,h) = T\Big(\frac{1}{1+t}h\Big) - T\Big(-\frac{t}{1+t}h\Big),$$

$h \in L, 0 \leq t \leq 1$. Then $\Phi(0,\cdot) = T$, while $\Phi(1,h) = T(\frac{1}{2}h) - T(-\frac{1}{2}h)$ is odd, that is, $\Phi(1,-h) = -\Phi(1,h)$. We claim that

$$0 \notin \Phi(t, \partial B_r(0))$$

for all $0 \leq t \leq 1$. Indeed, otherwise we would have $T(\frac{1}{1+t}h) = T(-\frac{t}{1+t}h)$ for some $0 \leq t \leq 1$ and $||h|| = r$. Or rather,

$$h = \Big(\frac{1}{1+t}\phi(h) - \frac{t}{1+t}\phi(-h)\Big)x.$$

So $h = \rho x$ for some $\rho \neq 0$, with

$$\rho = \frac{1}{1+t}\phi(\rho x) - \frac{t}{1+t}\phi(-\rho x).$$

Now, if $\rho > 0$, we obtain

$$1 = \frac{1}{1+t}\phi(x) + \frac{t}{1+t}(-\phi(-x)),$$

while $\rho < 0$ gives

$$1 = -\frac{1}{1+t}\phi(-x) + \frac{t}{1+t}\phi(x),$$

using the fact that ϕ is positively homogeneous. But

$$\phi(x) \quad = \langle x - d_+P(x)x, Px\rangle = \langle x, Px\rangle,$$

$$\phi(-x) = \langle -x - d_+P(x)(-x), Px\rangle = -\langle x, Px\rangle,$$

since $d_+P(x)x = d_+P(x)(-x) = 0$ is our special situation. Indeed, it is always true that $d_+P(x)(\lambda(x - Px)) = 0$, and here $x = 2Px$. So we end up with $1 = \langle x, Px\rangle = \langle(0,-2),(0,-1)\rangle = 2$, a contradiction, showing that $0 \notin \Phi(t, \partial B_r(0))$ for any t.

Let $B_s(0)$ be a ball such that

$$\Phi(t, \partial B_r(0)) \cap B_s(0) = \emptyset$$

for all $0 \le t \le 1$. For any $y \in B_s(0)$, the homotopy invariance of the degree function now shows

$$\deg(T, B_r(0), y) = \deg(\Phi(t,\cdot), B_r(0), (1-t)y)$$

$$= \deg(\Phi(1,\cdot), B_r(0), 0).$$

The latter value, however, is odd by Borsuk's Theorem, so $B_s(0) \subset T(B_r(0))$. This shows that T is open at 0, and positive homogeneity shows T is surjective. For the arguments concerning the degree we refer the reader to Deimling [**12**, p. 21 and p. 23]. □

Our next step is to use Theorem 3.3, which shows that, under the conditions of Lemma 4.2, P_{C^o} is now directionally Gâteaux differentiable at $(0,-2)$, hence $\sigma_{C^o} = \mu_C$ is twice Mosco differentiable at (0,-1). This leads to the following

Lemma 4.3. *Under the conditions of Lemma* 4.2, *suppose f is also Lipschitz smooth at* 0. *If P_C is directionally Gâteaux differentiable at* $(0,-2)$, *then f*

is twice differentiable and twice Mosco differentiable at 0. *Conversely, if* f *is twice Mosco differentiable at* 0 (*and hence by Lipschitz smoothness also twice differentiable*), *then* P_C *is directionally Gâteaux differentiable at* $(0,-2)$.

Proof. a) First assume that P_C is directionally Gâteaux differentiable at $(0,-2)$. It follows from Lemma 4.2 that μ_C is twice Mosco differentiable at $(0,-1) = (0, f(0))$ with respect to $(0,-1) \in \partial\mu_C(0,-1)$. We are going to argue that μ_C is also Lipschitz smooth at (0,-1). Suppose this has been shown. Then, by Proposition 3.1, μ_C is twice differentiable at (0,-1) in the sense of (2.5). Now observe that near (0,-1), ∂C coincides with the graph of f. The fact that second order differentiability of μ_C at $(0,-1) = (0, f(0))$ implies second order differentiability of f at 0, and vice versa now follows from the formula

$$\Delta_{\mu_C,(0,-1),(0,-1),t_n}(h,s) = \gamma_n \, \Delta_{f,0,0,t_n}(\gamma_n^{-1}h), \tag{4.3}$$

where $\gamma_n = \mu_C(t_n h, -1 + t_n s) \to \mu_C(0,-1) = 1$. So it remains to show that μ_C is Lipschitz smooth at (0,-1). To prove this, recall that in a Hilbert space, Lipschitz smoothness of f at 0 is equivalent to the following geometric condition: There exists a ball B touching the epigraph of f at $(0, f(0))$ which is entirely contained in epi f. Let us assume, then, that in the above situation, $B = \{(z,\alpha) : ||z||^2 + (\alpha + 1 - \varepsilon)^2 \leq \varepsilon^2\}$ is contained in epi f. Recalling the definition of the set C, we may also assume that $B \subset C$. We have to show that, for some $\delta > 0$, the ball

$$B_\delta = \{(z,\alpha,\beta) : ||z||^2 + (\alpha + 1 - \delta)^2 + (\beta - 1 + \delta)^2 \leq \delta^2\}$$

is contained in epi(μ_C). Assume the contrary and find $(z_\delta, \alpha_\delta, \beta_\delta) \in B_\delta$ such that $\beta_\delta < \mu_C(z_\delta, \alpha_\delta)$. But notice that $\beta_\delta \to 1, \alpha_\delta \to -1, z_\delta \to 0$ as $\delta \to 0^+$, so $(\frac{z_\delta}{\beta_\delta}, \frac{\alpha_\delta}{\beta_\delta})$ is eventually contained in B, hence in C. This means $\mu_C(\frac{z_\delta}{\beta_\delta}, \frac{\alpha_\delta}{\beta_\delta}) \leq 1$, and so $(z_\delta, \alpha_\delta, \beta_\delta) \in$ epiμ_C, a contradiction which completes the first half of the proof.

b) Let us now assume that f is twice Mosco differentiable at 0 with respect to $0 \in \partial f(0)$, and with second Mosco derivative q, say. We have to show that μ_C is twice Mosco differentiable at $(0,-1)$ with respect to $(0,-1) \in \partial\mu_C(0,-1)$, and it will turn out that the second Mosco derivative is given by $Q(h,s) = q(h)$. Indeed, we have to check conditions (α) and (β) of Mosco convergence for the second difference quotient (4.3) of μ_C. Now observe that, due to the Lipschitz smoothness of f at 0, the second Mosco derivative q is also a second derivative in the usual sense, so $\Delta_{f,0,0,t_n} \to q$ pointwise for any fixed $t_n \to 0^+$, and hence the right hand side of (4.3) converges because of $\gamma_n \to 1$. This proves condition (α). As for (β), fix $(h_n, s_n) \rightharpoonup (h,s)$, then

(4.3) gives $\liminf \triangle_{\mu_C,(0,-1),(0,-1),t_n}(h_n,s_n) \geq q(h) = Q(h,s)$, as required. This proves the result. □

For a closed convex set C in H, the outer parallel set $C_{[\varepsilon]}$ at distance $\varepsilon > 0$ is defined as

$$C_{[\varepsilon]} = \{x \in H : d_C(x) \leq \varepsilon\}.$$

Suppose that locally, the boundary of C is represented as the graph of a continuous convex function f. We ask for a function f_ε locally representing the graph of $C_{[\varepsilon]}$. By the mere definition of $C_{[\varepsilon]}$, we find that the conjugates of f and f_ε must be related by

$$f_\varepsilon^*(z) = f^*(z) + \frac{\varepsilon}{\sqrt{1+||z||^2}}. \tag{4.4}$$

Conjugation therefore gives us

$$f_\varepsilon = f \,\square\, -\sqrt{\varepsilon^2 - ||\cdot||^2}. \tag{4.5}$$

Notice that f_ε is a $C^{1,1}$ function; in particular, it is everywhere Lipschitz smooth. With these preparations we are now in the position to give the *Proof of Theorem* 4.1. Involving an affine change of coordinates, we may assume that $f(0) = 0, 0 \in \partial f(0)$. We have to show that $P_{\text{epi}\, f}$ is directionally Gâteaux differentiable at $(0,\alpha), \alpha < 0$, if and only if f is twice Mosco differentiable at 0 with respect to $0 \in \partial f(0)$.

Let f_ε be the convex $C^{1,1}$ function representing the outer parallel surface of epi f in a neighbourhood of the point $(0,-\varepsilon) = (0, f_\varepsilon(0))$. As a first step a), we will show that directional differentiability of $P_{\text{epi}\, f}$ at $x = (0,\alpha)$, $\alpha < 0$, is equivalent to f_ε being twice differentiable at 0 for all ε small enough.

For this, first observe that the projections $P_{\text{epi}\, f}$ and $P_{\text{epi}\, f_\varepsilon}$ are related by the formula

$$P_{\text{epi}\, f_\varepsilon}(h) = P_{\text{epi}\, f}(h) + \varepsilon \frac{P_{\text{epi}\, f}(h) - h}{||P_{\text{epi}\, f}(h) - h||},$$

$h \in H \setminus \text{epi}\, f_\varepsilon$. Let us now for simplicity fix $\varepsilon = 1$, and suppose $x = (0,-2)$.

This means that the situation of Lemmas 4.2, 4.3 is met with f_1 and $P_{\text{epi}\, f_1}$ playing the roles of f and $P_{\text{epi}\, f}$ there. So the differentiability of $P_{\text{epi}\, f_1}$ at $(0,-2)$ implies second order differentiability of f_1 at 0, since f_1 is Lipschitz smooth, whence we may infer that f_1, and similarly every f_ε, is twice differentiable at 0. The converse follows from the second part of Lemma 4.3, i.e., if any one of the f_ε is twice Mosco differentiable at 0, then $P_{\text{epi}\, f_\varepsilon}$, and hence $P_{\text{epi}\, f}$, is directionally differentiable at $x = (0,-2)$. This completes the proof of a).

In a second step b), we now establish the following

Lemma 4.4. *Under the above circumstances, the following statements are equivalent:*

(1) *f is second order Mosco-differentiable at 0 with respect to $0 \in \partial f(0)$;*

(2) *For all $\varepsilon > 0$ sufficiently small, f_ε is twice differentiable at 0;*

(3) *For all $\varepsilon > 0$ sufficiently small, f_ε is twice Mosco differentiable at* 0.

Proof. As the f_ε are Lipschitz smooth, statement (3) immediately implies (2) by Proposition 3.1. So the implications (1) $\Rightarrow$ (3) and (2) $\Rightarrow$ (1) remain to be checked.

We need to relate $\triangle_{f,0,0,t}$ and $\triangle_{f_\varepsilon,0,0,t}$. From (4.4) and (4.5) we infer that

$$\triangle_{f_\varepsilon,0,0,t} = \triangle_{f,0,0,t} \,\square\, \triangle_{\phi_\varepsilon,0,0,t}, \tag{4.6}$$

where $\phi_\varepsilon(h) = -\sqrt{\varepsilon^2 - ||h||^2}$. Hence the second order difference quotient of ϕ_ε at 0 with respect to $0 = \nabla\phi_\varepsilon(0)$ is

$$\triangle_{\phi_\varepsilon,0,0,t}(h) = \frac{||h||^2}{\varepsilon + \sqrt{\varepsilon^2 - t^2||h||^2}}.$$

Now it is clear that (1) implies (3), for $\triangle_{f,0,0,t} \xrightarrow{m} q$ implies $\triangle_{f,0,0,t} \,\square\, \triangle_{\phi_\varepsilon,0,0,t} \xrightarrow{m} q \,\square\, \frac{1}{2\varepsilon}||\cdot||^2$, which implies statement (3) by formula (4.6). Finally, to see that (2) and (1) are in fact equivalent, we calculate the following estimate

$$\left|\frac{1}{2\varepsilon}||h||^2 - \triangle_{\phi_\varepsilon,0,0,t}(h)\right| = \frac{t^2||h||^4}{(\sqrt{\varepsilon^2 - t^2||h||^2} + \varepsilon)(2\varepsilon^2 + 2\varepsilon\sqrt{\varepsilon^2 - t^2||h||^2})} = \mathrm{O}(t^2)||h||^4, \tag{4.7}$$

with $\mathrm{O}(t^2)$ being uniform over all $t^2||h||^2 \leq \varepsilon$. From this we infer the estimate

$$\left|\left(\triangle_{f,0,0,t}(k) + \frac{1}{2\varepsilon}||h-k||^2\right) - (\triangle_{f,0,0,t}(k) + \triangle_{\phi_\varepsilon,0,0,t}(h-k))\right| \leq \mathrm{O}(t^2)||h-k||^4,$$

which shows $(\triangle_{f,0,0,t} \,\square\, \frac{1}{2\varepsilon}||\cdot||^2)(k) - (\triangle_{f,0,0,t} \,\square\, \triangle_{\phi_\varepsilon,0,0,t})(k) \to 0 (t \to 0^+)$. Hence statement (2) is equivalent to pointwise convergence of all the infimal convolutions $\triangle_{f,0,0,t} \,\square\, \frac{1}{2\varepsilon}||\cdot||^2$, which by Attouch's Theorem is equivalent to Mosco convergence of $\triangle_{f,0,0,t}$. This completes the proof. □

We conclude this section with the following Definition, which is justified by Theorem 4.1.

Definition 4.1 Let C be a closed convex set with nonempty interior. A point $x \in \partial C$ is called a point of second order Mosco smoothness with respect to the outer normal vector $v \in N_C(x)$ if, in a neighbourhood of x, the boundary of C may be represented as the graph of a continuous convex function $f : H[x] \to \mathbb{R}$, where $H[x] = \{h \in H : \langle v, h\rangle = \langle v, x\rangle\}$, such that f is twice Mosco differentiable at $x \in H[x]$ with respect to $0 \in \partial f(x)$.

Remark. Combining Theorem 4.1 and Corollary 3.4 shows that in the case where the corresponding second order Mosco derivative of f above is a generalized quadratic form, we might have introduced second order Mosco smoothness of C at $x \in \partial C$ by saying that the gauge μ_C is twice Mosco differentiable at x with respect to $\langle v, x\rangle^{-1}v \in \partial\mu_C(x)$. Notice, however, that our method of proof does not show whether this equivalency also holds when the derivative of the projection P_C at the corresponding point $x + v$ fails to be linear.

5. Fréchet Derivatives.

In the previous sections we have been dealing with directional Gâteaux derivatives of the operator P. In this paragraph we give a more compact account of the situation for directional Fréchet derivatives. This requires another notion of variational convergence which is known as Attouch-Wets convergence. Let us briefly recall the definition.

Let C, D be closed convex sets in H. The *excess* of C and D is defined as $\text{ex}(C, D) = \sup\{d(x, D) : x \in C\}$. For $\rho > 0$ let $C_\rho = C \cap B(0, \rho)$. The bounded Hausdorff distance of C and D is then defined as

$$\text{haus}_\rho(C, D) = \max\{\text{ex}(C_\rho, D), \text{ex}(D_\rho, C)\}.$$

Now the sequence (f_n) of lower semi-continuous proper convex functions is said to be Attouch-Wets convergent to the limit f, noted $f_n \overset{aw}{\to} f$, if

$$\text{haus}_\rho(\text{epi} f_n, \text{epi} f) \to 0 \quad (n \to \infty)$$

for all ρ sufficiently large. Notice that $f_n \overset{aw}{\to} f$ implies $f_n \overset{m}{\to} f$. We refer the reader to [**6**] and [**5**] for more information on this type of convergence. The following parallel of Theorem 3.3 was essentially proved in [**20**]. It shows that Attouch-Wets convergence plays the same role for Fréchet differentiability of the metric projection P as Mosco convergence does for Gâteaux differentiability.

Theorem 5.1. *Let C be a closed convex set in H, P_C its metric projection, σ_C its support function, d_C its distance function. Then the following statements are equivalent:*

(1) *σ_C is second order Attouch-Wets differentiable at x with respect to $y \in \partial\sigma_C(x)$, $y = P_C(x + y)$. That is, $\triangle_{\sigma_C, x, y, t} \overset{aw}{\to} q$ for some q;*

(2) *P_C is directionally Fréchet differentiable at $z = x + y$, $y = P_C z$, $x = z - P_C z$;*

(3) *d_C is second order Attouch-Wets differentiable at $z = x + y$;*

(4) *d_C is strongly second order differentiable at z, that is $\triangle_{d_C,x+y,x,t}$ converges uniformly on bounded sets (as $t \to 0^+$).*

With this result at hand, we may obtain an analogue of Theorem 4.1 using essentially the same pattern of reasoning.

Theorem 5.2. *Let f be a continuous convex function on H, and let (z, α) be a point not contained in the epigraph of f. Let $(y, f(y)) = P_{epif}(z, \alpha)$. Then the following are equivalent:*

(1) *P_{epif} is directionally Fréchet differentiable at (z, α);*

(2) *f is second order Attouch-Wets differentiable at y with respect to $(f(y) - \alpha)^{-1}(z - y) \in \partial f(y)$;*

(3) *For $\varepsilon > 0$ small enough, f_ϵ is strongly second order differentiable at y_ε, where $y_\varepsilon = y + \varepsilon\left(\|z - y\|^2 + (f(y) - \alpha)^2\right)^{-1/2}(z - y)$.*

Remark. Attouch Wets convergence is somehow related to uniform convergence on bounded sets as Mosco convergence is related to pointwise convergence. There is, however, one major difference. Namely, while uniform convergence on bounded sets implies Attouch-Wets convergence, pointwise convergence does not imply Mosco convergence. As shown in Example 7.4, this is so even when convergence of second order difference quotients is considered.

6. Applications.

In this section we apply our result to relate the differentiability properties of P_C and the smoothness type of the boundary of C.

Theorem 6.1. *Suppose P_C is directionally Gâteaux differentiable at $x \in H \backslash C$. Then C is second order Gâteaux smooth at $P_C x$ if and only if a ball contained in C touches ∂C at $P_C x$.*

Proof. By Theorem 4.1, any convex function f locally representing the boundary of C in a neighbourhood of $P_C x$ is twice Mosco differentiable at the corresponding point in its domain. But a ball touching ∂C at $P_C x$ from within means that f is also Lipschitz smooth there, and hence f is twice differentiable by Proposition 3.1. Almost needless to say, the second derivative is a quadratic form iff $d_+ P(x)$ is linear. □

We may add here that for C bounded and containing 0 in its interior, we may recast second order smoothness of C at $P_C x$ by saying that the gauge μ_C is twice differentiable and twice Mosco differentiable at $P_C x$.

Theorem 6.2. *Let $x \in H \backslash C$. The following are equivalent:*

(1) *P_C is directionally Gâteaux differentiable at x;*

(2) *For all $\varepsilon > 0$ small enough, the outer parallel sets $C_{[\varepsilon]}$ are second order smooth at $P_{C_{[\varepsilon]}}(x)$.*

Proof. (1) implies (2) by Theorem 5.1. The converse is implicit in the proof of Theorem 4.1. □

For directional Fréchet derivatives we obtain the analogous results:

Theorem 6.3. *Suppose P_C is directionally Fréchet differentiable at $x \in H \setminus C$. Then C is second order Fréchet smooth at $P_C x$ if and only if a ball contained in C touches ∂C at $P_C x$.*

Clearly, the Fréchet analogue of Theorem 6.2 is also true, that is, directional Fréchet differentiability of P_C at x corresponds to strong second order smoothness of the outer parallel bodies at the points $P_{C_{[\varepsilon]}} x$. As a consequence of this we obtain the following result, stated by Fitzpatrick and Phelps [**14**] under the stronger assumption that the boundary of C be of class C^1. Notice that this is a localized version of the principal result of Holmes [**17**].

Corollary 6.4. *Let C be closed convex with 0 in its interior. Suppose the boundary of C is second order Fréchet smooth at z, which means that μ_C is strongly second order differentiable at z. Then P_C is Fréchet differentiable at every x with $z = P_C x$.*

Proof. Use Theorem 5.2 and the fact that uniform convergence on bounded sets implies Attouch-Wets convergence. □

In view of this result, it is natural to ask whether, similarly, second order Gâteaux smoothness of C at $z \in \partial C$ implies Gâteaux differentiability of P_C at x, $P_C x = z$. Surprisingly, the answer is in the negative, as shown by Example 7.4. In fact, our theory says that second order Gâteaux smoothness alone is not sufficient, since second order Mosco smoothness of the boundary at z is needed, and the latter does not follow from the first in general. This does not seem to be true even when the boundary of C is everywhere Lipschitz smooth, i.e., if at every boundary point, a ball touches ∂C from within. What is surprising, in the somewhat pathological situation of Example 7.4, with $x \in H \setminus C$, $z = P_C x$, the set C will be second order Gâteaux smooth at z, but as a consequence of Lemma 4.4, at least one of the outer parallel sets $C_{[\epsilon]}$ will fail to be second order Gâteaux smooth at the corresponding point $z_\epsilon = P_{C_{[\epsilon]}} x$.

Let us now ask for conditions under which, dually to Theorems 6.1 and 6.3, the differentiability of P_C at $x \in H \setminus C$ implies the second order differentiability of the support function σ_C at $x - P_C x$. We have the following

Theorem 6.5. *Let C be a bounded closed and convex set with nonempty interior. Suppose P_C is directionally Gâteaux (Fréchet) differentiable at $x \in H \setminus C$. Then σ_C is second order differentiable (rsp. strongly second order differentiable) at $x - P_C x$ if and only if a ball containing C touches ∂C at $P_C x$.*

Proof. By Theorem 3.3 rsp. Theorem 5.1, directional Gâteaux (rsp. Fréchet) differentiability of P_C at x is equivalent to second order Mosco (rsp. Attouch-Wets) differentiability of σ_C at $x - P_C x$. Now Proposition 3.1 tells that the missing condition to ensure second order differentiability is Lipschitz smoothness of σ_C at $x - P_C x$. Notice that the same is true for Attouch-Wets via strong second order differentiability (see [**20**]). Now we use a result of Fabian [**13**, Prop. 2.2], which translates Lipschitz smoothness of σ_C at $x - P_C x$ into a dual condition: C is Lipschitz exposed at $P_C x$ by $x - P_C x$. But notice that in a Hilbert space, and for a bounded set C, Lipschitz exposedness of C at $P_C x$ by $x - P_C x$ and with constant $c > 0$ is the same as saying that the ball with radius $\frac{c}{2}$ and centre on the ray $P_C x + \lambda(P_C x - x)$, $\lambda > 0$ lies locally between the boundary of C and its tangent hyperplane at $P_C x$. By increasing the radius of the ball, we obtain the statement of the Theorem. □

Remark. Notice that in Hilbert space, the Lipschitz exposed points of a bounded closed convex set are precisely the farthest points. Here $x \in C$ is called a farthest point if, for some y, $\|x - y\| = \sup\{\|z - y\| : z \in C\}$. This coincidence is no longer true even in ℓ_p for $p \neq 2$.

We end this section with some analytic parallels to Theorems 6.1 and 6.5.

Theorem 6.6. *Let C be bounded closed and convex with nonempty interior. Suppose P is directionally Gâteaux differentiable at $x \in H \setminus C$. Then σ_C is second order differentiable at $x - Px$ if and only if*

$$\frac{\|P(x+h) - Px\|}{\|h\|} \leq c < 1 \tag{6.1}$$

for some $0 < c < 1$, $\delta > 0$ *and all* $\|h\| \leq \delta$. *In particular, if* $dP(x)$ *exists, then* $\|dP(x)\| < 1$.

Proof. We have to show that condition (6.1) is equivalent to Lipschitz smoothness of σ_C at $x - Px$. Now our analysis in [**9**, §7] shows that the latter is equivalent to the following condition:

$$\langle P(x+h) - Px, h\rangle \leq c\|h\|^2 \tag{6.2}$$

for some $0 < c < 1$, $\delta > 0$ and all $\|h\| \leq \delta$. Now recall that the projection operator is firmly non-expansive, that is, we have the estimates

(6.3)
$$\|P(x+h) - Px\|^2 \leq \langle P(x+h) - Px, h\rangle \leq \|P(x+h) - Px\|\|h\|,$$

(cf. for instance [**17**]). This shows (6.1) and (6.2) being equivalent. □

Corollary 6.7. *Let* C *be bounded closed convex with nonempty interior. Suppose* P *is Fréchet differentiable at* $x \in H \setminus C$. *Then* σ_C *is strongly second order differentiable at* $x - P_x$ *if and only if* $\|P'(x)\| < 1$.

Theorem 6.8. *Let* C *be closed and convex with nonempty interior, and suppose* P *is directionally Gâteaux differentiable at* $x \in H \setminus C$. *Then* C *is second order Gâteaux smooth at* Px *if and only if*

$$\frac{\|P(x+h) - Px\|}{\|h\|} \geq c > 0 \tag{6.4}$$

for some $0 < c < 1$, $\delta > 0$, *and all* $\|h\| \leq \delta$, $h \in H[x] = \{k \in H : \langle k, x - Px\rangle = 0\}$.

Proof. We may assume that $C \subset H \times \mathbb{R}$ and that the boundary of C is locally represented as the graph of a convex function $f : H \to \mathbb{R}$, with say $Px = (0, f(0)) = (0, 1)$, $x = (0, 0)$. This means $0 \in \partial f(0)$. We have to show that statement (6.4) is equivalent to f being Lipschitz smooth at 0. Notice that in our local coordinates, the hyperplane $H[x]$ consists of the vectors of the form $h = (v, 0)$, $v \in H$. Now let $P(x+h) = (w, f(w))$ for some $w \in H$, then we have the relation $v = w + f(w)\partial f(w)$, where $\partial f(w)$ stands for some subgradient of f at w. This gives

(6.5)
$$\frac{\|P(x+h) - Px\|^2}{\|h\|^2} = \left(\frac{\|w\|}{\|w + f(w)\partial f(w)\|}\right)^2 + \left(\frac{\|f(w) - f(0)\|}{\|w + f(w)\partial f(w)\|}\right)^2.$$

Now suppose first that f is Lipschitz smooth at 0. By Proposition 2.1 in [**13**], this means that the ratio $\|\partial f(w)\|/\|w\|$ is bounded on $\|w\| \leq \delta$, say, and this clearly means that the first term on the right hand side of (6.5) is uniformly bounded away from 0 for all $\|w\| \leq \delta$. Now observe that $\|v\| \geq \|w\|$ since P is non-expansive, so statement (6.4) is verified.

Conversely, suppose (6.5) is uniformly bounded away from 0 on $\|v\| \leq \delta$. Then all the ratios $\|\partial f(w)\|/\|w\|$ related to vectors v as above will be bounded by some constant. Choosing $\delta' > 0$ such that $\|w\| < \delta'$ implies $\|v\| < \delta$, we see that this estimate will then hold for all $\|w\| < \delta'$, and hence f is Lipschitz smooth at 0 by Fabian's result. □

Corollary 6.9 (cf.[**14**],[**17**]). *Let $x \in H \setminus C$. Then C is second order Fréchet smooth at Px if and only if P is Fréchet differentiable at x and $P'(x)$ is invertible on $H[x]$.*

Proof. Indeed, for a Fréchet derivative, invertibility of $P'(x)$ on $H[x]$ is just statement (6.4). Use Theorem 5.2 for the rest. □

7. Examples.

In this section we give some examples explaining our results. The first two are elementary though instructive.

Example 1. Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function satisfying $f(0) = 1$, $f'(0) = 0$. For simplicity we may assume that f is everywhere differentiable, but this is not essential. Consider the differentiability of the metric projection P onto the epigraph epi$f$ of f at the point $x = (0,0)$. We fix the direction $h = (1,0)$. Let $s \in \mathbb{R}$, then $P(x + t(s)h) = (s, f(s))$ for $t(s) = s + f(s)f'(s)$. Now the difference quotient of P at $(0,0)$ is

$$\frac{P(x+t(s)h) - Px}{t(s)} = \left(\frac{s}{s + f(s)f'(s)}, \frac{f(s) - f(0)}{s + f(s)f'(s)}\right),$$

which converges (for $s \to 0^+$) if and only if the ratio $f'(s)/s$ has a limit which may be either finite or ∞. This is precisely the same as saying that f has a one-sided second order Mosco derivative q at 0. Notice that the cases $\text{dom}(q) = \{0\}$, $\text{dom}(q) = [0,\infty)$, $\text{dom}(q) = (-\infty, 0]$ and $\text{dom}(q) = \mathbb{R}$ are all possible. Clearly f fails to have a second order Mosco derivative at 0 iff $f'(s)/s$ has different accumulation points. If we take $f(s) = 1 + |s|^{3/2}$, we get the standard example where P is differentiable at $(0,0)$, but the boundary is not second order smooth at $(0,1)$. This also explains Theorem 6.1, since no circle can touch the graph of $f(s) = 1 + |s|^{3/2}$ at $(0,1)$ from above. On the other hand, the circle $B((0,2),1)$ touches the graph of f at $(0,1)$ and lies locally between the graph of f and its tangent $y = 1$, so epi$f$ is Lipschitz exposed at $(0,1)$ by $(0,1)$, whence the support function is second order differentiable at $(0,1)$.

Example 2. Let us now consider $f(x) = |x|$ on the real line at $x = 0$. Then the projection P onto the epigraph C of f is differentiable at every point in the interior of the normal cone $N_C(0,0)$ of C at 0, that is, at points of the form (x,y), $|x| < y$. So f must be second order Mosco differentiable at 0 with respect to the corresponding subgradients v, $|v| < 1$. Since f is not even differentiable at 0, we need to look at its conjugate in order to understand what this means. Now f^* equals 0 on $[-1,1]$, and ∞ outside this interval. By duality, f is second order Mosco differentiable at 0 with respect to $v \in \partial f(0)$ iff f^* is second order Mosco differentiable at v with respect to $0 \in \partial f^*(v)$, and the latter statement clearly makes perfect sense for $|v| < 1$. As for the boundary values $|v| = 1$ we have to consider directional second order derivatives, then P is still directionally differentiable at points (x,x) rsp. $(x,-x)$.

Example 3. Define $f : \ell_2 \to \mathbb{R}$ by

$$f(x) = \sup_{n \in \mathbb{N}} \left(\frac{x_1 + x_n}{n} - \frac{1}{n^2} \right). \tag{7.1}$$

Then $f(0) = 0$, $f \geq 0$ and $\nabla f(0) = 0$. We show that f is second order differentiable at 0, that is, $\frac{1}{t}\partial f(th)$ converges weakly, but fails to converge in norm. Nevertheless, f is second order Mosco differentiable at 0, which shows that Proposition 3.2, (1) $\Rightarrow$ (2) is no longer true without the Lipschitz assumption on ∇f.

First observe that f is Lipschitz smooth at 0. Indeed, an elementary calculation shows that $f(x) \leq \|x\|^2$ for all $\|x\| \leq 1/2$. Next we claim that the second order difference quotient at 0 converges pointwise, i.e.,

$$\triangle_{f,0,0,\frac{1}{k}}(h) = \frac{f(\frac{1}{k}h)}{\frac{1}{k^2}} \to \frac{1}{4}h_1^2 =: q(h)$$

as $k \to \infty$. Due to the Lipschitz smoothness of f at 0, it suffices to check this for the finite sequences h (see [**9**, §5] for this argument), and this is easily done. Also, the sequence $t_k = \frac{1}{k}$ is representative here for all sequences $t_k \to 0^+$, (see [**2**] for this argument).

According to [**9**, §2], pointwise convergence of the second order difference quotient is equivalent to $\frac{1}{t}(\partial f(th) - \nabla f(0)) \rightharpoonup \frac{1}{2}e^1$ weakly for every fixed h.

We show that this difference quotient fails to converge in norm here. Indeed, take $h = (\frac{1}{n})$, then

$$\frac{k}{2k-2}(e^1 + e^{2k-2}) \in \frac{\partial f(\frac{1}{k}h)}{\frac{1}{k}},$$

which converges weakly but not in norm to $\frac{1}{2}e^1$. This follows by checking that for $x = \frac{1}{k}h$, the maximum in (7.1) is attained at $n = 2k-2$.

Let us finally show that $\Delta_{f,0,0,\frac{1}{k}}$ also converges to q in the Mosco sense. Since condition (α) is already clear, we have to check condition (β). Let $h^k \rightharpoonup h$. Fix $\varepsilon_k \to 0^+$ such that $\varepsilon_k k \to \infty$. Suppose $\|h^k\| \le M < \infty$, and let $h_1 \neq 0$ for simplicity. Now let $\delta_k = \frac{M}{\sqrt{\varepsilon_k k}} \to 0$. We define

$$I(k) = \{n \in \mathbb{N} : h_n^k \le -\delta_k\}.$$

Then $M^2 \ge \|h^k\|^2 \ge \sum_{n \in I(k)} (h_n^k)^2 \ge \delta_k^2 \text{card} I(k)$, so card $I(k) \le \frac{M^2}{\delta_k^2} = \varepsilon_k k < 2\varepsilon_k k$. Therefore, as the interval $J(k) = [2k/h_1^k - \varepsilon_k k, 2k/h_1^k + \varepsilon_k k]$ has length $2\varepsilon_k k$, we may pick $n(k) \in J(k) \setminus I(k)$. But then

$$\begin{aligned} k^2 f\left(\frac{1}{k}h^k\right) &\ge \frac{h_1^k + h_{n(k)}^k}{kn(k)} - \frac{1}{n(k)^2} \ge k^2\left(\frac{h_1^k - \delta_k}{k(2k/h_1^k - \varepsilon_k k)} - \frac{1}{(2k/h_1^k + \varepsilon_k k)^2}\right) \\ &= \frac{h_1^k - \delta_k}{2/h_1^k - \varepsilon_k} - \frac{1}{(2/h_1^k + \varepsilon_k)^2} \to \frac{(h_1)^2}{4} = q(h), \end{aligned}$$

which proves condition (β).

Example 4. Define $f : \ell_2 \to \mathbb{R}$ by

(7.2)
$$\begin{aligned} f(x) &= \sup_{n \in \mathbb{N}} \left(\frac{\frac{6}{\pi^2}(x_1 + \frac{1}{2}x_2 + \cdots + \frac{1}{n}x_n) + \frac{1}{\sqrt{n}}(x_{n+1} + \cdots + x_{2n})}{n} - \frac{1}{n^2} \right) \\ &=: \sup_{n \in \mathbb{N}} f_n(x). \end{aligned}$$

As in Example 3, $f(0) = 0$, $f \ge 0$, $\nabla f(0) = 0$. Again an elementary calculation shows that f is Lipschitz smooth at 0. Therefore second order differentiability of f at 0 follows since $\Delta_{f,0,0,\frac{1}{k}}(h) = k^2 f(\frac{1}{k}h)$ converges for every finite sequence h. We show that f is not second order Mosco differentiable at 0. Fix $h = (\frac{1}{n})$. Then $\Delta_{f,0,0,\frac{1}{k}}(h) \to \frac{1}{4} = q(h)$. This may be seen by checking that for $x = \frac{1}{k}h$, assymptotically the sup in (7.2) is attained for $n = 2k$. Now let us define a sequence $h^k \rightharpoonup h$ by

$$h^k = \left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{2k}, -\frac{1}{\sqrt{k}}, \ldots, -\frac{1}{\sqrt{k}}, 0, 0, \ldots\right),$$

with $-\frac{1}{\sqrt{k}}$ occurring $2k$ times. We prove that $\liminf k^2 f(\frac{1}{k}h^k) \leq \frac{3}{16} < \frac{1}{4}$, which shows that condition (β) of Mosco convergence is violated.

First consider the terms $k^2 f_n(\frac{1}{k}h^k)$ with $n \leq k$. The maximum possible contribution here comes from $n = k$, giving

$$k^2 f_k\left(\frac{1}{k}h^k\right) \leq \frac{6}{\pi^2}\Big(\frac{\pi^2}{6} + \frac{1}{k}\Big) + \frac{1}{\sqrt{k}}\log\frac{2k}{k+1} - 1 \to 0.$$

Next consider $k \leq n \leq 2k$, that is, $n = \alpha k$ for $1 \leq \alpha \leq 2$, say. Then

$$k^2 f_n\left(\frac{1}{k}h^k\right) = k^2\Big(\frac{1 + \mathrm{o}(1) - \alpha^{-1/2}(2\alpha - 2)}{\alpha k^2} - \frac{1}{\alpha^2 k^2}\Big)$$
$$\doteq \frac{\alpha - \alpha^{1/2}(2\alpha - 2) - 1}{\alpha^2} < 0,$$

so these indices n do not contribute to the supremum (7.2).

Next consider $n = \alpha k$, $2 \leq \alpha \leq 4$. Here we have

$$k^2 f_n\left(\frac{1}{k}h^k\right) = k^2\left(\frac{\frac{6}{\pi^2}\sum_{j=1}^{2k}\frac{1}{j^2} - \frac{6}{\pi^2}\sum_{j=2k}^{\alpha k}\frac{1}{j\sqrt{k}} - \sum_{j=\alpha k}^{4k}\frac{1}{k\sqrt{\alpha}}}{\alpha k^2} - \frac{1}{\alpha^2 k^2}\right)$$
$$\leq \frac{1 + \frac{6}{2k\pi^2} - \frac{6}{\pi^2\sqrt{k}}\log\frac{\alpha k}{2k} - \frac{(4-\alpha)k}{k\sqrt{\alpha}}}{\alpha} - \frac{1}{\alpha^2} < 0,$$

eventually, so these indices do not contribute either. Finally consider $n = \alpha k$ with $\alpha \geq 4$. Here we have

$$k^2 f_n\left(\frac{1}{k}h^k\right) = \frac{1 + \mathrm{o}(1) - \frac{1}{\sqrt{k}}\log 2}{\alpha} - \frac{1}{\alpha^2} \doteq \frac{\alpha - 1}{\alpha^2},$$

and since $\alpha \geq 4$, this has maximum value $\frac{3}{16}$, as claimed. This completes our argument.

Letting C denote the epigraph of f, we see that C is second order Gâteaux smooth at $(0,0)$, but P_C is not directionally differentiable at $(0,-1)$, since C fails to be second order Mosco smooth at $(0,0)$. As a consequence of Lemma 4.4, some of the outer parallel sets $C_{[\varepsilon]}$ therefore fails to be second order Gâteaux smooth at $(0,-\varepsilon)$. This means that the Gâteaux analogue of Corollary 6.4 is false.

References

[1] T. Abatzoglu, *The metric projection on C^2 manifolds in Banach spaces.* J. Approx. Theory, **26** (1979), 204 - 211.

[2] N. Aronszajn,*Differentiability of Lipschitz mappings between Banach spaces,* Studia Math., **LVII**, (1976), 147-190.

[3] H. Attouch, *Familles d'opérateurs maximaux monotones et mesurabilité,* Ann. Mat. Pura Appl., **120** (1979), 35-111.

[4] H. Attouch, *Variational Convergence for Functions and Operators,* Pitman 1984.

[5] H. Attouch, J.L. Ndoutoume, M. Théra, *On the equivalence between epi-convergence of sequences of functions and graph convergence of their derivatives,* to appear.

[6] H. Attouch, R.J.-B. Wets, *Isometries for the Legendre-Fenchel transform,* Trans. Amer. Math. Soc., **296** (1986), 33-60.

[7] V.I. Berdyshev, *Differentiability of the metric projection in normed spaces.* Collection: Approximation of functions by polynomials and splines, **150**, 58-71. Akad. Nauk. SSSR, Ural. Nauchn. Tsentr., Sverdlovsk. (1985).

[8] V.I. Berdyshev, *Metric projection on the class $H(\Omega)$.* Dokl. Akad. Nauk. SSSR, **266** (1982).

[9] J.M. Borwein, D. Noll, *Second order differentiability of convex functions on Banach spaces,* Trans. Amer. Math. Soc., **342** (1994), 43-81.

[10] H. Busemann, *Convex surfaces,* Interscience Publ., New York, 1955.

[11] J.P.R. Christensen, *Measure theoretic zero sets in infinite dimensional spaces and applications to differentiability of Lipschitz mappings II,* Coll. Anal. Funct., Bordeux (1973), 29-39.

[12] K. Deimling, *Nonlinear Functional Analysis*, Springer Verlag 1985.

[13] M. Fabian, *Lipschitz smooth points of convex functions and isomorphic characterization of Hilbert spaces,* Proc. London Math. Soc., **51** (1985), 113-126.

[14] S. Fitzpatrick, R.R. Phelps, *Differentiability of the metric projection in Hilbert space,* Trans. Amer. Math. Soc., **270** (1982), 483-501.

[15] A. Haraux, *How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities.* J. Math. Soc. Japan, **29** (1977), 615 - 631.

[16] J.B. Hiriart-Urruty, *At what points is the projection mapping differentiable?* Amer. Math. Monthly, **89** (1982), 456-458.

[17] R.B. Holmes, *Smoothness of certain metric projections on Hilbert space,* Trans. Amer. Math. Soc., **184** (1973), 87-100.

[18] K. Malanowski, *Differentiability with respect to parameters of solutions to convex programming problems,* Math. Programming, **33** (1985), 352-365.

[19] F. Mignot, *Contrôle dans les inéquations variationelles elliptiques,* J. Functional Anal., **22** (1976), 130-185.

[20] D. Noll, *Graphical methods in first and second order differentiability theory of integral functionals,* J. Set-valued Analysis, **2** (1994), 241-258.

[21] R.R. Phelps, *Openness of the metric projection in certain Banach spaces,* J. Approx. Theory, **42** (1984), 70-79.

[22] R.R. Phelps, *The gradient projection method using Curry's steplength,* SIAM J. Control Opt., **24** (1986), 692-699.

[23] R.T. Rockafellar, *Second order optimality conditions in nonlinear programming obtained by way of epi-derivatives,* Math. Oper. Res., **14**, 462-484.

[24] R.T. Rockafellar, *Generalized second derivatives of convex functions and saddle functions,* Trans. Amer. Math. Soc., **322** (1990), 51-77.

[25] A. Shapiro, *On differentiability of the metric projection in* $\mathbb{R}^n$. *Boundary case.* Proc. Amer. Math. Soc., **99** (1987), 123-128.

[26] A. Shapiro, *Directional differentiability of metric projections onto moving sets at boundary points.* J. Math. Anal. Appl., **131** (1988), 392-403.

[27] J. Sokolovski, *Shape sensitivity analysis of boundary optimal control problems for parabolic systems.* SIAM J. Control Optim., **26** (1988), 763-787.

[28] J. Sokolowski, *Sensitivity analysis of control constrained optimal control problems for distributed parameter systems.* SIAM J. Control Opt., **25** (1987), 1542-1556.

[29] E. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory,* Contrib. to Nonlin. Funct. Anal., 27 Math. Res. Center, Univ. of Wisconsin, AP (1971), 237-424.

[30] E. Zarantonello,*L'algebre des projecteurs coniques.* Analyse Convexe et Ses Applications, Lecture Notes in Econon. and Math. Syst. 102, Springer Verlag.

Received December 10, 1992. Research supported by Deutsche Forschungsgemeinschaft (DFG) and Natural Sciences and Engineering Research Council of Canada (NSERC). This paper was written while the author was a visitor at the University of Waterloo.

DEPARTMENT OF COMBINATORICS AND OPTIMIZATION
UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO, CANADA
N2L 3G1
AND
UNIVERSITÄT STUTTGART, MATHEMATISCHES INSTITUT B
PFAFFENWALDRING 57, 70550 STUTTGART, BR DEUTSCHLAND