TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 183, September 1973

SMOOTHNESS OF CERTAIN METRIC PROJECTIONS ON HILBERT SPACE

BY

RICHARD B. HOLMES

ABSTRACT. A study is made of differential properties of the distance function and the metric projection defined by a closed convex subset of Hilbert space. The former mapping is also considered within the context of more general Banach spaces.

Introduction. This paper contains a contribution to the study of best approximation out of a closed convex set K belonging to a Hilbert space X. As any such set is a Chebyshev set, there is defined on X a map $P_K: X \to K$ which assigns to each element of X its best approximation (nearest point) in K. The map P_K is generally called the *metric projection* of X on K. P_K is well known to be a nonexpansive map on X. In the present work we shall be concerned with differential properties of P_K, under appropriate smoothness assumptions about the boundary of K.

By considering such a simple set as a line segment, it is clear that P_K may fail to possess a two-sided directional derivative at certain points in X. More surprisingly, J. Kruskal [5] recently constructed a convex subset of R^3 whose associated metric projection failed to have a one-sided directional derivative at some points of R^3. On the other hand, due to the aforementioned Lipschitz continuity of P_K, the classical theorem of Rademacher and Stepanoff [3, p. 216] guarantees that P_K is almost everywhere (Fréchet) differentiable when X is finite dimensional. (This observation, incidentally, yields an affirmative answer to the question of Kruskal at the bottom of p. 697 of [5].) However, when X is infinite dimensional, there is no guarantee that Lipschitzian maps are differentiable (cf. [8, pp. 91–92]).

The main result to be presented below is (roughly) that if for some $x \in X$ the boundary of K is of class C^{p+1} near $P_K(x)$, then P_K is of class C^p on a neighborhood of the ray normal to K at $P_K(x)$ (hence in particular on a neighborhood of x). The proof ultimately reduces to an application of the implicit function theorem. We further obtain various properties of the differential $DP_K(x)$, and in par-

Received by the editors March 1, 1972 and, in revised form, October 30, 1972.

AMS (MOS) subject classifications (1970). Primary 41A65; Secondary 46B99, 46C10.

Key words and phrases. Best approximation, distance function, metric projection, Hilbert space, implicit function theorem, flat point.

ticular characterize those cases where this operator is a constant multiple of the orthogonal projection of X onto the tangent space to K at $P_K(x)$. It turns out, for example, that $DP_K(x)$ equals this orthogonal projection exactly when $P_K(x)$ is a flat point of K (definition below).

The two sections containing the aforementioned results are preceded by two other sections of a preliminary nature. The first of these collects and summarizes generally known facts about approximation in Hilbert space. The second is devoted to a discussion of what the assertion "the boundary of K is of class C^p near y" should mean. Two natural definitions are given, one intrinsic and one extrinsic, and shown to be equivalent.

Under the hypothesis of the main result indicated above, namely that the boundary of K be of class C^{p+1} near $P_K(x)$, it also follows that the function "distance to K" is of class C^{p+1} on a neighborhood of x. Now it is known (cf. § 1 below) that this distance function is always of class C^1, regardless of the boundary behavior of K. We extend this result in the Appendix to other Banach spaces, provided they are endowed with a differentiable norm; thus it is seen that a quadratic norm is not essential for smoothness of the distance function.

1. **Preliminaries.** Let K be a closed convex subset of the real Hilbert space X. We shall assume that K contains a core point relative to its closed affine hull Y, so that after a translation, we may suppose that K absorbs this hull. Of course such an assumption entails no loss of generality when X is finite dimensional. Now, letting P_Y be the orthogonal projection of X on Y, we have

$$P_K = (P_K \mid Y) \circ P_Y.$$

Since our interest is in smoothness properties of P_K, and since P_Y is a bounded linear (hence C^∞) mapping, we may as well assume that $X = Y$.

We now introduce for consideration several convex functions on X:

$$\delta_K(x) = \begin{cases} 0, & \text{if } x \in K, \\ +\infty, & \text{if } x \notin K; \end{cases}$$

$$d_K(x) = \text{distance from } x \text{ to } K; \qquad e_K = \tfrac{1}{2} d_K(\cdot)^2;$$

$$w = \tfrac{1}{2}\|\cdot\|^2; \quad \psi_K = w - e_K; \quad \rho_K(x) = \inf\{t > 0\colon\ x \in tK\}.$$

The first and last of these functions are called respectively the *indicator* and the *gauge* of K.

Our first observation is that the functions d_K and e_K are of class C^1 on the open set $X \setminus K$ (all derivatives are to be understood in the Fréchet sense, unless otherwise noted). This fact seems to have first been established by Moreau

[6, p. 286], for the function e_K, as a special case of the differentiability of functions of the form $f \oplus w$, the inf-convolution of a convex function f with w (in the present case we have $e_K = \delta_K \oplus w$). A direct proof of the differentiability of e_K, based on Moreau's argument, also appears in [4, p. 64]. The formulas for the respective gradients which finally emerge are

$$\nabla e_K = I - P_K, \qquad \nabla d_K = \operatorname{sgn} \circ (I - P_K), \tag{1.1}$$

where $\operatorname{sgn}(x) \equiv x/\|x\|$, $x \neq \theta$. The first formula shows that e_K is actually of class C^1 on all of X.

We note that the smoothness of the distance function d_K does not depend on any special boundary behavior of K; convexity alone is adequate for first order smoothness (but not, as we have seen, for higher order smoothness). We show in the Appendix that the validity of this fact does not require a Hilbert norm: it holds in any Banach space with a differentiable norm.

From (1.1) we see that

$$P_K = \nabla \psi_K, \tag{1.2}$$

using that $I = \nabla w$. Now various authors (e.g., Asplund [1], Zarantonello [9]) have noted that the function ψ_K is convex, although again this fact goes back to Moreau's paper [6], where it is shown that every "prox" mapping is the gradient of a convex function. Because of its interest and eventual importance to us, we shall next outline an alternative derivation of the fact that P_K is the gradient of a convex function; in so doing we shall also record several other facts which will be needed later.

To best approximate x from K is to solve the ordinary convex program

$$\min\{w(x-z): \rho_K(z) \leq 1\}.$$

Recalling the Kuhn-Tucker conditions [7, pp. 42–44], [4, p. 34], we find that there exists a number $\lambda > 0$ and a subgradient $\phi \in \partial\rho_K(P_K(x))$ such that

$$\lambda\phi = \nabla w(x - P_K(x)) = x - P_K(x). \tag{1.3}$$

Recalling further that the gauge ρ_K is positively homogeneous and subadditive, we see that

$$\partial\rho_K(y) = \{\phi \in X^* : \phi \leq \rho_K, \phi(y) = \rho_K(y)\}. \tag{1.4}$$

(Whenever convenient, we are, of course, identifying X and X^* in the usual manner.)

Next, letting $\langle \cdot, \cdot \rangle$ be the inner product on X, and fixing $y \in K$, we utilize (1.3) and (1.4) to obtain

$$\lambda\phi(y) \le \lambda\rho_K(y) \Rightarrow \langle x - P_K(x), y\rangle$$
$$\le \lambda\rho_K(y) \le \lambda = \lambda\phi(P_K(x)) = \langle x - P_K(x), P_K(x)\rangle,$$

that is,

$$0 \le \langle x - P_K(x), P_K(x) - y\rangle, \quad y \in K, \tag{1.5}$$

which is the well-known Bourbaki-Cheney-Goldstein inequality for $P_K(X)$. This inequality in turn easily leads to the monotonicity inequality for the metric projection

$$\|P_K(x) - P_K(y)\|^2 \le \langle P_K(x) - P_K(y), x - y\rangle, \tag{1.6}$$

$\forall\, x, y \in X$. Finally, (1.2) and (1.6) together entail the convexity of ψ_K. (It is important to emphasize that there is no circularity in this argument. That is, while the usual proofs of (1.1) and hence (1.2) depend on (1.5), our proof of (1.2) proceeds by treating it as a special case of (A.1), which is established independently of (1.5).). We also note the following local growth estimates for ψ_K:

$$0 \le \psi_K(x + y) - \psi_K(x) - \langle P_K(x), y\rangle \le w(y), \quad \forall x, y \in X.$$

Let us recall that if y is a boundary point of K (notation: $y \in \partial K$), the *normal cone* $N(y, K)$ to K at y consists of all $z \in X$ for which $\langle \cdot, z\rangle$ attains its maximum value over K at y. As an immediate consequence of (1.5) we find for the P_K-inverse image of $P_K(x)$:

$$P_K^{-1}(P_K(x)) = P_K(x) + N(P_K(x), K), \tag{1.7}$$

$\forall\, x \in X \backslash K$. The cone defined by (1.7) is just a ray emanating from $P_K(x)$ exactly when the latter is a smooth point of K. Indeed, in this case the ray is given by $\{P_K(x) + t(x - P_K(x)) \colon t \ge 0\}$, as follows from (1.5).

To conclude this section we assemble a few facts concerning the gauge ρ_K. Under our hypotheses on K this is a continuous convex function on X (continuity follows from the assumption that K is an absorbing set and the completeness of X which entails that K must then be a neighborhood of the origin).

Lemma 1. *Let y be a boundary point of K. Then $\rho_K(y) > 0$. Suppose also that ρ_K is differentiable at y. Then $\nabla\rho_K(y) \ne \theta$. Furthermore, the (open) normal ray $\{y + t\nabla\rho_K(y) \colon t > 0\}$ lies exterior to K, and projects via P_K onto y.*

Proof. If $\rho_K(y) = 0$, then K contains the ray $[0, \infty)y$; this together with the fact that K is a convex neighborhood of the origin implies that y is interior to K, a contradiction. From this in turn, and the relation

$$\rho_K(y) = \langle y, \nabla\rho_K(y)\rangle \tag{1.8}$$

(because ρ_K is positively homogeneous), we see that $\nabla\rho_K(y) \neq \theta$. Next, combining these two observations with (1.4), we have for any $t > 0$

$$\begin{aligned} 1 = \rho_K(y) &< \rho_K(y) + t\|\nabla\rho_K(y)\|^2 \\ &= \langle y + t\nabla\rho_K(y), \nabla\rho_K(y)\rangle \leq \rho_K(y + t\nabla\rho_K(y)). \end{aligned}$$

Finally, the last assertion of the lemma follows from (1.4) and (1.5).

2. **Smoothness of the boundary.** We continue with the assumptions on the convex set K made in the previous section. Let $y \in \partial K$. A reasonable definition of the statement "the boundary ∂K is of class C^p near y" is that the gauge ρ_K should be of class C^p on a neighborhood of y. However, for this definition to be consistent, we must then resolve the following question. The set K is not assumed to have a center of symmetry. Thus when K was translated to the origin, so that ρ_K could be defined, we simply picked an arbitrary core point $k_1 \in K$, and replaced K by $K - k_1$. Let us call the resulting gauge ρ_1. If some other core point $k_2 \in K$ had been chosen instead, leading to another gauge ρ_2, how are the differentiability properties of ρ_2 related to those of ρ_1? In particular, if $y \in \partial K$, and ρ_1 is of class C^p on a neighborhood of $y - k_1$, can we assert that ρ_2 is also of class C^p on a neighborhood of $y - k_2$?

One way to settle such questions is to consider another more intrinsic criterion for C^p-smoothness of ∂K near y. Namely, we can ask that some (relative) neighborhood of y in ∂K be a C^p-submanifold of X, appropriately modeled on some hyperplane in X. More precisely, we shall require that a neighborhood of y in ∂K be the range of a C^p-embedding defined on an open set in some Hilbert space, which is also an immersion of a special kind. The next result gives the details and establishes the equivalence of the two definitions, thereby in particular resolving affirmatively the question raised in the preceding paragraph. Let us agree in advance that if T is a bounded linear map between Hilbert spaces, having a closed range, then the *corank* of T is the codimension of its range (with respect to the target Hilbert space).

Theorem 1. *Let f be a continuous real-valued function on the Hilbert space X.*

(a) *Assume that $f(x_0) = 1$, that $\nabla f(x_0) \neq \theta$, and that f is of class C^p $(p \geq 1)$ on an x_0-neighborhood. Then there is a θ-neighborhood V in a Hilbert space Z, and a C^p-embedding $r\colon V \to X$ such that $f(r(v)) = 1$, $r(\theta) = x_0$, and $r'(v)$ is left invertible with corank unity, for every $v \in V$.*

(b) *Assume that* f *is a gauge, that* V, X, *and* r *are defined as in* (a), *that* $r'(v)$ *is left invertible for* $v \in V$, *and that* corank $(r'(\theta)) = 1$. *Then* f *is of class* C^p *on an* x_0*-neighborhood in* X.

Before giving the proof, let us remark that if in part (a) it is assumed that X is finite dimensional, then any Hilbert space Z which meets the conditions of (a) must have dimension one less than X, so that the corank condition is automatically fulfilled.

Proof of Theorem 1. Let Z be the tangent space $\{\nabla f(x_0)\}^{\perp}$, and define a function $F: Z \times R^1 \to R^1$ by

$$F(z, \alpha) = f(x_0 + z + \alpha \nabla f(x_0)) - 1.$$

We can apply the implicit function theorem to obtain a θ-neighborhood $V \subset Z$ and a C^p-map $g: V \to R^1$ such that $g(\theta) = 0$ and

$$f(x_0 + v + g(v)\nabla f(x_0)) = 1.$$

We define $r: V \to X$ by

$$r(v) = x_0 + v + g(v)\nabla f(x_0).$$

If $z \in Z$ and $\|z\| = 1$, we find for $v \in V$

$$\|r'(v) \cdot z\|^2 = \|z\|^2 + |g'(v) \cdot z|^2 \|\nabla f(x_0)\|^2 \geq \|z\|^2 = 1,$$

so that $r'(v)$ is left invertible. Since $g'(\theta) = \theta$, we have $r'(\theta) = I_Z$, so that corank $(r'(\theta)) = 1$. That we then have corank $(r'(v)) = 1$ for $v \in V$ follows from the next lemma, after possibly replacing V by a smaller θ-neighborhood in Z.

Lemma 2. *Let* X *and* Z *be Hilbert spaces, and let* A *be a continuous map from an open set* V *in some Banach space to the space* $[Z, X]$ *of all bounded linear maps from* Z *to* X. *Then* $\{v \in V: A(v)$ *is left invertible with corank unity*$\}$ *is open.*

Proof. It is clear that $\{v \in V: A(v)$ is left invertible$\}$ is open; call it V_1. For $v \in V_1$, let $A^+(v)$ be the pseudo-inverse of $A(v)$, that is,

$$A^+(v) = (A^*A)^{-1}A^*(v) \in [X, Z].$$

These maps are surjective and A^+ is continuous on V_1. Suppose that there is $v_0 \in V_1$ for which corank $(A(v_0)) = 1$. Then we must find a v_0-neighborhood on which nullity $(A^+(v)) = 1$. First we show that there is a v_0-neighborhood on which nullity $(A^+(v)) \leq 1$. If not, there would be a sequence $v_n \to v_0$ with nullity $(A^+(v_n)) \geq 2$, and consequently a sequence $\{x_n\}$ of unit vectors in

$\ker(A^+(v_n)) \cap \ker(A^+(v_0))^\perp$. But then

$$0 < \inf_n \|A^+(v_0) \cdot y_n\| = \|(A^+(v_0) - A^+(v_n)) \cdot y_n\| \longrightarrow 0,$$

a contradiction. It remains to show that there is a v_0-neighborhood on which nullity $(A^+(v))$ is exactly one. If not, there would be a sequence $v_n \longrightarrow v_0$ with $A^+(v_n)$ injective. Let x be a unit vector for which $A^+(v_0) \cdot x = \theta$. Then

$$\begin{aligned} 0 = \|A^+(v_0) \cdot x\| &\geq \|A^+(v_n) \cdot x\| - \|(A^+(v_0) - A^+(v_n)) \cdot x\| \\ &\geq \|A^+(v_n)^{-1}\|^{-1} - o(1) = \|A(v_n)\|^{-1} - o(1) \\ &\longrightarrow \|A(v_0)\|^{-1} > 0, \end{aligned}$$

a contradiction. q.e.d.

At this point the proof of part (a) of Theorem 1 is complete, and we now begin work on part (b). Let us first note that, in the notation of Lemma 2, if the map A is of class C^p on V, and is also left invertible there, then there is a C^p-choice of left inverse for $A(v)$, namely $A^+(v)$. Thus, in addition to the hypotheses of (b), we may assume that corank $(r'(v)) = 1$ throughout V, and that there is a C^{p-1}-left inverse for $r'(v)$, say $q(v)$.

Lemma 3. *For $v \in V$, the hyperplane $H_v = r(v) + \text{range}(r'(v))$ is the tangent hyperplane to the surface $f^{-1}(1)$ at the point $r(v)$, in the sense that*

$$d_{H_v}(r(v_n)) = o(\|r(v_n) - r(v)\|),$$

as $v_n \longrightarrow v$. Consequently, H_v is the unique supporting hyperplane to the convex body $\{x \in X : f(x) \leq 1\}$ at the point $r(v)$.

We next define a C^{p-1}-map $L: V \longrightarrow [X, Z \times R^1]$ by

$$L(v) : x \mapsto (q(v) \cdot x, \langle x, r(v)\rangle).$$

These maps $L(v)$ are bijective. To see this we recall that the maps $q(v)$ are surjective and have nullity 1. We must also recognize that $r(v)$ cannot be orthogonal to $\ker(q(v))$. Because, if it were, then $r(v) \in \text{range}(r'(v))$, so that f would vanish on the tangent hyperplane $r(v) + \text{range}(r'(v))$, in contradiction to Lemma 3. This argument demonstrates that $L(v)$ is bijective and hence invertible. Consequently we can define a C^{p-1}-map $g: V \longrightarrow X$ by

$$g(v) = L(v)^{-1}(\theta, 1).$$

The proof of part (b) of Theorem 1 is now completed by verifying that $\nabla f(\lambda r(v)) = g(v)$, for $\lambda > 0$. q.e.d.

3. **Smoothness of the metric projection.** This section contains a proof and some discussion of the following theorem.

Theorem 2. *Let K be a closed convex body in the Hilbert space X, and suppose that ∂K is of class C^{p+1} near the boundary point y. Then there is a neighborhood of the open normal ray to K at y on which d_K is of class C^{p+1} and P_K is of class C^p.*

Remark. We emphasize the restriction to the *open* normal ray. It is not possible to improve the theorem to include the assertion that d_K and/or P_K have the indicated smoothness on a neighborhood of y. As a simple example, let K be the unit ball of X, so that ∂K is of class C^∞. Here we have

$$d_K = \begin{cases} 0, & \|x\| \le 1, \\ \|x\| - 1, & \|x\| \ge 1, \end{cases}$$

and

$$P_K = \begin{cases} I, & \|x\| \le 1, \\ \operatorname{sgn}(x), & \|x\| \ge 1. \end{cases}$$

Clearly, d_K (and consequently P_K) is not of class C^1 near any unit vector. More generally, Zarantonello [9, p. 300] has considered the differentiability of P_K at boundary points. He proves that there is a *nonlinear* "conical differential" for P_K at the point $y \in \partial K$, this differential being in fact the metric projection onto the supporting half-space to K at y (under the assumption that K is a convex body). Thus, regardless of the smoothness of K near y, P_K cannot be (Fréchet) differentiable at y. From the proof of Theorem 2 below it follows that d_K cannot be of class C^1 near y.

Proof of Theorem 2. As usual we assume that K has been translated so as to form a neighborhood of the origin. From § 1, the normal ray at y has the description $\{y + t\nabla\rho_K(y): t > 0\}$. We fix an x on this ray (so that $y = P_K(x)$ by Lemma 1) and prove that P_K is of class C^p near x; in view of (1.1) this will also establish the assertion about d_K.

Letting V be a y-neighborhood on which ρ_K is of class C^{p+1}, we introduce the function $F: X + V \to X$,

$$F(u, v) = u - v - d_K(u)\operatorname{sgn}(\nabla\rho_K(v)). \tag{3.1}$$

This function is of class C^p on $U \times V$ where U is any x-neighborhood on which d_K is of class C^p; in turn, this latter condition follows from (1.1) if P_K is of class C^{p-1} on U. Our interest in the function F lies in the fact that

$$F(u, v) = 0 \quad \Leftrightarrow \quad v = P_K(u). \tag{3.2}$$

Indeed, if $v = P_K(u)$, then by (1.3) we have a positive λ such that $\lambda \nabla \rho_K(v) = u - P_K(u)$; hence

$$\lambda = d_K(u)/\|\nabla \rho_K(v)\|,$$

and $F(u, v) = 0$. Conversely, suppose that $F(u, v) = 0$ but $v \notin K$. Let $K' = \{z \in X: \rho_K(z) \leq \rho_K(v)\}$. This is a neighborhood of K and so $d_{K'}(u) < d_K(u)$. But we have from (1.3) and Lemma 1 that $v = P_{K'}(u)$, so that the contradiction $d_{K'}(u) = \|w - v\| = d_K(u)$ results. Hence $v \in K$ and equals $P_K(u)$, so that (3.2) is justified.

We are going to apply the implicit function theorem to F in the case $p = 1$. The case for arbitrary p will follow by induction from (1.1) and the remarks in the preceding paragraph. Thus what remains to be shown is that the partial derivative $D_2F(x, y)$ is an automorphism of X, where x and $y = P_K(x)$ are fixed. Making use of the chain rule, we find

$$D_2F(x, y) = -I - d_K(x) D \operatorname{sgn}(\nabla \rho_K(y)) \circ \nabla^2 \rho_K(y) \equiv -(I + T). \tag{3.3}$$

The next lemma will reveal the structure of the operator T. To formulate it, we introduce Q, the orthogonal projection of X onto the tangent space $M = \{\nabla \rho_K(y)\}^{\perp} \equiv \{x - y\}^{\perp}$ at y, and the positive semidefinite operator $S = \nabla^2 \rho_K(y)$.

Lemma 4. *The operator T defined by* (3.3) *is a positive multiple of $Q \circ S$.*

Proof. First we differentiate the function $\operatorname{sgn}(\cdot)$ to obtain

$$D \operatorname{sgn}(u) \cdot v = (\|u\| v - \langle \operatorname{sgn}(u), v\rangle u)/\|u\|^2,$$

valid for all $u \neq \theta$ and all v in X. Then we can express T by

$$T(z) = \frac{d_K(x)}{\|\nabla \rho_K(y)\|^2} (\|\nabla \rho_K(y)\| S(z)) - \langle \operatorname{sgn}(\nabla \rho_K(y)), S(z)\rangle \nabla \rho_K(y))$$

$$= \frac{d_K(x)}{\|\nabla \rho_K(y)\|} (S(z) - \langle \operatorname{sgn}(\nabla \rho_K(y)), S(z)\rangle \operatorname{sgn}(\nabla \rho_K(y)))$$

$$= \frac{d_K(x)}{\|\nabla \rho_K(y)\|} (S(z) - (I - Q) \circ S(z)) = \frac{d_K(x)}{\|\nabla \rho_K(y)\|} Q \circ S(z) \equiv cQ \circ S(z). \qquad \text{q.e.d.}$$

We now complete the proof of Theorem 2 by showing that $I + T$ is bijective on X. Suppose that $(I + T)(z) = \theta$ for some $z \in X$, that is $z + cQ \circ S(z) = \theta$. Then $z \in M$ (the tangent space to K at y), and so

$$0 \leq \langle S(z), z\rangle = \langle S(z), Q(z)\rangle$$
$$= \langle Q \circ S(z), z\rangle = \langle -c^{-1}z, z\rangle = -c^{-1}\|z\|^2,$$

whence $z = \theta$ and $(I + T)$ is injective.

Finally, we show $(I + T)$ is surjective. We note first that the restriction $T \mid M$ is positive semidefinite. Since M is invariant under $(I + T)$ it then follows that $(I + T) \mid M$ is an automorphism of M. Now let $n = \operatorname{sgn}(\nabla\rho_K(y))$, and observe that $n \notin M$ while $T(n) \in M$. Hence $p \equiv (I + T)(n) \notin M$. Expressing any $z \in X$ as $\alpha p + m$, for some scalar α and $m \in M$, we set

$$m' = ((I + T) \mid M)^{-1}(m),$$

and see that

$$(I + T)(\alpha n + m') = \alpha p + m \equiv z.$$

The proof of Theorem 2 is now complete.

4. The differential of the metric projection. In this section we discuss some properties of the differential $DP_K(x)$, under the hypotheses of Theorem 2, and the notation of § 3. We begin by evaluation of the partial derivative $D_1F(x, y)$. Using (1.1) and (1.3) we find

$$D_1F(x, y) = I - \langle \operatorname{sgn}(x - P_K(x)), \cdot\rangle \operatorname{sgn}(\nabla\rho_K(y))$$
$$= I - \langle \operatorname{sgn}(\nabla\rho_K(y)), \cdot\rangle \operatorname{sgn}(\nabla\rho_K(y)) = Q.$$

Consequently, we obtain the formula

$$DP_K(x) = (I + T)^{-1} \circ Q = (I + cQ \circ S)^{-1} \circ Q, \tag{4.1}$$

where c is the constant $d_K(x)/\|\nabla\rho_K(y)\|$ (using Lemma 4).

Although not apparent immediately from (4.1), it follows from (1.2), the convexity of ψ_K, and the nonexpansiveness of P_K, that $DP_K(x)$ is a selfadjoint positive semidefinite operator on X of norm at most unity. We observe from (4.1) that $DP_K(x)$ can never vanish. However, its norm can be arbitrarily small, as we see by taking K to be the unit ball in X, and computing that $DP_K(x) = \|x\|^{-1}Q$. We can also see from (4.1) and the results of § 3 that the tangent space M reduces $DP_K(x)$ and that $DP_K(x) \mid M$ is an automorphism.

Our final results characterize relations of the form $DP_K(x) = \mu Q$ (where $0 <$

$\mu \leq 1$) in terms of the behavior of the operator $S \equiv \nabla^2 \rho_K(y)$, introduced in § 3. It is convenient to begin with a lemma.

Lemma 5. *The range of S is orthogonal to y, and hence $S(y) = \theta$.*

Proof. For all $u \neq \theta$ we have the Euler relation $\rho_K(u) = \langle u, \nabla \rho_K(u) \rangle$ (cf. (1.8)). We differentiate both sides of this equation to obtain

$$\langle v, \nabla \rho_K(u) \rangle = \langle u, \nabla^2 \rho_K(u) \cdot v \rangle + \langle v, \nabla \rho_K(u) \rangle,$$

for all $v \in X$. Therefore, setting $u = y$ yields $\langle y, S(v) \rangle = 0$; the second assertion of the lemma is immediate from the selfadjointness of S. q.e.d.

Let us agree to say that y is a *flat point* of K if $S = 0$.

Theorem 3. *We have $DP_K(x) = Q$ if and only if $y \equiv P_K(x)$ is a flat point of K.*

Proof. The condition is clearly sufficient since

$$S = 0 \Rightarrow T = 0 \Rightarrow (I + T) \circ Q = Q \Rightarrow Q = (I + T)^{-1} \circ Q = P_K(x),$$

using (4.1). Conversely, if $DP_K(x) = Q$, then by (4.1) again, we have $T \mid M = Q$. Then by Lemma 4, $Q \circ S \mid M = 0$. Hence the range of $S \mid M$ lies in the null space of Q, that is, in $\operatorname{span}(\nabla \rho_K(y))$. But, by Lemma 5, this range also lies in $(y)^\perp$. We claim that this entails $S \mid M = 0$. Because,

$$\operatorname{span}(\nabla \rho_K(y)) \cap (y)^\perp = M^\perp \cap (y)^\perp = (\operatorname{span}\{y, M\})^\perp,$$

and this orthogonal complement will be $\{\theta\}$ provided that $y \notin M$. If we did have $y \in M$ then it would follow from (1.8) that

$$0 = \langle y, \nabla \rho_K(y) \rangle = \rho_K(y);$$

this equation, however, is a contradiction to the first assertion of Lemma 1. So we have $S \mid M = 0$. But since $S(y) = \theta$ by Lemma 5, and since we just showed $y \notin M$, it follows that $S = 0$. q.e.d.

Lastly, we consider the equation $DP_K(x) = \mu Q$ for $0 < \mu < 1$. Again we wish to characterize this possibility in terms of the behavior of S. Since we know that S annulls a subspace complementary to M (namely, $\operatorname{span}(y)$), it will suffice to describe the behavior of the restriction $S \mid M$.

Theorem 4. *We have $DP_K(x) = \mu Q$ (where $0 < \mu < 1$) if and only if there exist $\lambda > 0$ and a tangent vector $m' \in M$ such that $S \mid M = \lambda I - \langle \cdot, m' \rangle \nabla \rho_K(y)$. Further, $m' = \theta$ if and only if $P_K(x)$ is a multiple of x (which is to say that at x metric projection coincides with radial projection on K).*

Proof. (*Necessity*) We assume that $DP_K(x) = \mu Q$, for some μ, $0 < \mu < 1$. From (4.1) we obtain

$$\lambda Q = Q \circ S \circ Q, \qquad \lambda = (1 - \mu)/c\mu > 0.$$

In particular,

$$\lambda m = Q \circ S(m), \quad \forall m \in M \tag{4.2}$$

(which shows, incidentally, that $S \mid M$ is invertible). Now (4.2) implies the existence of $\phi \in M^*$ such that $S(m) = \lambda m - \phi(m)\nabla\rho_K(y)$, so we may take m' to be the Riesz representer of ϕ. Next, by Lemma 5, we can write, for any $m \in M$,

$$0 = \langle S(m), y\rangle = \langle \lambda m, y\rangle - \langle m, m'\rangle\langle y, \nabla\rho_K(y)\rangle, \tag{4.3}$$

and the coefficient of $\langle m, m'\rangle$ here is not zero by Lemma 1. Thus if $m' = \theta$ then $y \in M$; this means that $P_K(x)$ $(\equiv y)$ is a multiple of $x - P_K(x)$, and hence of x. Conversely, if $P_K(x)$ is a multiple of x, the preceding steps are reversible, so we have $y \in M^\perp$ and then (4.3) entails $m' = \theta$.

(*Sufficiency*) We now assume that we have a $\lambda > 0$ and $m' \in M$ for which $S(m) = \lambda m - \langle m, m'\rangle \nabla\rho_K(y)$, $\forall\, m \in M$. From this follows equation (4.2). Now define

$$\mu = 1/(1 + c\lambda),$$

where c is the constant appearing in (4.1). Then

$$(1 - \mu)Q = c\mu Q \circ S \circ Q = \mu T \circ Q,$$

$$Q = \mu(Q + T \circ Q) = \mu(I + T) \circ Q,$$

$$\mu Q = (I + T)^{-1} \circ Q = DP_K(x),$$

and this completes the proof.

APPENDIX

Smoothness of the distance function in Banach spaces. In this final section we let X be a Banach space whose norm will be assumed to be appropriately smooth. Our object is to discuss the smoothness of the distance function d_K where K is a given convex Chebyshev set in X. We show that this question can be resolved by appeal to general facts about the differentiability of convex functions.

Lemma 6. *Let f and g be real-vauled functions on X such that $f \le g$, f (resp. g) is continuous (resp. differentiable) at $x \in X$, and f is convex. Then f is differentiable at x and $\nabla f(x) = \nabla g(x)$.*

Proof. If ϕ is any subgradient of f at x, our hypotheses imply $\phi \le \nabla g(x)$

and so $\phi = \nabla g(x)$. That is, $\partial f(x) = \{\nabla g(x)\}$. This implies the conclusion of the lemma using the formula of Moreau and Pshenichnii (cf. [4, pp. 27–28]). q.e.d.

Let us note that in Lemma 6 the assumed differentiability of g may be either of Fréchet or Gateaux type and the same type will then be obtained by f. In the following theorem we shall understand differentiation to be in the Fréchet sense, but the result remains valid in the case of Gateaux differentiability as well.

Theorem 5. *Let K be a convex Chebyshev subset of the Banach space X, with metric projection P_K.*

(a) *If for some $x \in X \setminus K$ the norm on X is differentiable at $x - P_K(x)$, then the distance function d_K is differentiable at x.*

(b) *If the norm on X is differentiable on $X \setminus \{\theta\}$ (i.e., X is a "strongly smooth" Banach space), then d_K is of class C^1 on $X \setminus K$.*

Proof. To establish part (a) we simply apply Lemma 6 with $f = d_K$ and $g(z) = \|z - P_K(x)\|$. Now, as for part (b), our assumption together with the result of part (a) show us that the convex function d_K is differentiable throughout the open set $X \setminus K$. But now we can appeal to a result of Asplund and Rockafellar [2, p. 461] to conclude that d_K is actually of class C^1 on $X \setminus K$. q.e.d.

Note that the formula for ∇d_K implied by Lemma 6 is

$$\nabla d_K(x) = G(x - P_K(x)), \tag{A.1}$$

where G is the gradient of the norm on X, and that the continuity of ∇d_K in (b) does not depend on the continuity of P_K (which may of course be lacking).

Acknowledgement. The author is pleased to thank the referee of an earlier draft of this paper for his helpful and constructive criticism, particularly in regard to the material in the Appendix.

REFERENCES

1. E. Asplund, *Chebyshev sets in Hilbert space*, Trans. Amer. Math. Soc. 144 (1969), 235–240. MR 40 #6238.

2. E. Asplund and R. Rockafellar, *Gradients of convex functions*, Trans. Amer. Math. Soc. 139 (1969), 443–467. MR 39 #1968.

3. H. Federer, *Geometric measure theory*, Die Grundlehren der math. Wissenschaften, Band 153, Springer-Verlag, New York, 1969. MR 41 #1976.

4. R. Holmes, *A course on optimization and best approximation*, Springer-Verlag, Berlin and New York, 1972.

5. J. Kruskal, *Two convex counterexamples: a discontinuous envelope function and a nondifferentiable nearest-point mapping*, Proc. Amer. Math. Soc. 23 (1969), 697–703. MR 41 #4385.

6. J. Moreau, *Proximité et dualité dans un espace Hilbertian*, Bull. Soc. Math. France 93 (1965), 273–299. MR 34 #1829.

7. R. Rockafellar, *Convex functions, monotone operators and variational inequalities*, Theory and Applications of Monotone Operators, (Proc. NATO Advanced Study Institute, Venice, 1968), Edizioni "Oderisi", Gubbio, 1969, pp. 35–65. MR 41 #6028.

8. M. Vaĭnberg, *Some questions of the differential calculus in linear spaces*, Uspehi Mat. Nauk 7 (1952), no. 4 (50), 55–102. (Russian) MR 14, 384.

9. E. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory*, Contributions to Nonlinear Functional Analysis (E. Zarantonello, Ed.), Academic Press, New York and London, 1971, pp. 237–424.

DEPARTMENT OF MATHEMATICS, PURDUE UNIVERSITY, LAFAYETTE, INDIANA 47907