

NONLINEAR APPROXIMATION IN UNIFORMLY SMOOTH BANACH SPACES

BY

EDWARD R. ROZEMA AND PHILIP W. SMITH⁽¹⁾

ABSTRACT. John R. Rice [*Approximation of functions*. Vol. II, Addison-Wesley, New York, 1969] investigated best approximation from a nonlinear manifold in a finite dimensional, smooth, and rotund space. The authors define the curvature of a manifold by comparing the manifold with the unit ball of the space and suitably define the "folding" of a manifold. Rice's Theorem 11 extends as follows: **Theorem.** *Let X be a uniformly smooth Banach space, and $F: R^n \rightarrow X$ be a homeomorphism onto $M = F(R^n)$. Suppose $\nabla F(a)$ exists for each a in X , ∇F is continuous as a function of a , and $\nabla F(a) \cdot R^n$ has dimension n . Then, if M has bounded curvature, there exists a neighborhood of M each point of which has a unique best approximation from M . A variation theorem was found and used which locates a critical point of a differentiable functional defined on a uniformly rotund space Y . [See M. S. Berger and M. S. Berger, *Perspectives in nonlinearity*, Benjamin, New York, 1968, p. 58ff. for a similar result when $Y = R^n$.] The paper is concluded with a few remarks on Chebyshev sets.*

Introduction. Recently several papers have appeared concerning the approximation from nonlinear manifolds embedded in a Banach space (for example, Wulbert [10], [11] and Rice [6]). This paper extends Rice's results [6], which were obtained in a finite dimensional setting, to infinite dimensional Banach spaces. In the course of this investigation a variational theorem was discovered concerning the location of critical points of a functional.

Our manifold \mathcal{F} is given by a function $F: Y \rightarrow X$ which satisfies the following conditions:

(A) $Y = R^n$, n -dimensional Euclidean space.

(B) F is a homeomorphism onto its image \mathcal{F} (where \mathcal{F} is given the topology induced by X).

Received by the editors October 16, 1972 and, in revised form, April 12, 1973.

AMS (MOS) subject classifications (1970). Primary 41A50.

Key words and phrases. Approximation, nonlinear approximation, nonlinear functional analysis, uniformly smooth Banach space.

⁽¹⁾ The work on this paper was done while the writers were graduate students at Purdue University, Lafayette, Indiana. They were both supported by David Ross University Fellowships.

Copyright © 1974, American Mathematical Society

(C) The Fréchet derivative $\nabla F(\alpha)$ exists at every $\alpha \in Y$ and is continuous as a function of α .

(D) The dimension of $\nabla F(\alpha) \cdot Y$ is n for each $\alpha \in Y$.

The first section below is devoted to showing that the manifold does not "fold" back on itself excessively. The second section gives a definition of the curvature of a manifold in a Banach space by comparing the manifold with the unit ball of the space. In §3, we give our variational theorem which locates a critical point of a differentiable functional defined on a uniformly rotund space. In the final section we show that, if \mathcal{F} has bounded curvature, F satisfies (A) through (D), and X is (UF), then there exists a neighborhood of \mathcal{F} each point of which has exactly one best approximation. Since we assume that X is (UF), there is little overlap with the papers of Wulbert which deal with $C(T)$. We conclude with a few remarks concerning Chebyshev sets.

The authors would like to thank Rice for his encouragement during the writing of this paper, and Mr. Phillip Schmidt for his help in the presentation of the results in §3.

1. Notation, lemmas, and the folding of \mathcal{F} . Let X be a Banach space, θ the zero element of X , $M \subset X$, $x \in X$, and $r > 0$. Then

$$B(x, r) \equiv \{y \in X; \|x - y\| \leq r\},$$

$$B^0(x, r) \equiv \{y \in X; \|x - y\| < r\},$$

$$\partial B(x, r) \equiv \{y \in X; \|x - y\| = r\},$$

$$\text{dist}(x, M) \equiv \inf \{\|x - y\|; y \in M\},$$

$$P_M(x) \equiv \{y \in M; \|x - y\| = \text{dist}(x, M)\}.$$

The elements of $P_M(x)$ are called best approximations to x from M and the set-valued map $x \rightarrow P_M(x)$ is called the metric projection of X onto M .

If $x \neq \theta$, then $\text{sgn } x = x/\|x\|$; set $\text{sgn } \theta \equiv \theta$. If M is a subspace of X , set $S(M) \equiv \partial B(\theta, 1) \cap M$.

We introduce a notion of the "angle" or "deviation" $\delta(M, N)$ of the subspace M from the subspace N , which has been used in various forms by many authors:

$$\delta(M, N) \equiv \sup_{x \in S(M)} \text{dist}(x, S(N)).$$

This function is clearly not symmetric although the function

$$\Delta(M, N) \equiv \max \{\delta(M, N), \delta(N, M)\}$$

is a metric and was studied by Gohberg and Markus [5]. For this, related ideas, and references see the book of Singer [7, p. 161]. If M and N are linear varieties (i.e., $M - m$ is a subspace for any $m \in M$), we define the deviation $\delta(M, N)$ of M from N to be

$$\delta(M, N) \equiv \delta(M - m, N - n) + \inf_{x \in M, y \in N} \|x - y\|,$$

where $m \in M$ and $n \in N$. It is clear that we get the same subspaces $M - m$ and $N - n$ for any choice of m and n ; hence $\delta(M, N)$ is well defined. We define $\Delta(M, N)$ as above.

If x and y are in X , then the line determined by $l(x, y)$ is

$$l(x, y) \equiv \{z \in X; z = tx + (1 - t)y, t \in R^1\}.$$

Let Y be another Banach space, D an open subset of Y , and $y \in D$. Suppose $G: Y \rightarrow X$. Then G is Fréchet differentiable at y if there exists a continuous linear map $\nabla G(y) \in L(Y, X)$ such that

$$(1.1) \quad \lim_{t \rightarrow 0} \frac{G(y + tb) - G(y)}{t} = \nabla G(y) \cdot b$$

uniformly for $b \in S(Y)$. The linear map $\nabla G(y)$ is called the Fréchet derivative of G at y .

A Banach space X is called (UF) if the norm of X is Fréchet differentiable on $X \setminus \{\theta\}$ and the convergence in (1.1) is uniform for (y, b) in $S(X) \times S(X)$.

Throughout the remainder of this paper we will assume that $F: Y \rightarrow X$ satisfies conditions (A) through (D) of the introduction.

We will try to consistently use small Greek letters for elements of R^n , t for an element of R^1 , and lower case Latin letters for elements of X . We will write, for each $\alpha \in R^n$,

$$T(F(\alpha)) \equiv F(\alpha) + \nabla F(\alpha) \cdot R^n$$

and call $T(F(\alpha))$ the tangent plane to \mathcal{F} at $F(\alpha)$. If $x \in P_{T(x)}(y)$, then we write $l(x, y) \perp T(x)$. Note that, even if $l(u, v) = l(x, y)$, we will refrain from writing $l(u, v) \perp T(x)$ unless $u \in P_{T(x)}(v)$. A point $y \in \mathcal{F}$ is said to be a local best approximation to $x \in X$ if there exists a neighborhood U of y such that $y \in P_{U \cap \mathcal{F}}(x)$. (Wulbert uses a neighborhood U of $F^{-1}(y)$ and assumes that $y \in P_{F(U) \cap \mathcal{F}}(x)$, but this amounts to the same thing since F is a homeomorphism.)

The following lemma connects the approximation properties of \mathcal{F} to its tangent planes, is well known and not too difficult to prove (see, e.g. [10, p. 385]).

Lemma 1.1. *If x is a local best approximation to y from \mathcal{F} , then $l(x, y) \perp T(x)$.*

The next lemma interprets, for a function $G: Y \rightarrow X$, the uniform continuity of ∇G in terms of the uniform convergence of (1.1).

Lemma 1.2. *Let X and Y be Banach spaces and $G: Y \rightarrow X$. Suppose ∇G is uniformly continuous on bounded subsets of Y . Then, for any bounded subset Ω of Y ,*

$$(1.2) \quad \sup_{\omega \in \Omega, \|\beta\| \leq t} \|G(\omega + \beta) - G(\omega) - \nabla G(\omega) \cdot \beta\| = o(t).$$

Proof. Since Ω is bounded, so is $\Gamma \equiv \Omega + B(\theta, t)$. By [4, p. 156], we have

$$\|G(\omega + \beta) - G(\omega) - \nabla G(\omega) \cdot \beta\| \leq t \sup_{\alpha \in B(\omega, t)} \|\nabla G(\omega) - \nabla G(\alpha)\|.$$

Since ∇G is uniformly continuous on Γ , we may conclude that (1.2) holds.

Remark 1. The function F satisfies the conditions in Lemma 1.2: any bounded subset E of R^n is precompact and ∇F is continuous on R^n , therefore ∇F is uniformly continuous on E . Another intended application of Lemma 1.2 is to functions involving the norm (see §3).

Remark 2. It can be shown, as a corollary to Lemma 1.2, that for each compact subset Ω of R^n ,

$$\sup_{x \in F(\Omega), z \in T(x) \cap B(x, t)} \text{dist}(z, \mathcal{F}) = o(t).$$

This is the analog of Theorem 11-4 in [6]. Hypothesis (D) plays a significant role in the proof of this fact.

The next lemma interprets the continuity of ∇F in terms of the convergence of the tangent planes $T(x)$.

Lemma 1.3. Suppose $\{x_n\}_{n=1}^\infty \subset \mathcal{F}$ and $x_n \rightarrow x$. Then, as $n \rightarrow \infty$,

$$(1.3) \quad \delta(T(x), T(x_n)) \rightarrow 0$$

and

$$(1.4) \quad \delta(T(x_n), T(x)) \rightarrow 0.$$

Furthermore, if $l(x, y) \perp T(x)$, then

$$(1.5) \quad \delta(l(x, y), T(x)) \geq 1.$$

Proof. The inequality (1.5) is immediate from the definitions. To prove (1.3) and (1.4), we may assume that $x = \theta = F(\theta)$ and $\alpha_n = F^{-1}(x_n)$.

For (1.3) it is clearly enough to show that $\delta(T(x) - x, T(x_n) - x_n) \rightarrow 0$. Consider, for any $\beta \in \partial B(\theta, 1)$ in R^n ,

$$(1.6) \quad \|\text{sgn}(\nabla F(\alpha_n) \cdot \beta) - \text{sgn}(\nabla F(\theta) \cdot \beta)\|.$$

Since ∇F is continuous and $\alpha_n \rightarrow \theta$, it follows via hypothesis (D) on F that the difference in (1.6) converges uniformly to 0 as $n \rightarrow \infty$. Since, in X ,

$$\partial B(\theta, 1) \cap T(\theta) = \bigcup_{\beta \in B(\theta, 1)} \text{sgn}(\nabla F(\theta) \cdot \beta),$$

it follows that

$$\delta(T(x) - x, T(x_n) - x_n) \rightarrow 0.$$

Expression (1.4) follows in a similar manner.

In [6], Rice did not assume that the manifold \mathcal{F} was given by a function F and so he did not define the tangent plane $T(z)$ at $z \in \mathcal{F}$ as we have. Rather, he defined $T(z)$ as the set of limit points of lines $l(x_n, y_n)$ generated by elements x_n and y_n in \mathcal{F} which converge to z . Our next lemma verifies that if we assume \mathcal{F} is given by F , then our tangent planes agree with the ones defined by Rice.

Lemma 1.4. *Let $\{x_n\}$ and $\{y_n\}$ be sequences in \mathcal{F} so that $x_n \rightarrow z$, $y_n \rightarrow z$, and $x_n \neq y_n$. Then, as $n \rightarrow \infty$,*

$$\delta(l(x_n, y_n), T(z)) \rightarrow 0.$$

Proof. We may normalize the problem by setting $z = \theta$, $F^{-1}(z) = \theta$, $F^{-1}(x_n) = \alpha_n$, and $F^{-1}(y_n) = \beta_n$. It is clearly enough to show that

$$\delta(l(x_n, y_n) - y_n, T(\theta)) \rightarrow 0.$$

By a well-known version of the mean value theorem [1, p. 219] we have, because F is continuously Fréchet differentiable,

$$F(\alpha_n) - F(\beta_n) \in \overline{\text{co}} \bigcup_{\gamma \in [\alpha_n, \beta_n]} \nabla F(\gamma) \cdot (\alpha_n - \beta_n)$$

where $[\alpha_n, \beta_n]$ is the line segment joining α_n to β_n and " $\overline{\text{co}}$ " denotes the closed convex hull. Thus, for each n we have

$$F(\alpha_n) - F(\beta_n) = \lim_{i \rightarrow \infty} \sum_{j=1}^i t_{ij}^{(n)} \nabla F(\gamma_{ij}^{(n)}) \cdot (\alpha_n - \beta_n)$$

where $t_{ij}^{(n)} \geq 0$, $\sum_{j=1}^i t_{ij}^{(n)} = 1$, and $\gamma_{ij}^{(n)} \in [\alpha_n, \beta_n]$. Set

$$A_{in} = \sum_{j=1}^i t_{ij}^{(n)} \nabla F(\gamma_{ij}^{(n)}).$$

Since $\|F(\alpha_n) - F(\beta_n)\| \neq 0$, we obtain

$$(1.7) \quad \text{sgn}(F(\alpha_n) - F(\beta_n)) = \lim_{i \rightarrow \infty} \text{sgn} A_{in} (\alpha_n - \beta_n).$$

Since $\nabla F(\theta) \cdot R^n$ has dimension n and ∇F is continuous, it can readily be shown that for some $\eta > 0$ and all n large enough,

$$(1.8) \quad \|F(\alpha_n) - F(\beta_n)\| / \|\alpha_n - \beta_n\| \geq \eta > 0$$

and

$$(1.9) \quad \|\nabla F(\theta) \cdot (\alpha_n - \beta_n)\| / \|F(\alpha_n) - F(\beta_n)\| \rightarrow 1.$$

Since ∇F is continuous

$$\sup_{\gamma \in [\alpha_n, \beta_n]} \|\nabla F(\gamma) - \nabla F(\theta)\| \equiv \epsilon_n \rightarrow 0$$

as $n \rightarrow \infty$. Thus

$$(1.10) \quad \begin{aligned} & \|\operatorname{sgn}(F(\alpha_n) - F(\beta_n)) - \operatorname{sgn}(\nabla F(\theta) \cdot (\alpha_n - \beta_n))\| \\ & \leq \|\operatorname{sgn}(F(\alpha_n) - F(\beta_n)) - \operatorname{sgn}(A_{in}(\alpha_n - \beta_n))\| \\ & \quad + \|\operatorname{sgn}(A_{in}(\mu_{in})) - \operatorname{sgn} \nabla F(\theta) \cdot \mu_{in}\| \end{aligned}$$

where $\mu_{in} = (\alpha_n - \beta_n)/\|A_{in}(\alpha_n - \beta_n)\|$. We will let $i \rightarrow \infty$, so we can ignore the first term on the right-hand side of (1.10). Since $\|A_{in}(\mu_{in})\| = 1$, the second term in (1.10) is bounded above by

$$\|A_{in}(\mu_{in}) - \nabla F(\theta) \cdot (\mu_{in})\| + |1 - \|\nabla F(\theta) \cdot (\mu_{in})\||.$$

Letting $i \rightarrow \infty$, we obtain

$$(1.11) \quad \begin{aligned} & \|\operatorname{sgn}(F(\alpha_n) - F(\beta_n)) - \operatorname{sgn}(\nabla F(\theta) \cdot (\alpha_n - \beta_n))\| \\ & \leq \epsilon_n \frac{\|\alpha_n - \beta_n\|}{\|F(\alpha_n) - F(\beta_n)\|} + \left| 1 - \frac{\|\nabla F(\theta) \cdot (\alpha_n - \beta_n)\|}{\|F(\alpha_n) - F(\beta_n)\|} \right|. \end{aligned}$$

Thus (1.8) and (1.9) imply that the right-hand side of (1.11) approaches 0 as $n \rightarrow \infty$. Hence, as $n \rightarrow \infty$,

$$\delta(l(x_n, y_n) - y_n, T(\theta)) \rightarrow 0$$

since $l(x_n, y_n) - y_n = \{w \in X; w = t(x_n - y_n), t \in R^1\}$.

The folding of a manifold at a point is intended to be a measure of the amount that a manifold "turns back on itself". Before making the definition, we introduce another notation which is used here and in the sequel.

Notation. If C is a compact subset of X , then

$$(1.12) \quad \mathcal{C}(C) = \{\Gamma \subset X; \Gamma \supset C \text{ and } \Gamma \text{ is compact and connected}\}.$$

Definition. The folding $\operatorname{fld}(\mathcal{F}, x)$ of \mathcal{F} at $x \in \mathcal{F}$ is

$$\operatorname{fld}(\mathcal{F}, x) \equiv \sup\{t_0 \in R^1; B(x, t) \cap \mathcal{F} \in \mathcal{C}(x) \text{ for each } t \leq t_0\}.$$

Note that $B(x, \operatorname{fld}(\mathcal{F}, x)) \cap \mathcal{F}$ need be neither compact nor connected. The following theorem establishes an interesting property of the folding.

Theorem 1.5. *For every $x \in \mathcal{F}$ there exists an $\epsilon \equiv \epsilon(x) > 0$ such that $\operatorname{fld}(\mathcal{F}, y) \geq \epsilon$ whenever $y \in B(x, \epsilon) \cap \mathcal{F}$.*

Proof. Since F is a homeomorphism and R^n is locally compact, there exists an $\epsilon_0 > 0$ so that $\mathcal{F} \cap B(x, \epsilon_0)$ is compact. Suppose the conclusion of the theorem does not hold. Then there is a sequence $x_n \rightarrow x$, $x_n \in \mathcal{F} \cap B(x, \epsilon_0/2)$, so that $\mathcal{F} \cap B(x_n, \lambda_n)$ is disconnected for some $\lambda_n \leq \|x_n - x\|$. Let E_n be a component of $\mathcal{F} \cap B(x_n, \lambda_n)$ which does not contain x_n . Since $E_n \subset \mathcal{F} \cap B(x, \epsilon_0)$, E_n is compact. Furthermore, if we view E_n as a subset of \mathcal{F} , then the boundary of E_n with respect to \mathcal{F} must be contained in $\partial B(x_n, \lambda_n)$. It follows from the compactness of E_n that there exists a $y_n \in E_n$ which satisfies $\|y_n - x_n\| = \text{dist}(x_n, E_n)$ and hence y_n is a local best approximation from \mathcal{F} to x_n . Lemma 1.1 then implies that $l(y_n, x_n) \perp T(y_n)$. Clearly both $\{x_n\}$ and $\{y_n\}$ converge to x , so Lemma 1.4 tells us that $\delta(l(y_n, x_n), T(x)) \rightarrow 0$ as $n \rightarrow \infty$. On the other hand, Lemma 1.3 yields $\Delta(T(x), T(y_n)) \rightarrow 0$ as $n \rightarrow \infty$. But these last two expressions are clearly incompatible with $l(y_n, x_n) \perp T(y_n)$.

2. Curvature. The notion of curvature in R^n , when endowed with the Euclidean norm, is rather standard. When we attempt to study curvature in other spaces, however, the proper definition is not obvious. For our purposes, we wish to define the curvature of the manifold \mathcal{F} by comparing it with spheres in the Banach space. To do this we wish to define the "normals" to the surface at a point $x \in \mathcal{F}$ in terms of the notion of perpendicularity already introduced in §1, a notion that seems best suited to the needs of approximation theory. We then wish to look at spheres whose centers are in the set of normals and which contain x and another point z on \mathcal{F} . To describe the radius of curvature of the manifold, we see what happens to the radii of the above spheres when the points z of \mathcal{F} approach x , while keeping the center of the sphere on the line determined by the normal. In formulating the definition of curvature below, we are indebted to the discussion in [6].

Let $x \in \mathcal{F}$. We define the set of normals, $\mathcal{N}(x)$, of \mathcal{F} at x , as

$$\mathcal{N}(x) = \{y \in X; l(x, y) \perp T(x)\}.$$

If $z \in \mathcal{F}$, then, for any $y \in \mathcal{N}(x)$, there may or may not be a point $y(z)$ on $l(x, y)$ such that $\partial B(y(z), \|y(z) - x\|)$ contains z . If there is no such point, set $\rho(x, y, z) = \infty$. If such points exist, set $\rho(x, y, z) = \inf \|y(z) - z\|$. Note that if X is rotund (or, in other words, strictly convex) then the point $y(z)$, when it exists, is unique. We may now give the definition of curvature.

Definition. The radius of curvature of \mathcal{F} at x in the direction y , $y \in \mathcal{N}(x)$, written $\rho(x, y)$, is

$$\rho(x, y) \equiv \inf_{z_i \rightarrow x} \liminf_{i \rightarrow \infty} \rho(x, y, z_i).$$

The radius of curvature $\rho(x)$ of \mathcal{F} at x is

$$\rho(x) = \inf_{y \in \mathcal{N}(x)} \rho(x, y).$$

The curvature $\sigma(x)$ of \mathcal{F} at x is

$$\sigma(x) = \rho(x)^{-1}.$$

If the unit sphere of X is somewhat oddly "curved", peculiar properties may occur. For example, if $S(X)$ contains a line segment $[x_1, x_2]$, then it can easily be seen that $l(x_1, x_2)$ has infinite curvature at $\frac{1}{2}(x_1 + x_2)$. Pathological behavior this extreme cannot occur if X is rotund; however, Rice has shown that, in two-dimensional L^p space, $p > 2$, the Euclidean circle defined by $x^2 + y^2 = 1$ has infinite curvature at the point $(0, 1)$. This may not be too dismaying if one keeps in mind that the two dimensional L^p sphere, $1 < p < 2$, has infinite curvature at $(1, 0)$ in two dimensional L^2 space. For more discussion of curvature see [6, pp. 190–198].

The following lemma may be used to obtain an elementary upper bound on the radius of curvature.

Lemma 2.1. *If $x \in \mathcal{F}$, $l(y, x) \perp T(x)$, and x is accumulation point of $\mathcal{F} \cap B(y, \|x - y\|)$, then*

$$(2.1) \quad \rho(x, y) \leq \|x - y\|.$$

Proof. Pick $z_i \rightarrow x$ such that $\|z_i - y\| \leq \|x - y\|$. It is then clear that $\rho(x, z_i)$ is found by moving from y toward x . Thus $\rho(x, z_i) \leq \|x - y\|$. The definition of $\rho(x, y)$ immediately implies (2.1).

3. A variational theorem. Suppose $y \notin \mathcal{F}$. A best approximation $F(\alpha_0)$ to y from \mathcal{F} solves the following problem:

Find $\alpha_0 \in R^n$ minimizing the functional $G(\alpha) = \|y - F(\alpha)\|$.

If G is differentiable, then it is natural to look for *critical points* of G , i.e., points $\alpha \in R^n$ such that $\nabla G(\alpha) = 0$. In fact, during the proof of our main theorem, we had occasion to search for other critical points: points on sets Γ which joined two best approximations. Intuitively, it seems that the point farthest from y in Γ will yield a critical point of G if Γ is, in some sense, the smallest set containing the two best approximations. It is with this in mind that we introduce the following terminology (derived from the language of nonlinear functional analysis).

Let Y be a Banach space and $K \subset Y$ be compact. Let $\mathcal{C}(K)$ be, as in §1, the family of all compact connected subsets of Y which contain K . (Note that $\overline{\text{co}}(K) \in \mathcal{C}(K)$ so $\mathcal{C}(K) \neq \emptyset$.) It is customary to introduce a family \mathcal{D} of functions from Y into Y , called *deformations*, which carry elements of $\mathcal{C}(K)$ into elements

of $\mathcal{C}(K)$; we let \mathcal{D} be all continuous functions $f: Y \rightarrow Y$ such that $f(K) = K$. Suppose $G: Y \rightarrow R^1$. We say that $L_0 \in \mathcal{C}(K)$ is *extremal* for G if

$$\sup_{\alpha \in L_0} G(\alpha) = \inf_{L \in \mathcal{C}(K)} \sup_{\alpha \in L} G(\alpha).$$

If $\alpha^* \in L_0$ is such that

$$G(\alpha^*) = \inf_{L \in \mathcal{C}(K)} \sup_{\alpha \in L} G(\alpha),$$

then α^* is called a *minimax point* of G with respect to $\mathcal{C}(K)$. The set of all minimax points in L_0 is denoted by L_0^* . In order to show that there are critical points of G among the minimax points, we assume not and then, by moving along the gradient of G in the steepest descent, we deform L_0 to a set $L \in \mathcal{C}(K)$ such that

$$\sup_{\alpha \in L} G(\alpha) < \sup_{\alpha \in L_0} G(\alpha),$$

contrary to the assumption that L_0 is extremal for G .

In order to choose the direction of descent, we introduce the duality map J' from Y' , the continuous linear dual of Y , defined by sending a point $y' \in Y'$ into the set of points $J'(y') = \{y \in Y; \|y\| = \|y'\| \text{ and } \langle y, y' \rangle = \|y'\|^2\}$. This function is single valued iff Y is rotund and reflexive. It is known that a necessary and sufficient condition for the continuity of J' is that the norm of Y' be Fréchet differentiable [3, Corollary 4.12]. In order to choose the direction of descent continuously, we will therefore assume that the norm of Y' is Fréchet differentiable.

With this terminology, we state and prove the following variational theorem.

Theorem 3.1. *Suppose the norm of Y' is Fréchet differentiable $G: Y \rightarrow R^1$ is Fréchet differentiable and ∇G is uniformly continuous on closed bounded subsets of Y . Let K be a compact subset of Y and suppose there exists an extremal element L_0 of $\mathcal{C}(K)$. If*

$$d_1 \equiv \sup_{\alpha \in K} G(\alpha) < \sup_{\alpha \in L_0} G(\alpha) \equiv d_2,$$

then there exists an $\alpha^ \in L_0^*$ which is a critical point of G .*

Proof. Suppose not. Since ∇G is continuous and L_0^* is compact, there exist numbers $\eta > 0$ and $d, d_1 < d < d_2$, such that $\|\nabla G(\beta)\| \geq \eta$ for all $\beta \in L_0 \cap G^{-1}([d, \infty)) \equiv N$. Let $N_1 \equiv L_0 \cap G^{-1}([(d + d_2)/2, \infty))$. Define $H: Y \rightarrow Y$ by $H = -J' \circ \nabla G$. Then H is continuous on N and, for every $\alpha \in Y$,

$$\langle H(\alpha), \nabla G(\alpha) \rangle = -\|\nabla G(\alpha)\|^2.$$

By Lemma 1.2, for $\alpha \in N$

$$G(\alpha + tH(\alpha)) = G(\alpha) + \nabla G(\alpha) \cdot tH(\alpha) + o(t) \leq G(\alpha) - t\eta^2 + o(t)$$

where $o(t)$ is independent of $\alpha \in N$. Choose t^* so that $o(t) < \eta^2 t/2$ for $0 \leq t \leq t^*$. By defining

$$g(\alpha) = \frac{\text{dist}(\alpha, L_0 \setminus N)}{\text{dist}(\alpha, N_1) + \text{dist}(\alpha, L_0 \setminus N)},$$

we obtain a continuous function $g: Y \rightarrow [0, 1]$ such that $g(L_0 \setminus N) = 0$ and $g(N_1) = 1$. Then $f: Y \rightarrow Y$, defined by $f(\alpha) = \alpha + g(\alpha)t^*H(\alpha)$, is in \mathcal{D} ; hence $f(L_0) \in \mathcal{C}(K)$. Now, for $\alpha \in L_0 \setminus N$, $G(f(\alpha)) = G(\alpha)$. For $\alpha \in N$,

$$\begin{aligned} G(f(\alpha)) &= G(\alpha + g(\alpha)t^*H(\alpha)) \\ &= G(\alpha) + \nabla G(\alpha) \cdot g(\alpha)t^*H(\alpha) + o(t^*g(\alpha)) \\ &\leq G(\alpha) - \eta^2 t^* g(\alpha) + \eta^2 t^* g(\alpha)/2 = G(\alpha) - \eta^2 t^* g(\alpha)/2. \end{aligned}$$

Thus

$$\begin{aligned} \sup_{\alpha \in f(L_0)} G(\alpha) &\leq \max \left\{ \sup_{\alpha \in L_0 \setminus N_1} G(\alpha), \sup_{\alpha \in N_1} G(\alpha) - \frac{\eta^2 t^*}{2} \right\} \\ &\leq \max \{ (d + d_2)/2, d_2 - \eta^2 t^*/2 \} < d_2. \end{aligned}$$

This is impossible since L_0 is extremal.

This theorem takes the place of Lemma 11.11 in [6] and obviates the need of Lemma 11.5.

This result can be considered as a generalization of Rolle's theorem, for if we specialize to the case $Y = \mathbb{R}^1$, then Rolle's theorem can be used to obtain the critical point. A result very similar to ours can be found in [2, p. 60] for $Y = \mathbb{R}^n$. The Bergers assume that G satisfies the growth condition $G(\alpha) \rightarrow \infty$ as $\|\alpha\| \rightarrow \infty$. This allows one to assert the existence of an extremal set L_0 . However, in our application of this result, we will not be assuming this growth condition.

4. Best approximation from manifolds. We will extend a result of Rice which was stated for a closed manifold \mathcal{F} in a finite dimensional, rotund, and smooth space X . We have managed to remove the rotundity condition on X , but, when X is infinite dimensional, we must assume that X is (UF), i.e. the norm is Fréchet differentiable and the limit (1.1) is uniform for y and b in $S(X) \times S(X)$. In finite dimensions, this is equivalent to the assumption that X is smooth. We have not assumed that \mathcal{F} is boundedly compact (or even closed), though it is locally compact, of course. (Recall that a set is boundedly compact if its intersection with

any closed sphere is compact.) It would be interesting to see to what extent the conditions of local compactness and uniform differentiability of the norm could be relaxed.

We say that $\mathcal{U} \subset X$ is a neighborhood of \mathcal{F} if \mathcal{U} is a neighborhood of each $x \in \mathcal{F}$. This does not imply, of course, that $\partial\mathcal{U}$ is bounded away from \mathcal{F} .

Theorem 4.1. *Suppose X is (UF) and the curvature of \mathcal{F} is uniformly bounded on compact subsets of \mathcal{F} . Then there exists a neighborhood \mathcal{U} of \mathcal{F} such that each point of \mathcal{U} has a unique best approximation from \mathcal{F} .*

Proof. Let $x_0 \in \mathcal{F}$. By Theorem 1.5, the folding $\text{fld}(\mathcal{F}, x_0) \equiv \phi_0 > 0$. Let $t_0 < \phi_0$, $t_0 > 0$, and let σ be a bound on the curvature $\sigma(x)$ for each x in $\mathcal{F} \cap B(x_0, t_0)$. Set $\epsilon_0 \equiv \min\{1/3\sigma, t_0/4\}$. We claim that if $y \in B^0(x_0, \epsilon_0)$, then y has a unique best approximation from \mathcal{F} . Suppose, to the contrary, that $C \equiv \{x_1, x_2\} \subset P_{\mathcal{F}}(y)$. Clearly $\|y - x_i\| \leq \|y - x_0\| < \epsilon_0$ for $i = 1, 2$. Hence, by definition of the folding ϕ_0 , $\tilde{\Gamma} \equiv \mathcal{F} \cap B(x_0, 2\|y - x_0\|)$ is compact, connected and contains C . Setting $\mathcal{C} = \{\Gamma \subset \mathcal{F}; \Gamma \text{ is compact, connected and } C \subset \Gamma\}$, we see that

$$d \equiv \inf_{\Gamma \in \mathcal{C}} \sup_{y \in \Gamma} \|x - y\| \leq \sup_{x \in \tilde{\Gamma}} \|x - y\| < 3\epsilon_0.$$

Since $B(y, 3\epsilon_0) \subset B(x_0, t_0)$ there is a compact connected component Γ_i of $\mathcal{F} \cap B(y, d + 1/i)$ which contains C , for i sufficiently large (say $i \geq l$). Since each Γ_i is compact, connected and $\Gamma_i \supset \Gamma_{i+1}$, it follows that $\Gamma_0 \equiv \bigcap_{i \geq l} \Gamma_i$ is compact, connected, contains C and $d = \sup_{x \in \Gamma_0} \|x - y\|$.

Suppose that $\text{dist}(y, C) = d$. Since $\text{dist}(y, \mathcal{F}) = \text{dist}(y, C)$, it follows from Lemma 1.1 that, for $x^* \in \Gamma_0^* \equiv \{x \in \Gamma_0; \|x - y\| = d\}$, we have $\ell(y, x^*) \perp T(x^*)$.

If $\text{dist}(y, C) < d$, then we will find $x^* \in \Gamma_0^*$ such that $\ell(y, x^*) \perp T(x^*)$ by applying the variational Theorem 3.1 in the following way: let $Y = R^n$ (with the Euclidean norm), $G(\alpha) = \|y - F(\alpha)\|$, $K = F^{-1}(C)$. Clearly, if $L \in \mathcal{C}(K)$, then $F(L) \in \mathcal{C}$. Thus $L_0 \equiv F^{-1}(\Gamma_0)$ satisfies $L_0 \in \mathcal{C}(K)$ and

$$\sup_{\alpha \in K} G(\alpha) < \sup_{\alpha \in L_0} G(\alpha) = \inf_{L \in \mathcal{C}(K)} \sup_{y \in L} G(y).$$

Since $G(\alpha) \geq \text{dist}(y, \mathcal{F}) > 0$, we know that the norm is uniformly Fréchet differentiable on $G(B)$, where B is any bounded subset of R^n . Thus ∇G satisfies the hypothesis of Theorem 3.1. That theorem then yields an $\alpha^* \in L_0^*$ such that $\nabla G(\alpha^*) = 0$. Clearly, $F(\alpha^*) \equiv x^* \in \Gamma_0^*$. Using the chain rule on G , we obtain for any $\beta \in R^n$,

$$0 = \langle \beta, \nabla G(\alpha^*) \rangle = \langle \nabla F(\alpha^*) \cdot \beta, \nabla \| \cdot \| (F(\alpha^*) - y) \rangle.$$

Thus

$$\nabla F(\alpha^*) \cdot R^n \subset \ker \nabla \| \cdot \| (F(\alpha^*) - y).$$

Using the well-known connection between the derivative of the norm and the supporting hyperplanes to spheres we see that $T(x^*)$ is contained in the supporting hyperplane to $B(y, \|y - x^*\|)$ at x^* . Thus $l(y, x^*) \perp T(x^*)$.

Now note that $\|x^* - x_0\| \leq \|x^* - y\| + \|y - x_0\| < 3\epsilon_0 + \epsilon_0 \leq t_0$ so that $\sigma(x^*) \leq \sigma$. Since $\Gamma_0 \subset \mathcal{F} \cap B(y, \|y - x^*\|)$ it follows that x^* is an accumulation point of $\mathcal{F} \cap B(y, \|y - x^*\|)$. Thus we may apply Lemma 2.1 to obtain

$$(4.1) \quad 1/\sigma \leq \rho(x^*, y) \leq \|y - x^*\| < 3\epsilon_0 \leq 1/\sigma.$$

This contradiction establishes that each $y \in B^0(x_0, \epsilon_0)$ has a unique best approximation from \mathcal{F} . Setting $\mathcal{U} = \bigcup_{x_0 \in \mathcal{F}} B^0(x_0, \epsilon_0)$ yields the theorem.

Rice has obtained the value $\epsilon_0 = \min\{1/3\sigma, \phi_0/2\}$, which is correct if \mathcal{F} is boundedly compact (as Rice assumes) and σ is taken to be the bound of the curvature on $\mathcal{F} \cap B(x_0, 2\phi_0)$.

It is natural to ask whether every element of X has a unique best approximation from \mathcal{F} (in which case \mathcal{F} is called Chebyshev). It is clear that \mathcal{F} cannot be Chebyshev unless it is closed. An argument of Rice [6, p. 203] can easily be adapted to the infinite dimensional case to show that if \mathcal{F} is contained in any Banach space X , then \mathcal{F} closed and convex implies that \mathcal{F} is a linear variety. If we assume that X is smooth, then results of Vlasov, Efimov and Stechkin imply that if \mathcal{F} is Chebyshev and boundedly compact, then \mathcal{F} is convex [7, p. 368]. These two statements yield the following theorem.

Theorem 4.2. *If X is smooth and \mathcal{F} is boundedly compact and Chebyshev, then \mathcal{F} is a linear variety.*

A result of Vlasov [8] lets us remove the hypothesis that \mathcal{F} is boundedly compact if we assume that X is uniformly rotund and smooth.

The folding of a manifold is also connected to the property of being Chebyshev.

Theorem 4.3. *If X is a Banach space, \mathcal{F} is boundedly compact, and $\text{fld}(\mathcal{F}, x)$ is bounded above at some point $x \in \mathcal{F}$, then \mathcal{F} is not Chebyshev.*

Proof. Suppose \mathcal{F} is Chebyshev. Since \mathcal{F} is boundedly compact, it follows that the metric projection onto \mathcal{F} is continuous. Thus, by a result of Wulbert [9], we can conclude that the intersection of every open sphere with \mathcal{F} is connected. For each $x \in \mathcal{F}$ and $t > 0$,

$$B(x, t) \cap \mathcal{F} = \bigcap_{\epsilon > 0} \text{cl}(B^0(x, t + \epsilon) \cap \mathcal{F});$$

hence $B(x, \iota) \cap \mathcal{F}$ is connected. This contradicts the assumption that $\text{fld}(\mathcal{F}, x)$ is bounded above.

This theorem yields an indirect proof of the fact that for the manifolds \mathcal{F} studied by Wulbert in [10], $\text{fld}(\mathcal{F}, x) = \infty$ for each x in \mathcal{F} : since Wulbert shows that the manifolds in [10] are boundedly compact and Chebyshev, it follows that $\text{fld}(\mathcal{F}, x) = \infty$. A proof that the curvature of such manifolds is zero would indicate that Theorem 4.2 may be true without the assumption that X is (UF).

If X is uniformly rotund, then Vlasov [8] has shown that if $P_{\mathcal{F}}(x)$ is connected for each $x \in \mathcal{F}$, then the intersection of \mathcal{F} with any closed ball is connected. Thus in a uniformly rotund space, the assumption that \mathcal{F} is boundedly compact may be removed in Theorem 4.3.

REFERENCES

1. V. I. Averbuh and O. G. Smoljanov, *Differentiation theory in linear topological spaces*, Uspehi Mat. Nauk 22 (1967), no. 6 (138), 201–260 = Russian Math. Surveys 22 (1967), no. 6, 201–258. MR 36 #6933.
2. M. S. Berger and M. S. Berger, *Perspectives in nonlinearity. An introduction to nonlinear analysis*, Benjamin, New York, 1968. MR 40 #4971.
3. D. F. Cudia, *The geometry of Banach spaces. Smoothness*, Trans. Amer. Math. Soc. 110 (1964), 284–314. MR 29 #446.
4. J. Dieudonné, *Foundations of modern analysis*, Pure and Appl. Math., vol. 10, Academic Press, New York, 1960. MR 22 #11074.
5. I. C. Gohberg and A. S. Markus, *Two theorems on the gap between subspaces of a Banach space*, Uspehi Mat. Nauk 14 (1959), no. 5 (89), 135–140. (Russian) MR 22 #5880.
6. J. R. Rice, *Approximation of functions. Vol. II. Nonlinear and multivariate theory*, Addison-Wesley, Reading, Mass., 1969. MR 39 #5989.
7. I. Singer, *Best approximation in normed vector spaces by elements of vector subspaces*, Editura Academiei Republicii Socialiste România, Bucharest, 1967; English transl., Die Grundlehren der math. Wissenschaften, Band 171, Springer-Verlag, New York and Berlin, 1970. MR 38 #3677; 42 #4937.
8. L. P. Vlasov, *Čebyšev sets and some generalizations of them*, Mat. Zametki 3 (1968), 59–69. (Russian) MR 37 #3329.
9. D. E. Wulbert, *Continuity of metric projections*, Trans. Amer. Math. Soc. 134 (1968), 335–341. MR 38 #472.
10. ———, *Uniqueness and differential characterization of approximations from manifolds of functions*, Amer. J. Math. 93 (1971), 350–366. MR 45 #4036.
11. ———, *Nonlinear approximation with tangential characterization*, Amer. J. Math. 93 (1971), 718–730. MR 45 #4037.

DEPARTMENT OF MATHEMATICS, PURDUE UNIVERSITY, HAMMOND, INDIANA 46323

DEPARTMENT OF MATHEMATICS, TEXAS A & M UNIVERSITY, COLLEGE STATION, TEXAS 77843 (Current address of P. W. Smith)

Current address (E. R. Rozema): Department of Mathematics, University of Tennessee, Chattanooga, Tennessee 37401