3/17/65
b

DIFFERENTIABILITY IN GENERAL ANALYSIS

A THESIS

Presented to

The Faculty of the Graduate Division

by

Sanford Martin Wiener

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Applied Mathematics

Georgia Institute of Technology

May, 1966

DIFFERENTIABILITY IN GENERAL ANALYSIS

Approved:

Chairman

Date approved by Chairman: May 17, 1966

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Dr. M. Z. Nashed, my thesis adviser, for his expert guidance in the preparation of this thesis. I also wish to thank the other two members of my committee, Dr. E. R. Immel and Dr. H. A. Gersch, for their careful reading of the manuscript. Finally, I feel special thanks are due my parents, without whose sacrifices this thesis could not have been possible.

FOREWORD

In order to avoid unnecessary repetition, it seems appropriate to collect here some of the symbolism and terminology used throughout this thesis.

The symbols X, Y, and Z will designate normed linear spaces, while A and B will denote open subsets of X and Y, respectively. The symbol E_n will denote the n-dimensional Euclidean space of n-tuples of real numbers. The symbol 0 will be used to denote all zero elements, regardless of the space in which they lie. The precise meaning of each 0 in any particular expression should be clear from the context.

The term scalar will mean real number and the term functional will mean an operator whose range is a subset of the real line. The terms integrable and summable will be used interchangeably to mean that the integral of a function over some set is finite.

If $f\colon X \to Y$ and $g\colon X \to E_1$, then the symbolism

$$f(t) = o\big(g(t)\big) \quad \text{as} \quad t \to 0$$

will mean

$$\lim_{t\to 0} \frac{f(t)}{g(t)} = 0,$$

where the limit is understood in the sense of the norm in Y.

TABLE OF CONTENTS

Page

Chapter

SUMMARY

This study is concerned with several notions of differentiability which are useful in the theory of operators on normed linear spaces. In addition to the familiar notions of Fréchet and Gâteaux differentials [33], the notions of quasi-differential [8] and strong differential [23] are defined in Chapter I. Various implication relationships among these differentials are obtained and sufficient conditions for the existence of these differentials are established. In particular it is shown that if f is Lipschitzian and has a linear Gâteaux variation, then f is quasi-differentiable. It is further shown that the quasi-differential satisfies the chain rule property. A theorem of Esser and Shisha [11] concerning strong differentials of real-valued functions is generalized to normed spaces in what appears to be a new theorem. It is also shown that a continuous Fréchet differential is a strong differential.

In Chapter II a detailed proof of a theorem due to Rademacher [27] is presented, a corollary of which asserts that every Lipschitzian operator mapping E_n into E_1 is Fréchet differentiable almost everywhere. A generalization of Rademacher's result, due to Rado and Reichelderfer [28], proves the same result when the range is E_m. The question

concerning a similar statement for infinite dimensional spaces remains open.

Chapter III is devoted to a brief discussion of differentiability of convex functionals on normed spaces. This topic is the subject of recent investigations [5,25] which would fall beyond the scope of this study. The main result of this chapter is that a convex functional possesses a one-sided Gâteaux variation at each point of its domain. It is further shown that this one-sided Gâteaux variation is itself a convex functional.

In the last chapter, the concept of smooth operator is developed as a generalization of smooth function, a topic studied by Zygmund and others [32,37]. The differentiability properties of smooth operators are investigated and a new theorem, which provides an immediate connection with results on Gâteaux differentiability from Chapter I, is proved. The final result, a collection of several previous results, shows that a continuous, convex, and smooth function is differentiable.

CHAPTER I

SOME NOTIONS OF DIFFERENTIABILITY

The purpose of this chapter is to define several notions of differentiability in general analysis and to study some of their properties. The notions of Gâteaux variation and differential, Fréchet differential, and strong Fréchet differential are defined for operators and implication relationships among these are studied. Sufficient conditions for the existence of these differentials are also established.

Definition 1.1. Let J be an open interval of the real line. A mapping φ: $J \to X$ is said to have a first derivative at the point $t_0 \in J$ if

$$\lim_{t \to t_0} \frac{\varphi(t)-\varphi(t_0)}{t-t_0} \triangleq \varphi'(t_0) \tag{1}$$

exists, where the limit is understood in the sense of the norm in X.

It is evident that whenever $\varphi'(t_0)$ exists, it is a unique element of X.

Definition 1.2. Let f: $A \to Y$ and let x_0 and h be fixed elements of A and X, respectively. Since A is an open set, there exists an interval $J = (-\delta,\delta)$ such that if $t \in J$, then $x_0 + th \in A$. If for each $h \in X$

$$\frac{d}{dt} f(x_0+th) \Big|_{t=0} \triangleq Vf[x_0; h]$$

exists, then f is said to have a Gâteaux variation or to be weakly differentiable at x_0. If it exists, $Vf[x_0; h]$ is called the Gâteaux variation (weak differential, first variation, directional derivative) of f at x_0 with increment h.

If f has a Gâteaux variation at x_0, then $Vf[x_0; \cdot]$ is a unique operator mapping X into Y.

Theorem 1.1. If f has a Gâteaux variation at x_0, then $Vf[x_0; \cdot]$ is homogeneous of degree one; that is, for any scalar α, $Vf[x_0; \alpha h]$ exists and

$$Vf[x_0; \alpha h] = \alpha\, Vf[x_0; h]$$

for all $h \in X$.

Proof: If $\alpha = 0$, the proof is trivial. If $\alpha \neq 0$, then

$$Vf[x_0; \alpha h] = \lim_{t\to 0} t^{-1}[f(x_0+t\alpha h) - f(x_0)]$$

$$= \alpha \lim_{\tau\to 0} \tau^{-1}[f(x_0+\tau h) - f(x_0)]$$

where $\tau = \alpha t$.

Therefore,

$$Vf[x_0;\ \alpha h] = \alpha\ Vf[x_0;\ h].$$

In the remainder of this chapter the task of determining when a useful approximation to $f(x_0+h) - f(x_0)$ exists and the accuracy of such an approximation will be investigated. Some insight can be gained through a thorough examination of the problem in the one dimensional case.

Let J be an open interval and let $x_0 \epsilon J$. Suppose that $f\colon\ J \to E_1$. If $f'(x_0)$ exists, then $f(x_0+h) - f(x_0)$ can be approximated for small h by $f'(x_0)h$. The nature of this approximation is well known. In fact let

$$R[x_0;\ h] \triangleq f(x_0+h) - f(x_0) - f'(x_0)h$$

and let $df[x_0;\ h] \triangleq f'(x_0)h$. Then

$$f(x_0+h) - f(x_0) = df[x_0;\ h] + R[x_0;\ h] \tag{3}$$

where $R[x_0;\ h] = o(\,|\,h\,|\,)$ as $h \to 0$.

The following special properties of $df[x_0;\ h]$ are important:

(A) $df[x_0;\ h]$ is linear in h; that is,

$$df[x_0;\ \alpha_1 h_1+\alpha_2 h_2] = \alpha_1\ df[x_0;\ h_1] + \alpha_2\ df[x_0;\ h_2]$$

for each pair of scalars α_1, α_2 and elements h_1, h_2.

(B) $df[x_0; h]$ is bounded in h; that is, there exists a finite constant M, independent of h, such that

$$| df[x_0; h] | \leq M | h |$$

for all real h.

(C) $df[x_0; h]$ is continuous in h; that is, if $\lim_{n\to\infty} h_n = h$, then

$$\lim_{n\to\infty} df[x_0; h_n] = df[x_0; h].$$

(D) If f is defined in some neighborhood of x_0, g is defined in some neighborhood of $y_0 = f(x_0)$, and if $df[x_0; \cdot]$ and $dg[y_0; \cdot]$ exist, then defining $h(t) = g[f(t)]$, it follows that $dh[x_0; \cdot]$ exists and

$$dh[x_0; \cdot] = dg\Big[y_0; df[x_0; \cdot]\Big].$$

This is called the chain rule property.

(E) If $df[x_0; \cdot]$ exists, then f is continuous at x_0.

(F) The condition

$$| f(x_0+h) - f(x_0) - df[x_0; h] | = o(| h |) \text{ as } h \to 0$$

is satisfied.

The objective now is to use conditions A through F as a basis for defining generalizations of the classical differential, which apply to operators on normed spaces.

The Gâteaux variation may in general fail to have any of the properties A through F. One natural approach seems to be to add these properties to $Vf[x_0; \cdot]$ either one at a time or in groups, and to investigate the resulting concepts.

Definition 1.3. Suppose that f has a Gâteaux variation $Vf[x_0; \cdot]$ at $x_0 \in A$. If $Vf[x_0; \cdot]$ is continuous and linear, then it is denoted by $Df[x_0; \cdot]$ and f is said to be Gâteaux differentiable at x_0 with Gâteaux derivative $Df[x_0; \cdot]$. The Gâteaux derivative, when it exists, is an operator mapping X into Y. For $h \in X$, $Df[x_0; h]$ is called the Gâteaux differential of f at x_0 with increment h.

By defining Gâteaux differentiability in this way, it is evident that a Gâteaux differential is a Gâteaux variation and that if both exist, they must be the same. This further implies uniqueness of the Gâteaux differential whenever it exists.

The next two theorems deal with necessary and sufficient conditions for $Vf[x_0; \cdot]$ to be a Gâteaux derivative.

Theorem 1.2. Let $f\colon A \to Y$ with $x_0 \in A$. A necessary and sufficient condition for $Vf[x_0; \cdot]$ to be a Gâteaux derivative is that there exists a continuous, linear operator $L[x_0; \cdot]\colon H \to Y$ such that for each $h \in H$,

$$f(x_0+h) - f(x_0) = L[x_0; h] + R[x_0; h] \tag{4}$$

where

$$H = \{h \in X \mid x_0 + h \in A\}$$

and

$$\| R[x_0;\ th] \| = o(|t|) \quad \text{as } t \to 0. \tag{5}$$

If such an operator L exists, it is unique and for $h \in H$,

$$L[x_0;\ h] = Vf[x_0;\ h] = Df[x_0;\ h]. \tag{6}$$

<u>Proof</u>: (Necessity) Suppose that $Vf[x_0;\ \cdot]$ is a Gâteaux derivative. Then $Vf[x_0;\ \cdot]$ is continuous and linear and by Definition 1.2,

$$\lim_{t \to 0} \| t^{-1}[f(x_0+th) - f(x_0)] - Vf[x_0;\ h] \| = 0$$

for each $h \in X$.

Thus

$$\lim_{t \to 0} \| t^{-1}\{f(x_0+th) - f(x_0) - Vf[x_0;\ th]\} \| = 0. \tag{7}$$

Now define

$$R[x_0;\ th] \triangleq f(x_0+th) - f(x_0) - Vf[x_0;\ th].$$

Then (7) becomes

$$\| R[x_0;\ th] \| = o(|t|) \quad \text{as } t \to 0. \tag{8}$$

Thus the representation (4) exists where

$$L[x_0;\ h] = Vf[x_0;\ h].$$

It follows that L is unique and (6) is satisfied.

(Sufficiency) Suppose that the representation (4) exists. If h is replaced by th in (4), it follows that

$$t^{-1}[f(x_0+th) - f(x_0)] = L[x_0;\ h] + t^{-1}\ R[x_0;\ th]. \quad (9)$$

Thus, in view of (5), f has a Gâteaux variation at x_0 and furthermore

$$Vf[x_0;\ h] = L[x_0;\ h]$$

for all $h \in H$. This proves that $L[x_0;\ \cdot]$ is unique and (6) is satisfied since L is linear and continuous.

Definition 1.4. The operator $F\colon\ A \rightarrow Y$ satisfies at the point $x_0 \in A$ a weak Lipschitz condition if for each $h \in X$ there exists $\delta(h) > 0$ such that if $|t| < \delta(h)$, then

$$\| F(x_0+th) - F(x_0) \| \leq C \| th \| ,$$

where C is a positive constant independent of h.

Theorem 1.3. (See [34].) Suppose that F has a Gâteaux variation $VF[x_0;\ \cdot]$ at $x_0 \in A$. Then $VF[x_0;\ \cdot]$ is linear and continuous if and only if F satisfies the following two conditions:

1. F satisfies a weak Lipschitz condition at x_0
2. $\| \Delta^2_{th_1,th_2} F(x_0) \| = o(|t|)$ as $t \rightarrow 0$

where

$$\Delta^2_{h_1,h_2} F(x_0) \triangleq F(x_0+h_1+h_2) - F(x_0+h_1) - F(x_0+h_2) + F(x_0).$$

Proof: (Necessity) Suppose that $VF[x_0;\ \cdot]$ is linear and continuous. Then $VF[x_0;\ \cdot]$ is bounded;

$$\| VF[x_0;\ \cdot] \| = M < \infty.$$

Let h be a fixed element in X. Then by Definition 1.2

$$\lim_{t \to 0} \left\| t^{-1}\left[F(x_0+th) - F(x_0)\right] \right\| = \| VF[x_0;\ h] \| < (M+1) \| h \|.$$

Thus, there exists $\delta(h) > 0$ such that if $0 < |t| < \delta$, then

$$\left\| t^{-1}\left[F(x_0+th) - F(x_0)\right] \right\| < (M+1) \| h \|. \quad (10)$$

But (10) implies that

$$\| F(x_0+th) - F(x_0) \| < (M+1) \| th \| \quad \text{if } 0 < |t| < \delta(h).$$

Thus F satisfies a weak Lipschitz condition at x_0.

By linearity of $VF[x_0;\ \cdot]$, it follows that for $h_1, h_2 \in X$,

$$VF[x_0;\ h_1+h_2] - VF[x_0;\ h_1] - VF[x_0;\ h_2] = 0.$$

But by Definition 1.2,

$$VF[x_0;\ h_1+h_2] - \alpha_1 = t^{-1}\left[F(x_0+th_1+th_2) - F(x_0)\right],$$

$$VF[x_0;\ h_1] - \alpha_2 = t^{-1}\left[F(x_0+th_1) - F(x_0)\right],$$

and

$$VF[x_0;\ h_2] - \alpha_3 = t^{-1}\left[F(x_0+th_2) - F(x_0)\right]$$

where

$$\lim_{t\to 0} \| \alpha_i \| = 0, \qquad i = 1,2,3.$$

Now

$$\left\| t^{-1}\left[F(x_0+th_1+th_2) - F(x_0+th_1) - F(x_0+th_2) + F(x_0)\right] \right\|$$

$$= \| VF(x_0;\ h_1+h_2) - \alpha_1 - VF(x_0;\ h_1) + \alpha_2 - VF(x_0;\ h_2) + \alpha_3 \|$$

$$= \| \alpha_2 + \alpha_3 - \alpha_1 \|$$

$$\leq \| \alpha_1 \| + \| \alpha_2 \| + \| \alpha_3 \|.$$

Thus

$$\| \Delta^2_{th_1,th_2} \| = o(\,|\,t\,|\,) \quad \text{as } t \to 0.$$

This completes the proof of necessity.

(Sufficiency) Suppose first that F satisfies a weak Lipschitz condition at x_0. Then for $h \in X$, there exists $\delta(h) > 0$ such that

$$\left\| t^{-1}\left[F(x_0+th) - F(x_0)\right] \right\| \leq C \| h \| \quad \text{if } 0 < |\,t\,| < \delta(h). \tag{11}$$

Then take the limit as $t \to 0$ in (11) and it follows that

$$\| VF[x_0;\ h] \| \leq C \| h \| \tag{12}$$

for each $h \in X$. Thus $VF[x_0;\ \cdot\,]$ is a bounded operator since C is independent of h.

By an argument identical to the one used in the proof of necessity, it is clear that for $h_1, h_2 \in X$,

$$\| VF[x_0;\ h_1+h_2] - VF[x_0;\ h_1] - VF[x_0;\ h_2] \|$$

$$\leq \left\| t^{-1}\left[F(x_0+th_1+th_2) - F(x_0+th_1) - F(x_0+th_2) + F(x_0)\right] \right\|$$

$$+ \| \alpha_1 - \alpha_2 - \alpha_3 \|$$

$$\leq \frac{1}{|t|} \left\| \Delta^2_{th_1, th_2} F(x_0) \right\| + \| \alpha_1 - \alpha_2 - \alpha_3 \|. \qquad (13)$$

Now since

$$\| \Delta^2_{th_1, th_2} F(x_0) \| = o(\,|t|\,) \quad \text{as } t \to 0,$$

the right-hand side in (13) has limit 0 as $t \to 0$, by the way in which the α_i were defined. Thus

$$\| VF[x_0;\ h_1+h_2] - VF[x_0;\ h_1] - VF[x_0;\ h_2] \| = 0.$$

This proves that $VF[x_0;\ \cdot]$ is additive but $VF[x_0;\ \cdot]$ is also homogeneous by Theorem 1.1. Thus $VF[x_0;\ \cdot]$ is linear and bounded. This completes the proof of sufficiency.

In general, properties E and F are not satisfied by the Gâteaux differential. Note that Theorem 1.2 provides a condition similar to property F but not as strong since the limit there is essentially taken along a fixed direction. This restriction can be removed by assuming the limit in (7) is uniform in h for $\|h\| = 1$. More precisely we have

Theorem 1.4. Let f be Gâteaux differentiable at x_0. Define

$$R[x_0;\ h] \triangleq f(x_0+h) - f(x_0) - Df[x_0;\ h].$$

Then the condition

$$\| R[x_0;\ th] \| = o(\ |t|\ ) \quad \text{as } t \to 0 \tag{14}$$

uniformly in h for $\| h \| = 1$ is necessary and sufficient for the condition

$$\| R[x_0;\ h] \| = o(\|h\|) \quad \text{as } \|h\| \to 0. \tag{15}$$

Proof: Suppose that (14) holds. Then given $\varepsilon > 0$, there exists $\delta > 0$ such that if $0 < |t| < \delta$, then

$$\frac{1}{|t|} \| R[x_0;\ th] \| < \varepsilon \tag{16}$$

for all h with $\|h\| = 1$. Choose any element $k \in X$ with $0 < \|k\| < \delta$. Then

$$k = \|k\| \frac{k}{\|k\|} \triangleq t'h'$$

where $0 < |t'| < \delta$ and $\|h'\| = 1$. Thus relation (16) holds for $t = t'$, $h = h'$. Hence

$$\frac{1}{\|k\|} \| R[x_0;\ k] \| < \varepsilon. \tag{17}$$

But (17) holds for all k with $0 < \|k\| < \delta$, so that (15) holds. This proves sufficiency.

(Necessity) Suppose condition (15) holds. Then given $\varepsilon > 0$, there exists $\delta > 0$ such that if $0 < \|h\| < \delta$, then

$$\frac{\|R[x_0;\ h]\|}{\|h\|} < \varepsilon. \tag{18}$$

Now consider the subset of X whose elements are of the form tk where $\|k\| = 1$ and $|\ t\ | < \delta$. Clearly (18) must be valid for all these elements. So if $0 < |\ t\ | < \delta$, then

$$\frac{\|R[x_0;\ tk]\|}{|\ t\ |} < \varepsilon$$

for all k, with $\|k\| = 1$. Hence condition (14) holds and the proof is complete.

The following definition takes advantage of Theorem 1.4.

Definition 1.5. If f is Gâteaux differentiable at x_0 and if condition (15) holds, then f is said to be Fréchet differentiable at x_0. The bounded linear operator

$$df[x_0;\ \cdot] \triangleq Df[x_0;\ \cdot]$$

is then called the Fréchet derivative of f at x_0 and for $h \in X$, the element $df[x_0;\ h]$ is called the Fréchet differential of f at x_0 with increment h.

Fréchet differentiability clearly implies Gâteaux differentiability but the converse is in general false.

Example 1.1. Let

$$f(z_1,z_2) = \begin{cases} z_1 + z_2 + \dfrac{z_1^3 z_2}{z_1^4 + z_2^2} & \text{if } (z_1,z_2) \neq (0,0) \\ 0 & \text{if } (z_1,z_2) = (0,0). \end{cases}$$

Consider, for $h = (h_1,h_2) \neq (0,0)$

$$\lim_{t\to 0} \frac{f(0+th_1,0+th_2) - f(0,0)}{t}$$

$$= \lim_{t\to 0} \frac{f(th_1,th_2)}{t}$$

$$= \lim_{t\to 0} \left[h_1 + h_2 + \frac{th_1^3 h_2}{t^2 h_1^4 + h_2^2} \right]$$

$$= h_1 + h_2.$$

If $(h_1,h_2) = (0,0)$, then the limit is 0 so that $Vf[(0,0); \cdot]$ exists and

$$Vf[(0,0); h] = h_1 + h_2$$

for all $h \in E_2$. The operator $Vf[(0,0); \cdot]$ is linear and continuous so that f is Gâteaux differentiable at (0,0) and

$$Df[(0,0); h] = h_1 + h_2.$$

But if

$$R[(0,0);\ h] \triangleq f(h) - f(0) - Df[(0,0);\ h],$$

then

$$R[(0,0);\ h] = \frac{h_1^3 h_2}{h_1^4 + h_2^2}$$

and

$$\lim_{\|h\|\to 0} \frac{|R[(0,0);\ h]|}{\|h\|} = \lim_{\|h\|\to 0} \left| \frac{h_1^3 h_2}{\left(h_1^4 + h_2^2\right)\left(h_1^2 + h_2^2\right)^{1/2}} \right| .$$

But if this limit is evaluated along the path $h_2 = h_1^2$, then

$$\lim_{\substack{\|h\|\to 0\\ h_2 = h_1^2}} \frac{|R[(0,0);\ h]|}{\|h\|} = \frac{1}{2}.$$

This proves that f is not Fréchet differentiable at (0,0).

At this point it is clear that the Fréchet derivative satisfies properties A,B,C, and F. It turns out that properties D and E are also satisfied. The proof of D is a corollary of Theorem 1.9, a more general result. Property E is the subject of

Theorem 1.5. If f is Fréchet differentiable at x_0, then f is continuous at x_0.

Proof: Fréchet differentiability implies that

$$f(x_0+h) - f(x_0) = df[x_0;\ h] + R[x_0;\ h]$$

$$= df[x_0;\ h] + \frac{\|h\|}{\|h\|} R[x_0;\ h].$$

Thus

$$\lim_{h\to 0} [f(x_0+h) - f(x_0)] = 0$$

since $df[x_0;\ \cdot]$ is continuous at 0 and

$$\lim_{h\to 0} \frac{1}{\|h\|} R[x_0;\ h] = 0.$$

The definition of Fréchet differential provides sufficient conditions for its existence. Another set of sufficient conditions is given by

Theorem 1.6. If f has a Gâteaux derivative in some neighborhood of x_0 and if $Df[x;\ \cdot]$ is continuous in x at $x = x_0$, then $df[x_0;\ \cdot]$ exists and

$$df[x_0;\cdot] = Df[x_0;\ \cdot].$$

Proof: (See [33].)

Two additional notions of operator differentiability will now be presented. Each has some interesting connections with those differentials already introduced.

Definition 1.6. (See [8].) Let f be a continuous mapping of A into Y and let $x_0 \in A$. The mapping f is said to be quasi-differentiable at x_0 if there exists a linear mapping $u(\cdot)$: $X \to Y$ such that for any continuous mapping g: $[0,1] \to A$, for which $g(0) = x_0$ and $g'(0^+)$ exists, the mapping

$$H(t) \triangleq f[g(t)]$$

is differentiable (Definition 1.1) at $t = 0^+$ and

$$H'(0^+) = u\Big(g'(0^+)\Big).$$

The mapping $u(\cdot)$ is called the quasi-derivative of f at x_0.

A connection between quasi-differentiability and the notions already introduced is readily available in

Theorem 1.7. If f is quasi-differentiable at $x_0 \in A$, then f is Gâteaux differentiable at x_0.

Proof: Suppose that $u(\cdot)$ is the quasi-derivative of f at x_0. Let $h \in X$. Then since A is open, there exists a number $\delta > 0$ such that if $0 \leq t \leq \delta$, then $x_0 + th \in A$. Define the mapping g as follows:

$$g(t) = \begin{cases} x_0 + th & (0 \leq t \leq \delta), \\ g(\delta) & (\delta < t \leq 1). \end{cases}$$

Then g: $[0,1] \to A$, $g(0) = x_0$, g is continuous on $[0,1]$, and furthermore $g'(0^+) = h$. Thus

$$\frac{d}{dt} f[g(t)]\bigg|_{t=0^+} = u(h).$$

Now let $\varepsilon > 0$ be given. Then there exists a number r, $0 < r < \delta$, such that if $0 < t < r$, then

$$\left\| t^{-1}\left[f[g(t)] - f(x_0) \right] - u(h) \right\| < \varepsilon. \qquad (19)$$

But if $t < \delta$, then

$$g(t) = x_0 + th$$

so this substitution can be made in (19) and hence if $0 < t < r$, then

$$\left\| t^{-1}\left[f(x_0+th) - f(x_0) \right] - u(h) \right\| < \varepsilon. \qquad (20)$$

However (20) implies that f has a Gâteaux variation at x_0 and that

$$Vf[x_0; h] = u(h).$$

But $u(\cdot)$ is linear by definition and continuous [8], so that $u(\cdot)$ is the Gâteaux derivative of f at x_0.

Corollary 1.8. If f is quasi-differentiable, the quasi-derivative is unique.

The converse of Theorem 1.7 is false and the mapping of Example 1.1 provides a counterexample. In that example,

$$Df[(0,0); h] = h_1 + h_2.$$

Suppose that f is quasi-differentiable at (0,0) with quasi-derivative $u(\cdot)$. Then

$$u(h) = h_1 + h_2$$

by Theorem 1.7. Now define

$$g(t) = (t,t^2) \qquad \text{for } t \in [0,1].$$

Then

$$g(0) = (0,0) \qquad \text{and } g'(0^+) = (1,0).$$

But

$$H(t) = f[g(t)] = t^2 + \frac{3}{2}t$$

so that

$$H'(0^+) = \frac{3}{2}$$

but

$$u\left[g'(0^+)\right] = 1 + 0 = 1$$

and this is a contradiction. Hence f is not quasi-differentiable at (0,0).

The following theorem shows that the quasi-derivative satisfies the chain rule property (D).

Theorem 1.9. Suppose that $f\colon A \to Y$. Let $x_0 \in A$ and $y_0 = f(x_0) \in B \subset Y$. If f is quasi-differentiable at x_0 and $g\colon B \to Z$ is quasi-differentiable at y_0, then the composite mapping $h(\cdot) = g[f(\cdot)]$ is quasi-differentiable at x_0 and furthermore

$$Dh[x_0;\ \cdot] = Dg\Big[y_0;\ Df[x_0;\ \cdot]\Big].$$

Proof: Let $A_0 \subset A$ be an open neighborhood of x_0 such that $h\colon\ A_0 \to Z$. Now let φ be a continuous mapping of $[0,1]$ into A_0, such that $\varphi(0) = x_0$ and $\varphi'(0^+)$ exists. Then φ maps $[0,1]$ into A and since f is quasi-differentiable at x_0, the mapping P defined by

$$P(t) = f[\varphi(t)]$$

is differentiable at $t = 0^+$ and

$$P'(0^+) = Df[x_0;\ \varphi'(0^+)].$$

But $P(t)$ maps $[0,1]$ into B_0, where $B_0 \subset B$ and B_0 is an open neighborhood of y_0. Furthermore,

$$P(0) = f[\varphi(0)] = f(x_0) = y_0$$

and

$$P'(0^+) = Df[x_0;\ \varphi'(0^+)]$$

exists, so letting

$$Q(t) = g[P(t)],$$

then the quasi-differentiability of g at y_0 implies that $Q'(0^+)$ exists and

$$Q'(0^+) = Dg[y_0;\ P'(0^+)]$$
$$= Dg\Big[y_0;\ Df[x_0;\ \varphi'(0^+)]\Big].$$

But

$$Q(t) = g\Big[f[\varphi(t)]\Big] = h[\varphi(t)]$$

and since $Dg\Big[y_0;\ Df[x_0;\ \cdot\,]\Big]$ is linear, this proves that h is quasi-differentiable at x_0 and that

$$Dh[x_0;\ \cdot\,] = Dg\Big[y_0;\ Df[x_0;\ \cdot\,]\Big].$$

The following theorem, important in its own right, also implies that the Fréchet derivative satisfies property (D), a claim made earlier but not proved.

<u>Theorem 1.10</u>. If f is Fréchet differentiable at $x_0 \in A$, then f is quasi-differentiable at x_0 and the two differentials are identical.

<u>Proof</u>: Let g be a continuous mapping of [0,1] into A, for which $g(0) = x_0$ and $g'(0^+)$ exists. For $t \in [0,1]$, let

$$H(t) = f[g(t)]$$

and consider

$$t^{-1}[H(t) - H(0)] = t^{-1}\Big[f[g(t)] - f[g(0)]\Big]$$
$$= t^{-1}\Big\{df[x_0;\ g(t) - g(0)]$$
$$+ Rf[x_0;\ g(t) - g(0)]\Big\}$$

since f is Fréchet differentiable at $g(0) = x_0$. But $df[x_0;\ \cdot]$ is homogeneous, so

$$t^{-1}[H(t) - H(0)] = df[x_0;\ t^{-1}\{g(t) - g(0)\}] + t^{-1}\, Rf[x_0;\ g(t) - g(0)]. \tag{21}$$

Now

$$\lim_{t\to 0^+} df[x_0;\ t^{-1}\{g(t) - g(0)\}] = df[x_0;\ g'(0^+)]$$

since $df[x_0;\ \cdot]$ is continuous. Consider

$$t^{-1}\, \| Rf[x_0;\ g(t) - g(0)] \| = \frac{\|Rf\ x_0;\ g(t) - g(0)\ \|}{\|g(t) - g(0)\|} \cdot \frac{\|g(t) - g(0)\|}{t}$$

and let $\varepsilon > 0$ be given. Then there exists $\delta > 0$ such that if $0 < \|k\| < \delta$, then

$$\frac{\|Rf[x_0;\ k]\|}{\|k\|} < \varepsilon. \tag{22}$$

But g is continuous so given $\delta > 0$, there exists $r > 0$ such that

$$\| g(t) - g(0) \| < \delta \tag{23}$$

if $0 < t < r$. Combining (22) and (23)

$$\frac{\|Rf[x_0;\ g(t) - g(0)]\|}{\|g(t) - g(0)\|} < \varepsilon$$

if $0 < t < r$. This proves that

$$\lim_{t\to 0^+} \frac{\|Rf[x_0;\ g(t) - g(0)]\|}{\|g(t) - g(0)\|} = 0$$

and since $\|g'(0^+)\|$ is finite, it follows that

$$H'(0^+) = \lim_{t\to 0^+} t^{-1}[H(t) - H(0)] = df[x_0;\ g'(0^+)]\ .$$

Hence f is quasi-differentiable at x_0 and the quasi-derivative of f at x_0 is $df[x_0;\ \cdot]$.

The following theorem gives sufficient conditions for the existence of the quasi-derivative in terms of the Gâteaux variation.

Theorem 1.11. Let $f\colon\ A \to Y$ and suppose that f satisfies a Lipschitz condition on A; that is, there exists a constant M such that for each pair of points $x, y \in A$,

$$\|f(x) - f(y)\| \leq M\|x-y\|.$$

Suppose also that f has a linear Gateaux variation $Vf[x_0;\ \cdot]$ at $x_0 \in A$. Then f is quasi-differentiable at x_0.

Proof: Let g be a continuous mapping of $[0,1]$ into A, for which $g(0) = x_0$ and $g'(0^+)$ exists. For $t \in [0,1]$, let

$$H(t) = f[g(t)].$$

Then

$$\begin{aligned} &\| t^{-1}[H(t) - H(0)] - Vf[x_0; g'(0^+)] \| \\ \leq\ &\| t^{-1}\{f[g(t)] - f[x_0 + tg'(0^+)]\} \| \\ &+ \| t^{-1}\{f[x_0+tg'(0^+)] - f(x_0)\} - Vf[x_0; g'(0^+) \| \\ \leq\ &Mt^{-1} \| g(t) - g(0) - tg'(0^+) \| \\ &+ \| t^{-1}\{f[x_0+tg'(0^+)] - f(x_0)\} - Vf[x_0; g'(0^+)] \|. \end{aligned}$$

Now let $\varepsilon > 0$ be given. There exists $\delta_1 > 0$ such that

$$\| t^{-1}\{f[x_0+tg'(0^+)] - f(x_0)\} - Vf[x_0; g'(0^+)] \| < \varepsilon/2$$

if $0 < t < \delta_1$ and there exists $\delta_2 > 0$ such that

$$t^{-1} \| g(t) - g(0) - tg'(0^+) \| < \varepsilon/2M$$

if $0 < t < \delta_2$. Hence if

$$0 < t < \min(\delta_1, \delta_2) \triangleq \delta \quad ,$$

then

$$\| t^{-1}[H(t) - H(0)] - Vf[x_0; g'(0^+)] \| < \varepsilon.$$

This proves that

$$H'(0^+) = Vf[x_0; g'(0^+)],$$

and since $Vf[x_0; \cdot]$ was assumed to be linear, then f is quasi-differentiable at x_0 and $Vf[x_0; \cdot]$ is the quasi-derivative of f at x_0.

Corollary 1.12. Under the hypotheses of Theorem 1.11, f is Gâteaux differentiable at x_0.

Until now, none of the notions of differential has been stronger than the Fréchet differential. However Leach in [23] has defined a differential which implies the Fréchet differential.

Definition 1.7. Let f: $A \to Y$ and let $x_0 \in A$. If there exists a continuous linear operator $dF^*[x_0; \cdot]$: $X \to Y$ such that for each $\varepsilon > 0$, there exists $r > 0$ where

$$\| f(y) - f(z) - dF^*[x_0; y-z] \| \leq \varepsilon \| y-z \|$$

for each pair of points y,z with $\|y-x_0\| \leq r$, $\|z-x_0\| \leq r$, then f is said to be strongly differentiable at x_0 and the operator $dF^*[x_0; \cdot]$ is called the strong derivative of f at x_0.

Clearly if f is strongly differentiable at x_0, then f is Fréchet differentiable at x_0.

The next theorem is a generalization of results obtained by Esser and Shisha [11] for real-valued functions of real variables. This theorem does not seem to appear in the literature.

Let F: $A \to Y$. Let D' denote the subset of A on which F is Fréchet differentiable and let D^* denote the subset of A on which F is strongly differentiable. Then clearly $D^* \subset D'$ and in general either or both of these sets may be empty. If $x_1 \in D'$, then the symbol $dF[x_1; \cdot]$ will denote the Fréchet derivative of F at x_1 and if $x_2 \in D^*$, then $dF^*[x_2; \cdot]$ will denote the strong derivative of F at x_2.

Theorem 1.13. If F is strongly differentiable at $a \in A$, then for each $h \in X$,

$$\lim_{\substack{x \to a \\ x \in D^*}} dF^*[x; h] = \lim_{\substack{x \to a \\ x \in D'}} dF[x; h] = dF^*[a; h] = dF[a; h]$$

whenever both limits are meaningful.

Proof: Let $\varepsilon > 0$ be given. Since F is strongly differentiable at a, there exists $\delta > 0$ such that $\|x_1-a\| < \delta$, $\|x_2-a\| < \delta$, and $x_1 \neq x_2$ together imply that $x_1, x_2 \in A$ and also that

$$\| F(x_2) - F(x_1) - dF^*[a; x_2-x_1] \| \leq \varepsilon \|x_2-x_1\|. \quad (24)$$

Let $x \in D'$ where $\|x-a\| < \frac{\delta}{2}$. Then x is a permissible value for x_1 in (24). Now let h be a fixed nonzero element of X. Then if

$$| \tau | < \frac{\delta}{2\|h\|} ,$$

it follows that

$$\| x+\tau h-a \| \leq \| x-a \| + \| \tau h \| < \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

and hence $x + \tau h$ is a permissible value for x_2 in (24) provided

$$|\tau| < \frac{\delta}{2\|h\|}.$$

Thus if

$$\|\tau h\| < \frac{\delta}{2}$$

then

$$\| F(x+\tau h) - F(x) - dF^*[a; \tau h] \| \leq \varepsilon \| \tau h \|. \tag{25}$$

But since $dF^*[a; \cdot]$ is homogeneous, (25) is equivalent to

$$\|\tau^{-1}[F(x+\tau h) - F(x)] - dF^*[a; h] \| \leq \varepsilon \| h \| \tag{26}$$

from which it follows that

$$\| VF[x; h] - dF^*[a; h] \| \leq \varepsilon \| h \|. \tag{27}$$

But $x \in D'$ so that $dF[x; h]$ exists and

$$VF[x; h] = dF[x; h].$$

Thus

$$\| dF[x; h] - dF^*[a; h] \| \leq \varepsilon \| h \| \tag{28}$$

if $x \in D'$ and

$$\| x-a \| < \frac{\delta}{2}.$$

This proves that

$$\lim_{\substack{x \to a \\ x \in D'}} dF[x;\ \cdot] = dF^*[a;\ \cdot]$$

since

$$\| dF[x;\ \cdot] - dF^*[a;\ \cdot] \| = \sup_{h \neq 0} \left\{ \frac{\|dF[x;\ h] - dF^*[a;\ h]\|}{\|h\|} \right\}.$$

Thus for each $h \in X$,

$$\lim_{\substack{x \to a \\ x \in D'}} dF[x;\ h] = dF^*[a;\ h]. \tag{29}$$

Now assuming that $\lim_{\substack{x \to a \\ x \in D^*}} dF[x;\ h]$ exists,

$$\lim_{\substack{x \to a \\ x \in D^*}} dF^*[x;\ h] = \lim_{\substack{x \to a \\ x \in D^*}} dF[x;\ h]$$

since $dF[\cdot;\ \cdot]$ and $dF^*[\cdot;\ \cdot]$ are identical on D^*. But now it follows that

$$\lim_{\substack{x \to a \\ x \in D^*}} dF[x;\ h] = \lim_{\substack{x \to a \\ x \in D'}} dF[x;\ h] \tag{30}$$

since $D^* \subset D'$. The combination of (29) and (30) gives the desired result.

The following theorem provides a sufficient condition for a Fréchet derivative to be a strong derivative.

Theorem 1.14. Let $a \in A$ and suppose that F is Fréchet differentiable in some neighborhood N of a. If $dF[x; \cdot]$ is continuous in x at $x = a$, then F is strongly differentiable at a.

Proof: Let $\varepsilon > 0$ be given and let $x_1 \in N$, $x_2 \in X$. Let

$$F_x'(\cdot) = dF[x; \cdot]$$

and consider

$$\begin{aligned} & \| F(x_2) - F(x_1) - F_a'(x_2-x_1) \| \\ \leq \; & \| F(x_2) - F(x_1) - F_{x_1}'(x_2-x_1) \| \\ & + \| F_{x_1}'(x_2-x_1) - F_a'(x_2-x_1) \| \\ \leq \; & \| F(x_2) - F(x_1) - F_{x_1}'(x_2-x_1) \| \\ & + \| F_{x_1}'(\cdot) - F_a'(\cdot) \| \, \| x_2-x_1 \| \end{aligned} \tag{30}$$

since a Fréchet derivative is a bounded operator.

Now since $F_x'(\cdot)$ is continuous in x at $x = a$, there exists a number $\delta'' > 0$ such that if $\|x_1-a\| < \delta''$, then $x_1 \in N(a)$ and furthermore

$$\| F_{x_1}'(\cdot) - F_a(\cdot) \| < \tfrac{\varepsilon}{2}. \tag{31}$$

Thus (30) and (31) both are valid if $\|x_1-a\| < \delta''$.

But F is Fréchet differentiable at x_1, so there exists $\delta' > 0$ such that $\|x_2-x_1\| < \delta'$ implies that

$$\| F(x_2) - F(x_1) - F'_{x_1}(x_2-x_1) \| < \frac{\varepsilon}{2} \| x_2-x_1 \|. \quad (32)$$

Thus if $\|x_1-a\| < \delta''$ and $\|x_2-x_1\| < \delta'$, then (31) and (32) together imply that

$$\| F(x_2) - F(x_1) - F'_a(x_2-x_1) \| \leq \frac{\varepsilon}{2} \| x_2-x_1 \| + \frac{\varepsilon}{2} \| x_2-x_1 \|$$

$$= \varepsilon \| x_2-x_1 \|. \quad (33)$$

Now let

$$\delta = \frac{1}{2} \min\{\delta', \delta''\}$$

and suppose that

$$\| x_1-a \| < \delta, \ \| x_2-a \| < \delta. \quad (34)$$

Then

$$\| x_2-x_1 \| < 2\delta \leq \delta'.$$

Thus (34) implies (33) and the theorem is proved.

Since strong differentiability implies Fréchet differentiability, one might expect that strong differentiability at a point would imply more than just continuity at that point.

Theorem 1.15. (See [22].) If F is strongly differentiable at x_0, then F satisfies a Lipschitz condition in some neighborhood of x_0.

Proof: By definition there exists $r > 0$ such that

$$\| F(y) - F(z) - dF^*[x_0; y-z] \| \leq \| y-z \| \qquad (35)$$

if y and z are in

$$N(x_0; r) = \{x \in X \mid \|x-x_0\| \leq r\}.$$

Now by the triangle inequality, (35) implies that

$$\| F(y) - F(z) \| \leq \| y-z \| + M \| y-z \|$$

where

$$\| dF^*[x_0; \cdot] \| \leq M.$$

Thus

$$\| F(y) - F(z) \| \leq (M+1) \| y-z \|$$

for all $y,z \in N(x_0; r)$.

A discussion of the next topic has been purposely postponed until now to avoid needless confusion.

When dealing with arbitrary normed linear spaces, it must be pointed out that the differentiability properties of an operator may depend on which norms are used in the range or domain spaces. This can best be illustrated with an example.

Example 1.2. Let $C'[0,1]$ be the linear space of all real-valued continuous functions x defined on $[0,1]$ such that x' is continuous and $x(0) = x(1) = 0$. Now for $x \in C'[0,1]$, let

$$f(x) = \int_0^1 \varphi\Big(t, x(t), x'(t)\Big)dt$$

where φ is continuous and has continuous second partial derivatives with respect to each of its arguments. If we let

$$\| x \|_1 = \sup_{0 \leq t \leq 1} | x(t) | ,$$

then $Vf[x; h]$ exists and

$$Vf[x; h] = \int_0^1 \{\varphi_x h + \varphi_{x'} h'\} dt .$$

Furthermore $Vf[x; h]$ is linear in h but not necessarily continuous in h. So using $\|\cdot\|_1$, only a linear Gâteaux variation is implied for f.

However, letting

$$\| x \|_2 = \sup_{0 \leq t \leq 1} \{ |x(t)|, |x'(t)| \}$$

it can be shown that f is Fréchet differentiable at each $x \in C'[0,1]$. This of course implies that f is a continuous functional with respect to $\|\cdot\|_2$.

This problem, however, does not exist when both the range and domain of the operator are finite dimensional spaces. It is well known that all norms in a finite dimensional space are equivalent. The problem also does not exist when equivalent norms are used in infinite dimensional spaces.

CHAPTER II

DIFFERENTIABILITY OF LIPSCHITZIAN OPERATORS

This chapter is concerned primarily with the differentiability properties of Lipschitzian operators.

A brief introduction to Lipschitzian operators is presented for the sake of completeness. Then in a series of theorems, some significant known results are presented. These results are known only for operators whose domains and ranges are subsets of the finite dimensional Euclidean spaces $\{E_n\}$. It is of course out of the question to expect theorems implying differentiability everywhere for such operators, for counterexamples are well known.

Very strong results, however, of the almost everywhere type are known. The main result of this chapter is a theorem of Rademacher [27], a corollary of which states that Lipschitzian operators mapping E_n into E_1 are Fréchet differentiable almost everywhere. Rado and Reichelderfer [28] have extended Rademacher's result, although under slightly more restrictive hypotheses.

The question of differentiability of arbitrary Lipschitzian operators remains unanswered.

Definition 2.1. Let $Q \subset X$ and let $f\colon\ Q \to Y$. If there exists a finite positive constant M such that for each pair of points $x, y \in Q$ the inequality

$$\| f(y) - f(x) \| \leq M \| x-y \|$$

holds, then f is said to satisfy a Lipschitz condition on Q or f is said to be Lipschitzian on Q. If $M < 1$, then f is called a contraction on Q.

Clearly if f is Lipschitzian on Q, then f is uniformly continuous on Q.

The following results concerning differentiability of Lipschitzian operators, whose domain and range are both subsets of E_1, are well known.

Theorem 2.1. Let $f: [a,b] \rightarrow E_1$. Suppose that on [a,b] f satisfies a Lipschitz condition. Then f is Fréchet differentiable almost everywhere on [a,b].

Proof: Since f satisfies a Lipschitz condition on [a,b], then f is absolutely continuous on [a,b]. This implies that $f'(x)$ exists for almost all $x \in [a,b]$. (See [29].)

The following theorem is a generalization of Theorem 2.1 in two ways. First the domain space is allowed to be E_n and second the Lipschitz assumption is weakened slightly. However in Theorem 2.3, where the range space is allowed to be E_m, the stronger Lipschitz hypothesis is again used.

Theorem 2.2. (See Rademacher, [27].) Let f be a continuous mapping of G into E_1, where G is a bounded open subset of E_n. For each fixed $z \in G$, let

$$W_f[z;\ \rho] \triangleq \sup\left\{ \frac{|\,f(z+h) - f(z)\,|}{\|h\|} \;\middle|\; h \in E_n \text{ and } 0 < \|h\| < \rho \right\}. \tag{1}$$

Let

$$L_f(z) \triangleq \lim_{\rho \to 0^+} W_f[z;\ \rho]. \tag{2}$$

If $L_f(z)$ is finite for each $z \in G$ and if $L_f(z)$ is Lebesgue summable over G, then f is Fréchet differentiable almost everywhere on G.

Proof: Since the proof of this theorem is rather lengthy, we shall proceed by making a few observations and introducing some functions which will be crucial to the proof. We will then break up the proof into a sequence of lemmas.

We first remark that $W_f[z;\ \rho]$ is nonnegative and furthermore $W_f[z;\ \rho]$ decreases as ρ decreases. Thus the limit in (2) exists but this limit may not be finite in general.

It is clear that f is measurable on G since f is continuous on G. Thus $W_f[z;\ \rho]$ and $L_f(z)$ are also measurable functions.

Let e_j be the vector in E_n which has j^{th} component 1 and all others 0. Then if

$$x = (x_1, x_2, \ldots, x_n) = (x_i) \in G,$$

then

$$x + ke_i = (x_1, x_2, \ldots, x_{i-1}, x_i+k, x_{i+1}, \ldots, x_n)$$

for real k. Using this notation, the vector

$$(x_1, x_2, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n)$$

can be written simply $x + (t-x_i)e_i$.

Now let $x = (x_i) \in G$ and define the <u>i^{th} principal upper partial derivate</u> of f at x, $\overline{D}_i f(x)$, as follows:

$$\overline{D}_i f(x) = \lim_{\rho \to 0^+} \sup\left\{\frac{f(x+ke_i) - f(x)}{k} \,\middle|\, 0 < |k| < \rho\right\} \tag{3}$$

for $i = 1, 2, \ldots, n$.

It will now be shown that $\overline{D}_i f(x)$ is finite for each $x \in G$. Let

$$S = \left\{\frac{f(x+ke_i) - f(x)}{k} \,\middle|\, 0 < |k| < \rho\right\},$$

$$S_1 = \left\{\frac{|f(x+ke_i) - f(x)|}{|k|} \,\middle|\, 0 < |k| < \rho\right\},$$

and

$$S_2 = \left\{\frac{|f(x+h) - f(x)|}{\|h\|} \,\middle|\, h \in E_n \text{ and } 0 < \|h\| < \rho\right\}.$$

Then by well known properties of suprema,

$$|\sup S| \leq \sup S_1 \leq \sup S_2,$$

so that

$$- \sup S_2 \leq \sup S \leq \sup S_2. \tag{4}$$

Furthermore, since sup S, sup S_1 and sup S_2 all decrease as ρ decreases, the limit can be taken in (4) to get

$$- \lim_{\rho \to 0^+} \sup S_2 \leq \lim_{\rho \to 0^+} \sup S \leq \lim_{\rho \to 0^+} \sup S_2. \tag{5}$$

But (5) is equivalent to

$$- L_f(x) \leq \overline{D}_i f(x) \leq L_f(x) \tag{6}$$

and (6) implies that

$$| \overline{D}_i f(x) | \leq L_f(x). \tag{7}$$

This shows that $\overline{D}_i f(x)$ is a finite number for each $x \in G$ and for $i = 1,2,\ldots,n$. Furthermore $\overline{D}_i f(x)$ is a measurable function on G and (7) implies that $\overline{D}_i f(x)$ is also summable over G.

From now on, whenever any statement is made which concerns the subscript i, it is to be understood that i can be any integer from 1 to n, inclusive, unless statements to the contrary are made.

Similarly the <u>i^{th} principal lower partial derivate</u> of f at x, $\underline{D}_i f(x)$, is defined as in (3) except that "sup" is replaced by "inf". Furthermore finiteness, measurability, and summability also hold for $\underline{D}_i f(x)$.

The first objective is to show that the n partial derivatives of f exist simultaneously almost everywhere on G.

Let $y = (y_1, y_2, \ldots, y_{n-1})$ be a fixed point in E_{n-1}. For each real t define

$$y_t^i \triangleq (y_1, y_2, \ldots, y_{i-1}, t, y_i, \ldots, y_{n-1}).$$

Next define

$$E_y^i = \left\{ x \in G \mid x = y_t^i \quad \text{for some real } t \right\}.$$

Furthermore define

$$T_y^i = \left\{ t \in E_1 \mid y_t^i \in G \right\}.$$

Only those sets E_y^i which are nonempty need be considered.

Define functions g_y^i as follows:

$$g_y^i(t) = f(y_t^i)$$

for each $t \in T_y^i$. Then for those y and t such that $\frac{d}{dt}\, g_y^i(t)$ exists, it is clear that the i^{th} partial derivative of f, written f_i, also exists at y_t^i. The plan is to show that this is the case for "almost all" y_t^i, where "almost all" will be made precise shortly.

Now since $E_y^i \subset G$, then $\overline{D}_i f$ is summable over E_y^i. Hence, by a theorem of Fubini [Appendix, Theorem 1], the integral

$$I_i(y) = \int_{E_y^i} \overline{D}_i f(y_t^i)dt$$

is finite for almost all y; that is, except possibly for an n-1 dimensional null set, since $y \in E_{n-1}$.

Now let y be such that $| I_i(y) | < \infty$. Then since $\overline{D}_i f(x)$ is finite for each $x \in E_y^i$, it follows from another theorem of Fubini [Appendix, Theorem 2] that for almost all y, the function g_y^i is absolutely continuous on T_y^i. But this implies that for almost all y, $\frac{d}{dt}\, g_y^i(t)$ exists for almost all t.

However

$$\frac{d}{dt}\, g_y^i(t) = f_i\left(y_t^i\right)$$

and hence for almost all y, $f_i\left(y_t^i\right)$ exists for almost all t. This conclusion is valid for i = 1,2,...,n, but the null sets on which the conclusion fails may depend on i.

Now define a subset P_i as follows:

$$P_i = \left\{x \in G \mid \overline{D}_i f(x) \neq \underline{D}_i f(x)\right\}.$$

The set P_i is clearly measurable. The first objective will be achieved when it is shown that P_i is a null set in E_n.

So for $y \in E_{n-1}$ such that E_y^i is not empty, let

$$P_y^i = P_i \cap E_y^i$$

and let

$$S_y^i = \left\{t \in T_y^i \;\middle|\; y_t^i \in P_y^i\right\}.$$

Then S_y^i is that subset of T_y^i for which $f_i\left(y_t^i\right)$ does not exist. Hence S_y^i must be a null set in E_1 for almost all y, by previous results.

At this point, another theorem of Fubini [Appendix, Theorem 3] can be used to conclude that P_i itself is a null set in R_n. But P_i represented the subset of G on which f_i failed to exist, so that f_i exists almost everywhere on G. By a standard argument, it follows that all f_i exist simultaneously almost everywhere on G. Let

$$K = \{x \in G \mid f_i(x) \text{ exists for } i = 1,2,\ldots,n\}.$$

Then $mK = mG$.

Now let $\lambda > 0$ be given. The next objective is to show that there exists a subset E_λ of G such that

$$mE_\lambda > mG - 3\lambda$$

and f is Fréchet differentiable on E_λ. This will be accomplished by a sequence of lemmas.

<u>Lemma 2.3</u>. There exists a subset H of K and a finite constant $M(\lambda)$ such that

$$mH > mK - \lambda = mG - \lambda$$

and for $z \in H$,

$$L_f[z] < M(\lambda).$$

Proof of Lemma 2.3. By hypothesis, $mG < \infty$ so that $mK < \infty$, and since L_f is summable over K, then

$$\lim_{n\to\infty} m\{z \in K \mid L_f[z] \geq n\} = 0,$$

by a well known continuity theorem for descending sequences of sets. Then there exists an integer $M(\lambda)$ such that

$$m\{z \in K \mid L_f[z] \geq M(\lambda)\} < \lambda.$$

Let

$$H = K - \{z \in K \mid L_f(z) \geq M(\lambda)\}.$$

Then

$$mH = mK - m\{z \in K \mid L_f(z) \geq M(\lambda)\} \geq mK - \lambda = mG - \lambda$$

and for $z \in H$,

$$L_f(z) < M(\lambda).$$

Lemma 2.4. Let H be as in Lemma 2.3. Then there exists a subset S of H such that

$$mS > mK - 2\lambda = mG - 2\lambda$$

and the n + 1 limits

$$\lim_{\rho\to 0^+} W_f[x;\ \rho] = L_f(x)$$

and

$$\lim_{h\to 0} \frac{f(x+he_i) - f(x)}{h} = f_i(x)$$

hold uniformly on S.

Proof of Lemma 2.4: The lemma follows from n + 1 applications of the Egoroff theorem [31] which is valid since $mK < \infty$. At each stage a subset of measure smaller than $\frac{\lambda}{n+1}$ is extracted from H until at the $(n+1)^{st}$ stage, the subset S emerges with $mS > mH-\lambda$.

Lemma 2.5. Let S be as in Lemma 2.4 and let $M(\lambda)$ be as in Lemma 2.3. Then there exists a positive number $\rho_0(\lambda)$ such that if $0 < \rho < \rho_0(\lambda)$, then

$$W_f[z;\ \rho] < 2M(\lambda)$$

for all $z \in S$.

Proof of Lemma 2.5: By Lemma 2.4,

$$\lim_{\rho\to 0^+} W_f[z;\ \rho] = L_f(z)$$

holds uniformly for $z \in S$. So given $M(\lambda) > 0$, there exists a number $\rho_0(\lambda) > 0$ such that if $0 < \rho < \rho_0(\lambda)$, then for all $z \in S$,

$$| W_f[z; \rho] - L_f(z) | < M(\lambda). \tag{8}$$

But for $z \in G$,

$$W_f[z; \rho] \geq L_f(z)$$

so that the absolute value is redundant in (8). Hence if $0 < \rho < \rho_0(\lambda)$, then for all $z \in S$,

$$W_f[z; \rho] < M(\lambda) + L_f(z) < 2M(\lambda)$$

by Lemma 2.3. This completes Lemma 2.5.

Now let $\varepsilon > 0$ be given. Then by Lemma 2.4, there exists a number $h_0(\lambda; \varepsilon) > 0$ such that for all $x \in S$,

$$\left| \frac{f(x+he_i) - f(x)}{h} - f_i(x) \right| < \varepsilon \tag{9}$$

if $0 < | h | < h_0(\lambda; \varepsilon)$. It should be clear how one number $h_0(\lambda; \varepsilon)$ can be chosen to make (9) valid for all $i = 1,2,\ldots,n$.

The uniform convergence of Lemma 2.4 also implies that each f_i is continuous on S. But since S is a measurable set of positive measure, there exists a compact subset T of S such that

$$mT > mS - \lambda > mG - 3\lambda. \tag{10}$$

(See [6, p. 288, Theorem 12].) But since T is compact, f_i is uniformly continuous on T. Hence there exists $\overline{h}(\varepsilon) > 0$ such that if $x, y \in T$ and $\|x-y\| < \overline{h}(\varepsilon)$, then

$$| f_i(x) - f_i(y) | < \varepsilon. \tag{11}$$

Now for $z \in T$ and $r > 0$, define

$$S[z; r] \triangleq \left\{ x = (x_i) \in G \mid z_i - \frac{r}{2} \leq x_i \leq z_i + \frac{r}{2}, \quad i = 1,2,\ldots,n \right\}.$$

Then clearly

$$m\{S[z; r]\} = r^n.$$

<u>Lemma 2.6</u>. Let

$$E_\lambda = \left\{ z \in T \mid \lim_{r \to 0} \frac{m\{T \cap S[z; r]\}}{r^n} = 1 \right\}.$$

Then $mE_\lambda = mT$.

<u>Proof of Lemma 2.6</u>: Define the set function ν as follows:

$$\nu R = m(R \cap T)$$

for each $R \subset G$. Then ν is additive and absolutely continuous with respect to m. Furthermore by [26, vol 2, p. 97],

$$\lim_{r \to 0} \frac{\nu\{S[z;\ r]\}}{r^n}$$

exists for almost all $z \in G$.

Now let χ_T be the characteristic function of T. Then if $R \subseteq G$,

$$\int_R \chi_T = m(R \cap T) = \nu R.$$

Since χ_T is summable over G, it follows from [26, vol 2, p. 102] that

$$\lim_{r \to 0} \frac{\nu\{S[z;\ r]\}}{r^n} = \chi_T(z)$$

for almost all $z \in T$. Thus

$$\lim_{r \to 0} \frac{\nu\{S[z;\ r]\}}{r^n} = 1$$

for almost all $z \in T$. But

$$\nu\{S[z;\ r]\} = m\{T \cap S[z;\ r]\}$$

and hence

$$\lim_{r \to 0} \frac{m\{S[z;\ r] \cap T\}}{r^n} = 1$$

for almost all $z \in T$. This proves that $mE_\lambda = mT$ and Lemma 2.6 is complete.

Now Lemma 2.6 implies that, given any fixed point $z \in E_\lambda$ and integer $N > 1$, there exists $r_N(z) > 0$ such that if $0 < r < r_N(z)$, then

$$\frac{m\{S[z;\ r] \cap T\}}{r^n} > 1 - \frac{1}{N^n}. \tag{12}$$

But $mT < \infty$ and $mT = mE_\lambda$ so that

$$m\{(T-E_\lambda) \cap S[z;\ r]\} = 0$$

which implies that

$$m\{T \cap S[z;\ r]\} = m\{E_\lambda \cap S[z;\ r]\}.$$

Thus (12) implies that if $0 < r < r_N(z)$, then

$$\frac{m\{E_\lambda \cap S[z;\ r]\}}{r^n} > 1 - \frac{1}{N^n}.$$

Now choose $r_0 > 0$ so that

$$r_0 < \tfrac{1}{2} \min\{h_0(\varepsilon;\ \lambda),\ \overline{h}(\varepsilon),\ r_N(z)\}. \tag{13}$$

Then choose $y = (y_i) \in S[z;\ 2r_0]$ with the following properties:

(i) $y_i = z_i + h_1 \cos \alpha_i$ where $h_1 > 0$ and

$$\sum_{i=1}^{n} \cos^2\alpha_i = 1$$

(ii) $0 < \|y-z\| < \frac{N-1}{N} r_0$

(iii) $0 < \|y-z\| < \dfrac{(N-1)\rho_0(\lambda)}{\sqrt{n}}$.

Then by (i), $\|y-z\| = h_1$, and by (ii),

$$0 < 2h_1 < \frac{2(N-1)r_0}{N}$$

so that by (13),

$$0 < \frac{2Nh_1}{N-1} < 2r_0 < r_N(z).$$

Now by (12),

$$\frac{m\left\{E_\lambda \cap S\left[z;\ \frac{2Nh_1}{N-1}\right]\right\}}{\left(\frac{2Nh_1}{N-1}\right)^n} > 1 - \frac{1}{N^n}$$

and hence

$$m\left\{E_\lambda \cap S\left[z;\ \frac{2Nh_1}{N-1}\right]\right\} > \frac{2^n h_1^n (N^n-1)}{(N-1)^n} \ . \tag{14}$$

Lemma 2.7. From the definition of y, it follows that

$$S\left[y;\ \frac{2h_1}{N-1}\right] \subset S\left[z;\ \frac{2Nh_1}{N-1}\right]$$

Proof: Let

$$t = (t_i) \in S\left[y; \frac{2h_1}{N-1}\right].$$

Then for $i = 1,2,\ldots,n$

$$y_i - \frac{h_1}{N-1} \leq t_i \leq y_i + \frac{h_1}{N-1} \tag{15}$$

but $\|y-z\| = h_1$ so that

$$| y_i - z_i | \leq h_1$$

or

$$z_i - h_1 \leq y_i \leq z_i + h_1. \tag{16}$$

Now combining (15) and (16), it follows that

$$z_i - \frac{Nh_1}{N-1} \leq t_i \leq z_i + \frac{Nh_1}{N-1}$$

and this implies that $t \in S\left[z; \frac{2Nh_1}{N-1}\right]$. This completes Lemma 2.7.

Now

$$mS\left[y; \frac{2h_1}{N-1}\right] = \frac{2^n h_1^n}{(N-1)^n}$$

so that

$$m\left\{E_\lambda \cap S\left[z;\ \frac{2Nh_1}{N-1}\right]\right\} + mS\left[y;\ \frac{2h_1}{N-1}\right] > \frac{2^n h_1^n (N^n - 1)}{(N-1)^n} + \frac{2^n h_1^n}{(N-1)^n}$$

$$= \frac{2^n h_1^n N^n}{(N-1)^n} = mS\left[z;\ \frac{2Nh_1}{N-1}\right].$$

Thus the sum of the measures of two subsets of $S\left[z;\ \frac{2Nh_1}{N-1}\right]$ is greater than the measure of $S\left[z;\ \frac{2Nh_1}{N-1}\right]$ itself. This is possible only if

$$\left\{E_\lambda \cap S\left[z;\ \frac{2Nh_1}{N-1}\right]\right\} \cap S\left[y;\ \frac{2h_1}{N-1}\right]$$

has positive measure. So let $u = (u_i) \neq y$ be a point in this intersection. Then $u \in S\left[y;\ \frac{2h_1}{N-1}\right]$ and hence

$$\| u-y \| = \left[\sum_{i=1}^{n} (u_i - y_i)^2\right]^{1/2} \leq \left[\sum_{i=1}^{n} \left(\frac{h_1}{N-1}\right)^2\right]^{1/2} = \frac{\sqrt{n}\ h_1}{N-1}. \tag{17}$$

We are now ready to show that f is Fréchet differentiable at z, where z is any fixed point in E_λ.

To this end, consider the difference quotient

$$\frac{f(y) - f(z)}{h_1}.$$

By addition and subtraction of proper terms, this quotient can be written as the sum of n + 1 terms, two of which must be estimated separately and the remaining n - 1 terms can be estimated by considering only a typical one. So we write

$$\begin{aligned}\frac{f(y) - f(z)}{h_1} = \frac{f(y) - f(u)}{h_1} &+ \frac{f(u) - f(z_1,u_2,\dots,u_n)}{h_1} \\ &+ \frac{f(z_1,u_2,\dots,u_n) - f(z_1,z_2,u_3,\dots,u_n)}{h_1} \\ &+ \cdots \\ &+ \frac{f(z_1,z_2,\dots,z_{n-2},u_{n-1},u_n) - f(z_1,\dots,z_{n-1},u_n)}{h_1} \\ &+ \frac{f(z_1,\dots,z_{n-1},u_n) - f(z)}{h_1}. \qquad (18)\end{aligned}$$

The first and last terms are the ones which must be estimated separately.

By Lemma 2.5, (17) and (iii),

$$\left| \frac{f(y) - f(u)}{h_1} \right| = \frac{|f(y) - f(u)|}{\|y-u\|} \frac{\|y-u\|}{h_1}$$

$$\leq W_f\left[u; \frac{\sqrt{n}\ h_1}{N-1}\right] \frac{\sqrt{n}\ h_1}{(N-1)h_1} \leq 2M(\lambda) \frac{\sqrt{n}}{N-1} .$$

Thus

$$\frac{f(y) - f(u)}{h_1} = \eta_0$$

where

$$|\eta_0| \leq \frac{2 \sqrt{n}\, M(\lambda)}{N-1}. \qquad (19)$$

Now let $h_2 > 0$ and $\beta_1, \beta_2, \ldots, \beta_n$ be constants such that

$$u_i = z_i + h_2 \cos \beta_i \qquad (20)$$

and

$$\sum_{i=1}^{n} \cos^2 \beta_i = 1.$$

Then

$$\| u-z \| = h_2 \qquad (21)$$

and this will be used in the estimation of the n - 1 "middle terms" of the expansion (18). A typical one of these terms has the form

$$\frac{f(z_1, z_2, \ldots, z_{i-1}, u_i, u_{i+1}, \ldots, u_n) - f(z_1, \ldots, z_i, u_{i+1}, \ldots, u_n)}{h_1}.$$

Note: For i = 1, the above form is not exactly correct since there should be no z_i at all in the first term of the numerator. This, however, does not affect the proof.

Using (20), it follows that

$$\frac{f(z_1,\ldots,z_{i-1},u_i,\ldots,u_n) - f(z_1,\ldots,z_i,u_{i+1},\ldots,u_n)}{h_1}$$

$$= \frac{h_2 \cos \beta_i}{h_1} \frac{[f(z_1,\ldots,z_{i-1},u_i,\ldots,u_n) - f(z_1,\ldots,z_i,u_{i+1},\ldots,u_n)]}{u_i - z_i} \tag{22}$$

but

$$h_2 = \| u-z \| \le \| u-y \| + \| y-z \| \le \frac{\sqrt{n}\ h_1}{N-1} + h_1 \tag{23}$$

$$\le \left(\frac{\sqrt{n}}{N-1} + 1\right) \frac{N-1}{N}\ r_0$$

by (ii) and hence

$$h_2 \le \left(1 + \frac{\sqrt{n}-1}{N}\right) r_0 < 2r_0 < \begin{cases} h_0(\lambda;\ \varepsilon) \\ \bar{h}(\varepsilon) \end{cases} \tag{24}$$

if N is chosen sufficiently large. Now by (20) and (24),

$$| u_i - z_i | \le h_2 < \begin{cases} h_0(\lambda;\ \varepsilon) \\ \bar{h}(\varepsilon) \end{cases}$$

so that by (9), it must be true that

$$\left| \frac{f(z_1,\ldots,z_{i-1},u_i,\ldots,u_n)-f(z_1,\ldots z_i,u_{i+1},\ldots,u_n)}{u_i - z_i} - f_i(z_1,\ldots,z_i,u_{i+1},\ldots,u_n)\right| < \varepsilon . \qquad (25)$$

Furthermore since

$$\| (z_1,\ldots,z_i,u_{i+1},\ldots,u_n) - z \| \leq \| u-z \| = h_2 < \overline{h}(\varepsilon)$$

it follows from (11) that

$$| f_i(z_1,\ldots,z_i,u_{i+1},\ldots,u_n) - f_i(z) | < \varepsilon . \qquad (26)$$

Putting (25) and (26) together, it follows that

$$\frac{f(z_1,\ldots,z_{i-1},u_i,\ldots,u_n) - f(z_1,\ldots,z_i,u_{i+1},\ldots,u_n)}{u_i - z_i} = f_i(z) + \eta_1 \qquad (27)$$

where $|\eta_1| < 2\varepsilon$.

Thus (27) gives an estimate of the last factor in (22).

The complete estimate of (22) requires an estimate of $\frac{h_2}{h_1}$ and of $\cos \beta_i$.

From (23),

$$h_2 - h_1 \leq \frac{\sqrt{n}\ h_1}{N-1}$$

and from similar reasoning it is also true that

$$h_1 - h_2 \leq \frac{\sqrt{n}\ h_1}{N-1}$$

so that

$$| h_2 - h_1 | \leq \frac{\sqrt{n}\ h_1}{N-1}$$

which implies that

$$\left| \frac{h_2}{h_1} - 1 \right| \leq \frac{\sqrt{n}}{N-1} \tag{28}$$

and hence

$$\frac{h_2}{h_1} = 1 + \eta_2 \text{ where } | \eta_2 | \leq \frac{\sqrt{n}}{N-1} . \tag{29}$$

Now to estimate $\cos \beta_i$, consider

$$| \cos \beta_i - \cos \alpha_i | = \left| \frac{u_i - z_i}{h_2} - \frac{y_i - z_i}{h_1} \right|$$

$$= \left| \frac{h_1 u_i - h_1 z_i - h_2 y_i + h_2 z_i}{h_1 h_2} \right| . \tag{30}$$

But by (29), $h_2 = h_1(1 + \eta_2)$ and this can be substituted into (30) to give

$$| \cos \beta_i - \cos \alpha_i | = \left| \frac{h_1 u_i - h_1 z_i - h_1(1+\eta_2) y_i + h_1(1+\eta_2) z_i}{h_1 h_2} \right|$$

$$= \left| \frac{u_i - y_i - \eta_2 y_i + \eta_2 z_i}{h_2} \right|$$

$$\leq \frac{|u_i - y_i|}{h_2} + \frac{|\eta_2|\,|z_i - y_i|}{h_2}$$

$$\leq \frac{\|u-y\|}{h_2} + \frac{|\eta_2|\,\|z-y\|}{h_2}.$$

Thus

$$|\cos \beta_i - \cos \alpha_i| \leq \frac{\sqrt{n}\ h_1}{(N-1)h_2} + |\eta_2|\ \frac{h_1}{h_2}. \tag{31}$$

In order to proceed from (31), an estimate of $\frac{h_1}{h_2}$ is needed. So from (28),

$$1 - \frac{\sqrt{n}}{N-1} \leq \frac{h_2}{h_1} \leq 1 + \frac{\sqrt{n}}{N-1}$$

and hence

$$h_1\left(1 - \frac{\sqrt{n}}{N-1}\right) \leq h_2 \leq h_1\left(1 + \frac{\sqrt{n}}{N-1}\right). \tag{32}$$

But for N sufficiently large, the quantity $\dfrac{1}{1 - \frac{\sqrt{n}}{N-1}}$ is positive so (32) can be divided by $\dfrac{1}{1 - \frac{\sqrt{n}}{N-1}}$ to give

$$0 < h_1 \leq \frac{h_2}{1 - \frac{\sqrt{n}}{N-1}} \leq h_1\ \frac{1 + \frac{\sqrt{n}}{N-1}}{1 - \frac{\sqrt{n}}{N-1}}\ . \tag{33}$$

Now divide inequality (33) by h_2 to get

$$0 < \frac{h_1}{h_2} \le \frac{1}{1 - \frac{\sqrt{n}}{N-1}} . \tag{34}$$

The right-hand term of (33) is of no further use.

Using (34) in (31) it follows that

$$|\cos \beta_i - \cos \alpha_i| \le \left(\frac{\sqrt{n}}{N-1} + |\eta_2|\right) \cdot \frac{1}{1 - \frac{\sqrt{n}}{N-1}} \le \frac{\frac{2\sqrt{n}}{N-1}}{1 - \frac{\sqrt{n}}{N-1}} = \frac{2\sqrt{n}}{N-1-\sqrt{n}}$$

which implies that

$$\cos \beta_i = \cos \alpha_i + \eta_3 \tag{35}$$

where

$$|\eta_3| < \frac{2\sqrt{n}}{N-1-\sqrt{n}} .$$

Now putting (35), (29), and (27) together, the left-hand side of (22) is just

$$(f_i(z) + \eta_1)(1+\eta_2)(\cos \alpha_i + \eta_3) \tag{36}$$

where

$$|\eta_1| < 2\varepsilon, \ |\eta_2| < \frac{\sqrt{n}}{N-1} \quad \text{and} \quad |\eta_3| < \frac{2\sqrt{n}}{N-1-\sqrt{n}} .$$

We now expand (36) to get

$$f_i(z)\cos\alpha_i + f_i(z)\left[\eta_2\cos\alpha_i + \eta_3(1+\eta_2)\right] + \eta_1\left[1+\eta_2\right]\left[\cos\alpha_i + \eta_3\right] \tag{37}$$

and using (37) and (22), it follows that

$$\left| \frac{f(z_1,\ldots,z_{i-1},u_i,\ldots,u_n) - f(z_1,\ldots,z_i,u_{i+1},\ldots,u_n)}{h_1} - f_i(z)\cos\alpha_i \right|$$

$$\leq |f_i(z)| \left[\frac{\sqrt{n}}{N-1}(1) + \frac{2\sqrt{n}}{N-1-\sqrt{n}}\left(1+\frac{\sqrt{n}}{N-1}\right)\right] + 2\varepsilon\left[1+\frac{\sqrt{n}}{N-1}\right]\left[1+\frac{2\sqrt{n}}{N-1-\sqrt{n}}\right]. \tag{38}$$

Choose N so large that (24) and (33) are valid and also $\frac{\sqrt{n}}{N-1-\sqrt{n}} < \varepsilon$. Then having ε and N, r_0 can be chosen so that if

$$0 < h_1 < \min\left\{\frac{N-1}{N} r_0,\ \frac{\rho_0(\lambda)(N-1)}{\sqrt{n}}\right\},$$

the left-hand side of (38) is less than or equal to

$$|f_i(z)|\,[\varepsilon+2\varepsilon(1+\varepsilon)] + 2\varepsilon[1+\varepsilon][1+2\varepsilon]$$

$$= |f_i(z)|\,[3\varepsilon+2\varepsilon^2] + 2\varepsilon + 6\varepsilon^2 + 4\varepsilon^3. \tag{39}$$

This clearly implies that the left-hand side of (38) equals

$$f_i(z)\cos\alpha_i + \theta_i(h_1;\ \alpha_i) \tag{40}$$

where $|\theta_i| \to 0$ as $h_1 \to 0$ uniformly in α_i.

This completes the estimate of the middle $(n-1)$ terms of the expansion (18). We now consider the last term in (18):

$$\frac{f(z_1,\ldots,z_{n-1},u_n) - f(z)}{h_1} = \frac{[f(z_1,\ldots,z_{n-1},u_n) - f(z)]}{u_n - z_n}\,\frac{(u_n-z_n)}{h_1}$$

$$= \left(\frac{f(z_1,\ldots,z_{n-1},u_n) - f(z)}{u_n - z_n}\right)\left(\frac{h_2\cos\beta_n}{h_1}\right). \tag{41}$$

But (29) is still valid and (35) surely holds if $i = n$. Furthermore by the same type of argument which led to (25), it follows that

$$\left|\frac{f(z_1,\ldots,z_{n-1},u_n) - f(z)}{u_n - z_n} - f_n(z)\right| < \varepsilon \tag{42}$$

and hence

$$\frac{f(z_1,\ldots,z_{n-1},u_n) - f(z)}{u_n - z_n} = f_n(z) + \eta_4 \quad \text{where} \quad |\eta_4| < \varepsilon.$$

By reasoning similar to that which led to (40), it follows that

$$\frac{f(z_1,\ldots,z_{n-1},u_n) - f(z)}{h_1} = f_n(z)\cos\alpha_n + \theta_n(h_1,\alpha_n) \quad (44)$$

where $|\theta_n| \to 0$ as $h_1 \to 0$ uniformly in α_n.

Now combining all of the intermediate estimates, we have

$$\frac{f(y) - f(z)}{h_1} = \sum_{i=1}^{n} f_i(z)\cos\alpha_i + \sum_{i=1}^{n} \theta_i(h_1\alpha_i) + \eta_0$$

$$= \sum_{i=1}^{n} f_i(z)\cos\alpha_i + \Theta(h_1,\alpha_1,\ldots,\alpha_n) \quad (45)$$

where $|\Theta| \to 0$ as $h_1 \to 0$ uniformly in all α_i.

To complete the proof let

$$\lambda_i = h_1\cos\alpha_i. \quad (46)$$

Then if $\lambda = (\lambda_i)$, it follows that $y = z + \lambda$ and $\|\lambda\| = h_1$. Then (45) becomes

$$\frac{f(z+\lambda) - f(z)}{\|\lambda\|} - \sum_{i=1}^{n} f_i(z)\cos\alpha_i = \Theta'(\|\lambda\|)$$

or equivalently, using (46)

$$\frac{f(z+\lambda) - f(z) - \sum_{i=1}^{n} f_i(z)\lambda_i}{\|\lambda\|} = \Theta'(\|\lambda\|)$$

where

$$|\Theta'(\|\lambda\|)| \to 0 \quad \text{as} \quad \|\lambda\| \to 0.$$

This proves that f is Frechet differentiable at z since the quantity $\sum_{i=1}^{n} f_i(z)\lambda_i$ is well known to be continuous and linear in λ. But since z was any vector in E_λ, then f is Fréchet differentiable everywhere on E_λ for each $\lambda > 0$.

Now let $\lambda_n = \frac{1}{n}$ for $n \geq 1$ and consider the corresponding sequence $\{E_{\lambda_n}\}$. Then

$$mE_{\lambda_n} > mG - 3/n \tag{47}$$

and f is Fréchet differentiable on E_{λ_n} for each $n \geq 1$. Let

$$E^* = \bigcup_{n=1}^{\infty} E_{\lambda_n}.$$

Then f is Fréchet differentiable on E*. Now since $E^* \subset G$, it must follow that $mG = mE^* + \delta$ where $\delta > 0$. But $mE_{\lambda_n} \leq mE^*$ for each $n \geq 1$ so that

$$mG - mE_{\lambda_n} \geq mG - mE^* = \delta > 0$$

for each $n \geq 1$. Thus $mE_{\lambda_n} \leq mG - \delta$ for each $n \geq 1$. This clearly contradicts (47) and thus f is Fréchet differentiable almost everywhere on G.

Corollary 2.8. Let f: $G \rightarrow E_1$ where G is a bounded open subset of E_n. Suppose that f satisfies a Lipschitz condition on G. Then f is Fréchet differentiable almost everywhere on G.

Proof: Suppose that for $x,y \in G$,

$$| f(x) - f(y) | \leq M \| x-y \|.$$

Then for $h \in E_n$ such that $f(x+h)$ is defined, it follows that

$$\frac{| f(x+h) - f(x) |}{\|h\|} \leq M.$$

Hence,

$$W_f[x; \rho] \leq M$$

and

$$L_f[x] \leq M.$$

Thus $L_f[x]$ is finite for each $x \in G$ and furthermore $L_f[x]$ is summable over G. Thus the conclusion of Theorem 2.1 follows.

Theorem 2.9. (See [28].) Let D be a bounded open subset of E_n and let f: $D \rightarrow E_m$. If f satisfies a Lipschitz condition on D, then f is Fréchet differentiable almost everywhere on D.

Proof: Since all norms in a finite dimensional space are equivalent, the norm

$$\|x\| = \max_i | x_i |$$

will be used in this proof for its computational advantages.

Let $f = (f_1, f_2, \ldots, f_m)$ and let $x = (x_i)$, $y = (y_i) \in D$. By hypothesis there exists a finite constant $M > 0$ such that

$$\| f(x) - f(y) \| \leq M \| x-y \|.$$

Then for all $i = 1,2,\ldots,n$,

$$| f_i(x) - f_i(y) | \leq \|f(x) - f(y)\| \leq M \| x-y \|.$$

This proves that each f_i is Lipschitzian on D, with Lipschitz constant M. But each f_i is real-valued and hence Corollary 2.8 applies. Thus for each $i = 1,2,\ldots,n$, $df_i[x; \cdot]$ exists for almost all $x \in D$. By a standard argument, it can be shown that all $df_i[x; \cdot]$ exist simultaneously almost everywhere on D.

Let x_0 be a point in D such that $df_i[x_0; \cdot]$ exists for $i = 1,2,\ldots,n$. Then define a mapping L from E_n into E_m as follows:

$$L(h) = (df_i[x_0; h]), \quad \text{for} \quad h \in E_n.$$

Now let $\varepsilon > 0$ be given and consider

$$\frac{\|f(x_0+h) - f(x_0) - L(h)\|}{\|h\|} = \frac{\max_i\{|f_i(x_0+h) - f_i(x_0) - df_i[x_0; h]|\}}{\|h\|}$$

$$= \frac{|f_j(x_0+h) - f_j(x_0) - df_j[x_0; h]|}{\|h\|} \tag{48}$$

for some j, $1 \leq j \leq m$. But since f_j is Fréchet differentiable at x_0, there exists a $\delta > 0$ such that if $0 < \|h\| < \delta$, then the right-hand side of (48) is less than ε. Thus if $0 < \|h\| < \delta$, then

$$\frac{\|f(x_0+h) - f(x_0) - L(h)\|}{\|h\|} < \varepsilon.$$

Furthermore

$$\| L(h) \| = \max_i\{ | df_i[x_0; h] |\} \leq K \| h \|$$

since each df_i is a bounded operator. Hence f is Fréchet differentiable at x_0 and $df[x_0; h] = L(h)$ for each $h \in E_n$. But by the choice of x_0, this implies that f is Fréchet differentiable almost everywhere on D.

This completes the discussion of differentiability of Lipschitzian operators on finite dimensional spaces. The question of differentiability of Lipschitzian operators on arbitrary normed spaces remains open.

CHAPTER III

DIFFERENTIABILITY OF CONVEX FUNCTIONALS ON NORMED SPACES

Throughout the literature there seems to be several isolated results in the field of differentiability of convex functionals. For recent results on subdifferentiability of convex functions, the reader may refer to Brønsted and Rockafellar [5] and Moreau [25].

The strongest results are obtained when the domain of the convex functional is a subset of the real line. If the domain is allowed to be an open subset of an arbitrary normed linear space, weaker results have been found.

The first part of this chapter is devoted to some introductory definitions which help to keep this thesis substantially self-contained.

Some known results on differentiability and convexity are presented. Then several of these results are combined to form an apparently new theorem.

Definition 3.1. Let $\mathcal{L}$ be a linear space and let S be a subset of $\mathcal{L}$. The set S is called convex if for each pair of points $x,y, \in S$ and real number λ, with $0 \leq \lambda \leq 1$, the point $\lambda x + (1-\lambda)y$ is also in S.

Note that the whole space $\mathcal{L}$ and the empty set Φ are convex sets.

Definition 3.2. Let $\mathfrak{L}$ be a linear space and let S be a convex subset of $\mathfrak{L}$. A functional f with domain S is said to be convex on S if for each pair of points x,y in S and real number λ, with $0 \leq \lambda \leq 1$, the following condition holds:

$$f[\lambda x + (1-\lambda)y] \leq \lambda f(x) + (1-\lambda)f(y). \tag{1}$$

In Definition 3.2, the set S was assumed to be convex so that the point $\lambda x + (1-\lambda)y$ would automatically fall in the domain of f. If this assumption is dropped, the concept of a convex functional can still be made meaningful.

Definition 3.3. Let $\mathfrak{L}$ be a linear space and let S be any subset of $\mathfrak{L}$. A functional f with domain S is said to be convex if for each pair of points x,y in S and real number λ, with $0 \leq \lambda \leq 1$, such that $\lambda x + (1-\lambda)y \in S$, the condition (1) holds.

Note that Definition 3.3 designates f as "convex", and not as "convex on S".

Theorem 3.1. [26, vol. 2] Let f be a continuous convex functional defined on the finite real interval (a,b). Then f satisfies a Lipschitz condition on each closed subinterval of (a,b).

Proof: Let $c \in (a,b)$. Then the ratio

$$\frac{f(x) - f(c)}{x - c} \tag{2}$$

is an increasing function of x for $x \neq c$. (See [26, vol. 2].)

Now let $[p,q]$ be any closed subinterval of (a,b). Let x,y be fixed points of $[p,q]$ and let $m = \frac{q+b}{2}$. Then since $q < b$, it follows that

$$y \leq q < m < b. \tag{3}$$

Thus

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(m) - f(x)}{m - x} < \frac{f(m) - f(x)}{m - q} \tag{4}$$

and by a triangle inequality,

$$\frac{f(y) - f(x)}{y - x} < \frac{|f(m)| + |f(x)|}{m - q} \;. \tag{5}$$

But f is continuous on $[p,q]$ and hence bounded on $[p,q]$ so that

$$\frac{f(y) - f(x)}{m - x} < \frac{|f(m)| + F}{m - q} = K \tag{6}$$

where K is a constant independent of x and y.

Now let

$$n = \frac{p+a}{2}.$$

Then

$$0 < n < p \leq \begin{Bmatrix} x \\ y \end{Bmatrix} \leq q < m < b$$

and hence

$$\frac{f(n) - f(x)}{n - x} \leq \frac{f(y) - f(x)}{y - x}. \tag{7}$$

But f is bounded below on the closed interval $[n,m]$ and hence there exists some finite number B such that $f(t) \geq B$ for $t \in [n,m]$. Thus

$$f(n) - B \geq f(n) - f(x)$$

and since $n - x < 0$, it follows that

$$\frac{f(n) - B}{n - x} \leq \frac{f(n) - f(x)}{n - x}. \tag{8}$$

Now since

$$0 > n - p \geq n - x,$$

it follows that

$$\frac{1}{n-p} \leq \frac{1}{n-x} < 0 \tag{9}$$

and since

$$f(n) - B \geq 0$$

it follows also that

$$\frac{f(n) - B}{n - p} \leq \frac{f(n) - B}{n - x}. \tag{10}$$

Now combine (7), (8), (10) to get

$$L = \frac{f(n) - B}{n - p} \leq \frac{f(y) - f(x)}{y - x} \tag{11}$$

where L is a constant independent of x and y.

So let

$$M = \max\{ |L| , |K| \}$$

and then from (6) and (11) follows

$$\left| \frac{f(y) - f(x)}{y - x} \right| \leq M, \tag{12}$$

where M depends only on p,q,a, and b and not on x or y. Thus (12) must hold for each x,y in [p,q] and hence f satisfies a Lipschitz condition on [p,q]. But [p,q] was an arbitrary closed subinterval of (a,b) and the theorem is proved.

The conclusion of Theorem 3.1 is the best that can be obtained in the sense that the conclusion "f satisfies a Lipschitz condition on (a,b)" is false. Consider

Example 3.1. Let

$$f(x) = -\sqrt{1-x^2} \quad \text{for} \quad |x| < 1.$$

Then

$$f''(x) = (1-x^2)^{-3/2}$$

and for $|x| < 1$,

$$f''(x) > 0.$$

Hence by a known result [12, p. 24], f is convex on (-1,1).

Now suppose that f satisfies a Lipschitz condition on (-1,1). Then since f is differentiable on (-1,1), f′ must be bounded on (-1,1). But f′ is clearly not bounded on (-1,1). This is a contradiction and hence f does not satisfy a Lipschitz condition on the entire interval (-1,1).

Theorem 3.2. Under the hypotheses of Theorem 3.1, f is differentiable almost everywhere on (a,b).

Proof: Let

$$I_n = \left[a + \frac{1}{n}, b - \frac{1}{n}\right] \quad \text{for} \quad n \geq 1.$$

Then by Theorems 3.1 and 2.1, f is differentiable almost everywhere on I_n for each integer $n \geq 1$. The result now follows from the identity

$$\bigcup_{n=1}^{\infty} I_n = (a,b)$$

and the well known fact that the union of a countable collection of null sets is a null set.

The next theorem is a special case of a more general result to be established later. However, to prove that more general result now would only interrupt the present sequence of theorems. Thus proofs for both the special case and the more general result will be given. This theorem has a slightly stronger result than that of Theorem 3.2.

Theorem 3.3. Let Φ be a real valued function, continuous and convex on the real interval $[a,b]$. Then Φ has finite right-hand and left-hand derivatives at each point of (a,b) and furthermore the subset of $[a,b]$ on which f is not differentiable is countable.

Proof: The proof begins with two lemmas.

Lemma 3.4. Let $x,y,z \in [a,b]$, with $x < y < z$. Then

$$\frac{\Phi(y) - \Phi(x)}{y - x} \leq \frac{\Phi(z) - \Phi(x)}{z - x} .$$

Proof of Lemma 3.4: There exists some t, $0 < t < 1$, such that

$$y = (1-t)x + tz.$$

Then by convexity,

$$\begin{aligned}\Phi(y) &\leq (1-t)\Phi(x) + t\Phi(z)\\ &\leq \Phi(x) + t[\Phi(z) - \Phi(x)]\end{aligned}$$

and hence

$$\Phi(y) - \Phi(x) \leq t[\Phi(z) - \Phi(x)].$$

But $y-x > 0$ and

$$y - z = t(z-x)$$

so that

$$\frac{\Phi(y) - \Phi(x)}{y - x} \leq \frac{\Phi(z) - \Phi(x)}{z - x} . \tag{13}$$

Lemma 3.5. Let $w < x < y < z$ where x,y,z are as in Lemma 3.4 and $w \in [a,b]$. Then

$$\frac{\Phi(x) - \Phi(w)}{x - w} \leq \frac{\Phi(z) - \Phi(y)}{z - y} .$$

Proof of Lemma 3.5: Let

$$y = \alpha x + (1-\alpha)z$$

where $0 < \alpha < 1$. Then proceeding exactly as in Lemma 3.4, it follows that

$$\frac{\Phi(z) - \Phi(y)}{z - y} > \frac{\Phi(z) - \Phi(x)}{z - x}. \tag{14}$$

But this result must also hold when x is replaced by w, y by x, and z by y. Thus

$$\frac{\Phi(y) - \Phi(x)}{y - x} \geq \frac{\Phi(y) - \Phi(w)}{y - w} . \tag{15}$$

Now using Lemma 3.4 with the points w,x, and y it follows that

$$\frac{\Phi(x) - \Phi(w)}{x - w} \leq \frac{\Phi(y) - \Phi(w)}{y - w}. \qquad (16)$$

Thus by (13), (14), (15), and (16) we have

$$\frac{\Phi(x) - \Phi(w)}{x - w} \leq \frac{\Phi(z) - \Phi(y)}{z - y}. \qquad (17)$$

Now to arrive at the conclusion of the theorem, let x be an interior point of [a,b]. Choose two fixed points w_1 and y_1 in [a,b] such that $w_1 < y_1 < x$ and let h be a positive number but small enough so that $x + h \in [a,b]$. Then since

$$w_1 < y_1 < x < x + h,$$

it follows from Lemma 3.5 that

$$\frac{\Phi(y_1) - \Phi(w_1)}{y_1 - w_1} \leq \frac{\Phi(x+h) - \Phi(x)}{h}. \qquad (18)$$

Thus the difference quotient

$$\frac{\Phi(x+h) - \Phi(x)}{h}$$

is bounded below for $h > 0$ and Lemma 3.4 further implies that the quotient is an increasing function of h for $h > 0$. Thus by a well known result,

$$\lim_{h\to 0^+} \frac{\Phi(x+h) - \Phi(x)}{h}$$

exists and hence Φ has a finite right-hand derivative at x.

Now let h and k be negative numbers with $h < k < 0$. Then

$$x + h < x + k < x$$

and by (14) with proper change of symbols,

$$\frac{\Phi(x+k) - \Phi(x)}{k} \geq \frac{\Phi(x+h) - \Phi(x)}{h}.$$

Thus if h is negative, the quotient

$$\frac{\Phi(x+h) - \Phi(x)}{h}$$

is increasing.

Now let h be negative and so small that $x + h \in [a,b]$. Then choose fixed $w_2, y_2 \in [a,b]$ such that

$$x + h < x < y_2 < w_2.$$

Using Lemma 3.5, it follows that

$$\frac{\Phi(x+h) - \Phi(x)}{h} \leq \frac{\Phi(w_2) - \Phi(y_2)}{w_2 - y_2}.$$

Hence the quotient is bounded above for all $h < 0$ and thus

$$\lim_{h \to 0^-} \frac{\Phi(x+h) - \Phi(x)}{h}$$

exists. Thus Φ has a finite left-hand derivative at x, and hence Φ has both finite right- and left-hand derivatives at each interior point of [a,b]. A proof that the set on which Φ' fails to exist is countable can be found in [17, p. 197].

The function Φ in the previous theorem need not have finite one-sided derivatives at a or b. (See Example 3.1.) This completes the section on real valued functions of a single real variable.

In the next section, the domain of the functional will be a subset of a normed linear space.

Definition 3.4. Let f be a functional defined on an open subset A of a normed linear space X and let $x \in X$. If

$$\lim_{t\to 0^+} t^{-1}[f(x+th) - f(x)] \triangleq Vf_+[x; h]$$

exists for each $h \in X$, then f is said to have a one-sided Gâteaux variation at x.

Theorem 3.6. If f has a one-sided Gâteaux variation at $x \in A$, then $Vf_+[x; \cdot]$ is positively homogeneous; that is, for each scalar $\alpha \geq 0$,

$$Vf_+[x; \alpha h] = \alpha Vf_+[x; h]$$

for all $h \in X$.

Proof: (See Theorem 1.1.)

Clearly if $Vf_+[x;\ \cdot]$ exists and if

$$Vf_+[x;\ h] = -Vf_+[x;\ -h]$$

for each $h \in X$, then $Vf[x;\ \cdot]$ exists.

Theorem 3.7. Let $f\colon\ A \to E_1$. If f is convex, then $Vf_+[x;\ \cdot]$ exists for each $x \in A$.

Proof: Let $x \in A$, $y \in X$ and keep them fixed throughout the proof. Since A is open, there exists some positive constant λ_0 such that both $x - \lambda_0 y$ and $x + \lambda_0 y$ are members of A. Let $h = \lambda_0 y$. Then there exists a positive constant $\lambda_1 < 1$ such that $x + \lambda h \in A$ if $0 < \lambda < \lambda_1$.

So let λ be such that $0 < \lambda < \lambda_1$ and write x in the following way:

$$x = \frac{\lambda}{1+\lambda}\ (x-h) + \frac{1}{1+\lambda}\ (x+\lambda h).$$

Since x, $x - h$, $x + \lambda h$ are all in A, Definition 3.3 implies that

$$f(x) \leq \frac{\lambda}{1+\lambda}\ f(x-h) + \frac{1}{1+\lambda}\ f(x+\lambda h)$$

which, upon rearrangement, becomes

$$f(x) - f(x-h) \leq \lambda^{-1}[f(x+\lambda h) - f(x)]. \qquad (19)$$

Similarly since x, $x + h$, and

$$(1-\lambda)x + \lambda(x+h) = x + \lambda h \in D,$$

it follows that

$$f(x+\lambda h) \leq (1-\lambda)f(x) + \lambda f(x+h)$$

which becomes

$$\lambda^{-1}[f(x+\lambda h) - f(x)] \leq f(x+h) - f(x). \tag{20}$$

Now combining (19) and (20) we have

$$f(x) - f(x-h) \leq \lambda^{-1}[f(x+\lambda h) - f(x)] \leq f(x+h) - f(x) \tag{21}$$

for all λ where $0 < \lambda < \lambda_1$. Thus the difference quotient in (21) is bounded.

To see that this difference quotient is also a monotonically increasing function of λ, let $\lambda' = t\lambda$ where $0 \leq t \leq 1$. Then from (20) it follows that

$$\begin{aligned}(\lambda')^{-1}[f(x+\lambda' h) - f(x)] &= (t\lambda)^{-1}[f(x+t\lambda h) - f(x)] \\ &= \lambda^{-1}[t^{-1}\{f[x+t(\lambda h) - f(x)\}] \\ &\leq \lambda^{-1}[f(x+\lambda h) - f(x)]. \end{aligned} \tag{22}$$

Then (21) and (22) imply that

$$\lim_{\lambda \to 0^+} \lambda^{-1}[f(x+\lambda h) - f(x)]$$

exists. This proves that $Vf_+[x;\ h]$ exists. To show that $Vf_+[x;\ y]$ exists, consider

$$\lambda^{-1}[f(x+\lambda y) - f(x)] = \lambda^{-1}\left[f\left(x + \frac{\lambda}{\lambda_0} h\right) - f(x)\right]$$

$$= (\lambda_0)^{-1}\left(\frac{\lambda}{\lambda_0}\right)^{-1}\left[f\left(x + \frac{\lambda}{\lambda_0} h\right) - f(x)\right].$$

Then

$$\lim_{\lambda\to 0^+} \lambda^{-1}[f(x+\lambda y) - f(x)] = (\lambda_0)^{-1} Vf_+[x; h].$$

Thus $Vf_+[x; y]$ exists and

$$Vf_+[x; y] = (\lambda_0)^{-1} Vf_+[x; h].$$

Since x,y were arbitrary, this proves that $Vf_+[x; \cdot]$ exists for each $x \in A$.

Theorem 3.8. Under the hypotheses of Theorem 3.7, the functional $Vf_+[x; \cdot]$ is convex on X for each $x \in A$.

Proof: Let $x \in A$ and let $y,z \in X$. Since A is open, there exists a positive constant λ^* such that if $0 < \lambda < \lambda^*$, then $x + 2\lambda y$, $x + 2\lambda z$, and $x + \lambda(y+z)$ are in A. Then by convexity of f, for $0 < \lambda < \lambda^*$,

$$f[\tfrac{1}{2}(x+2\lambda y) + \tfrac{1}{2}(x+2\lambda z)] \leq \tfrac{1}{2}f(x+2\lambda y) + \tfrac{1}{2}f(x+2\lambda z) \quad (23)$$

since

$$\tfrac{1}{2}(x+2\lambda y) + \tfrac{1}{2}(x+2\lambda z) = x + \lambda(y+z) \in A.$$

Thus from (23), if $0 < \lambda < \lambda^*$, then

$$2f[x+\lambda(y+z)] \leq f(x+2\lambda y) + f(x+2\lambda z).$$

Now subtract $2f(x)$ from both sides and multiply the result by $(2\lambda)^{-1}$. Thus if $0 < \lambda < \lambda^*$, then

$$\lambda^{-1}\{f[x+\lambda(y+z)] - f(x)\} \leq (2\lambda)^{-1}[f(x+2\lambda y) - f(x)] + (2\lambda)^{-1}[f(x+2\lambda z) - f(x)]. \tag{24}$$

Then take $\lim_{\lambda \to 0^+}$ in (24) and it follows that

$$Vf_+[x;\ y+z] \leq Vf_+[x;\ y] + Vf_+[x;\ z]. \tag{25}$$

It is clear that (25) together with Theorem 3.6 imply the result. In fact a stronger result is implied. This proof shows that if α, β are any two nonnegative real numbers, not necessarily with $\alpha + \beta = 1$, it still follows that

$$Vf_+[x;\ \alpha y+\beta z] \leq \alpha Vf_+[x;\ y] + \beta Vf_+[x;\ z].$$

CHAPTER IV

SMOOTH OPERATORS

In this chapter a topic apparently unexplored beyond the one-dimensional level is presented. This is the concept of the smooth operator, generalized from the definition of smooth function in E_1, used by Zygmund [37] and others.

Discussion concerning relationships among smooth operators and the more well known classes of operators is presented, including counterexamples wherever appropriate.

Then results concerning differentiability of smooth operators are given, including an original theorem which provides an immediate connection with Theorem 1.4.

The known one-dimensional results are not dwelled upon for they are readily available in the literature. (See [32] and [37].)

Definition 4.1. Let $f: A \to Y$ and let $x \in A$. The operator f is said to be smooth at x if

$$\| f(x+h) + f(x-h) - 2f(x) \| = o(\|h\|) \quad \text{as} \quad h \to 0. \tag{1}$$

The term "smooth" is a technical term and no connection whatsoever with the literal meaning should be assumed.

Example 4.1. Let

$$f(x) = \sum_{n=1}^{\infty} \frac{\cos 2^n x}{2^n \sqrt{n}} \quad \text{for} \quad x \in E_1.$$

This series converges absolutely and uniformly on E_1 since

$$\left| \frac{\cos 2^n x}{2^n \sqrt{n}} \right| \leq \frac{1}{2^n \sqrt{n}}.$$

Now fix $x \in E_1$ and consider

$$\lim_{h\to 0} \frac{f(x+h) + f(x-h) - 2f(x)}{h}$$

$$= \lim_{h\to 0} \sum_{n=1}^{\infty} \frac{(2 \cos 2^n x)(\cos 2^n h-1)}{h2^n \sqrt{n}}$$

$$= \sum_{n=1}^{\infty} \frac{2 \cos 2^n x}{2^n \sqrt{n}} \lim_{h\to 0} \left[\frac{\cos 2^n h-1}{h}\right],$$

making use of the uniform convergence. But for each $n \geq 1$,

$$\lim_{h\to 0} \left[\frac{\cos 2^n h-1}{h}\right] = 0$$

so that

$$\lim_{h\to 0} \frac{f(x+h) + f(x-h) - 2f(x)}{h} = 0.$$

This shows that f is smooth on E_1.

It is clear from Definition 4.1 that if f and g are smooth at x, then $\alpha f + \beta g$ is also smooth at x, for each pair of scalars α, β. Furthermore, a bounded linear operator is smooth at each point of its domain.

The function in Example 4.1 is typical of a wide class of smooth operators.

<u>Theorem 4.1</u>. Let $f\colon A \to Y$ and let $x \in A$. If f is Fréchet differentiable at x, then f is smooth at x.

<u>Proof</u>: Let $df[x; \cdot]$ denote the Fréchet derivative of f at x. Then

$$\| f(x+h) - f(x) - df[x; h] \| = o(\|h\|) \tag{2}$$

and

$$\| f(x-h) - f(x) - df[x; -h] \| = o(\|h\|) \tag{3}$$

as $h \to 0$. But

$$\begin{aligned} &\|f(x+h) + f(x-h) - 2f(x)\| \\ &= \|f(x+h) - f(x) - df[x; h] + df[x; h] + f(x-h) - f(x)\| \\ &\leq \|f(x+h) - f(x) - df[x; h]\| + \|f(x-h) - f(x) - df[x; -h]\| \\ &= o(\|h\|) + o(\|h\|) = o(\|h\|) \quad \text{as} \quad h \to 0. \end{aligned}$$

This proves that f is smooth at x.

The converse of Theorem 4.1 is false in general. Consider

Example 4.2. Let

$$f(x) = \begin{cases} \sin \frac{1}{x} & (x \neq 0), \\ 0 & (x = 0). \end{cases}$$

Then for $h \neq 0$,

$$\frac{f(0+h) + f(0-h) - 2f(0)}{h} = \frac{\sin \frac{1}{h} + \sin \frac{1}{-h}}{h} = 0$$

and hence f is smooth at $x = 0$, but f is neither differentiable nor continuous at $x = 0$.

The function defined in Example 4.2 is typical of a class of operators which are smooth at the origin. More precisely we have

Theorem 4.2. Let 0 be the zero element of X. Let $f\colon B \to Y$ where B is some neighborhood of 0 and suppose that for each $x \in B$,

$$f(x) = -f(-x).$$

Then f is smooth at $x = 0$.

Proof: The hypotheses imply that $f(0) = 0$ and hence

$$f(0+h) + f(0-h) - 2f(0) = 0.$$

This clearly implies the result.

The Example 4.2 shows that smoothness at a point is not sufficient to guarantee continuity at that point. It is also true that continuity is not sufficient to imply smoothness. To see this, consider

Example 4.3. Let $\mathcal{L}$ be a linear space and for each $x \in \mathcal{L}$, let

$$f(x) = \| x \|,$$

where $\|\cdot\|$ is any norm defined on $\mathcal{L}$. Then f is clearly continuous at $x = 0$, but for $h \in \mathcal{L}$ with $h \neq 0$,

$$\frac{f(0+h) + f(0-h) - 2f(0)}{\|h\|} = \frac{\|h\| + \|-h\|}{\|h\|} = 2$$

and hence f is not smooth at $x = 0$.

The norm functional f in Example 4.3 is clearly uniformly continuous and convex on $\mathcal{L}$ and furthermore, f satisfies a Lipschitz condition on $\mathcal{L}$.

Theorem 4.3. Let f be defined as in Example 4.3. Then f is not Fréchet differentiable at $x = 0$.

Proof: By Example 4.3, f is not smooth at $x = 0$ so that f is not Fréchet differentiable at $x = 0$, in view of Theorem 4.1.

However using Theorem 3.7, it follows that the norm functional f has a one-sided Gâteaux variation at each $x \in \mathcal{L}$, including $x = 0$.

If, however, the domain of f is further restricted to be a Hilbert space, then it is easy to show that a much stronger result follows.

Theorem 4.4. Let H be a Hilbert space with inner product $\langle \cdot , \cdot \rangle$. Then let f be the particular norm functional which is induced by $\langle \cdot , \cdot \rangle$; that is,

$$f(x) = \langle x, x \rangle^{\frac{1}{2}} = \| x \|$$

for each $x \in H$. Then f is Fréchet differentiable at each $x \in H$ where $x \neq 0$.

Proof: The proof follows easily from Definition 1.5 and Theorem 1.9 and will not be given here.

Thus far, Theorem 4.1 provides the only sufficient condition for smoothness. It is natural, therefore, to seek another sufficient condition which is not as strong as Fréchet differentiability. The following theorem, motivated partially by ideas presented in Theorem 1.4, provides this "weaker" sufficient condition.

Theorem 4.5. Let $x_0 \in X$. If

$$\| f(x_0+th) + f(x_0-th) - 2f(x_0) \| = o(|t|) \quad \text{as} \quad t \to 0$$

uniformly for all h with $\|h\| = 1$, then f is smooth at x_0.

Proof: The proof is almost identical to that used for Theorem 1.4, and will be omitted.

Definition 4.2. Let $H \subset X$ and suppose that $f: H \to Y$. Then f is said to be uniformly smooth on H if (1) holds uniformly for all $x \in H$.

Clearly if f is uniformly smooth on H, then f is smooth at each point of H. The converse is false.

Example 4.5: Let

$$f(x) = \begin{cases} \frac{1}{x} & (x \neq 0), \\ 0 & (x = 0). \end{cases}$$

If $x \neq 0$, then f is differentiable at x and hence smooth at x. Furthermore by Theorem 4.2, f is also smooth at $x = 0$. Thus f is smooth on E_1. Suppose now that

$$\lim_{h \to 0} \frac{f(x+h) + f(x-h) - 2f(x)}{h} = 0$$

uniformly for all $x \neq 0$. It is not necessary to consider $x = 0$ since the quotient above is automatically 0 in that case, and no limiting process is involved. Then for $\varepsilon = 2$, there exists a $\delta > 0$ such that if $0 < |h| < \delta$, then

$$\left| \frac{f(x+h) + f(x-h) - 2f(x)}{h} \right| < 2$$

for all $x \neq 0$.

But

$$\frac{f(x+h) + f(x-h) - 2f(x)}{h} = \frac{\frac{1}{x+h} + \frac{1}{x-h} - \frac{2}{x}}{h} = \frac{2h}{x(x^2-h^2)}$$

and hence if $0 < |h| < \delta$, then

$$\left| \frac{h}{x(x^2-h^2)} \right| < 1$$

for all real $x \neq 0$. But this is impossible since the choice $x = h = \delta/2$ violates the inequality. If, however, x is held fixed, it is easy to choose a δ small enough to avoid the trouble at $x = h$.

Theorem 4.6. Let $H \subset X$. If f is uniformly smooth on H, then for each fixed $z \in X$,

$$\| f(x+tz) + f(x-tz) - 2f(x) \| = o(|t|) \quad \text{as} \quad t \to 0$$

uniformly for all $x \in H$.

Proof: The theorem is an obvious consequence of Definition 4.2.

Corollary 4.7. Let $x_0 \in X$ and suppose that f is uniformly smooth on B_r where

$$B_r = \{y \in X \mid \|y-x_0\| < r\}. \tag{4}$$

Let u,v be two fixed elements of X, with $u \neq 0$. Then

$$\|f(x_0+tu+tv) + f(x_0+tu-tv) - 2f(x_0+tu)\| = o(|t|) \quad \text{as} \quad t \to 0.$$

Theorem 4.8. Let $x_0 \in X$ and suppose that f is uniformly smooth on B_r, where B_r is defined as in (4). Then for each pair of fixed elements $h_1, h_2 \in X$,

$$\| \Delta^2_{th_1,th_2} f(x_0) \| = o(| t |) \quad \text{as} \quad t \to 0.$$

Proof: Apply Corollary 4.7 with

$$u = \tfrac{1}{2}(h_1+h_2)$$

and $v = u$. Then

$$\|f(x_0+th_1+th_2) + f(x_0) - 2f(x_0+\tfrac{1}{2}th_1+\tfrac{1}{2}th_2)\| = o(| t |)$$

$$\text{as} \quad t \to 0. \quad (5)$$

Then apply Corollary 4.7 again but with

$$u = \tfrac{1}{2}(h_1+h_2)$$

and

$$v = \tfrac{1}{2}(h_2-h_1).$$

Then

$$\| f(x_0+th_2) + f(x_0+th_1) - 2f(x_0+\tfrac{1}{2}th_1+\tfrac{1}{2}th_2) \| = o(| t |)$$

$$\text{as} \quad t \to 0. \ (6)$$

Now combining (5) and (6), it follows that

$$\| f(x_0+th_1+th_2) - f(x_0+th_1) - f(x_0+th_2) + f(x_0) \| = o(|t|) \quad \text{as} \quad t \to 0$$

and hence

$$\| \Delta^2_{th_1,th_2} f(x_0) \| = o(| t |) \quad \text{as} \quad t \to 0.$$

We can now combine Theorem 4.8 with Theorem 1.3 to the following theorem:

Theorem 4.9. Let $x_0 \in A$ and suppose that f has a Gâteaux variation $Vf[x_0; \cdot]$ at x_0. Then $Vf[x_0; \cdot]$ is a Gâteaux derivative if the following two conditions are satisfied:

i. f satisfies a weak Lipschitz condition at x_0

ii. f is uniformly smooth in some neighborhood of x_0.

The remainder of this chapter deals primarily with known results in the area of differentiability and smoothness of real-valued functions of a single real variable.

Theorem 4.10. (See [37].) Let f be a real-valued function continuous and smooth on the interval [a,b], where a and b are not necessarily finite. Then f is differentiable on an everywhere dense subset of [a,b].

Proof: Let $x_0 \in (a,b)$. If f has a relative extremum at x_0, then there exists a number $\delta > 0$ such that

$$f(x_0+h) - f(x_0) \quad \text{and} \quad f(x_0-h) - f(x_0)$$

have the same sign if $0 < h < \delta$. But f is smooth at x_0 and hence

$$\frac{f(x_0+h) - f(x_0)}{h} + \frac{f(x_0-h) - f(x_0)}{h} = \frac{f(x_0+h) + f(x_0-h) - 2f(x_0)}{h}$$

approaches 0 as $h \to 0$. Thus

$$\lim_{h\to 0^+} \frac{f(x_0+h) - f(x_0)}{h} = 0 \tag{7}$$

and

$$\lim_{h\to 0^+} \frac{f(x_0-h) - f(x_0)}{h} = 0 \tag{8}$$

since both difference quotients have the same sign and their sum has limit 0. Clearly (7) and (8) imply that $f'(x_0)$ exists and furthermore $f'(x_0) = 0$.

The preceding is really a useful lemma in itself. It shows that if f is continuous and smooth on [a,b], then $f'(x_0) = 0$ for each relative extremum $x_0 \in (a,b)$.

Now let [a',b'] be a closed subinterval of [a,b] and let L be the unique linear mapping defined on [a',b'] such that

$$L(a') = f(a') \quad \text{and} \quad L(b') = f(b').$$

Define

$$G(x) \triangleq f(x) - L(x).$$

Then clearly G is smooth on $[a',b']$. Furthermore

$$G(a') = G(b') = 0$$

and hence since G is continuous on $[a',b']$, G must have a relative extremum at some point x_0 where $a' < x_0 < b'$. Thus by the first part of the proof,

$$G'(x_0) = 0.$$

But since $G(x) = f(x) - L(x)$ and $L(x)$ is known to be differentiable for each x, it follows that $f'(x_0)$ exists. Thus for each subinterval $[a',b']$ of (a,b), there is some point x_0 in (a',b') such that $f'(x_0)$ exists. The totality of such points x_0 clearly forms an everywhere dense subset of $[a,b]$. This completes the proof.

The interval $[a',b']$ in the previous theorem could very well be $[a,b]$ itself. In that event, an easy computation shows that

$$L'(x) = \frac{f(b) - f(a)}{b - a} \quad \text{for each} \quad x \in [a,b].$$

Then Theorem 4.10 guarantees that there is some point x_0, $a < x_0 < b$, such that

$$G'(x_0) = 0$$

and hence

$$f'(x_0) = L'(x_0) = \frac{f(b) - f(a)}{b - a}.$$

Thus the mean value theorem holds for f on [a,b].

Theorem 4.11. Let f be a continuous convex function defined on the interval (a,b) and let $x_0 \in (a,b)$. If f is smooth at x_0 and if f has both finite left- and right-hand derivatives at x_0, then f is differentiable at x_0.

Proof: Smoothness implies that

$$\lim_{h\to 0^+} \frac{f(x_0+h) + f(x_0-h) - 2f(x_0)}{h} = 0.$$

Thus

$$\lim_{h\to 0^+} \left\{\frac{f(x_0+h) - f(x_0)}{h} + \frac{f(x_0-h) - f(x_0)}{h}\right\} = 0. \qquad (9)$$

But the hypotheses imply that

$$\lim_{h\to 0^+} \frac{f(x_0+h) - f(x_0)}{h} \triangleq f'(x_0^+)$$

exists. Furthermore

$$\lim_{h\to 0^+} \frac{f(x_0-h) - f(x_0)}{h} = -\lim_{k\to 0^-} \frac{f(x_0+k) - f(x_0)}{k} = -f'(x_0^-)$$

exists. Hence from (9) it follows that

$$f'\left(x_0^+\right) - f'\left(x_0^-\right) = 0.$$

The conclusion is now obvious.

Corollary 4.12. If f is continuous, convex, and smooth on (a,b), then f is differentiable on (a,b).

The reader may refer to [32] and [37] for more extensive results on differentiability of smooth functions.

APPENDIX

THEOREMS CITED IN THE TEXT

Theorem 1. (See [30, p. 233, Theorem 21].) Let (X,A,μ) and (Y,B,ν) be two complete measure spaces and f an integrable function on $X \times Y$. Then for almost all $x \in X$, the function f_x defined by

$$f_x(y) = f(x,y)$$

is an integrable function on Y and for almost all $y \in Y$, the function f_y defined by

$$f_y(x) = f(x,y)$$

is an integrable function on X.

Theorem 2. (See [6, p. 597, Theorem 4].) Let f be a continuous finite-valued function defined on the interval J. Define

$$\limsup_{h \to 0^+} \frac{f(x+h) - f(x)}{h} \triangleq \overline{D}_+ f(x).$$

If $\overline{D}_+ f(x)$ is finite on J, except possibly for a countable subset of J, and if $\overline{D}_+ f(x)$ is summable over J, then f is absolutely continuous on J.

Theorem 3. (See [6, p. 628, Theorem 3].) Let G be a nonempty subset of E_{n+m} and let $P \in G$. Since

$$E_{n+m} = E_n \times E_m,$$

there exists $P_n \in E_n$ and $P_m \in E_m$ such that

$$P = (P_n, P_m).$$

Now for $P_n \in E_n$, define

$$B(P_n) \triangleq \{P_m \in E_m \mid P = (P_n, P_m) \in G\}.$$

If $B(P_n)$ is a null set in E_m for all P_n with the possible exception of a null set in E_n, then G is either nonmeasurable or a null set in E_{n+m}.

BIBLIOGRAPHY

1. Anselone, P. M., editor, *Nonlinear Integral Equations*, Madison, Wis.: United States Army Mathematics Research Center Publication No. 11, April, 1963.

2. Apostol, T. M., *Mathematical Analysis*, Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1957.

3. Bartle, R. G., *The Elements of Real Analysis*, New York, N. Y.: John Wiley and Sons, Inc., 1964.

4. Boas, R. P., Jr., "More About Quotients of Monotone Functions," *American Mathematical Monthly*, Vol. 72, No. 1, January 1965, p. 59.

5. Brøndsted, A. and R. T. Rockafellar, "On the Subdifferentiability of Convex Functions," *Proceedings of the American Mathematical Society*, Vol. 16, 1965, pp. 605-611.

6. Carathéodory, C., *Vorlesungen Uber Reele Funktionen*, New York, N. Y.: Chelsea Publishing Co., 1927.

7. Day, M. M., *Normed Linear Spaces*, Berlin: Springer-Verlag, 1958.

8. Dieudonne, J., *Foundations of Modern Analysis*, New York, N. Y.: Academic Press, 1960.

9. Ehrmann, H. H., "On Implicit Function Theorems and the Existence of Solutions of Nonlinear Equations," United States Army Mathematics Research Center, Technical Summary Report No. 343, August, 1962, p. 27.

10. Eggleston, H. G., *Convexity*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 47, Cambridge University Press, 1958.

11. Esser, M. and O. Shisha, "A Modified Differentiation," *American Mathematical Monthly*, Vol. 71, No. 10, October, 1964, pp. 904-06.

12. Fleming, W. H., *Functions of Several Variables*, Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1965.

13. Graves, L. M., *The Theory of Functions of Real Variables*, New York, N. Y.: McGraw-Hill Book Co., Inc., 1956.

14. ——————, "Topics in Functional Calculus," *Bulletin of the American Mathematical Society*, Vol. 41, 1935, pp. 641-662.

15. Halmos, P. R., *Measure Theory*, New York, N. Y.: Van Nostrand Co., Inc., 1950.

16. Hildebrandt, T. H. and L. M. Graves, "Implicit Functions in General Analysis," *Transactions of the American Mathematical Society*, Vol. 29, 1927, pp. 127-153.

17. Hille, E., *Analysis*, Vol. 1, New York, N. Y.: Blaisdell Publishing Co., 1964.

18. Hille, E. and R. S. Phillips, *Functional Analysis* and *Semigroups*, Providence, R. I.: American Mathematical Society, 1957.

19. Hyers, D. H., "Linear Topological Spaces," *Bulletin of the American Mathematical Society*, Vol. 51, 1945, pp. 1-21.

20. Kantorovich, L. V. and G. P. Akilov, *Functional Analysis in Normed Spaces*, New York, N. Y.: The Macmillan Co., 1964.

21. Krasnosel'skii, M. A. and Ia. B. Rutitskii, *Convex Functions and Orlicz Spaces*, New York, N. Y.: Gordon and Breach Publishers, Inc., 1961.

22. Lang, S., *Introduction to Differentiable Manifolds*, New York, N. Y.: John Wiley and Sons, Inc., 1962.

23. Leach, E. B., "A Note on Inverse Function Theorems," *Proceedings of the American Mathematical Society*, Vol. 12, 1961, pp. 694-697.

24. Lusternik, L. A. and V. J. Sobolev, *Elements of Functional Analysis*, New York, N. Y.: Gordon and Breach Publishers, Inc., 1961.

25. Moreau, J. J., "Étude Locale d'une Fonctionelle Convexe," Faculté des Sciences de Montpellier, Séminaires de Mathématiquis, 1963.

26. Natanson, I. P., *Theory of Functions of a Real Variable*, translated by Leo F. Boron, New York, N. Y.: Ungar Publishing Co., 1955.

27. Rademacher, H. "Partielle und Totale Differenzierbarkeit von Funktionen Mehrerer Variabeln und uber die Transformation der Doppelintegrale," *Mathematische Annalen*, Vol. 79, 1919, pp. 340-359.

28. Rado, T. and P. V. Reichelderfer, *Continuous Transformations in Analysis*, Berlin: Springer-Verlag, 1955.

29. Riesz, F. and B. Sz-Nagy, *Functional Analysis*, translated by Leo F. Boron, New York, N. Y.: Ungar Publishing Co., 1955.

30. Royden, H. L., *Real Analysis*, New York, N. Y.: Macmillan Co., 1963.

31. Saks, S., *Theory of the Integral*, Warsaw: 1937.

32. Stein, E. M. and A. Zygmund, "Smoothness and Differentiability of Functions," *Ann. Univ. Sci. Budapest, Eötvös Sect. Math.*, No. 3-4, 1960-61, pp. 295-307.

33. Vainberg, M. M., *Variational Methods for the Study of Nonlinear Operators*, translated by Amiel Feinstein, San Francisco, Calif.: Holden-Day, Inc., 1964.

34. Valentine, F. A., *Convex Sets*, New York, N. Y.: McGraw-Hill Book Co., Inc., 1964.

35. Zorn, M., "Derivatives and Fréchet Differentials," *Bulletin of the American Mathematical Society*, Vol. 52 1946, pp. 133-137.

36. Zygmund, A., *Trigonometric Series*, Vol. 1, Cambridge University Press, 1959.

37. ———, "Smooth Functions," *Duke Math Journal*, Vol. 12, 1945, pp. 47-76.