THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DIFFERENTIAL CALCULUS IN HILBERT SPACES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

ANTHONY EDWARD LABARRE, JR.

Norman, Oklahoma

1957

DIFFERENTIAL CALCULUS IN HILBERT SPACES

APPROVED BY

John B. Giever

Casper Goffman

[illegible]

T. K. Pan

J. J. Zinnes

DISSERTATION COMMITTEE

ACKNOWLEDGMENT

I wish to thank the Mathematics Department of the University of Oklahoma, in general, for its splendid efforts in my mathematical training. In particular, I want to thank Professors J. Giever and A. A. Grau for their encouragement and the valuable suggestions which they made during the preparation of this work; also to Professor C. Goffman, I want to express my indebtedness for his enthusiastic, clear, excellently-delivered lectures whose influence on my mathematical thinking is deep-rooted. Finally, it seems entirely appropriate to acknowledge the patience and kind understanding of my wife, Gloria, who has made the entire project worth-while.

TABLE OF CONTENTS

DIFFERENTIAL CALCULUS IN HILBERT SPACES

INTRODUCTION

The purpose of this work is to introduce, from a geometric point of view, a notion of differentiability for a function whose domain and range lie in Hilbert spaces, and to investigate some of the consequences of this definition.

Chapter I is purely background material and is included in order that the attention of the reader may be directed towards those results which play an important part in the theory developed. None of the material contained in this chapter is new. Proofs of the several theorems stated there will be found in the work of either von Neumann [1] or Riesz and Sz.-Nagy [2]. Of special importance to the later work are the notions of a projection, the Cartesian product of two Hilbert spaces, and the graph of a function.

In Chapter II a definition of differentiability for a function with domain and range contained in Hilbert spaces is offered and some equivalent forms of the definition are found. We give some examples of functions having the property required for differentiability and see that all of the important linear operators are differentiable in the introduced sense. In addition, all of the "intrinsic functions" of Hilbert space are differentiable. We then show that the "tangent", defined as the intersection of certain closed linear manifolds associated with a differentiable function at a point, is non-trivial in the presence of a very mild restriction on the domain of the function. We next show that the property of a function of being differentiable is preserved under both translation and unitary transformations of its graph space. We close the chapter by borrowing an idea of von Neumann's: we calculate the components of the projection $P_{G(A)} < x,y >$ of a vector $< x,y >$ in the Cartesian product $X \times Y$ on the graph $G(A)$ of a closed linear operator in terms of the operator A and its adjoint A^*.

The calculation of $P_{G(A)}$ is first applied in Chapter III to the important case of a functional and leads to a theorem having to deal with the approximability of a functional by a bounded linear functional. Then follows a close parallel of the elementary formulas of calculus dealing with the derivative of a sum, difference, product, and quotient with a calculus of functionals. We then show that for spaces with a real inner product the non-linear functional $||x||$ is differentiable.

The emphasis is next shifted to general functions. Using the investigation in the case of a functional as a guide, we obtain a theorem having to deal with the equivalence of the notion of differentiability, on the one hand, and the linear approximability of a function, on the other hand (a generalization of our earlier theorem concerned with functionals). This theorem is then used to show that differentiability is preserved under addition and to obtain an analog of the elementary theorem relative to the differentiability of a "composite function".

In Chapter IV we obtain conditions under which our concept of differentiability coincides with that due to Frechet, and draw a brief comparison of the two concepts.

DIFFERENTIAL CALCULUS IN HILBERT SPACES

CHAPTER I

DEFINITION AND BASIC PROPERTIES OF HILBERT SPACES; NOTATION

The Representation Theorem

A great deal of work on Hilbert spaces has been done since the importance of the concept to integral equation theory was first indicated by D. Hilbert in 1912 in [3]. Since that time, steady development of Hilbert spaces and their theory has greatly influenced other branches of analysis and physics (particularly quantum mechanics); indeed, some of the results have found such frequent application that we would be justified in labeling them "classical". Nevertheless, for the sake of completeness those theorems which we need are stated precisely. Unless otherwise indicated, proofs of these theorems can be found in [2]. In this section we recall the definition of Hilbert space, follow with some of the elementary consequences of this definition, and close with a representation theorem which emphasizes the transparency of Hilbert spaces.

A Hilbert space H is a linear space over the complex number field with a notion of an inner product which enables us to interpret H as a complete metric space. This interpretation is made possible by defining the distance $||x - y||$ between the pair of vectors x and y of H as follows:

$$||x - y|| = \sqrt{(x - y, x - y)}.$$

(We note the use of (x,y) for the inner product of x and y and recall that a metric (distance function) is required to have the properties: (i) $||x - y|| = ||y - x|| \geq 0$, (ii) $||x - y|| = 0$ is equivalent to $x = y$, (iii) $||x - y|| + ||y - z|| \geq ||x - z||$). In particular, we call the distance between x and 0, written $||x||$, the norm of x. Thus H becomes a topological space (even a topological group) and we define $\lim x_n = x$ for a sequence of vectors x_n and a vector x in H to mean that $\lim ||x_n - x|| = 0$. Continuity of a function f(x) on H with range in another such space is defined in the usual manner.

Some definitions and simple properties of H are now recalled for later reference; for convenience in referring to the theorems later, we number them.

We call a pair of vectors x,y orthogonal if $(x,y) = 0$. We say that the vector x is normalized if $||x|| = 1$. A set A of vectors of H is called orthogonal if every two distinct elements of A are orthogonal. A is normalized if each of its members is normalized. A is orthonormal if it is both orthogonal and normalized; A is complete in H if it is orthonormal and is not a proper subset of any orthonormal set in H.

I.1. The Pythagorean Theorem:

$$(x,y) = 0 \text{ implies } ||x + y||^2 = ||x||^2 + ||y||^2.$$

I.2. Bessel's Inequality:

If $\beta_1, \beta_2, \ldots, \beta_n$ is an orthonormal set of vectors of H, then $\sum_1^n |(x,\beta_j)|^2 \leq ||x||^2$ for every x in H. More generally, if A is any orthonormal set in H, then $(x,\beta) = 0$ for all β in A except at most a countable subset and $\sum_{\beta \varepsilon A} |(x,\beta)|^2$ not only has sense but converges to a number $\leq ||x||^2$.

I.3. Parseval's Theorem:

If x_β, $\beta \varepsilon A$, is complete, then $\sum_{\beta \varepsilon A} |(x,x_\beta)|^2 = ||x||^2$.

I.4. Riesz-Fischer Theorem:

If $\beta_1, \beta_2, \ldots$ is an orthonormal sequence in H, a necessary and sufficient condition that $\sum a_i \beta_i$ be convergent is that $\sum |a_i|^2$ be convergent.

With the usual definitions of Cauchy sequence, complete space, and separable space the following results:

I.6. Completion of Unitary Space:

A linear space S in which an inner product is defined (i.e., a unitary space) can be completed to a space T such that S is dense in T. The completion is unique (to within isomorphism) and the character of S with regard to separability and cardinality of a complete orthonormal set is preserved under the completion.

Let I be an arbitrary set of indices β. The space of all complex-valued functions (x_β), with domain I, $\beta \varepsilon I$, such that $x_\beta \neq 0$ for only a countable set of β's and $\sum |x_\beta|^2$ is finite, and in which the operations ax, x + y, and (x,y) are defined by

$$a(x_\beta) = (ax_\beta),$$
$$(x_\beta) + (y_\beta) = (x_\beta + y_\beta),$$
$$((x_\beta),(y_\beta)) = \sum_{\beta \varepsilon I} x_\beta \overline{y_\beta},$$

is called a space of type H_I.

I.7. The Representation Theorem:

Every Hilbert space H is isomorphic with a space of type H_I for some I.

In view of the result I.6. above, in the sequel we shall consider only complete spaces, i.e., Hilbert spaces. Except where explicitly stated, we do not assume separability. We use the letter H to denote a Hilbert space.

Projections

The notion of a projection in a Hilbert space is an abstraction of the familiar concept of a projection as found in Euclidean geometry; it is essential to our concept of differentiability introduced later. In this section we define a projection and state some of the well-known theorems relating to projections which we shall subsequently use.

We recall that a subset M of H is a linear manifold if it is closed under addition and multiplication by scalars. If a linear manifold is closed (in the topological sense, as a subset of H), we call M a closed linear manifold (c.l.m.).

I.8. The Unique Decomposition Theorem:

If M is a c.l.m. in H, every element $x \varepsilon H$ has a unique decomposition $x = x_1 + x_2$, where $x_1 \varepsilon M$ and $(x_1,x_2) = 0$.

We call x_1 the <u>projection of x on M</u> and we write $P_M x = x_1$. Thus a projection on a c.l.m. M is a function P_M with domain $D(P_M)$ a Hilbert space and range $R(P_M)$ a c.l.m.. Alternatively, we have the following criteria for a projection:

I.9. Projection Criteria:

A function E is a projection P_M if and only if:

(i) E is single-valued, linear, with $D(E) = H$,

(ii) $(Ex,y) = (x,Ey)$ for all x and y in H,

and (iii) $E^2 = E$, where E^2x means $E(Ex)$.

It is easy to prove and useful to know the following facts:

I.10. If $E = P_M$, then M is the set of all solutions of the equation $Ex = x$. $Ex = x$ as well as $\|Ex\| = \|x\|$ is characteristic for $x \varepsilon M$. $0 \leq (Ex,x) \leq \|x\|^2$ for all $x \varepsilon H$.

The set of all projections is a subset of a class of operators which plays an important role in the development of the theory of linear operators on Hilbert space. This is the class defined by the property of symmetry (property (ii) in I.9. above). We say that the linear operator A is <u>bounded</u> if there exists a constant C such that $\|Ax\| \leq C \|x\|$ for all $x \varepsilon D(A)$; the infimum of the set of all such C's is the <u>norm</u> $\|A\|$ of A. A bounded linear operator A is called <u>symmetric</u> if $(Ax,y) = (x,Ay)$ for all x and y in D(A).

I.11. Hellinger-Toeplitz Theorem:

Every linear symmetric operator defined on a Hilbert space is bounded.

In case the linear operator A is not bounded, we can still introduce the notion of a symmetric operator. We proceed as follows: if A is a linear operator whose domain is dense in H and if y is an element of H for which there exists an element y* of H such that $(Ax,y) = (x,y^*)$ for all $x \varepsilon D(A)$, it can be shown that y* is then uniquely determined by y. This enables us to define an operator A* by writing $A^*y = y^*$. A* is called the <u>adjoint</u> of A. We say that a linear operator A with domain dense in H is <u>symmetric</u> if A* is an extension of A (A* is an extension of A means that D(A) is a subset of D(A*) and A and A* coincide on D(A)). When A = A* we say that A is <u>self-adjoint</u>. It is for this class of operators that success was achieved in arriving at a "spectral decomposition".

<u>Construction of the Cartesian Product of two Hilbert Spaces</u>

The construction of what is known as the Cartesian product of two Hilbert spaces will play a fundamental role in the motivation and development of the theory to follow; it is suggested by the familiar construction of the Euclidean plane. Indeed, much of the terminology and notation has been lifted from elementary analytic geometry and calculus.

We first recall the construction of the Cartesian product of two Hilbert spaces. Let X and Y be two Hilbert spaces (not necessarily isomorphic). By the <u>Cartesian product</u> H = X x Y is meant the set of all elements of the form $< x,y >$ with $x \varepsilon X$, $y \varepsilon Y$ and the following definitions:

(1) $< x_1,y_1 > + < x_2,y_2 > = < x_1 + x_2, y_1 + y_2 >$,

(2) $a< x,y > = < ax, ay >$ for all complex numbers a,

(3) $(< x_1,y_1 >,< x_2,y_2>) = (x_1,x_2) + (y_1,y_2)$.

It is easy to show that with these definitions X x Y is a Hilbert space. Since the mapping which takes $x \varepsilon X$ into $< x,0 > \varepsilon H$ is obviously an isomorphism, we are justified in writing $X = [< x,0 >]$ and shall do so whenever convenient. Of course, we may also write $Y = [< 0,y >]$.

Another simple consequence of the construction of X x Y is the orthogonality of X and Y. More precisely, if $< x,y >$ is orthogonal to X, then $(< x_1,0 >,< x,y >) = 0$ or $(x_1,x) + (0,y) = 0$ for all $x_1 \varepsilon X$ which implies that $x = 0$ and hence $< x,y > \varepsilon Y$. This amounts to saying that the orthogonal complement of X, denoted by $\ominus X$, is contained in Y. On the other hand, $(< x,0 >,< 0,y >) = 0$ and therefore $\ominus X = Y$.

We shall also avail ourselves of the notion of the graph of a function. Let f(x) be a function having domain $D(f) \leq X$ and range R(f) contained in Y where X and Y are Hilbert spaces. By the <u>graph</u> G(f) of f we mean the set of all elements $< x,f(x) >$ where $x \varepsilon D(f)$. Thus the graph $G(f) \leq X \times Y$.

The notion of the graph of a function and the following results whose proofs can be constructed conveniently in terms of this notion are due to J. von Neumann [4]. Proofs of the next three theorems can be found in [2].

I.12. If A is a linear operator, the existence of A^{-1}, A^*, and $(A^{-1})^*$ imply the existence of $(A^*)^{-1}$ and $(A^*)^{-1} = (A^{-1})^*$.

I.13. If the closed linear operator A (this means that the graph G(A) is closed) has a domain which is dense in X, then the domain of A* is also dense in X and A** = (A*)* exists; in addition A** = A.

I.14. The Closed Graph Theorem:

If the closed linear operator A is defined everywhere in H, then A is bounded.

Functionals

Functions defined on Hilbert spaces but taking on only real or complex numbers as values are especially important; they are called functionals. Those functionals which are at the same time bounded have a particularly simple structure, as is indicated in the following result:

I.15. Representation of a Bounded Linear Functional:

If f(x) is a bounded linear functional on H, there exists a unique vector $u \varepsilon H$ such that $f(x) = (x,u)$ for all $x \varepsilon H$. Moreover, $|| f || = || u ||$.

We shall need later the fact that for linear functionals the notions of boundedness and closedness are equivalent. That boundedness implies closedness is clear since boundedness implies continuity which in turn implies closedness. On the other hand, suppose that the linear functional f(x) is not bounded. Then there exists a sequence $x_n \varepsilon D(f)$ with $|| x_n || = 1$ and $| f(x_n) |$ increasing without bound. Consider the sequence $y_n = x_n / | f(x_n) |$. Here we have $y_n \rightarrow 0$ whereas we may assume that $f(y_n) \rightarrow 1$ since there is a subsequence of y_n for which this is true. Thus $< 0,1 >$ is a limit point of the graph of f(x) but it does not belong to the graph since $f(x)$, being a linear functional, contains the point $< 0,0 >$. Consequently, $f(x)$ is not closed. This completes the proof of the statement.

I.16. For linear functionals the notions of boundedness and closedness are equivalent.

CHAPTER II

THE NOTION OF DIFFERENTIABILITY

Motivation and Definition

We have seen that every closed linear manifold (c.l.m.) M contained in a Hilbert space H affords a means of uniquely decomposing an arbitrary vector x in H as follows:

$$x = x_1 + x_2,$$

where $x_1 \varepsilon M$ and $x_2 \varepsilon \Theta M$, the orthogonal complement of M. The uniqueness of the decomposition enables us to define a single-valued operator $P_M x$, called the projection of x on M, by

$$P_M x = x_1 .$$

Now a closed linear manifold must contain the additive identity and consequently cannot play the role of the "tangent line to the curve f(x) at $x = x_0$" or the tangent plane to the surface f(x,y) at (x_0,y_0), for the tangent line or tangent plane need not contain the origin. It appears, therefore, that we shall require slightly more general geometric entities in order to subsume even these simples situations. The difficulty is superficial, however, and is readily overcome. Thus the notion of a "general line" (i.e., not not necessarily through the origin of coordinates) and a general plane finds its analog in the nature of a closed affine subspace. We observe that M, being a c.l.m. in H, and therefore a closed subgroup of the topological group H, induces a partition of H into cosets. The closed affine subspace

$$M + x^* = [m + x^* \mid m \varepsilon M]$$

(i.e., the set of all points of the form $m + x^*$ with m variable but in M and $x^* \varepsilon H$ fixed) is simply that coset of M which contains x^*. Geometrically, we may visualize the closed affine subspace $M + x^*$ as the "tangent line" for a suitable x^*. Actually, we need not adhere strictly to the geometric situation and we will, in effect, use M as the "tangent line", and for suitable x^* the affine subspace $M + x^*$ would be the actual "tangent".

Our definition of differentiability is motivated by the following simple geometrical observation: for a function of a single real variable f(x), to say that f(x) is differentiable at x_0 is equivalent to saying that there is a straight line M ("the tangent line") through the point $< x_0, f(x_0) >$ with the property that for any sequence x_n converging to x_0, the sequence of ratios of the lengths of the projections on M of the secant-chords joining the points $P_n = < x_n, f(x_n) >$ and $P_0 = < x_0, f(x_0) >$ to the lengths of the secant-chords $\overline{P_n P_0}$ themselves (i.e., the cosine of the angle between the segments $\overline{P_n P_0}$ and $P_M(\overline{P_n P_0})$) tends to 1. It is easy to establish this fact and we therefore omit the proof. The corresponding observation for the case of a function of two real variables f(x,y) may be carried out in the obvious manner.

There is yet another point which we must consider before offering a definition: it is suggested by the situation which occurs when we are concerned with the tangent line to a curve in three-dimensional Euclidean space E_3. When the given curve does have a tangent line at a given point, there is only one such. On the other hand, any plane containing this line is a proper c.l.m. (i.e., not all of E_3) for which the ratio considered in the preceding paragraph tends to 1. Note, however, that the tangent line is determined by the intersection of all the planes described.

Guided by the heuristic considerations above, it is suggested that we take the following supporting conditions for the notion of differentiability:

Let $f(x)$ be a single-valued function with domain $D(f) \leq X$, a Hilbert space, and range $R(f) \leq Y$, a Hilbert space. Let $G(f)$ be the graph of f, i.e., the set of all vectors of the form $< x, f(x) > \varepsilon X \times Y = H$, and suppose that $< x_0, f(x_0) > \varepsilon G(f)$ and that x_0 is a limit point of $D(f)$. We then make the following

DEFINITION II.1. <u>$f(x)$ is differentiable at x_0</u> means that there exists a c.l.m. $M \neq H$ such that if $x_n \to x_0$, $x_n \varepsilon D(f)$, then

$$\frac{(< x_n - x_0, f(x_n) - f(x_0) > , P_M < x_n - x_0, f(x_n) - f(x_0) >)}{|| < x_n - x_0, f(x_n) - f(x_0) > || \cdot || P_M < x_n - x_0, f(x_n) - f(x_0) > ||} \to 1.$$

The intersection T of all such c.l.m.'s is called the <u>tangent of f at x_0</u>.

It is clear from the definition that M need not be unique whereas T must be.

Since P_M is a projection, $(x, P_M x) = || P_M x ||^2$, and so we have the equivalent

DEFINITION II.2. <u>f(x) is differentiable at x_0</u> means that there exists a c.l.m. $M \neq H$, such that if $x_n \to x_0$, $x_n \varepsilon D(f)$, then

$$\frac{\| P_M < x_n - x_0, f(x_n) - f(x_0) > \|}{\| < x_n - x_0, f(x_n) - f(x_0) > \|} \to 1.$$

In order to shorten the notation, we shall frequently write X_n instead of $x_n - x_0$ and Y_n instead of $f(x_n) - f(x_0)$ and say $X_n \to 0$ implies

$$\frac{\| P_M < X_n, Y_n > \|}{\| < X_n, Y_n > \|} \to 1.$$

Since

$$\frac{\| P_M v \|}{\| v \|} = \left\| P_M \frac{v}{\| v \|} \right\|,$$

we conclude that our definition of differentiability can be put in the form:

DEFINITION II.3. <u>f(x) is differentiable at x_0</u> means that there exists a c.l.m. $M \neq H$, such that if $X_n \to 0$, $x_n \varepsilon D(f)$, then

$$\| P_M u_n \| \to 1,$$

where u_n denotes the unit vector in the direction of $< X_n, Y_n >$, i.e., the vector

$$\frac{< X_n, Y_n >}{\| < X_n, Y_n > \|}.$$

Because $\| P_M u_n \| \leq \| u_n \| = 1$, we have the slight variant of the above:

DEFINITION II.4. <u>f(x) is differentiable at x_0</u> means that there exists a c.l.m. $M \neq H$, such that every sequence $x_n \to x_0$ contains a subsequence $x_{N(n)} = x_N \to x_0$ for which $|| P_M u_N ||$ increases monotonically to 1.

In the sequel, when we speak of a subsequence x_K of the sequence x_k, it will be understood that, by definition, $x_K = x_{K(k)} = x_{K_k}$.

Associated with every c.l.m. M is its orthogonal complement $\ominus M$. Our notion of differentiability may be framed in terms of $\ominus M$ instead of M. To do this, we first note that $P_{\ominus M} u_k = (1 - P_M) u_k$ (where 1 denotes the identity mapping) and hence $|| P_M u_k || \to 1$ implies

$$|| P_{\ominus M} u_k ||^2 = || u_k - P_M u_k ||^2$$

$$= || u_k ||^2 - (u_k, P_M u_k) - (P_M u_k, u_k) + || P_M u_k ||^2$$

$$= || u_k ||^2 - || P_M u_k ||^2 \to 0.$$

Conversely, if we have a c.l.m. N such that $|| P_N u_k || \to 0$, then $|| P_{\ominus N} u_k || \to 1$. Therefore, we may write

DEFINITION II.5. <u>f(x) is differentiable at x_0</u> means that there exists a c.l.m. $N \neq H$, such that if $X_n \to 0$, $x_n \in D(f)$, then $|| P_N u_k || \to 0$, where u_k is the unit vector in the direction of $< X_k, Y_k >$.

In addition, because $|| P_N u_k || \geq 0$, we have

DEFINITION II.6. <u>f(x) is differentiable at x_0</u> means that there exists a c.l.m. $N \neq H$, such that every sequence $X_n \to 0$, $x_n \in D(f)$, contains a subsequence $X_K \to 0$ such that $|| P_N u_k ||$ decreases monotonically to 0.

Our definition is essentially geometric in the sense that we have required the existence of a certain proper c.l.m. in the space of the graph G(f) of f in order that f have the property of being differentiable. Now, if one defines (as in [1]) a linear operator to be a function whose graph is a linear manifold and calls such an operator closed if its graph is closed, then there will be a one-to-one correspondence between closed linear manifolds in X x Y and closed linear operators with domain in X, range in Y. But all such linear operators will not be single-valued. As a matter of fact, it is not difficult to prove that the single-valued linear operators A(x) are those with the property that A(0) has the unique value 0. In order to avoid ambiguity, however, we will use the term linear operator to imply single-valuedness. We shall find it convenient later on to limit ourselves to such operators and for these our definition of differentiability specializes to

DEFINITION II.7. f(x) is differentiable at x_0 means that there exists a closed linear operator A(x) (single-valued) such that every sequence $x_k \to x_0$ contains a subsequence $x_K \to x_0$ such that either $|| P_{G(A)} u_K ||$ converges monotonically increasing to 1 or $|| P_{G(A)} u_K ||$ converges monotonically decreasing to 0, where u_K is the unit vector in the direction of $< X_K, Y_K >$.

Examples of Differentiable Functions

It is clear that every closed linear operator is differentiable at every limit point of its domain; also every bounded linear operator is differentiable.

By the intrinsic functions of Hilbert space, we mean those needed in the postulational characterization of Hilbert space, namely: addition, x + y; scalar multiplication, ax; inner product, (x,y). We now consider the differentiability of these functions.

Let $f < x,y > = x + y$. Then f is a function on $X \times X$ with values in X; its graph is a subset of X^3. It is easy to see that this graph is a c.l.m. and therefore $x + y$ is a differentiable function of $< x,y >$.

Next write $f < a,x > = ax$. The graph of f is then a subset of $C \times X \times X$, where C denotes the complex numbers. If we fix either the scalar a or the vector x, the resulting graph $G(f) = [< a,x,ax >]$ is easily seen (by virtue of the continuity of f in both variables) to be a c.l.m.. We therefore conclude that f is a differentiable function of each of its variables.

A similar argument proves that $f < x,y > = (x,y)$ is, for any one of x,y arbitrary but fixed, a differentiable function of the other. Note, however, that as in the case of the function immediately above, this is not a linear function of the pair $< x,y >$. Let us summarize these results by stating, somewhat inexactly, but nevertheless suggestively:

THEOREM II.1. The intrinsic functions of Hilbert space are differentiable.

As another class of differentiable operators, we mention the symmetric operators. To see this we take, for a given symmetric operator $A(x)$, the operator $A^*(x)$ as the required linear operator (recall that A^* is, for all A, closed). In particular, every self-adjoint operator is differentiable. Furthermore, from the fact that every closed linear operator A with a dense domain gives rise to the self-adjoint operator A^*A, we conclude that if A is closed, linear, with dense domain, then A^*A is differentiable. For example, since the operator $Ax(q) = ix'(q)$ with $x(0) = x(1) = 0$, defined in $L^2(0,1)$ for all absolutely continuous functions $x(q)$ with $x'(q) \in L^2(0,1)$ is linear, closed, and has a dense domain ([2]: it is also symmetric), it follows that it is differentiable. Furthermore, $A^*Ax(q) = -x''(q)$ is self-adjoint and hence differentiable. In this connection, we mention the interesting fact that $A^*y(q) = iy'(q)$ is defined for all $y(q)$ which are absolutely continuous with $y'(q) \in L^2$ and hence is an extension of A; if A is modified to A_c by requiring that $x(1) = cx(0)$ for some constant c, then it turns out that $A_c^* = A_c$, i.e., A_c is self-adjoint and therefore differentiable.

Besides the operator just considered, another important operator to quantum mechanics (see[5]) is defined for the set of all functions $x(q)$ with finite

$$\int_{-\infty}^{\infty} |x(q)|^2 dq \quad \text{and} \quad \int_{-\infty}^{\infty} q^2 |x(q)|^2 dq.$$

It is the operator

$$Ax = q \cdot x(q).$$

This operator is closed, and therefore differentiable. Thus we see that the main operators of quantum mechanics are differentiable (actually, in lieu of our $Ax(q) = ix'(q)$ discussed above, in quantum mechanics we are concerned with the related operator

$$Dx(q) = \frac{h}{2\pi i} x'(q)$$

defined for all $x(q)$ with finite

$$\int_{-\infty}^{\infty} |x(q)|^2 dq \quad \text{and} \quad \int_{-\infty}^{\infty} |x'(q)|^2 dq;$$

this operator D is self-adjoint).

Another class of differentiable functions are the functions of a self-adjoint operator, as defined by von Neumann [4] and Stone [6]. It can be shown [2] that for every real-valued continuous function w of the real variable q, and for every self-adjoint operator A,

$$w(A) = \int_{-\infty}^{\infty} w(q)\,dE_q,$$

in the sense of convergence in the norm of sums of the Stieltjes type, with E_q the spectral family associated with A. We may therefore write

$$(w(A)x,y) = \int_{-\infty}^{\infty} w(q)\,d(E_q x,y)$$

in the ordinary Stieltjes sense. This last relation may be taken as the definition of <u>the functions w of A</u>. Now, for separable Hilbert space, these functions w of A are characterized by the properties of being

1) closed and linear with domain dense in a separable Hilbert space, and

2) such that every bounded symmetric operator which commutes with A also commutes with w(A).

In view of property 1) every function of the self-adjoint linear operator A on a separable Hilbert space is differentiable.

The Tangent T

We have called the intersection of all c.l.m.'s M associated with the differentiability of f(x) at x_0, the tangent of f at x_0. Since the tangent T is the intersection of closed sets, it too must be closed. It is not empty, for certainly it contains the additive identity. We now raise the question as to the non-triviality of the tangent (i.e., we want to know when T is neither the whole space nor the c.l.m. whose sole element is 0).

Suppose that s_n is a sequence of graph-secants at x_0 which converges to $< 0,0 >$ (this means that $s_n = < X_n, Y_n > \to < 0,0 >$ where we use the notation introduced immediately after Definition II.2), and furthermore, that $\lim u_n$ exists (u_n is the unit vector in the direction of $< X_n, Y_n >$, as usual). If M is any c.l.m. such that $X_n \to 0$ implies $|| P_M u_n || \to 1$, then M must contain $\lim u_n$ if it exists; this can be proved by recalling that $|| P_M v || = || v ||$ is characteristic for $v \in M$, and then noting that continuity and linearity of P_M and continuity of $|| v ||$ yield $|| P_M(\lim u_n) || = || \lim P_M u_n || = \lim || P_M u_n || = 1 = \lim || u_n || =$ $|| \lim u_n ||$. Thus the tangent of f at x_0 - when it exists - is a non-empty c.l.m. which contains all elements of the form $\lim u_n$. Since every vector of this form is clearly a unit vector, the tangent will be a non-trivial c.l.m. whenever there exists a convergent sequence u_n. We now digress to prove a theorem which we shall need shortly.

A vector is determined by a magnitude and a direction. We ask: Is there, in Hilbert space, an analog for this statement? More precisely, suppose that it is known that the sequence of norms $|| x_n ||$ converges, and that the sequence of vectors x_n converges weakly to x: $(x_n, y) \to (x, y)$ for all y (this last condition is our precise rendering of a "direction"). Does it follow that x_n converges (strongly)? Indeed it does, and to the vector x, for if we take in the statement of the weak convergence y = x, we have $(x_n, x) \to (x, x)$, from which we obtain

$$(x, x_n) = \overline{(x_n, x)} \to \overline{(x, x)} = (x, x).$$

Hence

$$\lim ||x_n - x||^2 = \lim [\, -(x_n,x) - (x,x_n) + (x_n,x_n) + (x,x)\,],$$
$$= \lim [\, ||x_n||^2 - 2R(x_n,x) + ||x||^2\,],$$
$$= \lim [\, ||x_n||^2 - ||x||^2\,].$$

But since for any complete orthonormal sequence β_i we have $(x_n,\beta_i) \to (x,\beta_i)$ for all i, it follows that

$$||x_n||^2 = \sum_j |(x_n,\beta_j)|^2 \to \sum_j |(x,\beta_j)|^2 = ||x||^2.$$

Thus

$$\lim ||x_n - x||^2 = 0.$$

This proves

THEOREM II.2. If the sequence of norms $||x_n||$ converges and if x_n converges weakly to x, then x_n converges (strongly) to x.

One of our equivalent forms of the definition of differentiability (Definition II.4) is to the effect that from every sequence $x_n \to x_0$ one can extract a subsequence x_N such that the sequence of norms $||P_M u_N||$ converges: $\lim ||P_M u_N|| = 1$. Applying the above theorem, we obtain

THEOREM II.3. If f(x) is differentiable at x_0 and if u_N and M are a corresponding sequence of unit graph-secants and a c.l.m., respectively, such that $||P_M u_N||$ converges monotonically increasing to 1, and if u_N converges weakly, then u_N converges strongly (since $\lim ||u_N|| = 1$) and hence the tangent to f at x_0 is non-trivial.

If we restrict ourselves to functions whose domains contain one-dimensional linear manifolds, then we may omit the hypothesis as to weak convergence of u_n. More precisely, we select an arbitrary but fixed vector $h \neq 0$, and let t_n denote a sequence of real numbers with limit 0, and we assume that $f(x)$ is defined at $x_0 + t_n h$. Then

$$u_n = \frac{< t_n h,\ f(x_0 + t_n h) - f(x_0) >}{\| < t_n h,\ f(x_0 + t_n h) - f(x_0) > \|} .$$

It can be shown [9] that existence of the derivative of $f(x)$ (i.e., differentiability of $f(x)$) implies existence of the directional derivative:

$$\lim_{t_n \to 0} \frac{f(x_0 + t_n h) - f(x_0)}{t_n}$$

and so it is clear that whenever the domain of $f(x)$ contains a sequence of points of the form $x_0 + t_n h$, then the tangent is non-trivial. In order to pave the way for stating our result in simple form, we call a sequence of points of the form $x_0 + t_n h$ a sequence of <u>collinear</u> <u>points</u>. Then we have

THEOREM II.4. Every function $f(x)$ defined on a sequence of collinear points $x_0 + t_n h$ which converges to x_0 and differentiable there has a non-trivial tangent.

The condition that the function be defined on a sequence of collinear points is not a necessary condition for the non-triviality of the tangent as the simple example of the twisted cubic (i.e., the curve in Euclidean 3-space defined by the function $f< t,t^2 > = t^3$) shows.

<u>Invariance of Differentiability under Translation and Unitary Transformations</u>

We now raise the question: If $f(x)$ is differentiable at x_0 and a translation T of the graph G(f) of f is carried out, is the function whose graph is the set of all points $T < x,f(x) > = < x + a,f(x) + b >$ (where a,b are fixed vectors in X,Y, respectively) differentiable at $x_0 + a$?

Let us first determine how the function whose graph is $T < x,f(x) > = < x + a,f(x) + b >$ is defined. We denote this function by F(w). Now since $T < x,f(x) > = < x + a,f(x) + b >$ is in the graph G(F), it follows that $F(x+a) = f(x) + b$ or $F(w) = f(w - a) + b$. Consequently, the domain of F is $D(F) = D(f) + a$ and the range of F is $R(F) = R(f) + b$. In particular, since $x_0 + a$, with $x_0 \varepsilon D(f)$, is an element of D(F), it makes sense to inquire about the differentiability of F at $x_0 + a$. As a matter of fact, let $x_n' \to x_0 + a$; then

$$\frac{|| P_M < x_n'-(x_0+a),\ F(x_n')-F(x_0+a) > ||}{||< x_n'-(x_0+a),\ F(x_n')-F(x_0+a) > ||} = \frac{|| P_M < x_n'-a-x_0,\ f(x_n'-a)-f(x_0) > ||}{||< x_n'-a-x_0,\ f(x_n'-a)-f(x_0) > ||}$$

which has limit 1 by virtue of the differentiability of f(x) at x_0. This proves

THEOREM II.5. If f(x) is differentiable at $x = x_0$ and if $< a,b >$ is a fixed vector, then the function $F(x) = f(x-a) + b$ with $D(F) = D(f) + a$ and range $R(F) = R(f) + b$, if differentiable at $x = x_0 + a$. In short: Differentiability is preserved under translation.

As a special case, we take $a = -x_0$, $b = -f(x_0)$, and get the

<u>Corollary</u>: If f(x) is differentiable at x_0, then $F(x) = f(x+x_0)-f(x_0)$ is differentiable at $x = 0$. (We note also that $F(0) = 0$).

For any two vectors u and v of the Hilbert space H there is a topological mapping t of H onto itself which carries u into v, namely $t(x) = v - x + u$. This means that as far as topological properties of H are concerned, any two vectors are equivalent. Thus, for example, to prove that a Hilbert space is not locally compact at any point, it is sufficient to show that the identity 0 has no compact neighborhoods. This remark, together with the theorem above and its corollary, allow us to conclude that in studying the topological properties of differentiability of a function at a point, we may replace the function by one which is differentiable at the origin and study its topological properties. In particular, in studying continuity properties of a differentiable function, we may work with a function differentiable at the origin.

We now turn to the question of the behavior of a differentiable function under a unitary transformation of the space of its graph. More precisely, the question at hand is the following: If $f(x)$ is differentiable at x_0 and the unitary transformation U on X x Y is carried out (a unitary transformation is a transformation of H onto itself which preserves addition, scalar multiplication, and inner products), is the function whose graph is the set of all points $U < x, f(x) >$ differentiable at $U < x_0, f(x_0) >$?

By hypothesis then, there exists a c.l.m. M such that if $X_n \to 0$, then

$$\frac{|| P_M < X_n, Y_n > ||}{|| < X_n, Y_n > ||} \to 1.$$

Let $< X_n, Y_n > = P_M < X_n, Y_n > + P_{\ominus M} < X_n, Y_n >$. Then $U < X_n, Y_n > = UP_M < X_n, Y_n > +$

$UP_{\ominus M} < X_n, Y_n >$, by linearity of U. $(P_M v, P_{\ominus M} v) = 0$ implies $(UP_M < X_n, Y_n >$, $UP_{\ominus M} < X_n, Y_n >) = 0$, so that $P_{UM} U < X_n, Y_n > = UP_M < X_n, Y_n > \varepsilon U(M)$ (we observe that we are justified in speaking of P_{UM} since M is a c.l.m. and U preserves linearity and closedness). Therefore,

$$\frac{|| P_{UM} U < X_n, Y_n > ||}{|| U < X_n, Y_n > ||} = \frac{|| UP_M < X_n, Y_n > ||}{|| U < X_n, Y_n > ||} = \frac{|| P_M < X_n, Y_n > ||}{|| < X_n, Y_n > ||} ,$$

since U is norm-preserving. This proves

THEOREM II.6. If f(x) is differentiable at x_0 and U is a unitary transformation of X x Y, then the function whose graph is the set of all vectors $U < x, f(x) >$ is differentiable at $U < x_0, f(x_0) >$. In short: Differentiability is preserved under a unitary transformation.

<u>Representation of the Projection</u> $P_{G(A)}$

In this section we obtain a formula for computing the projection $P_{G(A)} < x,y >$ of $< x,y >$ on the graph G(A) of a linear operator A; later (Chapter III) we shall apply this to functionals and then use the results for functionals in investigating differentiable functions.

Let A(x) be a closed linear operator with domain D(A) dense in X, range R(A) contained in Y, and hence graph $G(A) \leq X \times Y$. If we define $U < x,y > = < -y,x >$, then because the adjoint A*(y) has the characterizing property

$$(Ax,y) = (x,A^*y) \text{ for all } x,$$

or

$$(< x,Ax >, U<y,A^*y>) = 0,$$

we see that G(A) and UG(A*) are complementary orthogonal subspaces. We therefore have the unique decompositions:

(1) $\langle x,0\rangle = \langle x',Ax'\rangle + \langle -A^*y',y'\rangle \; ; \; \langle 0,y\rangle = \langle x'',Ax''\rangle + \langle -A^*y'',y''\rangle$,

valid for arbitrary $x \varepsilon X$, $y \varepsilon Y$. These imply unique solutions of the corresponding systems of equations:

$$x = x' - A^*y' \qquad\qquad 0 = x'' - A^*y''$$
$$0 = Ax' + y' \qquad\qquad y = Ax'' + y''.$$

If we define

$$P_{11}x = x' \qquad\qquad P_{12}y = x''$$
$$\text{and}$$
$$P_{21}x = y' \qquad\qquad P_{22}y = y'',$$

it is clear that all the P_{ij} are linear operators, and the system of equations above can be written:

$$1 = P_{11} - A^*P_{21} \qquad\qquad 0 = P_{12} - A^*P_{22}$$
$$0 = AP_{11} + P_{21} \qquad\qquad 1 = AP_{12} + P_{22} ,$$

where 1 and 0 denote the identity and zero transformations,respectively. Returning to our decomposition (1) above, we have, because of orthogonality of the terms in the right members:

$$||x||^2 = ||\langle x,0\rangle||^2 = ||x'||^2 + ||Ax'||^2 + ||A^*y'||^2 + ||y'||^2$$
$$= ||P_{11}x||^2 + ||AP_{11}x||^2 + ||A^*P_{21}x||^2 + ||P_{21}x||^2,$$

and

$$||y||^2 = ||\langle 0,y\rangle||^2 = ||x''||^2 + ||Ax''||^2 + ||A^*y''||^2 + ||y''||^2$$
$$= ||P_{12}y||^2 + ||AP_{12}y||^2 + ||A^*P_{22}y||^2 + ||P_{22}y||^2,$$

from which we conclude that all of P_{11}, AP_{11}, A_{21}, A^*P_{21}, P_{12}, AP_{12}, P_{22}, A^*P_{22} are bounded by unity.

If we add the decompositions (1) and take into account the linearity of A and A*, we get for arbitrary $\langle x,y\rangle \varepsilon X \times Y$:

$$< x,y > = < x' + x'', A(x' + x'') > + < -A^*(y' + y''), y' + y'' > ;$$

it follows that

$$P_{G(A)} < x,y > = < P_{11}x + P_{12}y,\ A(P_{11}x + P_{12}y) >$$

and

$$P_{\ominus G(A)} < x,y > = < -A^*(P_{21}x + P_{22}y),\ P_{21}x + P_{22}y > .$$

We collect these results in a

THEOREM II.7. If we define

$$P_{11} = (1 + A^*A)^{-1},\ P_{21} = -AP_{11},\ P_{22} = (1 + AA^*)^{-1},\ P_{12} = A^*P_{22},$$

then the following relations hold:

$$P_{G(A)} < x,y > = < P_{11}x + P_{12}y,\ A(P_{11}x + P_{12}y) >$$

$$P_{\ominus G(A)} < x,y > = < -A^*(P_{21}x + P_{22}y),\ P_{21}x + P_{22}y > ,$$

where $A(x)$ is a closed linear operator with domain dense in X and range contained in Y. Furthermore, $|| P_{ij} || \leq 1$ for $i,j = 1,2$.

CHAPTER III

SOME PROPERTIES OF DIFFERENTIABLE FUNCTIONS

Throughout this chapter, differentiability shall be understood in the sense of Definition II.7. That is, when we say that a function is differentiable at x_0 we shall mean that there exists a closed linear operator A(x) (single-valued) such that every sequence $x_k \to x_0$ contains a subsequence $x_K \to x_0$ such that $|| P_{G(A)} u_K ||$ is monotonic increasing with limit 1, where u_K is the unit vector in the direction of $< X_K, Y_K >$. The chapter is divided into two parts: in part one we consider the notion of differentiability as applied to functionals, and in part two we study some properties of functions with domain and range both in arbitrary Hilbert spaces.

Part One: Functionals

Application of the Notion of Differentiability to a Functional

Let **A** be a bounded linear functional:

$$Ax = y, \ x \varepsilon H, \ y \text{ a complex number.}$$

According to the theorem on the representation of a bounded linear functional (I.14), there exists a unique vector u ε H such that

$$Ax = (x,u) \text{ for all } x \varepsilon H.$$

We now determine A*. Since its defining property is $(Ax,z) = (x,A^*z)$, it is evident that A* must operate on complex numbers and produce abstract vectors:

$$A^*z = v.$$

Then since the inner product of two complex numbers $(z,w) = z\bar{w}$, we easily calculate

$$A^*z = zu.$$

We next compute $P_{11} = (A^*A+1)^{-1}$. P_{11} operates on vectors and produces vectors, say

$$(A^*A+1)^{-1}x = y,$$

$$x = (y,u)u + y.$$

In order to solve this equation for y, we form the inner product of both members with u and find that

$$(y,u) = \frac{(x,u)}{1 + ||u||^2}.$$

Therefore,

$$y = x - \frac{(x,u)u}{1 + ||u||^2}.$$

Consequently

$$(A^*A+1)^{-1}x = x - \frac{(x,u)u}{1 + ||u||^2}.$$

Next we find that

$$P_{12}z = A^*(AA^*+1)^{-1}z = \frac{zu}{1 + ||u||^2}.$$

Evidently if we define $m = 1 + ||u||^2$ and $\Delta = (x,u) - z$, we may write in accordance with Theorem II.7:

$$P_{G(A)} < x,z > = \frac{1}{m} < mx-u\Delta,\ \Delta+mz >,$$

and

$$P_{\Theta G(A)} < x,z > = \frac{\Delta}{m} < u,-1 >.$$

An easy calculation shows that the inner product of $P_{G(A)} < x,z >$ and $P_{\Theta G(A)} < x,z >$ vanishes, thereby justifying the notation. It is easy to see that

$$\frac{||P_{G(A)} < x,z >||^2}{||< x,z >||^2} = 1 - \frac{\Delta\bar{\Delta}}{m(||x||^2 + |z|^2)}.$$

A necessary and sufficient condition that this ratio tend to 1 is that

$$\frac{\Delta\bar{\Delta}}{||x||^2 + |z|^2} \to 0.$$

This shows that a functional $f(x) = z$ is differentiable at $x = x_0$ if and only if there exists a bounded linear functional $A(x) = (x,u)$ such that $X_n \to 0$ implies

$$\frac{|(X_n,u) - Z_n|^2}{||X_n||^2 + |Z_n|^2} \to 0,$$

where we have put $Z_n = f(x_n) - f(x_0)$. If we divide numerator and denominator by $||X_n||$, we obtain

$$\frac{\left|\frac{(X_n,u)}{||X_n||} - \frac{Z_n}{||X_n||}\right|^2}{1 + \frac{|Z_n|^2}{||X_n||^2}} \to 0.$$

By the Schwarz inequality, $|(X_n,u)| \leq ||u|| \cdot ||X_n||$. Now consider the sequence $|Z_n|/||X_n||$. It is bounded; for otherwise there would be a subsequence of this sequence such that the ratio above would tend to 1 instead of 0. Consequently,

$$\frac{|(X_n,u) - Z_n|}{||X_n||} \to 0.$$

With this, we have proved a generalization of an elementary theorem hav- to deal with the characterization of the tangent plane to a surface. We state this result as

THEOREM III.1. A functional $f(x) = z$ is differentiable at $x = x_0$ if and only if there exists a bounded linear functional $L(x) = (x,u)$ such that the difference $[L(x) - L(x_0)] - [f(x) - f(x_0)]$ for $x \to x_0$ is a quantity of higher order than the distance $||x - x_0||$ between x and x_0.

The Calculus of Functionals

We have just seen that a functional $f(x) = z$ is differentiable at $x = x_0$ if and only if there exists a bounded linear functional $A(x) = (x,u)$ such that $X_n \to 0$ implies

$$\frac{|(X_n,u) - Z_n|}{||X_n||} \to 0,$$

where $X_n = x_n - x_0$, $Z_n = f(x_n) - f(x_0)$. The role of u suggests that it be called the <u>derivative of $f(x)$ at x_0</u> and that we write $f'(x_0) = u$. Now consider the bounded linear functional

$$\begin{aligned}(A_1 + A_2)x &= A_1x + A_2x,\\ &= (x,u_1) + (x,u_2),\\ &= (x, u_1 + u_2).\end{aligned}$$

We calculate

$$|(X_n, u_1 + u_2) - (Z_{1n} + Z_{2n})| = |(X_n,u_1) - Z_{1n} + (X_n,u_2) - Z_{2n}|$$
$$\leq |(X_n,u_1) - Z_{1n}| + |(X_n,u_2) - Z_{2n}|,$$

and each of these terms divided by $||X_n||$ tends to 0 with X_n. This proves

THEOREM III.2. If the functionals $f_1(x)$ and $f_2(x)$ are differentiable at $x = x_0$, then so is their sum $(f_1 + f_2)x = f_1(x) + f_2(x)$. Furthermore, if $f_1'(x_0) = u_1$ and $f_2'(x_0) = u_2$, then $(f_1 + f_2)'(x_0) = f_1'(x_0) + f_2'(x_0) = u_1 + u_2$.

As an immediate consequence, we have the

Corollary: If the functionals $f_1(x)$ and $f_2(x)$ are differentiable at x_0, then so is their difference and the operations of differentiation and subtraction may be interchanged:

$$(f_1 - f_2)'(x_0) = f_1'(x_0) - f_2'(x_0) = u_1 - u_2.$$

We now consider the functional $P(x) = f_1(x)\cdot f_2(x)$, and ask if differentiability of the factors implies differentiability of the product. As suggested by the classical theorem of which this may serve as a generalization, we consider

$$\overline{f_1(x_0)}u_2 + \overline{f_2(x_0)}u_1$$

as a candidate for the role of u in Theorem III.1 (the expression above is considered instead of $f_1(x_0)u_2 + f_2(x_0)u_1$ because of the property of the inner product: $(x,ay) = \overline{a}(x,y)$; it is understood that the u_i here have the same meaning as in the above theorem). Then differentiability of $P(x)$ is intimately connected with the behavior, as $x_n \to x_0$, of

$$|(x_n - x_0, \overline{f_1(x_0)}u_2 + \overline{f_2(x_0)}u_1) - f_1(x_n)f_2(x_n) + f_1(x_0)f_2(x_0)| =$$
$$|f_1(x_0)(x_n - x_0, u_2) - f_1(x_0)f_2(x_n) + f_1(x_0)f_2(x_0) + f_1(x_0)f_2(x_n) +$$

$f_2(x_0)f_1(x_0) + f_2(x_0)(x_n - x_0, u_1) - f_2(x_0)f_1(x_n) + f_2(x_0)f_1(x_n) -$

$f_2(x_0)f_1(x_0) - f_1(x_n)f_2(x_n) | \leq$

$|f_1(x_0)| \cdot |(x_n - x_0, u_2) - f_2(x_n) + f_2(x_0)| + |f_2(x_0)| \cdot |(x_n - x_0, u_1) -$

$f_1(x_n) + f_1(x_0)| + |f_1(x_0)| \cdot |f_2(x_n) - f_2(x_0)| + |f_1(x_n)| \cdot$

$|f_2(x_0) - f_2(x_n)|.$

Now, by virtue of the differentiability of $f_1(x)$ and $f_2(x)$ at x_0, the first two terms divided by $||x_n - x_0||$ tend to 0 as $x_n \to x_0$. Furthermore, if we assume that one of the $|f_i(x_n) - f_i(x_0)|$ is $o(||x_n - x_0||)$ (by which we mean that the ratio $|f_i(x_n) - f_i(x_0)| / ||x_n - x_0|| \to 0$ with $x_n - x_0 \to 0$) and the other is bounded (i.e., the other $f_i(x)$ is bounded), then the last two terms divided by $||x_n - x_0||$ also tend to 0 as $x_n \to x_0$. The explanation of the permissibility of flexibility in the assumptions lies in the fact that the last two terms in the inequality could have been written: $|f_2(x_0)| \cdot |f_1(x_n) - f_1(x_0)| + |f_2(x_n)| \cdot |f_1(x_0) - f_1(x_n)|$. This proves

THEOREM III.3. If the functionals $f_1(x)$ and $f_2(x)$ are differentiable at x_0, if $f_1'(x_0) = u_1$ and $f_2'(x_0) = u_2$, and if one of $|f_i(x_n) - f_i(x_0)|$ is $o(||x_n - x_0||)$ while the other $f_i(x)$ is bounded, then $P(x) = f_1(x) \cdot f_2(x)$ is differentiable at x_o with $P'(x_0) = \overline{f_1(x_0)}\, u_2 + \overline{f_2(x_0)}\, u_1$.

We now turn to the analog of the classical theorem concerned with differentiability of a quotient. Some aid is afforded by the following

Lemma. If $|f(x_n) - f(x_0)|$ is $o(||x_n - x_0||)$ and if $f(x)$ is differentiable at x_0 with $f(x_0) \neq 0$ and $f'(x_0) = u$, then $1/f(x)$ is differentiable at x_0 with derivative there $= - f'(x_0)/|f(x_0)|^2 = -u/|f(x_0)|^2$.

Proof: We consider

$$|(x_n - x_0, -u/|f(x_0)|^2) - 1/f(x_n) + 1/f(x_0)| =$$

$$|1/|f(x_0)|^2 (x_n - x_0, -u) - f(x_n)/|f(x_0)|^2 + f(x_0)/|f(x_0)|^2 +$$

$$f(x_n)/|f(x_0)|^2 - f(x_0)/|f(x_0)|^2 - 1/f(x_n) + 1/f(x_0)| \leq$$

$$[1/|f(x_0)|^2] \cdot |(x_n - x_0, -u) + f(x_n) - f(x_0)| + [1/|f(x_0)|^2] \cdot$$

$$|-f(x_n) + f(x_0)| + |f(x_n) - f(x_0)| / |f(x_n)f(x_0)| .$$

Since $f(x)$ is differentiable at x_0, the first term divided by $||x_n - x_0||$ tends to 0 , and the second and third terms tend to 0 because of the assumed condition that $|f(x_n) - f(x_0)|$ is $o(||x_n - x_0||)$. This proves the lemma.

We now apply the lemma to prove

THEOREM III.4. If $f(x)$ is differentiable at x_0 with $f(x_0) \neq 0$ and $f_1'(x_0) = u_1$ and if $|f_1(x_n) - f_1(x_0)|$ is $o(||x_n - x_0||)$ and if $f_2(x)$ is differentiable at x_0 and bounded with $f_2'(x_0) = u_2$, then

$$Q'(x_0) = \frac{f_1(x_0)u_2 - \overline{f_2(x_0)}u_1}{|f_1(x_0)|^2}$$

exists where $Q(x) = f_2(x)/f_1(x)$.

Proof: Theorem III.3 yields differentiability of the product $Q(x) = f_2(x)/f_1(x) = f_2(x)\cdot[1/f_1(x)]$. In particular,

$$Q'(x_0) = \overline{f_2(x_0)}[1/f_1(x)]' \big|_{x=x_0} + [1/\overline{f_1(x_0)}]f_2'(x_0),$$

$$= \overline{f_2(x_0)}[-u_1/|f_1(x_0)|^2] + [1/\overline{f_1(x_0)}]u_2 ,$$

$$= \frac{f_1(x_0)u_2 - \overline{f_2(x_0)u_1}}{|f_1(x_0)|^2} .$$

Differentiability of $||x||$

A very important non-linear functional defined on H is $||x||$. We ask if it is differentiable at $x_0 \neq 0$ (an obvious restriction). An examination of this function for Euclidean spaces suggests that we consider the functional

$$A(x) = \frac{(x,x_0)}{||x_0||} = (x,u), \text{ where } u = \frac{x_0}{||x_0||} ,$$

which is obviously linear and bounded. According to Theorem III.1, it is sufficient to prove that $x_n \to x_0$ implies

$$r = \frac{|(x_n,x_0) - ||x_n|| \cdot ||x_0|||}{||x_0|| \cdot ||x_n - x_0||} \to 0.$$

We do this as follows. Since $x_0 = ||x_0||u$, it follows that

$$r^2 = \frac{|(x_n,u) - ||x_n|||^2}{||x_n||^2 - 2||x_0||R(x_n,u) + ||x_0||^2} ,$$

where $R(x_n,u)$ = real part of (x_n,u). Since $x_n \to x_0$, we have $||x_n|| \to ||x_0||$ and $(x_n,y) \to (x_0,y)$ for all y. From the first condition, we may write

$$||x_n|| = ||x_0|| + \sqrt{2||x_0||}\,\Theta$$

where $\Theta \to 0$ as n becomes infinite. Thus

$$r^2 = \frac{|(x_n,u) - ||x_0|| - \sqrt{2||x_0||}\,\Theta|^2}{2||x_0||[-R(x_n,u) + ||x_0|| + \sqrt{2||x_0||}\,\Theta] + 2||x_0||\Theta^2} .$$

Next, using the fact that $(x_n,u) \to (x_0,u) = ||x_0||$, we write $(x_n,u) = ||x_0|| + \beta(n) + i\gamma(n)$ where $\beta, \gamma \to 0$ as n becomes infinite. Then

$$r^2 = \frac{|\beta + i\gamma - \sqrt{2||x_0||}\theta|^2}{2||x_0||[-\beta + \sqrt{2||x_0||}\theta] + 2||x_0||\theta^2} .$$

If we are dealing with a space with a real inner product, we have, on putting $\gamma = 0$ and $\delta = \sqrt{2||x_0||}\theta - \beta$, $r^2 = |\delta|^2/2||x_0||(\delta + \theta^2) \leq |\delta|^2/2||x_0||\delta \to 0$ as n becomes infinite. This proves

THEOREM III.5. In a real Hilbert space, the norm $||x||$ is a differentiable function of x except at the origin.

Part Two: Functions in General

Approximability of Differentiable Functions by Linear Operators

The theorem proved above regarding the differentiability of functionals (Theorem III.1) can be generalized. We have, since $P_M + P_{\ominus M} = 1$ and M and $\ominus M$ are orthogonal,

$$||P_{G(A)}<X_n,Y_n>||^2 + ||P_{\ominus G(A)}<X_n,Y_n>||^2 = ||<X_n,Y_n>||^2$$
$$= ||X_n||^2 + ||Y_n||^2.$$

So

$$\frac{||P_{G(A)}<X_n,Y_n>||^2}{||<X_n,Y_n>||^2} = 1 - \frac{||P_{\ominus G(A)}<X_n,Y_n>||^2}{||X_n||^2 + ||Y_n||^2}$$

$$= 1 - \frac{||\Delta||^2}{||X_n||^2 + ||Y_n||^2} \cdot \frac{||P_{\ominus G(A)}<X_n,Y_n>||^2}{||\Delta||^2}$$

$$= 1 - \frac{||\Delta||^2}{m(||X_n||^2 + ||Y_n||^2)} ,$$

where we have put $X_n = x_n - x_0$, $Y_n = f(x_n) - f(x_0)$, $\Delta = AX_n - Y_n$ and

$$m = \frac{||AX - Y||^2}{||P_{\ominus G(A)} < X,Y > ||^2} .$$

To be precise, m as defined above is a function of both X and Y whereas in the preceding statement in which it occurs, m is used to denote this function evaluated at X_n, Y_n. The fact which concerns us most at present, however, is that 1/m is bounded. To prove this, we first observe that since G(A) is a c.l.m., we can decompose $< X_n, Y_n >$ into its unique representation as an element of G(A) and $\ominus G(A)$, and similarly for $< 0, AX_n - Y_n >$; then by addition and because $< X_n, AX_n >$ belongs to G(A), it follows that

$$P_{\ominus G(A)} < X_n, Y_n > = -P_{\ominus G(A)} < 0, AX_n - Y_n > .$$

Hence we may write

$$m = \frac{||< 0, AX_n - Y_n >||^2}{||P_{\ominus G(A)} < 0, AX_n - Y_n >||^2} .$$

We see immediately that

$$\frac{1}{m} \leq ||P_{\ominus G(A)}||^2 \leq 1,$$

because $||P_{\ominus G(A)}||$ is, by definition, the supremum of the ratio of $||P_{\ominus G(A)} < x,y >||$ to $||< x,y >||$. This proves that the function f(x) is differentiable at x_0 if and only if there exists an operator A with G(A) a c.l.m. such that as $X_n \to 0$,

$$\frac{||\Delta||^2}{||X_n||^2 + ||Y_n||^2} = \frac{||AX_n - Y_n||^2}{||X_n||^2 + ||Y_n||^2} \to 0.$$

If we divide numerator and denominator by $||X_n||^2$, we obtain as numerator

$$\frac{(AX_n - Y_n, AX_n - Y_n)}{||X_n||^2} = \frac{(AX_n, AX_n) - (Y_n, AX_n) - (AX_n, Y_n) + (Y_n, Y_n)}{||X_n||^2}$$

$$= \frac{||AX_n||^2}{||X_n||^2} - \frac{(Y_n, AX_n) + (AX_n, Y_n)}{||X_n||^2} + \frac{||Y_n||^2}{||X_n||^2} .$$

Since $(AX_n, Y_n) + (Y_n, AX_n)$ is real,

$$\frac{(AX_n - Y_n, AX_n - Y_n)}{||X_n||^2} \geq \frac{||Y_n||^2}{||X_n||^2} - \frac{(Y_n, AX_n) + (AX_n, Y_n)}{||X_n||^2}$$

$$\geq \frac{||Y_n||^2}{||X_n||^2} - \frac{2\,||AX_n||\cdot||Y_n||}{||X_n||\cdot||X_n||}$$

(this last inequality following from the Schwarz inequality). If we write s_n instead of $||Y_n||/||X_n||$, then

$$\frac{||AX_n - Y_n||^2}{||X_n||^2 + ||Y_n||^2} \geq \frac{1 - (2\,||AX_n||)/s_n\,||X_n||}{1 + 1/s_n^2} .$$

Let us now assume that A is defined everywhere; then, by the closed graph theorem, A is bounded. Therefore, if for some sequence $X_n \to 0$ the corresponding sequence $||Y_n||/||X_n||$ were unbounded, there would be a subsequence X_N such that the last term of the inequality immediately above would tend to 1, which is impossible. Thus s_n is bounded and

$$\frac{||AX_n - Y_n||}{||X_n||} \to 0.$$

This completes the proof of

THEOREM III.6. The function f(x) is differentiable at $x = x_0$ if and only if there exists a closed linear operator A such that

$$\frac{\| AX_n - Y_n \|}{\| X_n \|^2 + \| Y_n \|^2} \to 0$$

as $X_n \to 0$. If A is defined everywhere, then in addition

$$\frac{\| AX_n - Y_n \|}{\| X_n \|} \to 0.$$

This theorem is the analog of a well-known classical theorem. It specifies the sense in which the linear operator A approximates the differentiable function f(x).

As an easy application of this result, we prove the differentiability of $f\langle a,x \rangle = ax$ considered as a function of two variables. Now $f\langle a,x \rangle$ is defined on C x X, where C denotes the system of complex numbers and X denotes a Hilbert space; its values lie in X. It is not a linear function of both variables simultaneously. In investigating the differentiability of $f\langle a,x \rangle$ at $\langle a_0,x_0 \rangle$, we are led to consider the operator

$$A\langle a,x \rangle = a_0 x + a x_0 ,$$

which is linear and continuous in both variables collectively. We next consider the ratio (for $a_n \to a_0$, $x_n \to x_0$) :

$$\frac{\| A\langle a_n,x_n \rangle - A\langle a_0,x_0 \rangle - a_n x_n + a_0 x_0 \|^2}{\| \langle a_n,x_n \rangle - \langle a_0,x_0 \rangle \|^2} = \frac{| a_n - a_0 |^2 \cdot \| x_n - x_0 \|^2}{| a_n - a_0 |^2 + \| x_n - x_0 \|^2}$$

$$\leq | a_n - a_0 |^2 + \| x_n - x_0 \|^2 \to 0.$$

We therefore conclude that ax is a differentiable function of $< a,x >$.

As another application of this result, we ask if a differentiable function is continuous. Now, if $X_n \to 0$, then for n sufficiently large, $||X_n|| < \varepsilon$ and

$$\frac{||AX_n - Y_n||^2}{||X_n||^2 + ||Y_n||^2} < \varepsilon.$$

Suppose that Y_n does not tend to 0. Then we may assume that $||Y_n||$ is bounded away from 0 (for we could find a subsequence with this property) and

$$\frac{\left|\left|\frac{AX_n}{||Y_n||} - \frac{Y_n}{||Y_n||}\right|\right|^2}{\frac{||X_n||^2}{||Y_n||^2} + 1} < \varepsilon$$

from which it follows that

$$(1) \qquad \frac{||AX_n - Y_n||^2}{||Y_n||^2} < \varepsilon_1$$

where $\varepsilon_1 \to 0$ as n becomes infinite. Now either $||Y_n||$ is bounded above or it isn't. If $||Y_n|| < B$, then

$$||AX_n - Y_n||^2 < B^2\varepsilon_1 ;$$

but $AX_n \to 0$ and so we have a contradiction. On the other hand, if $||Y_n||$ is unbounded, we might as well assume that $||Y_n|| > n$, for a subsequence can be selected with this property. From (1) above

$$\varepsilon_1 > \left|\left|\frac{AX_n}{||Y_n||} - \frac{Y_n}{||Y_n||}\right|\right|^2 \geq \left|\frac{||AX_n||}{||Y_n||} - \frac{||Y_n||}{||Y_n||}\right|^2,$$

but $AX_n \to 0$ and so again we have reached a contradiction. Thus $Y_n = f(x_n) - f(x_0) \to 0$. We have proved

THEOREM III.7. Every differentiable function is continuous.

Differentiability of the Sum of Differentiable Functions

We now prove an analog of the fact that differentiation is a linear operator in the simple case of several real (or complex) variables. Theorem III.6 will be used in realizing this goal.

Suppose that $f(x)$ is differentiable at x_0 and that A is a linear operator defined everywhere which serves to establish this. Then $X_n \to 0$ implies $||AX_n - Y_n|| / ||X_n|| \to 0$. Similarly, suppose that $g(x)$ is differentiable at x_0 with associated linear operator B. Thus $X_n \to 0$ implies $||BX_n - V_n|| / ||X_n|| \to 0$, where $V_n = g(x_n) - g(x_0)$ and the range $R(g) \leq Y$. Now consider the linear manifold $L = [< x, (A+B)x >]$. Let v_n be a Cauchy sequence of elements of L with limit v; put $v_n = < x_n, (A+B)x_n >$. Then each of the sequences x_n and $(A+B)x_n$ is a Cauchy sequence. Let $x_0 = \lim x_n$, $y_0 = \lim (A+B)x_n$. Then $v = < x_0, y_0 >$ and since we are assuming that A and B are defined everywhere (and therefore by the closed graph theorem are continuous), we find that $y_0 = \lim(A+B)x_n = \lim(Ax_n + Bx_n) = A(x_0) + B(x_0)$ and $v \varepsilon L$, meaning that L is a c.l.m.. But

$$\frac{||AX_n - Y_n||}{||X_n||} + \frac{||BX_n - V_n||}{||X_n||} \geq \frac{||(A+B)X_n - (Y_n + V_n)||}{||X_n||}$$

and we therefore have proved

THEOREM III.8. If $f(x)$ and $g(x)$ are differentiable at x_0 with associated linear operators A and B, respectively, and if A and B are defined everywhere, then $f(x) + g(x)$ is differentiable at x_0 with associated linear operator A+B. In short: Differentiability is preserved under addition.

The Composite Function Rule

Let $f(x)$ be differentiable at x_0 with associated linear operator $A(x)$. Suppose that $g(y)$ is differentiable at $y_0 = f(x_0)$ with associated linear operator $B(y)$. If the operators A and B are defined everywhere, then it is easy to verify that the operator BAx is also linear and bounded and its graph is a c.l.m.. By virtue of the differentiability hypothesis, $x_n \to x_0$ implies

$$(1) \qquad \frac{\|A(x_n) - A(x_0) - f(x_n) + f(x_0)\|}{\|x_n - x_0\|} \to 0$$

and $y_n \to y_0$ implies

$$(2) \qquad \frac{\|B(y_n) - B(y_0) - g(y_n) + g(y_0)\|}{\|y_n - y_0\|} \to 0.$$

Suppose that $x_n \to x_0$. Then (1) holds and since B is bounded,

$$(3) \frac{\|B[A(x_n)-A(x_0)-f(x_n)+f(x_0)]\|}{\|x_n - x_0\|} \le \frac{\|B\|\cdot\|A(x_n)-A(x_0)-f(x_n)+f(x_0)\|}{\|x_n - x_0\|} \to 0.$$

Now put $f(x_n) = y_n$. Then $f(x_0) = y_0$ and from (2) we get

$$(4) \qquad \frac{\|Bf(x_n)-Bf(x_0)-gf(x_n)+gf(x_0)\|}{\|f(x_n)-f(x_0)\|} \to 0.$$

If the secant-slopes are bounded (i.e., if $f(x)$ satisfies the Lipschitz condition: $[\ \|f(x_n) - f(x_0)\| / \|x_n - x_0\|\ ] < K$), then from (4) we may write

$$(5) \qquad \frac{\| Bf(x_n)-Bf(x_0)-gf(x_n)+gf(x_0) \|}{\| x_n - x_0 \|} \to 0.$$

Adding (3) and (5) and then using the triangle inequality for the norm, we get

$$\frac{\| BA(x_n)-BA(x_0)-gf(x_n)+gf(x_0) \|}{\| x_n - x_0 \|} \to 0.$$

This completes the proof of

THEOREM III.9. If $f(x)$ is differentiable at x_0 with associated linear operator $A(x)$ defined everywhere and if $f(x)$ satisfies the Lipschitz condition $[\ \| f(x_n)-f(x_0) \| / \| x_n - x_0 \|] < K$, and if $g(y)$ is differentiable at $y_0 = f(x_0)$ with associated linear operator $B(y)$ defined everywhere, then the composite function $gf(x)$ is differentiable at x_0 with associated linear operator $BA(x)$.

CHAPTER IV

CONNECTION WITH DIFFERENTIABILITY IN THE SENSE OF FRECHET

The Definition of Frechet

Perhaps the earliest concept of differentiation to be introduced in Functional Analysis was the Gateaux differential $Df(x,h)$:

$$Df(x,h) = \lim_{t \to 0} \frac{f(x+th) - f(x)}{t} ,$$

which is the same thing as the "variation" of the function f employed in the Calculus of Variations or the "directional differential of f in the direction h". Without concerning ourselves with the question as to the most general setting in which this definition of the Gateaux differential makes sense, let us note that it is meaningful if f is a function whose domain lies in a normed linear space X and whose range lies in another such space Y; note also how the transition of f from a vector-valued function of a vector x to a vector-valued function of a real number t is accomplished. The Gateaux differential is not satisfactory, however. A hint at its shortcomings is furnished by the simple example (see[7] also) of the surface

$$f(x,y) = \frac{xy}{\sqrt{x^2+y^2}}, \text{ if } x^2+y^2 \neq 0,$$

$$f(x,y) = 0, \text{ if } x^2+y^2 = 0,$$

which has the property that it has a directional derivative at the origin in every direction yet it has no tangent plane at the origin, i.e., it cannot be approximated by a linear function in the suitable way in the neighborhood of the origin. More generally, it is easy to see that every function f(x) homogeneous of the first degree and having f(0) = 0 has a Gateaux differential at the origin. This led mathematicians to impose further restrictions on a generalized derivative in an attempt to secure a more classic-approximating theory. In particular, the standard definition of a differential is due to Frechet [8]:

DEFINITION IV.1. A function f(x) on E to E', where E and E' are normed linear spaces, is said to be Frechet differentiable or F-differentiable at the point x_0 with the differential $df(x_0,h)$, if $df(x_0,h)$ is a linear continuous function of h on E to E' such that

$$\frac{\| f(x_0+h) - f(x_0) - df(x_0,h) \|}{\| h \|} \to 0$$

as $h \to 0$. If we set $df(x_0,h) = f'(x_0)h$ where $f'(x_0)$ denotes the linear continuous operator, we call $f'(x_0)$ the derivative of f at x_0.

It is not difficult to see that every function which is Frechet differentiable at a point also has a Gateaux differential at that point in every direction (see [9]). On the other hand, it can be proved [9] that if the Gateaux differential exists in a sphere $\| x - x_0 \| \leq r$ and is uniformly continuous in x and continuous in h, then the Frechet differential df(x,h) exists in this sphere and df(x,h) = Df(x,h). Furthermore, the following desirable properties for a differential [10] are implied by

F-differentiability of f:

(i) continuity of f,

(ii) validity of the composite function rule: if $k(x) = f(g(x))$ and dg and df exist, then dk exists and $dk(x,h) = df(y,dy)$ where $y = g(x)$ and $dy = dg(x,h)$,

(iii) linearity and continuity of $df(x,h)$ in h (this is assumed in the definition above),

(iv) $df(x,h)$ is a first order approximation to the difference $f(x+h) - f(x)$ when h is "close" to 0 (the sense in which this loosely formulated condition is made precise in the case of the Frechet differential is evident).

Comparison with F-Differentiability

In order to investigate the connection between F-differentiability and the present concept, we must generate conditions under which the two definitions may be applied simultaneously. Now normed linear spaces include Hilbert spaces and so the definition of Frechet applies in a more general space. Some normed linear spaces, however, may be considered as being imbedded in Hilbert spaces and to these our definition applies (actually, some of our equivalent forms of the definition for Hilbert space are meaningful in more general spaces). The problem of selecting the Hilbert spaces from among the normed linear spaces was solved by J. von Neumann and P. Jordan in [11]:

THEOREM IV.1. In a normed linear space S, a necessary and sufficient condition that an inner product (x,y) may be defined in such a way that

$||x|| = \sqrt{(x,x)}$, is that the parallelogram law hold:

$$||x + y||^2 + ||x - y||^2 = 2||x||^2 + 2||y||^2.$$

Furthermore, if (x,y) can be defined, this can be done in only one way.

Thus, of all the normed linear spaces, the validity of the parallelogram law characterizes the Hilbert spaces. Now suppose that $f(x)$ is a function defined on a normed linear space X with values in another such Y and that $f(x)$ is F-differentiable at $x_0 \varepsilon X$; in addition, suppose that the parallelogram law holds in X and Y. We raise the question: Is $f(x)$ differentiable in the sense introduced in this paper? Since $f(x)$ is F-differentiable, there exists a linear continuous function $A(x)$ on X to Y such that

$$\lim_{h \to 0} \frac{||f(x_0 + h) - f(x_0) - A(h)||}{||h||} = 0.$$

Then, according to Theorem III.6, $f(x)$ is differentiable in our sense. Conversely, if $f(x)$ is differentiable in the sense of Definition II.8 and if the closed linear operator A is defined everywhere, then on invoking Theorem III.6 again, we conclude that $f(x)$ is F-differentiable provided $f(x)$ is defined everywhere (as required by F-differentiability). [Actually, this requirement that $f(x)$ have all of the space as its domain could be relaxed somewhat]. Summarizing:

THEOREM IV.2. For Hilbert spaces, every function which is differentiable according to Frechet is differentiable in the present sense. Conversely, if a function is differentiable in the present sense, with an associated closed linear operator which is defined everywhere, then that function is also F-differentiable. In short: If we limit ourselves to functions and continuous linear operators defined everywhere, then Frechet differentiability and differentiability in the present sense are equivalent notions.

Our definition of differentiability could be broadly described as being geometric in nature, whereas the definition given by Frechet is an analytic one. The approximability of a differentiable function by a linear operator is a consequence of our definition, whereas it is the essence of the Frechet definition.

Another point of comparison of the two definitions centers around the problem of justifying the casting, in Hilbert spaces, of a theory of differentiability. In particular, in view of the fact that normed linear spaces are more general than Hilbert spaces, what does one have to gain by restricting the theory to Hilbert spaces? In partial answer to this question, we cite the simple example of the definition of a norm for vectors in the plane given by

$$(1) \quad \| \langle x,y \rangle \| = | x | + | y | ;$$

according to the result of von Neumann and Jordan given above (Theorem IV.1), this normed linear space is not a Hilbert space since we know that the definition

$$(2) \quad \| \langle x,y \rangle \|^2 = | x |^2 + | y |^2$$

yields a space in which the parallelogram law is valid and therefore this is the only definition which will give aHilbert space. Now, it is readily verified that the function defined by (1), namely $f(x,y) = | x | + | y |$ has many points of non-differentiability other than the origin, for example, it is not differentiable at points of the form $< 0,y >$. This example, together with Theorem III.5, proves

THEOREM IV.3. For normed linear spaces, it is not true that $||x||$ is a differentiable function of x except at the origin; this statement is true, however, for real Hilbert spaces.

BIBLIOGRAPHY

[1] J. von Neumann, Functional Operators II: The Geometry of Orthogonal Spaces. Princeton: Princeton University Press, 1951.

[2] F. Riesz and B. Sz-Nagy, Functional Analysis, Second Edition. New York: Frederick Ungar Publishing Co.,1955.

[3] D. Hilbert, Grundzüge einer Allgemeinen Theorie der Linearen Integralgleichungen. Leipzig: Teubner-Verlag, 1912.

[4] J. von Neumann, "Über adjungierte Funktionaloperatoren", Annals of Mathematics, 32 (1932), pp. 294-310.

[5] J. von Neumann, Mathematical Foundations of Quantum Mechanics. Princeton: Princeton University Press, 1955.

[6] M. H. Stone, Linear Transformations in Hilbert Space. New York: American Mathematical Society Colloquium Publications, 1932.

[7] R. Courant, Differential and Integral Calculus: Vol. II. New York: Interscience Publishers, Inc., 1937.

[8] M. Frechet, "La Differentielle dans l'Analyse Générale", Annales de l'École Normale Supérieur, 42 (1925), pp. 293-323.

[9] L. Ljusternik and W. Sobolew, Elemente der Funktionalanalysis. Berlin: Akademie-Verlag, 1955.

[10] D. H. Hyers, "Linear Topological Spaces", American Mathematical Society Bulletin, 51 (1945), pp. 1-21.

[11] J. von Neumann and P. Jordan, "On Inner Products in Linear Metric Spaces", Annals of Mathematics, 36 (1935), pp. 719-723.