[3] J. Chmielowski, *Ensembles déterminants pour les fonctions analytiques*, C. R. Acad. Sci. Paris 279 (1974), Sér. A, pp. 639–641.
[4] F. Leja, *Sur les points zéros des fonctions analytiques de plusieurs variables*, Ann. Soc. Polon. Math. 17 (1939), pp. 227–230.
[5] — *Funkcje zespolone*, Warszawa 1967 (J. Siciak, *Wstęp do teorii funkcji analitycznych wielu zmiennych*).
[6] B. Levi, *Sul teorema d'identità per le funzioni analitiche di più variabili*, Boll. Un. Mat. Ital. 13 (1934), pp. 1–5.
[7] — *Sur les ensembles des points qui ne peuvent pas être ensembles de zéros d'une fonction analytique de plusieurs variables*, C. R. Acad. Sci. Paris 198 (1934), pp. 1734–1736.
[8] P. Montel, *Leçons sur les familles des fonctions analytiques et leurs applications*, Paris 1927.
[9] W.E. Osgood, *Lehrbuch der Funktionentheorie II*, 1929.
[10] M. T. Viola, *Sur le théorème d'identité pour les fonctions holomorphes de plusieurs variables*, C. R. Acad. Sci. Paris 198 (1934), pp. 705–707.

INSTYTUT MATEMATYKI UNIWERSYTETU ŚLĄSKIEGO
KATOWICE

Received January 21, 1975 (938)

STUDIA MATHEMATICA, T. LVII. (1976)

Differentiability of Lipschitzian mappings between Banach spaces

by

N. ARONSZAJN (Lawrence, Kan.)

Table of contents

INTRODUCTION

In 1919 H. Rademacher [9] proved the theorem that for any Lipschitzian mapping of an open set $G \subset \boldsymbol{R}^m$ into $\boldsymbol{R}^n$ the differential (Stoltz-differential) exists a.e. in G. This theorem became a standard tool in analysis and it was obviously of interest to extend it to Lipschitzian mappings between Banach spaces. There were, however, two difficulties in obtaining such extension. The most basic and immediately obvious difficulty was the non-existence in Banach spaces (of infinite dimension) of any measure analogous to the Lebesgue measure. Hence the notion of "Almost everywhere" cannot be defined in the usual manner by a measure. The second difficulty refers to the notion of differential. The most commonly used notion of differential is the Fréchet differential. But this notion of differential cannot be used to extend Rademacher's theorem since there exist Lipschitzian mappings (even in Hilbert spaces) without a Fréchet differential anywhere(1). It was only in 1967 that the author was led to define the class $\mathfrak{A}$ of exceptional sets in a separable Banach space which could take the place of sets of Lebesgue measure 0 in finite dimensional spaces. Furthermore, by considering the differentials (or τ-differentials) slightly weaker than the Fréchet differentials, it was possible to extend the Rademacher's theorem to Lipschitzian mappings between two Banach spaces X, Y with some restrictions on the image space Y.

This work has not been published until now. Besides short talks on different occasions, the subject was developed more fully in a series of talks at the Conference on Evolution Equations and Functional Analysis, University of Kansas, June 28–July 18, 1970; and also at Queen's University, Kingston, Ontario, January 5–8, 1971; and lastly in Paris, Université VI, February–April, 1974.

During 1974 the author became acquainted with the recent work of E. Zarantonello [12] which led him to define, in a way similar to the one used to define $\mathfrak{A}$, a much smaller exceptional class $\mathfrak{A}^0$ which for convex functions or their generalizations give differentiability theorems much stronger than the one for Lipschitzian mappings. These considerations are also included in this paper.

The final topic of the present paper arises naturally if one asks under what conditions can we apply a partial integration $\int DT(x; u)\,d\mu(x) = -\int T(x)\,d\nu_u(x)$ to obtain a kind of Stokes-formula(2). We are far, at present, from being able to answer this question in general, even though in many concrete cases we know what the meaning of $d\nu_u(x)$ should be. However, a preliminary requirement to attack this problem is to clarify what the Borel measures μ are for which the integral $\int DT(x; u)\,d\mu(x)$ is well defined for every Lipschitzian mapping T. It is clear that they should be finite measures absolutely continuous rel. $\mathfrak{A}$ (i.e. such that every set $A \in \mathfrak{A}$, where $DT(x; u)$ may not exist, be of μ-measure 0). The investigation of such measures μ is done in the last chapter of the paper.

In 1972 and 1973 two papers appeared by P. Mankiewicz [6] and [7]. The main purpose of these papers was to obtain theorems to the effect that if there exists a Lipschitzian homeomorphism from X into Y, then

(1) See, for instance, Example 1, Section 3, Chapter II.

(2) $DT(x; u)$ is the differential of T at x (linear in u) and $\nu_u(x)$ should have the meaning of the differential of μ in direction u.

there exists a linear isomorphism transforming X into Y. For this purpose it was sufficient to show in the first paper that the differential exists on a dense set. The second paper goes a little further towards extending Rademacher's theorem. On the other hand the work is concerned with mappings between Fréchet spaces requiring an extension of the notion of Lipschitzian mappings to these spaces.

We should mention also a recent work by J. P. R. Christensen [3] with which we became acquainted through a preprint. Christensen does define a class of exceptional sets which he calls zero-sets and by using which he is able to extend Rademacher's theorem. However, these zero-sets form a considerably larger class than our class $\mathfrak{A}$. Hence his result is weaker(3).

We would like to mention here that in February, 1974, F. Mignot gave us a manuscript of his paper (unpublished as yet) where he obtains the result of Mankiewicz in the special case of Hilbert spaces. The interest of his paper lies essentially in the fact that he applies this result to investigate properties of the differential of the projection operator on convex sets in a Hilbert space and applies these properties to the study of variational inequalities.

In Chapter I we define in Section 1 the class $\mathfrak{A}$ and its most elementary properties. In Section 2 we give a review of known facts concerning biorthogonal systems and generalized bases in a form needed for our developments. In Section 3 we give some more intricate properties of the class $\mathfrak{A}$ and in Section 4 we use the construction leading to the class $\mathfrak{A}$ to define other exceptional classes of sets. Among these the smallest non-trivial is the class $\mathfrak{A}^0$, and we show that the theorem of E. Zarantonello, mentioned before, gets a much stronger content by using the class $\mathfrak{A}^0$.

Chapter II deals with differentiability. In Section 1 we start by defining different notions of differentials, give some theorems about differentiability before assuming the Lipschitzian character of the mapping and only at the end of this section do we introduce the Lipschitzian mappings and give some of their elementary properties. In Section 2 we give the main theorem which is based on a lemma due essentially to I. M. Gelfand [4](4). In Section 3 we give a few concrete counter examples in which the differentials are explicitly determined and thus allow an easy confrontation with results of different theorems from the preceding sections.

(3) Whereas our class $\mathfrak{A}$ is the intersection of all classes $\mathfrak{A}\{a_n\}$ (see Definition 2, Section 1, Chapter I) each of the classes $\mathfrak{A}\{a_n\}$ is a part of the class of Christensen's zero-sets.

(4) By using recent work on spaces with Radon–Nikodym property (see, for instance, H. B. Maynard [8]) we could somehow weaken our requirements concerning the range space as far as concerns the strong differential.

In Chapter III we give some applications of the preceding theorems. In Section 1 we consider mappings formed by composition of a Lipschitzian mapping with a compact linear mapping. We show that for such mappings we can skip restrictions on the image-space Y or obtain that actually the differential is a Fréchet differential. Section 2 deals with convex functions $G \to \mathbf{R}^1$, G open in X. Such functions have everywhere a Gateaux differential. Our main theorem is that they have a differential (i.e. linear) except in a set of $\mathfrak{A}^0$(5). In Section 3 we extend the last theorem to more general mappings than the convex functions. This is done in stages. We define first relatively convex mappings $G \to Y$ (relative to a convex cone in Y), then variably convex mappings, convexoid mappings, and finally locally convexoid mappings which are the largest class to which we are able to extend suitably the theorem of Section 2. We give several examples of mappings of the different classes. In Section 4 we illustrate different results of the paper in the example of the function $F_S(x)$ = distance from a point x to a closed subset S contained in X.

In Chapter IV we are investigating the class of measures which are absolutely continuous relative to our exceptional class $\mathfrak{A}$. In Section 1 we describe the structure of such measures. This allows us to reduce the problem of constructing such measures to the problem of constructing a suitable compatible sequence of cylindrical measures $\{\mu_n\}$ defined in the consecutive finite dimensional spaces $[e_1, \ldots, e_n]$ determined by a generalized basis $\{e_n\}$ in X. In Section 2 we review and restate, conforming to our needs, different properties of cylindrical measures which are essentially known. In Section 3 we construct by means of cylindrical measures a large class of Borel measures a.c. rel. $\mathfrak{A}$.

Before finishing this introduction we would like to express all our thanks to F. Mignot whose careful notes of our lectures in Paris, Université VI, were very instrumental in the preparation of the present paper.

CHAPTER I
THE EXCEPTIONAL CLASS $\mathfrak{A}$

1. Definition and elementary propierties of the class $\mathfrak{A}$. We consider a real separable Banach space X and, unless otherwise indicated, we consider only Borel subsets of X.

DEFINITION 1. A class $\mathfrak{B}$ of Borel subsets of X is called *exceptional* (or a class of exceptional subsets) if it has the following properties:

(5) It is of interest to compare our results with those of E. Asplund [1] who uses methods completely different from ours. It seems that our methods have no bearing on his results concerning Fréchet differentiability. As concerns differentiability (which he calls — as do many other mathematicians — the Gateaux differentiability) our result complements very essentially his and is new even in finite dimensional spaces.

a) $\mathfrak{B}$ is σ-additive.

b) $\mathfrak{B}$ is hereditary; i.e. if $B_1 \subset B$, $B \in \mathfrak{B}$, then $B_1 \in \mathfrak{B}$.

c) $\mathfrak{B}$ does not contain any open subsets of X(6).

DEFINITION 2. 1° For $0 \neq a \in X$ let $\mathfrak{A}(a) = \{A \subset X\colon \forall x \in X,\ A \cap (x + \mathbf{R}^1 a)$ is of Lebesgue measure 0 on the line $x + \mathbf{R}^1 a\}$.

2° For every sequence $\{a_n\} \subset X$ with $a_n \neq 0$, $\mathfrak{A}\{a_n\} = \{A \subset X\colon A = \bigcup_n A_n,\ A_n \in \mathfrak{A}(a_n)\}$.

3° $\mathfrak{A} = \mathfrak{A}_X = \bigcap \mathfrak{A}\{a_n\}$, the intersection being taken over all sequences $\{a_n\}$ complete in X.

It is clear that properties a), b) of exceptional classes are satisfied for classes $\mathfrak{A}(a)$, $\mathfrak{A}\{a_n\}$ and $\mathfrak{A}$. Property c) will be proved in Section 3. It is also immediate that $\{a_n'\} \subset \{a_n\}$ implies $\mathfrak{A}\{a_n'\} \subset \mathfrak{A}\{a_n\}$.

We will use the following notations: for $A \subset X$, $[A]$ denotes the linear span of A; this makes evident the meaning of symbols $[\{a_n\}]$, $[a_1, a_2, \ldots]$, $[a_1, \ldots, a_k]$. In particular, a sequence $\{a_n\}$ in X is complete if and only if $\overline{[\{a_n\}]} = X$.

DEFINITION 3. For S — a finite dimensional subspace of X — we let $\mathfrak{A}(S) = \{A \subset X\colon \forall x \in X,\ A \cap (x + S)$ is of Lebesgue measure zero in the hyperplane $x + S\}$.

PROPOSITION 1. *If S is a finite dimensional subspace of X, $S = [\{a_k\}]$, then $\mathfrak{A}(S) = \mathfrak{A}\{a_k\}$.*

Proof. It is clear from Definition 2, 2° that $\mathfrak{A}(S) \supset \mathfrak{A}\{a_k\}$. To prove the opposite inclusion let $\dim S = n$; if $n = 1$, then $\mathfrak{A}(S) = \mathfrak{A}(a_i)$. Assume that $\mathfrak{A}(S') \subset \mathfrak{A}\{a_k'\}$ has been proved for every S' of dimension $\leqslant n-1$. Choose $a_{k_1}, \ldots, a_{k_n}$ — a basis in S, $S = [a_{k_1}, \ldots, a_{k_n}]$. If $A \in \mathfrak{A}(S)$, then the characteristic function χ_A of A is a Borel function on S and $\chi_A(x + \xi a_{k_1})$ is a Borel function of (x, ξ). Let

$$A_1 = \Big\{x\colon x \in A,\ \int_{-\infty}^{\infty} \chi_A(x + \xi a_{k_1})\, d\xi = 0\Big\} = \bigcap_{N=1}^{\infty} \Big\{x \in A\colon \int_{-N}^{N} \chi_A(x + \xi a_{k_1})\, d\xi = 0\Big\}.$$

A_1 is a Borel subset of A, $A_1 \in \mathfrak{A}(a_{k_1})$ and $A' = A \setminus A_1$ has the property that $A' \cap (x + \mathbf{R}^1 a_{k_1}) \neq \emptyset$ implies that $A' \cap (x + \mathbf{R}^1 a_{k_1})$ is of positive measure. Since $A' \in \mathfrak{A}(S)$, it follows that $A' \in \mathfrak{A}(S')$ with $S' = [a_{k_2}, \ldots, a_{k_n}]$ and by induction hypothesis $A' \in \mathfrak{A}\{a_{k_2}, \ldots, a_{k_n}\}$. Hence $A = A_1 \cup A' \in \mathfrak{A}\{a_{k_1}, \ldots, a_{k_n}\} \subset \mathfrak{A}\{a_n\}$.

DEFINITION 4. We say that two sequences $\{a_m\}, \{b_n\} \subset X$ are equivalent, $\{a_m\} \sim \{b_n\}$, if each element of any one of them is a linear combination of elements of the other.

COROLLARY 1. *If $\{a_m\} \sim \{b_n\}$, then $\mathfrak{A}\{a_m\} = \mathfrak{A}\{b_n\}$.*

(6) Such a class can be considered in any metrizable topological space.

It is sufficient to show that, for every m, $\mathfrak{A}(a_m) \subset \mathfrak{A}\{b_n\}$. Since $a_m = \xi_1 b_{n_1} + \ldots + \xi_p b_{n_p}$ taking $S = [b_{n_1}, \ldots, b_{n_p}]$ we get $\mathfrak{A}(a_m) \subset \mathfrak{A}(S) = \mathfrak{A}\{b_{n_1}, \ldots, b_{n_p}\} \subset \mathfrak{A}\{b_n\}$.

Remark 1. The corollary will be applied to complete sequences $\{a_n\}$ in two extreme cases: 1° Removing consecutively from the sequence $\{a_n\}$ the elements which are linear combinations of the preceding ones, one arrives at a complete linearly independent sequence $\{a'_n\}$, $\{a'_n\} \sim \{a_n\}$. 2° By arranging into a sequence all linear combinations with rational coefficients of elements of $\{a_n\}$ one obtains a sequence $\{a''_n\}$ dense in X such that $\{a''_n\} \sim \{a_n\}$.

Remark 2. If X is finite dimensional, then for every complete sequence $\{a_n\} \subset X$ we have $X = [\{a_n\}]$ hence by Proposition 1 $\mathfrak{A}(X) = \mathfrak{A}\{a_n\}$ and $\mathfrak{A}_X = \mathfrak{A}(X)$ = class of sets of Lebesgue measure zero.

Remark 3. Classes $\mathfrak{A}\{a_n\}$ are invariant under translations and homotheties. The class $\mathfrak{A}$ is, in addition, invariant under linear continuous automorphisms of X.

Remark 4. If $\mathfrak{B}$ is an exceptional class then we say that a pointwise property holds exc. $\mathfrak{B}$ (except $\mathfrak{B}$) if it holds everywhere outside of a set of class $\mathfrak{B}$. By analogy with finite dimensional vector spaces we say that a property holds a.e. (almost everywhere) if it holds exc. $\mathfrak{A}$.

2. Complete biorthogonal systems and generalized bases. The considerations of this section are well known; since we could not find a suitable reference, we found it convenient to give here a short outline.

Two real vector spaces V, W form a (real) pairing $\langle V, W\rangle$ if we are given a bilinear real form $(v, w) \in V \times W \to \langle v, w\rangle$. We consider on V and W the topologies $\sigma(V, W)$, $\sigma(W, V)$ (induced by the pairing). We only consider proper pairings, i.e. those giving rise to Hausdorff topologies.

We assume that V is separable in the topology $\sigma(V, W)$; it follows then that W is also separable (in the topology $\sigma(W, V)$).

Two complete sequences $\{e_n\} \subset V$, $\{f_n\} \subset W$ form a *complete biorthogonal system* in the pairing if

$$\langle e_m, f_n\rangle = \delta_{mn}. \tag{1}$$

If this is the case, $\{e_n\}$ is a *generalized basis* in V and $\{f_n\}$ its dual basis in W.

The next proposition is based on an easy extension of the idea of Gram–Schmidt orthogonalization.

PROPOSITION 1. *Let $\{v_n\}$ be complete in V and $\{w_n\}$ be complete in W. There is a canonical way of constructing a biorthogonal system $\{e_m\}$, $\{f_n\}$ for the pairing $\langle V, W\rangle$ such that $\{e_m\} \sim \{v_m\}$ and $\{f_n\} \sim \{w_n\}$.*

We list now some of the properties of complete biorthogonal systems; the proofs are immediate.

I. If $\{e_m\}$ is a generalized basis in V (relative to the pairing $\langle V, W\rangle$), then for every n we have the direct decomposition

$$V = [e_1, \ldots, e_n] + \overline{[e_{n+1}, e_{n+2}, \ldots]},$$

defining a linear continuous projection P_n, $P_n(V) = [e_1, \ldots, e_n]$, $P_n(\overline{[e_{n+1}, e_{n+2}, \ldots]}) = 0$. The adjoint projection in W satisfies $P_n^*(W) = [f_1, \ldots, f_n]$, $P_n^*(\overline{[f_{n+1}, f_{n+2}, \ldots]}) = 0$.

II. The dual sequence $\{f_n\}$ is unique and can be obtained by considering the one dimensional subspace $([e_1, \ldots, e_{n-1}] + \overline{[e_{n+1}, e_{n+2}, \ldots]})^{\perp}$ of W and choosing from it the unique element f_n satisfying $\langle e_n, f_n\rangle = 1$.

IIIa. To every $v \in V$ there corresponds a formal Fourier series $v \sim \sum \langle v, f_n\rangle e_n$, the mapping $v \mapsto \{\langle v, f_n\rangle\}$ defines an injection of V into the vector space of all real sequences.

The formal Fourier series of $v \in V$ converges to v in $\sigma(V, [\{f_n\}])$ but in general not in $\sigma(V, W)$. In order that $\sum \xi_n e_n$ be a formal Fourier series of some $v \in V$ it is necessary and sufficient that the series be convergent to v in $\sigma(V, [\{f_n\}])$.

IIIb. If $v \in V$ and $w \in [\{f_n\}]$, then $(v, w) = \sum \langle v, f_n\rangle \langle e_n, w\rangle$, the series containing only a finite number of non-zero terms.

If $\{w_a\} \subset [\{f_n\}]$ is a net convergent to $w \in W$ in the topology $\sigma(W, V)$, then for every $v \in V$ $\langle v, w\rangle = \lim\limits_{a} \sum\limits_{n} \langle v, f_n\rangle \langle e_n, w_a\rangle$.

A generalized basis $\{e_n\}$ is a *weak Schauder basis* if for every $v \in V$ the Fourier series $\sum \langle v, f_n\rangle e_n$ is convergent to v in the topology $\sigma(V, W)$.

IV. The basis $\{e_n\}$ is a weak Schauder basis if and only if for every $v \in V$ and $w \in W$ the series $\sum \langle v, f_n\rangle \langle e_n, w\rangle$ converges to $\langle v, w\rangle$. It follows that if $\{e_n\}$ is a weak Schauder basis in V, then $\{f_n\}$ is a weak Schauder basis in W.

For a separable Banach space X we consider the canonical pairing $\langle X, X^*\rangle$ with $\langle x, x^*\rangle = x^*(x)$.

V. If $\{e_n\}$ is a generalized basis in a Banach space X, then the projections P_n are bounded (but not in general uniformly). If $\{e_n\}$ is a weak Schauder basis, then P_n are uniformly bounded and converge weakly to I.

Remark 1. The considerations of this section remain valid for complex vector spaces V and W when the pairing $\langle v, w\rangle$ is *hermitian bilinear* (more commonly called nowadays *sesquilinear*).

3. Class $\mathfrak{A}$ and supports of measures.

THEOREM 1. *Let $\{a_n\}$ be a complete sequence in the Banach space X. Then $\mathfrak{A}\{a_n\}$ does not contain non-empty open sets.*

Proof. Since $\mathfrak{A}\{a_n\}$ is translation invariant, it is enough to prove that no closed ball $\overline{B_R(0)}$, $R>0$, is in $\mathfrak{A}\{a_n\}$.

By Proposition 1, Section 2, there is a biorthogonal complete system $\{e_n, f_n\}$ for the pairing $\langle X, X^*\rangle$ such that $\{e_n\}\sim\{a_n\}$, hence $\mathfrak{A}\{e_n\}=\mathfrak{A}\{a_n\}$. Choose a sequence $\{\gamma_n\}$, $\gamma_n>0$, such that $\sum\gamma_n\|e_n\|=R$. Put

$$C=\{x\in X\colon x=\sum_1^\infty\gamma_n\xi_n e_n,\ 0\leqslant\xi_n\leqslant 1\}.$$

It is clear that C is compact and $C\subset\overline{B_R(0)}$. Since the series $x=\sum\limits_1^\infty\gamma_n\xi_n e_n$ converges in norm, it converges also in the weaker topology $\sigma(X,[\{f_n\}])$ hence by property IIIa, Section 2, it is the Fourier series of x and the correspondence $x\to\{\xi_n\}$ is a bijection of C onto the space of all sequences $\{\xi_n\}$ with $0\leqslant\xi_n\leqslant 1$, which becomes a homeomorphism if in the last space we take the topology of pointwise convergence.

Denote by $[0;\gamma_n e_n]$ the straight segment of all points $\xi_n\gamma_n e_n$, $0\leqslant\xi_n\leqslant 1$ and by $C^{(k)}$ the set of all $x\in C$ with $\xi_k=0$. We can write then the obvious identifications

$$C=\prod_1^\infty[0;\gamma_n e_n],\qquad C^{(k)}=\prod_{m\neq k}[0;\gamma_m e_m],\qquad C=C^{(k)}\times[0;\gamma_k e_k].$$

On the segment $[0;\gamma_n e_n]$ we take the unit measure $d\mu_n=d\xi_n$. On C we consider the product measure μ of all the measures μ_n, on $C^{(k)}$ the product measure $\mu^{(k)}$ of all the μ_n with $n\neq k$. Clearly $\mu(C)=1$. For any bounded Borel measurable function φ on C we get then by Fubini's theorem

$$\int_C\varphi d\mu=\int_{C^{(k)}}\int_{[0;\gamma_k e_k]}\varphi(x^{(k)}+\xi_k\gamma_k e_k)\,d\xi_k\,d\mu^{(k)}(x^{(k)})\qquad\text{for any } k=1,2,\ldots,$$

where $x^{(k)}$ is a variable point on $C^{(k)}$.

In particular if φ is the characteristic function of a set A, we get

$$\mu(A)=\int_C\varphi d\mu=\int_{C^{(k)}}\int_{[0;\gamma_k e_k]}\varphi(x^{(k)}+\xi_k\gamma_k e_k)\,d\xi_k\,d\mu^{(k)}(x^{(k)}).$$

It follows that if $A\in\mathfrak{A}(e_k)$, $\mu(A)=0$ and, if $A\in\mathfrak{A}\{e_k\}$, i.e $A=\bigcup A_k$, $A_k\in\mathfrak{A}(e_k)$, $\mu(A)\leqslant\sum\mu(A_k)=0$. Since $\mu(C)=1$, $C\notin\mathfrak{A}\{e_k\}$ and $\overline{B_R(0)}\supset C$ is not in $\mathfrak{A}\{e_k\}=\mathfrak{A}\{a_n\}$.

Remark 1. Theorem 1 shows that all the classes $\mathfrak{A}(a)$, $\mathfrak{A}\{a_n\}$ and $\mathfrak{A}$ have property c) of exceptional classes (see Section 1), hence they are exceptional classes.

THEOREM 2. *Let X be an infinite dimensional Banach space and μ be a σ-finite Borel measure on X. Then there is a complete sequence $\{e_n\}$ in X such that μ is concentrated on a set in $\mathfrak{A}\{e_n\}$.*

A Borel measure μ is said to be *concentrated* on A if $\mu(B)=\mu(B\cap A)$ for every Borel set B.

Proof of Theorem 2 depends on two lemmas, the second one is of some interest by itself.

LEMMA 1. *Let X be separable infinitely dimensional. There is a sequence $\{u_n\}\subset X^*$, $\|u_n\|=1$ such that $\{u_n\}$ converges to 0 in the weak*-topology.*

Proof. For a separable infinite dimensional space X the weak*-closure of the unit sphere $\{u\in X^*;\ \|u\|=1\}$ in X^* is the closed unit ball. Furthermore the latter is metrizable in the weak*-topology. Hence the assertion follows.

The weak*-convergence in X^* is denoted by $\rightharpoonup$.

LEMMA 2. *Let $\{u_n\}\subset X^*$, $\|u_n\|=1$, $u_n\rightharpoonup 0$. If A is a bounded subset of X such that $\langle x,u_n\rangle\to 0$ uniformly on A, then there exists $a\in X$ such that for every $x\in X$, $A\cap(x+\boldsymbol{R}a)$ consists of at most one point.*

Proof. We can assume that $A\subset\{x\colon\|x\|<1\}$. There is a subsequence $\{u_n'\}\subset\{u_n\}$ with the following properties:

α) for all n there is an $x_n'\in X$, $\|x_n'\|=1$, such that $\langle x_n',u_n'\rangle\geqslant 5/6$,

β) for every $x\in A$, $|\langle x,u_n'\rangle|<2^{-4n}$, $n=1,2,\ldots$,

γ) for $k=1,\ldots,n-1$, $|\langle x_k',u_n'\rangle|\leqslant 2^{-4n}$.

The sequence in question is constructed by induction: we choose for u_1' the first u_p such that $|\langle x,u_p\rangle|\leqslant 2^{-4}$ for all $x\in A$; by Hahn–Banach theorem there is an $x_1'\in X$, $\|x_1'\|=1$, such that $\langle x_1',u_1'\rangle\geqslant 5/6$. To construct u_n' with given u_k' and x_k', $k<n$, we use the uniform convergence of $\{u_p\}$ on $A\cup\{x_1',\ldots,x_{n-1}'\}$; there is $q>n$ such that with $u_n'=u_q$ β) and γ) are satisfied, x_n' is then chosen using the Hahn–Banach theorem again. Define $a=\sum 2^{-2k}x_k'$; then $a\neq 0$ or else $1/4=2^{-2}\|x_1'\|\leqslant\sum\limits_{k=2}^\infty 2^{-2k}\|x_k'\|=1/12$. Assume that there is an $x\in X$ such that $(x+\boldsymbol{R}a)\cap A\supset(y,y')$, $y-y'=\eta a$, $\eta>0$. Then

$$|\langle y'-y,u_n'\rangle|\leqslant|\langle y',u_n'\rangle|+|\langle y,u_n'\rangle|\leqslant 2^{-4n+1}\tag{1}$$

and on the other hand

$$|\langle y'-y,u_n'\rangle|=\eta|\langle a,u_n'\rangle|=\eta\Big|\sum_{n=1}^\infty 2^{-2k}\langle x_k',u_n'\rangle\Big|$$

$$\geqslant\eta 2^{-2n}\langle x_n',u_n'\rangle-\eta\sum_{k=1}^{n-1}2^{-2k}|\langle x_k',u_n'\rangle|-\eta\sum_{k=n+1}^\infty 2^{-2k}|\langle x_k',u_n'\rangle|,\tag{2}$$

$$\sum_{k=1}^{n-1}2^{-2k}|\langle x_k',u_n'\rangle|\leqslant\frac{2^{-4n}}{3},\qquad\sum_{k=n+1}^\infty 2^{-2k}|\langle x_k',u_n'\rangle|\leqslant\frac{2^{-2n}}{3}.$$

(2) implies

$$(3)\qquad |\langle y'-y,\, u_n'\rangle| \geqslant \eta\left(\tfrac{5}{6}2^{-2n}-\tfrac{2}{3}2^{-2n}\right) = \eta\frac{2^{-2n}}{6}.$$

Comparing (1) and (3) we get

$$2^{-4n+1} \geqslant 2^{-2n}\frac{\eta}{6}$$

showing that $\eta = 0$.

Remark 1. With the same hypotheses one has the following result. There is a complete sequence $\{a_n\} \subset X$ such that for every $x \in X$ and every n, $x+[a_1, \ldots, a_n] \cap A$ consists of at most one point.

Proof of Theorem 2. Since μ is σ-finite, we can assume that $X = \bigcup B_n$, $B_n \subset B_n(0)$, $B_n \subset B_{n+1}$, $\mu(B_n) < \infty$. Let u_n be a sequence as in Lemma 1. Then the functions $x \to \langle x, u_n\rangle$ are μ measurable and converge pointwise to 0 on X. By Egorov's theorem there are Borel sets A_m, $A_m \subset B_m$, $A_m \subset A_{m+1}$ such that $\mu(B_m \setminus A_m) < 2^{-m}$ and u_n converge uniformly to 0 on A_m. Lemma 2 assigns to each A_m an element $e_n \in X$ such that $A_m \in \mathfrak{A}(e_m)$, hence $\bigcup A_m \in \mathfrak{A}\{e_m\}$ and $\mu(X \setminus \bigcup A_m) = 0$ since $\mu(X \setminus \bigcup_1^\infty A_m) = \mu\left(\bigcup_N^\infty B_m - \bigcup_N^\infty A_m\right) \leqslant \mu\left(\bigcup_N^\infty (B_m \setminus A_m)\right) \leqslant \sum_N^\infty 2^{-m} = 2^{1-N}$.

Remark 2. Theorem 2 shows that there is no σ-finite Borel measure μ such that all the μ-null sets belong to $\mathfrak{A}$ (the set on which μ is concentrated belongs to $\mathfrak{A}\{e_n\}$, its complement $\notin \mathfrak{A}\{e_n\}$ is of μ-measure 0).

4. Class $\mathfrak{A}^0$. The procedure used in construction of the class $\mathfrak{A}$ can be used to define a great number of other exceptional classes. Let $\mathfrak{B}_{\boldsymbol{R}^1}$ be an exceptional class on $\boldsymbol{R}^1$ whose elements are Borel sets of Lebesgue measure zero. We assume that $\mathfrak{B}_{\boldsymbol{R}^1}$ is invariant under translations and homotheties. By analogy with Definition 2, Section 1, we have the following (X being separable Banach space).

Definition 1. 1° Let $0 \neq a \in X$, $\mathfrak{B}(a) = \{A \subset X\colon \forall x \in X, A \cap (x + \boldsymbol{R}^1 a) \in \mathfrak{B}_{\boldsymbol{R}^1}$ on the line $x + \boldsymbol{R}^1 a\}$.

2° For every sequence $\{a_n\}$, $a_n \neq 0$, $\mathfrak{B}\{a_n\} = \{A \subset X\colon A = \bigcup_n A_n,\ A_n \in \mathfrak{B}(a_n)\}$.

3° $\mathfrak{B} = \mathfrak{B}_X = \bigcap \mathfrak{B}\{a_n\}$, the intersection being taken over all sequences $\{a_n\}$ complete in X.

It is clear that $\mathfrak{B}(a) \subset \mathfrak{A}(a)$, $\mathfrak{B}\{a_n\} \subset \mathfrak{A}\{a_n\}$, $\mathfrak{B} \subset \mathfrak{A}$, that $\mathfrak{B}(a)$, $\mathfrak{B}\{a_n\}$, and $\mathfrak{B}$ are exceptional classes, $\mathfrak{B}(a)$ and $\mathfrak{B}\{a_n\}$ are invariant under translations and homotheties and $\mathfrak{B}$ is in addition invariant with respect to continuous linear automorphisms of X.

There are infinitely many choices of $\mathfrak{B}_{\boldsymbol{R}^1}$, one could take the classes of sets of measure 0 for all kinds of Hausdorff measures, the classes of sets of zero capacity, for all kinds of capacities, etc.

The smallest (non-trivial) class $\mathfrak{B}_{\boldsymbol{R}^1}$ consists of all countable subsets of $\boldsymbol{R}^1$; we denote this class by $\mathfrak{A}^0_{\boldsymbol{R}^1}$. The corresponding class $\mathfrak{A}^0_X$ is considerably smaller than $\mathfrak{A}_X$, also in finite dimensional spaces. To illustrate this remark we mention the following fact.

Let $(e_1, \ldots, e_n)$ be an arbitrary basis in $\boldsymbol{R}^n$ and on each of the axes $\boldsymbol{R}e_k$ choose an arbitrary Cantor set (a compact which does not contain any interval and any isolated point). Then $A = C_1 \times \ldots \times C_k$ does not belong to $\mathfrak{A}^0_{\boldsymbol{R}^n}$.

The proof is obtained by constructing on each C_k a positive measure μ_k without point masses. Then the product measure $\mu_1 \times \ldots \times \mu_n$ of A is positive whereas every set of $\mathfrak{A}^0\{e_k\}$ is of measure 0.

The class $\mathfrak{A}^0$ will be of interest in the study of differentials of convex functions and their generalizations (Chapter II, Section 1 and Chapter III, Section 2); it has found an interesting application in connection with a recent result of E. Zarantonello [12]. To explain this result we have to introduce certain notions.

For a multivalued function T, a point x in the domain of T is a point of unicity if $T(x)$ is a single point, it is a point of multiplicity if $T(x)$ consists of more than one point. A multivalued transformation $T\colon D \subset X \to X^*$ is monotone if $x_1, x_2 \in D$, $y_1 \in T(x_1)$, $y_2 \in T(x_2)$ imply $\langle x_1 - x_2, y_1 - y_2\rangle \geqslant 0$. The theorem of Zarantonello can then be stated as follows.

Theorem Z. *If T is a monotone transformation defined on an open set G of a separable Banach space X, then the set of points of multiplicity of T belongs to the class $\mathfrak{A}^0$.*

The original result of Zarantonello was less precise: the points of unicity form a dense subset of G. The theorem stated above can be proved by repeating without any changes the argument of Zarantonello.

Remark 1. For an arbitrary class $\mathfrak{B}_{\boldsymbol{R}^1}$ (subject to our conditions) it may happen that the property analogous to the one indicated in Remark 1, Section 1 does not hold. In such a case, in Definition 1.3° of the present section we may restrict the sequences $\{a_n\}$ to those dense in X — we obtain then a class $\hat{\mathfrak{B}}$ larger than $\mathfrak{B}$. We will have an opportunity to use the class $\hat{\mathfrak{A}}^0$ in Chapter II, Section 1.

Remark 2. Checking the proof of Theorem 2 in the preceding section, especially the use of Lemma 2, we notice that actually we prove a stronger statement obtained by replacing in Theorem 2 the class $\mathfrak{A}\{e_n\}$ by $\mathfrak{A}^0\{e_n\}$.

CHAPTER II

DIFFERENTIALS OF LIPSCHITZIAN MAPPINGS

1. Different notions of differentials. Elementary properties. We consider two Banach spaces X and Y, in Y a locally convex Hausdorff topology τ weaker than the strong topology. Furthermore, we consider a mapping T of an open subset $G \subset X$ into Y. In X we will consider only the strong topology.

For a fixed $x \in G$ and a vector $u \in X$ we form the differential quotient

$$\frac{T(x+\varrho u)-T(x)}{\varrho} \quad \text{for} \quad \varrho > 0. \tag{1}$$

It has a meaning for ϱ small enough.

DEFINITION 1. 1° If the differential quotient (1) converges rel. τ in Y for $\varrho \searrow 0$ we put

$$DT(x; u) = \lim_{\varrho \searrow 0} \frac{T(x+\varrho u)-T(x)}{\varrho}. \tag{2}$$

If necessary we specify $DT(x; u)$ rel. τ.

2° If $DT(x; u)$ rel. τ exists for all $u \in X$ it is a mapping of X into Y which will be denoted by $DT(x)$ (rel. τ) and called the *Gateaux τ-differential* of T at $x \in G$.

PROPOSITION 1. *If $DT(x)$ exists it is positively homogeneous of degree 1, i.e. $DT(x, au) = aDT(x; u)$ for all $a \geqslant 0$.*

The proof is immediate.

If $DT(x, u) = -DT(x; -u)$ for some $u \in X$ we say that $DT(x)$ is *antisymmetric at u*. If it is the case for all $u \in X$ we say that $DT(x)$ is *anti-symmetric*(7).

PROPOSITION 2. *If $DT(x)$ exists rel. τ and $DT(x; u)$ is continuous as a function of u (rel. any Hausdorff topology τ' on Y weaker than the strong topology) and if $DT(x)$ is anti-symmetric at each vector a_n of a sequence $\{a_n\}$ dense in X, then $DT(x)$ is anti-symmetric.*

The proof is immediate.

It is clear that for real functions f of a real variable ($X = Y = \boldsymbol{R}^1$) the anti-symmetry of $Df(x)$ is equivalent to the linearity of $Df(x; u)$ as function of u.

It is of importance to compare the notions introduced here with the usual notions in the theory of real functions of a real variable. If such a function f is defined on an open interval $I \subset \boldsymbol{R}^1$ one introduces a unit

(7) Many authors restrict the notion of Gateaux differentials to anti-symmetric differentials.

vector e in $\boldsymbol{R}^1$ defining the unity of measure and the orientation on $\boldsymbol{R}^1$. One defines, then, the right and left derivatives $f'_d(x)$ and $f'_g(x)$ by the formulas

$$f'_d(x) = \lim_{\varrho \searrow 0} \frac{f(x+\varrho e)-f(x)}{\varrho}, \qquad f'_g(x) = \lim_{\varrho \searrow 0} \frac{f(x-\varrho e)-f(x)}{-\varrho}.$$

If $f'_d(x)$ and $f'_g(x)$ exist and are equal their common value is the derivative $f'(x)$. It is clear that the existence of $f'_d(x), f'_g(x)$ means the existence of $Df(x)$ with $Df(x; e) = f'_d(x)$, $Df(x; -e) = -f'_g(x)$. Also, the existence of $f'(x)$ means the existence and linearity of $Df(x)$ with $Df(x; e) = -Df(x; -e) = f'(x)$.

For our next proposition we will need a general lemma on functions f: $\boldsymbol{R}^1 \to \boldsymbol{R}^1$ which is due (under stronger form) to W. Sierpiński, but for which a reference is difficult to reach. We will give a short proof of it.

LEMMA 1. *Let $f(x)$ be a real function defined on some open set $G \subset \boldsymbol{R}^1$. The set E of points of G, where $f'_d(x)$ and $f'_g(x)$ exist, but $f'_d(x) \neq f'_g(x)$ is at most enumerable.*

Proof. It is enough to show that $E \cap I$ is at most enumerable for any closed interval $I \subset G$. Suppose the contrary. Then there exists an $a > 0$ and a non-enumerable subset $E_1 \subset E \cap I$ such that for each $x \in E_1$ $|f'_d(x)-f'_g(x)| > a$. Put $\varepsilon = a/7$. It follows that there exists a non-enumerable $E_2 \subset E_1$ and $\varrho_0 > 0$ such that for all $x \in E_2$, $\varrho < \varrho_0$, $\varrho > 0$ we have

$$\left|\frac{f(x+\varrho e)-f(x)}{\varrho} - f'_d(x)\right| < \varepsilon \quad \text{and} \quad \left|\frac{f(x-\varrho e)-f(x)}{-\varrho} - f'_g(x)\right| < \varepsilon.$$

Since E_2 is non-enumerable, it contains a point x_0 which is a limit of its points $x_1 > x_0$. We can find, therefore, such a point, $x_1 = x_0+\varrho e$ such that $\varrho < \varrho_0/2$ and that the closed interval $[x_0; x_0+2\varrho e] \subset G$. We have then the equalities:

$$f(x_0+\varrho e)-f(x_0) = \varrho f'_d(x_0)+\varepsilon_0\varrho, \; f(x_0+2\varrho e)-f(x_0+\varrho e) = \varrho f'_d(x_1)+\varepsilon_1\varrho,$$

$$f(x_0+2\varrho e)-f(x_0) = 2\varrho f'_d(x_0)+\varepsilon_2 2\varrho, \; f(x_0)-f(x_0+\varrho e) = -\varrho f'_g(x_1)+\varepsilon_3\varrho,$$

where for all the ε_k we have $|\varepsilon_k| < \varepsilon$. From these equalities we get

$$f'_d(x_1) = f'_d(x_0)+2\varepsilon_2-\varepsilon_0-\varepsilon_1, \qquad f'_g(x_1) = f'_d(x_0)+\varepsilon_0+\varepsilon_3.$$

Finally, $|f'_d(x_1)-f'_g(x_1)| < 6\varepsilon < a$ a contradiction.

PROPOSITION 3. *Let T be a continuous mapping of an open set $G \subset X$ into $\boldsymbol{R}^1$, X being a separable Banach space. We suppose that for each $x \in G$,*

$DT(x)$ exists and is continuous (as function of u). Then $DT(x)$ is antisymmetric exc. $\hat{\mathfrak{A}}^0_X$ (8).

Proof. Let A be the set of x's, where $DT(x)$ is not anti-symmetric and consider any sequence $\{a_n\} \subset X$, $a_n \neq 0$, dense in X. Put $A_n = \{x \in G$: $DT(x)$ is not anti-symmetric at $a_n\}$. By Proposition 2, $A = \bigcup A_n$. By applying Lemma 1 we would have immediately that $A_n \in \mathfrak{A}^0(a_n)$ if we knew that A_n is a Borel set. To prove this it is enough to prove that for each ball $B_R(x_0) \subset B_{2R}(x_0) \subset G$ such that $|T(x)|$ is uniformly bounded on $B_{2R}(x_0)$, $A_n \cap B_R(x_0)$ is a Borel set (9). Consider the expression

$$\varphi(x, \varrho) = \left|\frac{T(x+\varrho a_n)+T(x-\varrho a_n)-2T(x)}{\varrho}\right|.$$

It is a continuous function on (x, ϱ) for $x \in B_R(x_0)$ and $0 < \varrho < R/\|a_n\|$. We have $\lim\limits_{\varrho\searrow 0}\varphi(x, \varrho) = |DT(x; a_n)+DT(x; -a_n)|$, hence $|DT(x; a_n)+$ $+DT(x; -a_n)|$ is a Borel function of x and the set where it is $\neq 0$ is a Borel set.

Thus A_n is a Borel set in $\mathfrak{A}^0(a_n)$ and $A = \bigcup A_n \subset \mathfrak{A}^0\{a_n\}$. Finally, $\{a_n\}$ being an arbitrary dense sequence in X, $A \in \hat{\mathfrak{A}}^0$.

THEOREM 1. *Let X be separable and Y have a Hausdorff locally convex topology τ_1 weaker than τ such that there exists $\{v_n\}$ dense in Y rel. τ_1. We assume that for $T\colon X \supset G \to Y$ the Gateaux τ-differential exists for each $x \in G$ and that $DT(x)\colon X \to Y$ is continuous rel. a Hausdorff topology on Y (which may vary with x). Then $DT(x)$ is anti-symmetric for all $x \in G$ except $\hat{\mathfrak{A}}^0$.*

Proof. Let A be the set of x's, where $DT(x)$ is not anti-symmetric, consider the dual Y^{τ_1} rel. τ_1. The assummed properties of the topology τ_1 imply that there is a $\{w_n\} \subset Y^{\tau_1}$ dense in Y^{τ_1} rel. the topology $\sigma(Y^{\tau_1}, Y)$. Since for the real-valued functions $T_n(x) = \langle T(x), w_n\rangle$ we have $DT_n(x; u)$ $= \langle DT(x; u), w_n\rangle$, it follows that $DT(x)$ is anti-symmetric if and only if all the $DT_n(x)$ are anti-symmetric. By applying Proposition 3 our theorem is proved.

We return now to real-valued mappings.

The mapping $T\colon X \to \boldsymbol{R}^1$ is called *convex* if for all convex combinations $u = \sum a_k u_k$, $a_k \geqslant 0$, $\sum a_k = 1$, we have

$$T(u) \leqslant \sum a_k T u_k. \tag{3}$$

If the opposite inequality is always true we say that T is concave. If T is concave, $-T$ is convex. T is concave and convex if and only if T is linear.

(8) The class $\hat{\mathfrak{A}}^0$ was introduced in Remark 1, Chapter I, Section 4.

(9) Such balls exist for each $x_0 \in G$ by continuity of T and an enumerable system of such balls covers G because of separability of X.

If T is positively homogeneous of degree 1, and T is convex, inequality (3) is true for every linear combination $u = \sum a_k u_k$ with positive coefficients.

PROPOSITION 4. *If X is separable and the mapping $T\colon X \to \boldsymbol{R}^1$ is positively homogeneous of degree 1, continuous and convex (or concave), and if T is anti-symmetric on a sequence $\{a_k\}$ complete in X, then T is linear.*

Proof. Clearly we can restrict ourselves to T convex. Take an arbitrary, finite linear combination of the a_k's, say $u = \sum \beta_k a_k$. By convexity and anti-symmetry we have

$$Tu = T\Big(\sum |\beta_k| (\operatorname{sgn}\beta_k) a_k\Big) \leqslant \sum |\beta_k| T(\operatorname{sgn}\beta_k a_k) = \sum \beta_k T a_k. \tag{4}$$

Similarly,

$$T(-u) \leqslant \sum \beta_k T(-a_k). \tag{4'}$$

Adding the extreme members of these inequalities, we obtain

$$0 = T(u-u) \leqslant T(u)+T(-u) \leqslant 0. \tag{4''}$$

It follows that all inequalities (4), (4′) and (4″) are actually equalities. In particular $T\big(\sum \beta_k a_k\big) = \sum \beta_k T(a_k)$. It follows that T is linear on the subspace $[\{a_k\}]$ which is dense in X. T being continuous is, therefore, linear on X.

THEOREM 2. *Let T be a mapping of an open set $G \subset X$ into $\boldsymbol{R}^1$. We assume that X is separable and that for each $x \in G$, $DT(x)$ exists, is continuous and convex or concave (depending on x). The set A of points of G, where $DT(x)$ is not linear belongs to $\mathfrak{A}^0$.*

Proof. Let $\{a_n\}$ be an arbitrary complete sequence in X, $a_n \neq 0$. By Proposition 4, $DT(x)$ is linear if and only if $DT(x)$ is anti-symmetric at each a_n. Therefore $A = \bigcup A_n$, where A_n is the set of $x \in G$ such that $DT(x)$ is not anti-symmetric at a_n. As in the proof of Proposition 3, one shows that A_n is a Borel set and $A_n \in \mathfrak{A}^0(a_n)$. Thus $A \in \mathfrak{A}^0\{a_n\}$. Since the $\{a_n\}$ is an arbitrary complete sequence in X, $A \in \mathfrak{A}^0$.

Remark 1. We could give a stronger form to Theorem 2 by not asking that $DT(x)$ satisfy the conditions of the theorem at every point $x \in G$. Then A should be the set of x's, where all the conditions are satisfied and where $DT(x)$ is not linear. The proof proceeds in the same fashion but we have to assume that the set, where $DT(x)$ does satisfy the conditions of the theorem is a Borel set.

We are going to consider now two properties of a Gateaux τ-differential:

(A) $DT(x)$ is linear.

(B) The τ-convergence in (2) is uniform in u on each compact.

Remark 2. In Theorem 2 we had a case where condition (A) is satisfied exc. $\mathfrak{A}^0$. But to be sure of condition (B) we will have to add certain hypotheses concerning the mapping T.

To abbreviate the terminology we will say that $DT(x)$ is *a τ-differential* if it satisfies (A) and (B). We will say that $DT(x)$ is a *differential* if it is a τ-differential τ being the norm topology on Y.

PROPOSITION 5. *Let T be a τ-continuous mapping of G into Y, G open in X.* a) *If $DT(x)$ rel. τ exists and condition* (B) *is satisfied then $DT(x; u)$ as function of u is continuous rel. τ.* b) *If $DT(x)$ rel. τ exists and conditions* (A) *and* (B) *are satisfied, then $DT(x)$ as function of u is a linear mapping of X into Y continuous in the strong topology of Y.*

Proof. a) We have to prove that if $v_n \to v$ in X, then $DT(x; v_n) \xrightarrow{\tau} DT(x; v)$. Since for every $\varrho > 0$, $\dfrac{T(x+\varrho u)-T(x)}{\varrho}$ is a τ-continuous function of u and on the compact set $\{v_n\} \cup (v)$ this function converges τ-uniformly to $DT(x; u)$, it follows that the limit on the set $\{v_n\} \cup (v)$ is τ-continuous which proves our assertion.

b) Now $DT(x; u)$ is a τ-continuous linear mapping of X into Y. Hence it is a closed linear mapping in the strong topology and by closed graph theorem, it is strongly continuous.

We may replace condition (B) by a stronger one:

(B′) The τ-convergence in (2) is uniform in u on every bounded set.

If (A) and (B′) are valid we will call $DT(x)$ a *Fréchet τ-differential.* If τ is the strong topology $DT(x)$ is called *Fréchet-differential.*

It is customary when dealing with differentials in Banach spaces to use Fréchet-differentials; they have many properties which differentials in our sense do not have. They are used to define transformations T of class C^1 in an open set $G \subset X$ which are transformations $T: G \to Y$ which are strongly continuous and bounded in G, and have a Fréchet-differential $DT(x)$ at each $x \in G$, $DT(x): G \to \mathscr{B}(X, Y)$ (10) being a strongly continuous bounded mapping.

The next theorem shows that by replacing in the last definition "Fréchet-differential" by "τ-differential" we do not change the notion of C^1-mappings.

THEOREM 3. *If $T: G \to Y$ is τ-continuous and uniformly bounded in G and at each point $x \in G$ has a τ-differential $DT(x)$ which as a mapping of $G \to \mathscr{B}(X, Y)$ is strongly continuous and uniformlyb ounded, then $T \in C^1$ in G.*

Proof. We have to show that at each $x \in G$, T is strongly continuous and the τ-differential $DT(x)$ is actually a Fréchet differential. It is enough to prove the second assertion of which the first is an obvious consequence. By our assumptions, for any $x \in G$ we can find a closed ball $\overline{B_R(x)} \subset G$, such that $\|DT(x+y)-DT(x)\| < \varepsilon(\|y\|)$ for $\|y\| < R$, and $\varepsilon(\varrho) \searrow 0$ *as* $\varrho \searrow 0$. For any $u \in X$ we write the Bochner integral:

$$(*) \qquad \int_0^\varrho DT(x+tu; u)\,dt, \qquad \varrho\|u\| < R.$$

This integral is a fortiori a τ-integral, and since the integrand is a continuous τ-derivative of the τ-continuous function $T(x+tu)$, the integral is equal to $T(x+\varrho u)-T(x)$. On the other hand the integral can be written for $\varrho < R/\|u\|$ as

$$\int_0^\varrho DT(x, u)\,dt + \int_0^\varrho [DT(x+tu; u)-DT(x, u)]\,dt = \varrho DT(x, u)+\varrho v(\varrho),$$
$$\|v(\varrho)\| \leqslant \varepsilon(\varrho\|u\|)\|u\|.$$

This proves condition (B′) with the strong topology and finishes the proof.

Our main concern in the present paper will be with Lipschitzian mappings T.

DEFINITION 2. If $T: G \to Y$, G open in X, we put

$$(3) \qquad M_T(G) = \sup_{\substack{x,y \in G \\ x \neq y}} \frac{\|T(x)-T(y)\|}{\|x-y\|}.$$

If $M_T(G)$ is finite, we say that T is *Lipschitzian* with (Lipschitz-) constant $M_T(G)$. We say that T is *locally Lipschitzian* in G if for every $x \in G$ there exists a ball $B_R(x) \subset G$, $R > 0$ such that T is Lipschitzian in $B_R(x)$. We put then

$$(3') \qquad M_T(x) = \inf_{\substack{R>0 \\ B_R(x) \subset G}} M_T(B_R(x)).$$

It is clear that $M_T(x)$ is an upper semi-continuous function in G. Since we are interested in differentiability, which is a local property, our results for Lipschitzian mappings will imply similar results for locally Lipschitzian mappings. From now on in the present chapter, unless otherwise stated, we will assume that T is a Lipschitzian mapping in G with constant $M = M_T(G) < \infty$.

We will continue our assumption that X is separable and restrict Y to be a conjugate space of a separable Banach space $^*Y$ (a pre-conjugate of Y). The topology τ on Y will be fixed as the topology $\sigma(Y, {}^*Y)$ (i.e., the weak*-topology relative to $^*Y$) or the strong topology on Y; in the last case we will usually omit the use of τ.

Since the pre-conjugate is not uniquely determined the topology τ will depend on the choice of $^*Y$; therefore, for each choice of Y we will consider the choice of $^*Y$ as fixed. These assumptions have consequences which will be constantly used, namely that any closed finite ball in Y

(10) $\mathscr{B}(X, Y)$ is the class of all linear strongly continuous mappings $X \to Y$ made into a Banach space with the usual norm.

is metrizable and compact in the topology τ and that the norm of y, $y \in Y$, is lower semi-continuous in the topology τ.

PROPOSITION 6. *The differential quotient* $\dfrac{T(x+\varrho u)-T(x)}{\varrho}$ *for fixed* x *and* ϱ *as mapping:* $\dfrac{1}{\varrho}(G-x) \to Y$ *is bounded by* $M\|u\|$ *and is Lipschitzian with constant* $\leqslant M$.

The proof is immediate.

PROPOSITION 7. *If* $DT(x)$ *rel.* τ *exists, then* (B) *is satisfied and* $DT(x; u)$ *as function of* u *is a Lipschitzian mapping:* $X \to Y$ *with constant* $\leqslant M$.

Proof. By Proposition 6 the mappings $\dfrac{T(x+\varrho u)-T(x)}{\varrho}$ for fixed x and variable ϱ are uniformly equicontinuous in norms; hence a fortiori with the weaker topology τ on Y. Therefore they are uniformly τ-convergent for $\varrho \searrow 0$ on every compact. On the other hand the lower semi-continuity of the norm in Y rel. τ assures the second assertion.

PROPOSITION 8. *If* $DT(x; u)$ *exists rel.* τ *(or the strong topology) on a dense sequence* $\{u_n\} \subset X$, *then* $DT(x)$ *exists rel.* τ *(or the strong topology).*

The proof follows from the general fact that if uniformly equicontinuous functions converge on a dense subset of a domain they converge in the whole domain if the range of all the functions is in a space with uniform complete topology. Here the functions are $\big(T(x+\varrho u)-T(x)\big)/\varrho$ and choosing the domain $\|u\| < R$, the images are in $\overline{B_{RM}(0)}$ which is complete in the norm topology and even compact in the τ-topology.

PROPOSITION 9. *For Lipschitzian mappings* T, $DT(x; u)$ *exists rel.* τ *(or strongly) if and only if* $k\left(T\left(x+\dfrac{1}{k}u\right)-T(x)\right)$ *converges rel.* τ *(or strongly) for integer* $k \nearrow \infty$.

Proof. Take any $\dfrac{1}{k+1} \leqslant \varrho < \dfrac{1}{k}$ then

$$\left\|\frac{T(x+\varrho u)-T(x)}{\varrho} - k\left(T\left(x+\frac{1}{k}u\right)-T(x)\right)\right\|$$

$$= \left\|\frac{T(x+\varrho u)-T\left(x+\frac{1}{k}u\right)}{\varrho} + \left(\frac{1}{\varrho}-k\right)\left(T\left(x+\frac{1}{k}u\right)-T(x)\right)\right\|$$

$$\leqslant \frac{M\left|\varrho-\frac{1}{k}\right|\|u\|}{\varrho} + \left(\frac{1}{\varrho}-k\right)M\frac{1}{k}\|u\| = 2M\|u\|\frac{1-\varrho k}{k\varrho}.$$

Since $1-\varrho k \leqslant \varrho$ we get that the norm of the difference is majorated by $2M\|u\|1/k$ which proves our assertion.

PROPOSITION 10. *For a Lipschitzian mapping* T *the set of* x's, *where* $DT(x; u)$ *does not exist (for fixed* u*) is a Borel set; the same is true for the set, where* $DT(x)$ *does not exist.*

Proof. In fact the first set is a set, where the sequence of continuous functions $k\left(T\left(x+\dfrac{1}{k}u\right)-T(x)\right)$ does not converge and the second set is the one, where for a fixed dense sequence $\{u_n\} \subset X$ at least for one n the sequence $k\left(T\left(x+\dfrac{1}{k}u_n\right)-T(x)\right)$ does not converge (we use here Proposition 8).

2. The main theorem.

LEMMA 1. *Consider a Lipschitzian mapping* T *of an interval* $I \subset R^1$ *into* Y. *Then* dT/dt *exists rel.* τ *for almost all* $t \in I$. *If* Y *is separable, then* dT/dt *exists strongly a.e.* (11)

Proof. For the first assertion take a dense sequence $\{v_n\}$ in ${}^*Y$ and consider $\langle T(t), v_n\rangle$. It is a real function of the real variable t, Lipschitzian with constant $\leqslant M\|v_n\|$. Hence $\dfrac{d}{dt}\langle T(t), v_n\rangle$ exists a.e. for each n; hence also, for all n's simultaneously. Thus almost everywhere there exists $\dfrac{dT(t)}{dt}$ determined by the set of relations $\dfrac{d}{dt}\langle T(t), v_n\rangle = \left\langle\dfrac{dT(t)}{dt}, v_n\right\rangle$ for all n. dT/dt is thus the τ-derivative of $T(t)$ for almost all t. It follows that dT/dt is a bounded Borel function rel. τ.

If Y is separable the Borel sets in Y rel. τ are the same as rel. the strong topology (12). Hence dT/dt is a Borel function in the strong topology too. We have the representation by a τ-integral

$$T(t_0+\varrho)-T(t_0) = \int_{t_0}^{t_0+\varrho}\frac{dT(t)}{dt}\,dt.$$

This integral is actually a Bochner integral and hence the derivative $dT(t)/dt$ is a strong derivative for almost all t.

COROLLARY 1. *Let* T *be a Lipschitzian mapping* $T\colon G \to Y$, G *open in* X. a) *The set where* $DT(x)$ *does not exist rel.* τ *is in* $\mathfrak{A}$. b) *If* Y *is separable the set where* $DT(x)$ *does not exist strongly is in* $\mathfrak{A}$.

(11) This Lemma is essentially well known but is usually considered in another context. That is why we give here a short proof of it.

(12) Since a closed ball in Y is closed (even compact) rel. τ.

Proof. We take any complete sequence $\{a_n\} \subset X$. We form the sequence $\{a''_n\}$ of all finite linear combinations of the a_n's with rational coefficients. This is a dense sequence in X with $\mathfrak{A}\{a''_n\} = \mathfrak{A}\{a_n\}$ (see Remark 1, Section 1, Chapter I). Considering T restricted to $G_{x,n} = G\cap(x+ +\boldsymbol{R}^1 a''_n)$ for any $x \in X$, $n = 1, 2, \ldots$, we apply Lemma 1 and obtain that for $x' \in G_{x,n}$, $DT(x; a''_n)$ exists a.e. Hence the set of x's where $DT(x; a''_n)$ does not exist is in $\mathfrak{A}(a''_n)$. It follows that the set, where $DT(x; a''_n)$ does not exist for at least one n, is a Borel set belonging to $\mathfrak{A}\{a''_n\} = \mathfrak{A}\{a_n\}$. In the complement of this set $DT(x; a''_n)$ exists for all n rel. τ or strongly depending on case a) or b) of the corollary. Therefore, by Proposition 8 of Section 1 the set where $DT(x)$ does not exist is in $\mathfrak{A}\{a_n\}$ for every complete sequence $\{a_n\}$. Hence it is in the class $\mathfrak{A}$.

The next lemma is an extension of the classical Rademacher's theorem when $X = \boldsymbol{R}^n$ and Y is a Banach space satisfying our assumption.

LEMMA 2. *Let T be a Lipschitzian mapping of an open set $G \subset \boldsymbol{R}^n$ into Y. Then a.e. $DT(x)$ exists rel τ and is a τ-differential.*

Proof. That $DT(x)$ exists rel. τ a.e. follows from Corollary 1 (since $\mathfrak{A}$ in $\boldsymbol{R}^n$ is the class of sets of Lebesgue measure 0). Consider a sequence $\{b_k\}$ dense in $^*Y$. It is enough to show that $\langle DT(x; u), b_k\rangle$ is linear in u for all k. At points x, where $DT(x)$ rel. τ exists we have obviously

$$\langle DT(x; u), b_k\rangle = DT_k(x; u), \quad \text{where} \quad T_k(x) = \langle T(x), b_k\rangle.$$

But $T_k\colon G \to \boldsymbol{R}^1$ is Lipschitzian and by the classical Rademacher's theorem has a.e. in G a linear differential. Our lemma follows then immediately.

Remark 1. One can give a very short proof of the classical Rademacher's theorem by regularizing the mapping $T\colon G \to \boldsymbol{R}^m$ and using Corollary 1.

THEOREM 1. (Main Theorem). *Let T be a Lipschitzian mapping of an open set $G \subset X$ into Y, X and Y satisfying our general requirements. Then*: a) *$DT(x)$ is a τ-differential exc. $\mathfrak{A}$,* b) *if Y is separable, then $DT(x)$ is a differential exc. $\mathfrak{A}$,* c) *if Y is reflexive (without being necessarily separable), then $DT(x)$ is also a differential exc. $\mathfrak{A}$.*

Proof. b) follows from a) by Corollary 1, b).

To prove a) consider any complete sequence $\{a_n\} \subset X$. By rational linear combinations of the a_n's we form the dense sequence $\{a''_n\} \subset X$. If $DT(x)$ exists it is Lipschitzian and hence it is linear if for any two vectors (a''_{n_1}, a''_{n_2})

$$(*) \qquad DT(x; a''_{n_1}+a''_{n_2}) = DT(x; a''_{n_1})+DT(x; a''_{n_2}).$$

Let $E = \{x \in G\colon DT(x) \text{ rel. } \tau \text{ exists}\}$; $G\setminus E \in \mathfrak{A}$. Let further $S_{n_1,n_2} = [a''_{n_1}, a''_{n_2}]$. Then the set of x's in E where $(*)$ does not hold is by Lemma 2 in $\mathfrak{A}(S_{n_1,n_2}) = \mathfrak{A}\{a''_{n_1}, a''_{n_2}\}$, and therefore the set where $(*)$ does not hold for at least one couple (n_1, n_2) is in $\mathfrak{A}\{a''_n\} = \mathfrak{A}\{a_n\}$. The sequence $\{a_n\}$ being an arbitrary complete sequence in X the assertion a) is thus proved.

To prove c) we remark that T is actually a mapping into $Y_1 \subset Y$, where Y_1 is the closed subspace of Y spanned by $T(G)$, $Y_1 = \overline{[T(G)]}$. Y_1 is reflexive and separable, therefore, we can apply part b).

Remark 2. In case a) of the main theorem the τ-differential existing exc. $\mathfrak{A}$ may depend on the choice of $^*Y$ which determines the topology τ. It is an open question if for two such choices of $^*Y$ the resulting τ-differentials will coincide exc. $\mathfrak{A}$. By the main theorem, b), this is actually the case when Y is separable.

Remark 3. It may happen that we have a Lipschitzian mapping $T\colon G \to Y$, where Y does not satisfy our requirements and we wish, however, to define a differential of T. We may achieve it by enlarging suitably the space Y. We remark first that we can always consider Y separable, otherwise, we will replace it by its closed subspace $\overline{[T(G)]}$. With Y assumed separable we take any sequence $\{y_n\}$ dense in Y and consider a corresponding sequence $\{y^*_n\} \subset Y^*$ (for $y \in Y$, y^* is any of the points of Y^*, existing by Hahn–Banach Theorem, which satisfy $\|y^*\| = \|y\|$ and $\langle y, y^*\rangle = \|y\|^2$). We consider then the closed subspace $Z = \overline{[\{y^*_n\}]} \subset Y^*$, where the closure is in the norm topology of Y^*. Then it is easy to see that Y is embedded isomorphically and isometrically in Z^*, hence T is a Lipschitzian mapping of G into Z^* and Z^* satisfies our requirements. Thus exc. $\mathfrak{A}$ $DT(x)$ exists now in Z^* and is a τ-differential where τ is the weak $*$-topology on Z^*. Obviously, depending on our choice of $\{y_n\}$ and $\{y^*_n\}$ we will have different representations of $DT(x)$.

3. Some counter-examples.

EXAMPLE I. We will construct here an example of a Lipschitzian mapping T of a separable Hilbert space $\mathscr{H}$ onto itself (which will be actually a Lipschitzian automorphism) such that each point $x \in \mathscr{H}$ has a differential $DT(x)$ in our sense and at no point $x \in \mathscr{H}$, $DT(x)$ is a Fréchet differential. Such examples are known([13]). We give here our example since it is more explicit and therefore, more illustrative.

Let $\{\lambda_j\}$ be a sequence of positive real numbers such that $\lambda_j \searrow 0$ and $\sum_{j=1}^{\infty} \lambda_j^2 = \infty$. Let further φ be a real valued function in $C^\infty(\boldsymbol{R}^1)$ such that

1° $\varphi(t)$ is increasing and anti-symmetric i.e. $\varphi(t) = -\varphi(-t)$.

2° $\varphi(t) = t$ for $0 \leqslant t \leqslant 1$, $\varphi(t) = t+2$ for $t \geqslant 2$.

3° For every $t \in \boldsymbol{R}^1$, $0 < M'_0 \leqslant \varphi'(t) \leqslant M'$, $|\varphi''(t)| \leqslant M''$, with M'_0, M' and M'' finite positive constants.

([13]) See, for instance, M. Sova [11].

In $\mathscr{H}$ choose an orthonormal basis $\{e_n\}$ and define

$$(1) \qquad \text{for} \quad x = \sum_1^\infty \xi_n e_n \in \mathscr{H}, \quad T(x) = \sum_1^\infty \lambda_n \varphi\left(\frac{\xi_n}{\lambda_n}\right) e_n.$$

It is clear that T is a Lipschitzian mapping with constant $\leqslant M'$ and that it is actually a Lipschitzian automorphism of $\mathscr{H}$ with T^{-1} having a Lipschitzian constant $\leqslant 1/M_0'$. Take an arbitrary $u \in \mathscr{H}$, $u = \sum_1^\infty \theta_n e_n$ and form the differential quotient

$$(2) \qquad \frac{T(x+\varrho u)-T(x)}{\varrho} = \sum_1^\infty \frac{\lambda_n}{\varrho}\left(\varphi\left(\frac{\xi_n+\varrho\theta_n}{\lambda_n}\right) - \varphi\left(\frac{\xi_n}{\lambda_n}\right)\right) e_n.$$

We may assume that $\varrho < 1$. The set of all positive integers N can be divided in two subsets $N_\varrho^{(1)}$ and $N_\varrho^{(2)}$;

$$N_\varrho^{(1)} = \{n \in N: \frac{\varrho^{1/2}|\theta_n|}{\lambda_n} \leqslant 1\} \quad \text{and} \quad N_\varrho^{(2)} = N \setminus N_\varrho^{(1)}.$$

It is clear that when $\varrho \searrow 0$, $N_\varrho^{(2)} \searrow \emptyset$. The coefficients in development (2) will be evaluated separately for $n \in N_\varrho^{(1)}$ and $n \in N_\varrho^{(2)}$. For $n \in N_\varrho^{(1)}$ we write

$$\frac{\lambda_n}{\varrho}\left(\varphi\left(\frac{\xi_n}{\lambda_n}+\frac{\varrho\theta_n}{\lambda_n}\right) - \varphi\left(\frac{\xi_n}{\lambda_n}\right)\right) = \frac{\lambda_n}{\varrho}\left(\frac{\varrho\theta_n}{\lambda_n}\varphi'\left(\frac{\xi_n}{\lambda_n}\right) + \left(\frac{\varrho\theta_n}{\lambda_n}\right)^2 a_n\right) = \theta_n\varphi'\left(\frac{\xi_n}{\lambda_n}\right) + \frac{\theta_n^2\varrho}{\lambda_n}a_n,$$

$$\text{where} \quad |a_n| \leqslant M''.$$

Since $\dfrac{\varrho^{1/2}|\theta_n|}{\lambda_n} \leqslant 1$, we get

$$\left|\frac{\lambda_n}{\varrho}\left(\varphi\left(\frac{\xi_n+\varrho\theta_n}{\lambda_n}\right) - \varphi\left(\frac{\xi_n}{\lambda_n}\right)\right) - \theta_n\varphi'\left(\frac{\xi_n}{\lambda_n}\right)\right| \leqslant \varrho^{1/2}|\theta_n|\, M''.$$

For $n \in N_\varrho^{(2)}$ we write

$$\left|\frac{\lambda_n}{\varrho}\left(\varphi\left(\frac{\xi_n+\varrho\theta_n}{\lambda_n}\right) - \varphi\left(\frac{\xi_n}{\lambda_n}\right)\right)\right| \leqslant \frac{\lambda_n}{\varrho}\,\frac{\varrho|\theta_n|}{\lambda_n} M' = |\theta_n|\, M'.$$

Writing the whole sum in (2) as $\sum_{N_\varrho^{(1)}} + \sum_{N_\varrho^{(2)}}$ it is now clear that when $\varrho \searrow 0$, $\sum_{N_\varrho^{(2)}}$ is converging strongly to 0 whereas $\sum_{N_\varrho^{(1)}}$ converges strongly to $\sum \theta_n \varphi'\left(\frac{\xi_n}{\lambda_n}\right) e_n$.

Thus $DT(x)$ exists in the strong topology and $DT(x; u) = \sum_1^\infty \theta_n\varphi'\left(\frac{\xi_n}{\lambda_n}\right) e_n$. T being Lipschitzian, condition (B) is automatically satisfied and (A) is obvious from the formula for $DT(x; u)$. Hence T has a differential in our sense everywhere.

This differential is not a Fréchet differential for any $x \in \mathscr{H}$. In fact, let $x = \sum_1^\infty \xi_n e_n$ and consider any $\varepsilon > 0$. Since $\lambda_i \searrow 0$ there exists j_0 such that $\lambda_j < \varepsilon/3$ for all $j > j_0$. Since $\sum_1^\infty \lambda_j^2 = \infty$ and $\sum_1^\infty |\xi_j|^2 < \infty$ there exist infinitely many $k > j_0$ such that $|\xi_k| \leqslant \frac{1}{2}\lambda_k$. For every such k

$$\left\|\frac{T(x+3\lambda_k e_k) - T(x)}{3\lambda_k}\right\| = \left|\frac{\varphi(3+\xi_k/\lambda_k) - \varphi(\xi_k/\lambda_k)}{3}\right|.$$

Since $3+(\xi_k/\lambda_k) > 2$ and $-1/2 \leqslant \xi_k/\lambda_k \leqslant 1/2$, by properties 1° and 2° of φ the last expression is equal to $\frac{1}{3}\left(3+\frac{\xi_k}{\lambda_k}+2-\frac{\xi_k}{\lambda_k}\right) = \frac{5}{3}$. This shows that on the bounded set of e_k's for the above chosen integers k the convergence of the differential quotient to $DT(x, e_k) = \varphi'\left(\frac{\xi_k}{\lambda_k}\right)e_k = e_k$ cannot be uniform.

In view of Theorem 3, Section 1 of the present chapter it follows that $DT(x)$ as a mapping of $\mathscr{H}$ into $\mathscr{B}(\mathscr{H}, \mathscr{H})$ cannot be strongly continuous in any open set of x's. In fact, it is easy to check that $DT(x)$ is not strongly continuous at any point x in $\mathscr{H}$.

Example I'. Using the same function φ and sequence $\{\lambda_n\}$ as in the previous example we can construct a real valued function f on the space l^1 as follows: for $x = (\xi_1, \xi_2, \ldots) \in l^1$ we put

$$f(x) = \sum_1^\infty \lambda_n \varphi\left(\frac{\xi_n}{\lambda_n}\right).$$

By completely analogous argument (as in Example I) we show that for $u = (\theta_1, \theta_2, \ldots) \in l^1$ we have $Df(x; u) = \sum \theta_n \varphi'(\xi_n/\lambda_n)$ is a differential in our sense but not in Fréchet sense.

Example II. We will consider now a mapping T of the interval $[0; 1] \subset R^1$ into $L^1[0; 1]$ defined as follows. For $0 \leqslant x \leqslant 1$ let χ_x be the characteristic function of the closed interval $[0; x]$ and put $T(x) = (t+1)\chi_x(t)$. This mapping is obviously Lipschitzian with Lipschitz constant 2. If $L^1[0; 1]$ were a conjugate space, by our main theorem this mapping T would have a strong differential for almost all x's. This is, however, impossible since the differential quotient obviously cannot converge strongly at any point x. This gives the shortest proof to our knowledge that $L^1[0; 1]$ is not a conjugate space.

Suppose that instead of considering our mapping T as a mapping into $L^1[0;1]$ we consider it as a mapping into the space $\mathfrak{M}[0;1]$ of all Borel finite measures on the interval. This is possible by identification of functions $f \in L^1[0;1]$ with measures $f dt \in \mathfrak{M}[0;1]$. This identification is actually an isometry if on $\mathfrak{M}[0;1]$ we take for norm $\|\mu\| =$ the total mass of μ. Since $\mathfrak{M}[0;1]$ is the conjugate space of the space $C^0[0;1]$ applying again our main theorem we have the existence a.e. of the τ-differential and, in fact, one sees immediately that for every $x \in (0;1)$ the differential quotient converges in the weak *-topology of $\mathfrak{M}$ to the measure $(x+1)\,\delta(x)\,u$, where $\delta(x)$ is the Dirac point measure.

This example shows that if the space Y is not separable in general we will not have a strong differential anywhere.

CHAPTER III

APPLICATIONS

1. Compositions of Lipschitzian mappings with linear compact operators. Let T: $G \to Y$ be a Lipschitzian operator. If we compose it with a compact linear operator we may expect to have some stronger properties of the differential or be able to abandon certain restrictions. We will consider two cases:

I. *Left composition.* We will skip the requirement that Y be a conjugate space but will maintain X separable. Let K be a compact linear operator, in the norm topology, of Y into Y_1 (Y_1 a Banach space). Hence KT: $G \to Y_1$ is a Lipschitzian mapping. We can proceed as in Remark 3, Section 3, Chapter II and replace Y_1 by $Y_2 = \overline{[KT(G)]} \subset Y_1$. Then, as in the remark, we construct a subspace $Z \subset Y_2^*$ such that Y_2 is embedded isomorphically and isometrically in Z^*. We have, therefore, a τ-differential $DKT(x)$ in Z^*, where τ is the weak*-topology of Z^*. Consider now for fixed $u \in X$ the set $S_{x,u}$ of the differential quotients

$$\frac{KT(x+\varrho u)-KT(x)}{\varrho} = K\left(\frac{T(x+\varrho u)-T(x)}{\varrho}\right).$$

This set is precompact in Y_2 (since the set of $\dfrac{T(x+\varrho u)-T(x)}{\varrho}$ is bounded in Y). Hence, the closure $\overline{S_{x,u}} \subset Y_2$ is compact in the norm-topology and, since the τ-topology is weaker, the two topologies on $\overline{S_{x,u}}$ coincide. Therefore, the τ-limit of the differential quotients is their strong limit, hence we get this statement:

THEOREM 1. *$DKT(x)$ is a differential exc. $\mathfrak{A}$.*

II. *Right composition.* We come back to our requirement that Y be a conjugate space of a separable Banach space. Consider a compact linear operator $K: X_1 \to X$, X_1 a separable Banach space. Then $G_1 = K^{-1}(K(X_1) \cap G)$ is an open subset of X_1 and TK: $G_1 \to Y$. By our main theorem of Section 2, Chapter II, $DTK(x_1)$ is a τ-differential exc. $\mathfrak{A}_{X_1}$. Consider the differential quotients

$$\frac{TK(x_1+\varrho u_1)-TK(x_1)}{\varrho} = \frac{T(Kx_1+\varrho Ku_1)-T(Kx_1)}{\varrho}.$$

One sees immediately that the existence of $DTK(x_1; u_1)$ is equivalent to the existence of $DT(Kx_1; Ku_1)$. If we take any bounded set of u_1's it is transformed into a precompact set of Ku_1's. Hence condition (B′) for TK is implied by condition (B) for T. The last condition being true wherever $DT(Kx_1)$ exists (see Proposition 7, Section 1, Chapter II), we obtain the following statement:

THEOREM 2. *$DTK(x_1)$ is a Fréchet τ-differential exc. $\mathfrak{A}_{X_1}$.*

Remark 1. If in Theorem 2 we add to our requirements that Y be either separable or reflexive, the conclusion will be that $DT(x_1)$ is a Fréchet differential exc. $\mathfrak{A}_{X_1}$.

Remark 2. K_1: $Y \to Y_1$ and K_2: $X_1 \to X$ satisfy the requirements of Theorem 1 or 2 respectively, Y being an arbitrary Banach space, we get that $DK_1TK_2(x_1)$ exists and is a Fréchet differential exc. $\mathfrak{A}_{X_1}$.

2. Convex functions. In Section 1, Chapter II, we introduced convex functions in a special case when they were defined on the whole Banach space X. We will now consider convex functions defined on an open set $G \subset X$.

DEFINITION 1. For any set $S \subset X$, we denote by $C[S]$ the convex span of S, i.e. the set of all convex combinations of elements of S; by $CC[S]$ we will denote the convex-cone span of S, i.e. the set of all finite linear combinations with positive coefficients of elements of S.

It is clear that $C[x_1, \ldots, x_n]$ is the $(n-1)$-dimensional simplex formed by the n points x_k and that $\overline{CC[S]}$ is the closed convex cone with vertex at 0 generated by elements of S.

DEFINITION 2. Function F: $G \to \mathbf{R}^1$ is called *convex* if for any finite sequence $(x_1, \ldots, x_n)$ such that $C[x_1, \ldots, x_n] \subset G$ and for any convex combination $x = \sum_1^n a_k x_k$ we have

$$F(x) \leqslant \sum a_k F(x_k). \tag{1}$$

Sometimes a more restricted definition of convexity is used, valid for an arbitrary domain of definition G which says that whenever $(x_1, \ldots, x_n)$

$\subset G$ and the convex combination $x = \sum a_k x_k \in G$ we have inequality (1). Classical properties of convex functions on finite dimensional spaces give immediately:

PROPOSITION 1. *For the convexity of F in G it is sufficient (and obviously necessary) that F be convex on each straight line segment contained in G.*

In an obvious way we can define "local convexity" but this notion is redundant because of the following proposition:

PROPOSITION 2. *If F is locally convex in G, then it is convex in G.*

Proof. By Proposition 1 we have to show that F is convex on any closed segment contained in G. By local convexity this segment can be covered by a finite number of open segments on which F is convex. Again by a classic property of convex functions on intervals, F, is convex on the whole closed segment.

PROPOSITION 3. *For a convex F in G the three following properties are mutually equivalent:*

a) *F is locally Lipschitzian,*

b) *F is continuous,*

c) *F is locally bounded.*

Proof. Implications a)$\Rightarrow$b) and b)$\Rightarrow$c) are obvious. To prove c)$\Rightarrow$a) consider for an arbitrary $x \in G$ a ball $B_{2R}(x) \subset G$, $R > 0$ such that $|F(y)| \leqslant C < \infty$ for $y \in B_{2R}(x)$. For any two points $x_1 \neq x_2$ in $B_R(x)$ consider

$$x_0 = \frac{x_1 + x_2}{2} \quad \text{and} \quad u = \frac{x_2 - x_1}{\|x_2 - x_1\|}.$$

On the straight line $x_0 + \boldsymbol{R}^1 u$ we have then the segment $[x_1 - Ru; x_2 + Ru]$ contained in $B_{2R}(x)$ and containing the points x_1, x_0 and x_2. The function $f(t) = F(x_0 + tu)$ is convex in the real variable t and bounded by C for

$$-R - \frac{\|x_1 - x_2\|}{2} \leqslant t \leqslant R + \frac{\|x_2 - x_1\|}{2}.$$

It follows that $f'_d(t)$ (the right derivative) exists everywhere in the interval, is an increasing function and we have

$$-2C \leqslant f(t_2) - f(t_1) = \int_{t_1}^{t_2} f'_d(t)\,dt \leqslant 2C.$$

By applying it to the two intervals

$$\left[-R - \frac{\|x_2 - x_1\|}{2}, \frac{-\|x_2 - x_1\|}{2}\right] \quad \text{and} \quad \left[\frac{\|x_2 - x_1\|}{2}, \frac{\|x_2 - x_1\|}{2} + R\right]$$

by mean value theorem we obtain that there exists t' and t'' in each of the intervals respectively such that $-2C/R \leqslant f'_d(t') \leqslant f'_d(t'') \leqslant 2C/R$.

Finally,

$$|F(x_2) - F(x_1)| = \left| \int_{-\|x_2 - x_1\|/2}^{\|x_2 - x_1\|/2} f'_d(t)\,dt \right| \leqslant \frac{2C}{R} \|x_2 - x_1\|.$$

Contrary to what happens in finite dimensional spaces a convex function in an infinite dimensional Banach space need not be in general continuous. From now on we will be assuming that the convex function $F(x)$ is continuous, hence, locally Lipschitzian.

THEOREM 1. *For a continuous convex function $F\colon G \to \boldsymbol{R}^1$ the Gateaux-differential $DF(x)$ exists everywhere in G. As function of u, $DF(x; u)$ is convex and Lipschitzian with a constant $\leqslant M_F(x)$ ([14]). $DF(x)$ is a differential in G exc. $\mathfrak{A}^0$.*

Proof. That $DF(x; u)$ exists for every $x \in G$, $u \in X$ reduces to the fact that a convex function of the real variable has a right derivative everywhere (where it is defined). That $DF(x; u)$ is convex follows from the fact that each differential quotient $\dfrac{F(x + \varrho u) - F(x)}{\varrho}$ is obviously convex as function of u. That $DF(x; u)$ is Lipschitzian with constant $\leqslant M_F(x)$ follows from Proposition 7, Section 1, Chapter II and the fact that $DF(x; u)$ is determined by F in arbitrarily small neighborhood of x. The last part of the theorem follows from Theorem 2, Section 1, Chapter II.

Remark 1. The last part of Theorem 1 could be proved in an interesting way by applying the theorem of Zarantonello (Theorem Z, Section 4, Chapter I). We use the fact that $DF(x; u)$ being convex and Lipschitzian is completely determined by its supporting linear functionals $\varphi(u)$, i.e. such that $\varphi(u) \leqslant DF(x; u)$ for all $u \in X$. These $\varphi(u)$ form a subset $\nabla F(x) \subset X^*$, called the (multivalued) gradient of F. $\nabla F(x)$ is obviously convex and weak *-closed. It is known that $\nabla F(x)\colon G \to X^*$ is a multi-valued monotone mapping ([15]). The mapping is single-valued at x if and only if $DF(x; u)$ is linear in u. Hence Theorem Z gives the last statement of our theorem.

EXAMPLE. An interesting application of Theorem 1 is obtained by taking $F(x) = \|x\|$. Here $G = X$. Obviously $F(x)$ is convex and Lipschitzian with constant 1. Thus the Gateaux differential $DF(x)$ exists for every x. It is certainly not a differential for $x = 0$. The set of points where $DF(x)$ is not a differential belongs to the class $\mathfrak{A}^0$. The fact that $DF(x)$ is a differential has the following simple geometric interpretation. We consider

([14]) See (3′), Section 1, Chapter II.

([15]) See H. Brezis, [2], p. 21, Example 2.1.4. In this reference the gradient is called sub-differential.

the closed ball $\overline{B_{\|x\|}(0)}$; we have $x \in \partial B_{\|x\|}(0)$ and $DF(x)$ is a differential (i.e. linear) if and only if there exists only one supporting hyper-plane (of co-dimension 1) at x to $\overline{B_{\|x\|}(0)}$.

3. Convex mappings and generalizations. If we want to extend the notion of convexity from real-valued functions to mappings between Banach spaces we have to consider in the image space Y a suitable convex cone Γ playing the role of positive numbers in $\boldsymbol{R}^1$. To define the notions suitably we will assume certain hypotheses. The space Y will again be required to be a conjugate space of a separable Banach space $^*Y$. In $^*Y$ we will consider a complete sequence $\{b_n\} \subset {}^*Y$ and set

(1) $$\Gamma = \{y \in Y;\ \langle y, b_n\rangle \geqslant 0 \text{ for all } n\}.$$

With τ the topology $\sigma(Y, {}^*Y)$ it is obvious that:

PROPOSITION 1. *Γ is a τ-closed convex cone with vertex 0 in Y not containing any straight line.*

Γ may be reduced to the element 0.

DEFINITION 1. A mapping $T: G \to Y$, G open in X, X separable, is said *rel. convex* (*rel.* Γ) if for every finite number of points $x_1, \ldots, x_n$ with $C[x_1, \ldots, x_n] \subset G$ and for each convex combination $x = \sum a_k x_k$:

(2) $$\sum a_k T x_k - T(x) \in \Gamma.$$

Most often we will skip the parenthesis (rel. Γ).

PROPOSITION 2. *T is rel. convex if and only if all the real-valued functions $\langle T(x), b_n\rangle$, $n = 1, 2, \ldots$, are convex.*

The proof is immediate.

Remark 1. When $Y = \boldsymbol{R}^1$ there are only two cones Γ available, namely the one of non-negative numbers and the other of non-positive numbers. In the first case we get the usual notion of convex functions, in the other case we get the notion of concave functions.

For rel. convex mappings it does not seem that a proposition analogous to Proposition 3 of Section 2 is valid. Therefore, instead of continuity we will assume that our mapping T is locally Lipschitzian, in abbreviation we will write T l.L. rel. convex.

THEOREM 1. *For a mapping $T: G \to Y$ which is l.L. rel. convex, the Gateaux τ-differential $DT(x)$ exists everywhere in G. As function of u $DT(x; u)$ is rel. convex and Lipschitzian with constant $\leqslant M_T(x)$. $DT(x)$ is a τ-differential in G exc. $\mathfrak{A}^0$. $DT(x+tu; u)$ is, as function of the real variable t, right τ-continuous.*

Proof. For the real-valued functions $T_n(x) = \langle T(x), b_n\rangle$, $DT_n(x)$ exists for every x (by Theorem 1, preceding section). Hence $\left\langle \frac{T(x-\varrho u) - T(x)}{\varrho}, b_n \right\rangle$ converges for $\varrho \searrow 0$ and for every n. Since by the locally Lipschitzian character of T for sufficiently small ϱ the differential quotients are uniformly bounded and $\{b_n\}$ is complete in $^*Y$, $\left\langle \frac{T(x+\varrho u) - T(x)}{\varrho}, b_n \right\rangle$ has a limit for every $b \in {}^*Y$. Hence, the differential quotient converges in τ-topology. The differential quotients as functions of u are rel. convex. The same is true of their τ-limit (we use here the fact that Γ is τ-closed). That $DT(x; u)$ is Lipschitzian with constant $\leqslant M_T(x)$ is obtained as in the theorem of the preceding section. That exc. $\mathfrak{A}^0$ $DT(x)$ is linear is reduced to the fact that for each $T_n(x) = \langle T(x), b_n\rangle$, $DT_n(x)$ is linear exc. $\mathfrak{A}^0$. Finally, considering the function $f(t) = DT(x+tu; u)$ we have that $\langle f(t), b_n\rangle = DT_n(x+tu; u)$. Since $DT_n(x+tu; u)$ is the right-derivative of the function $f_n(t) = T_n(x+tu)$ which is real-valued and convex of the real variable t we know that it is right-continuous. Again, using the fact that sequence $\{b_n\}$ is complete and that T is locally Lipschitzian, hence, $\|f(t)\| \leqslant M_T(x+tu)\|u\|$, $M_T(x+tu)$ being upper semi-continuous as function of t, we obtain finally the last statement of the theorem.

We will next enlarge the class of mappings for which our theorem is valid.

DEFINITION 2. A mapping $T: G \to Y$ is called *var. convex* (abbreviation for variably convex) if each point of G is contained in an open neighbourhood, where T is rel. convex relatively to a cone Γ varying in general with the neighborhood.

Since var. convex means locally rel. convex we obtain immediately:

THEOREM 1'. *For a mapping $T: G \to Y$ which is l.L. var. convex the Gateaux τ-differential $DT(x)$ exists everywhere in G. As function of u $DT(x; u)$ is rel. convex and Lipschitzian with constant $\leqslant M_T(x)$. $DT(x)$ is a τ-differential in G exc. $\mathfrak{A}^0$. $DT(x+tu; u)$ is, as function of the real variable t, right τ-continuous.*

To introduce more general mappings for which our theorem can be extended we have to introduce certain notions and notations.

If U is open in X we consider the space $\mathrm{Lip}(U, Y)$ of all Lipschitzian mappings $U \to Y$. This space is obviously linear and we can introduce there a norm by fixing any $x_0 \in U$ as follows: $\|T\| = \|T(x_0)\| + M_T$, where M_T is the Lipschitz constant of T in U. It is immediately checked that $\|T\|$ is actually a norm and that different choices of x_0 give equivalent norms. Furthermore, with this norm, $\mathrm{Lip}(U, Y)$ is a complete Banach space. In this Banach space consider the subset of all var. convex mappings.

The closed linear span of this subset forms a class of mappings which we will call *convexoid*. Finally we will consider the locally convexoid mappings. These mappings are obviously locally Lipschitzian. We get then the following theorem:

THEOREM 1″. *For locally convexoid mappings $T: G \to Y$ the Gateaux τ-differential $DT(x)$ exists everywhere in G. As function of u $DT(x; u)$ is convexoid and Lipschitzian with constant $\leqslant M_T(x)$. $DT(x)$ is a τ-differential in G exc. $\mathfrak{A}^0$. $DT(x+tu; u)$ is, as function of the real variable t, right τ-continuous.*

Proof. Since all the statements in the theorem are local, we can restrict ourselves to convexoid mappings. If T is a finite linear combination of Lipschitzian var. convex mappings the theorem follows immediately from Theorem 1′. Let now T be convexoid and T_n be finite linear combinations of var. convex Lipschitzian maps such that $\|T - T_n\|$ converges to 0. It follows that T_n is a Cauchy sequence in Lip (G, Y) and that for every $x \in G$, and $u \in X$, $\|D(T_m - T_n)(x; u)\|_Y \leqslant \|T_m - T_n\| \|u\|$ and $\|D(T_m - T_n)(x)\|_{\mathrm{Lip}(X,Y)} \leqslant \|T_m - T_n\|$. These inequalities allow us to prove all the statements of the theorem.

Remark 2. The theorems in this section present an improvement for the mappings considered in them on our main theorem since we have the existence of a τ-differential exc. $\mathfrak{A}^0$ which is a much smaller class than $\mathfrak{A}$. However, in our main theorem we were able to prove under special assumptions on Y (separability or reflexivity) that exc. $\mathfrak{A}$ $DT(x)$ was actually a differential. We cannot improve this kind of statement in our present cases since for a relatively convex Lipschitzian mapping, even if Y is a separable Hilbert space, we are not able to prove in general that $DT(x)$ is a differential exc. $\mathfrak{A}^0$.

EXAMPLES. I. *Rel. convex mappings.* a) Assume that the sequence $\{b_n\}$ is a generalized basis for $^*Y$. Therefore, there exists a dual generalized basis $\{a_n\}$ in Y forming with $\{b_n\}$ a biorthogonal system. Clearly, $y \in \Gamma$ means that all the Fourier coefficients $\langle y, b_n\rangle$ are non-negative. Consider now real-valued functions $\varphi_n(x)$, $x \in G$ which are convex and Lipschitzian. It is then clear that for a sequence $\{a_n\}$ of positive numbers converging fast enough to zero, $T(x) = \sum a_n \varphi_n(x) a_n$ is a mapping of G into Y which is Lipschitzian and rel. convex (rel. Γ).

b) Consider $X = Y = L^p(\Omega, d\mu)$, where $d\mu$ is a separable measure on the measure space Ω. X and Y are separable and for $1 < p < \infty$ they are reflexive, hence our requirements are satisfied. Clearly, $Y^* = {}^*Y = L^{p'}(\Omega, d\mu)$. Take in the measure ring $(\Omega, d\mu)$ a dense sequence of sets and the corresponding characteristic functions which we will denote by $b_n(\omega)$. The convex cone Γ_p corresponding to the sequence $\{b_n\}$ in $L^{p'}(\Omega, d\mu)$ is formed by all the non-negative functions in $L^p(\Omega, d\mu)$. We define the

projection $P: L^p \to \Gamma_p$ by putting for $x(\omega) \in L^p$, $P(x) = x^+(\omega)$ [16]. It is clear that P is rel. convex (rel. Γ_p) and Lipschitzian with constant 1. By Theorem 1, $DP(x)$ exists everywhere as a Gateaux τ-differential. We can give an explicit formula for it:

$$(3) \qquad DP(x; u) = \begin{cases} u(\omega) & \text{for} \quad \omega \in \Omega^+(x), \\ u^+(\omega) & \text{for} \quad \omega \in \Omega^0(x), \\ 0 & \text{for} \quad \omega \in \Omega^-(x), \end{cases}$$

where $\Omega^+(x)$, $\Omega^0(x)$ and $\Omega^-(x)$ are the sets where $x(\omega) > 0$, $= 0$ or < 0 respectively. One checks directly that $DP(x)$ is actually a strong Gateaux-differential and that it is a differential except when $\Omega^0(x)$ is of positive measure.

Even though our requirements for Y are not satisfied for $p = 1$, one checks directly that even in this case, (3) gives the Gateaux differential for every $x \in L^1(\Omega, d\mu)$. Hence, by Theorem 1 of the present section, in case $1 < p < \infty$; and by Theorem 2, Section 1, Chapter II, in case $p = 1$, the set where $DP(x)$ is not a differential belongs to $\mathfrak{A}^0$, i.e. the set of functions $x(\omega) \in L^p(\Omega, d\mu)$ which vanish on a set of positive measure belongs to $\mathfrak{A}^0$. The result is rather surprising and it seems rather difficult to prove it directly.

II. *Var. convex mappings.* a) It is trivial to construct a var. convex mapping $\boldsymbol{R}^1 \to \boldsymbol{R}^1$. Then if we have any separable Banach space X we can take any vector g, $0 \neq g \in X$ and any linear projection P of X onto the one-dimensional subspace $[g]$. On $[g]$ choose any var. convex function f into $\boldsymbol{R}^1$ and define for $x \in X$, $F(x) = f(Px)$. One checks immediately that F is locally Lipschitzian variably convex.

b) Suppose that we have N couples of Banach spaces X_n, Y_n, $n = 1, \ldots, N$, Y_n being conjugate of $^*Y_n$, X_n and $^*Y_n$ separable. We form the direct sums $X = \sum_1^{N\cdot} X_n$, $Y = \sum_1^{N\cdot} Y_n$ and $^*Y = \sum_1^{N\cdot} {}^*Y_n$. If we define on any of the direct sums, say X, $\|\sum_1^{N\cdot} x_k\| = (\sum_1^N \|x_k\|^2)^{1/2}$ we obtain that X and $^*Y$ are separable Banach spaces and Y is the conjugate space of $^*Y$.

Suppose now that we have locally Lipschitzian var. convex mappings $T_n: G_n \to Y_n$, G_n open in X_n. We can define then $G = \sum G_n$ — an open subset of X — and a mapping $T: G \to Y$ by putting for $x = x_1 \dot{+} x_2 \dot{+} \ldots \dot{+} x_N$, $x_n \in G_n$, $T(x) = \sum_1^N T_n(x_n)$. This mapping is obviously locally Lipschitzian and is var. convex. To verify the last statement we should cover G_n by

(16) For every real x, $x^+ = 0$ or x depending on whether $x < 0$ or $\geqslant 0$.

a countable number of neighborhoods $U_n^{(k)}$ in each of which T_n is convex relatively to a cone $\Gamma_n^{(k)}$. Then the open sets $U^{(k_1,\ldots,k_n)} = \sum_1^N U_n^{(k_n)}$ for all sequences $(k_1, \ldots, k_n)$, form an open covering of G and in each $U^{(k_1,\ldots,k_n)}$, T is convex relatively to the cone $\sum_1^{N\cdot} \Gamma_n^{(k_n)}$.

III. *Convexoid mappings*. Consider now an infinite sequence of couples of Banach spaces X_n, Y_n, Y_n being the conjugate of ${}^*Y_n$, X_n and Y_n being separable. We can then form the infinite direct sums $X = \sum_1^{\infty\cdot} X_n$, $Y = \sum_1^{\infty} Y_n$, ${}^*Y = \sum_1^{\infty\cdot} {}^*Y_n$ with similar choice of norm as in Example II, b)). We will consider now mappings T_n: $X_n \to Y_n$. To simplify our developments we will assume that $T_n(0) = 0$ and that in $\mathrm{Lip}(X_n, Y_n)$ and $\mathrm{Lip}(X, Y)$ we form the norm by choosing $x_0 = 0$. Suppose then that $\|T_n\| \searrow 0$. If we form $T = \sum_1^{\infty} T_n$ we notice that $\|\sum_{N+1}^{\infty} T_n\| \leqslant \|T_{N+1}\| \searrow 0$. If we assume that each T_n is var. convex we have that $\sum_1^{N\cdot} T_n$ is var. convex, at first in $\sum_1^{N\cdot} X_n$ (by Example II, b)), but we extend it to the whole of X by replacing all the T_n for $n > N$ by 0. It follows, therefore, that T is convexoid.

It is clear that if we assume that for $n > N_0$, T_n is rel. convex then T becomes a var. convex mapping of X into Y.

4. Distance from a point to a subset. We will now consider a separable Banach space X and a non-empty subset $S \subset X$ and consider the real valued function:

$$F_S(x) = \inf_{y \in S} \|x - y\|, \qquad x \in X. \tag{1}$$

It is clear that the value $F_S(x)$ will not change if we replace S by its strong closure $\bar{S}$. From now on we will therefore assume that S is *strongly closed*.

It is immediately checked that $F_S(x)$ is Lipschitzian with constant 1. Hence by our main theorem, Section 2, Chapter II, $DF_S(x)$ exists and is a differential a.e. and $DF_S(x; u)$ for all x exc. $\mathfrak{A}$ is a linear functional in u with norm $\leqslant 1$. Our aim in this section will be to establish certain geometric relations between the existence of the differential $DF_S(x)$ and geometric properties of the set S. The example of Section 2 is a special case of our present developments where $S = (0)$. It is clear that we will obtain more precise relationships if we assume some additional hypotheses about the space X and its norm, $\|x\|$, and also on the nature of the set S. Our analysis will be far from exhaustive, the aim being only to indicate some such relations.

As additional requirements on the space X we may assume that X is a conjugate space of some ${}^*X$ or that it is reflexive. Before we describe the requirements we may impose on the norm in X, we will introduce certain notations.

For every $x \in X$ we denote by $(x)^*$ the set of all $x^* \in X^*$ satisfying $\|x^*\| = \|x\|$, and $\langle x, x^*\rangle = \|x\|^2$. The mapping $x \to (x)^*$ will be called the *conjugate mapping*. The conjugate mapping is in general multi-valued and monotone (see Remark 1, Section 2, Chapter III). For every real a we have obviously $(ax)^* = a(x)^*$. If this mapping is single valued for every $x \in X$, we will say that the norm in X is *smooth*. The norm in X is smooth if and only if the unit ball $B_1(0)$ (and therefore any ball with positive radius) has a unique supporting hyperplane at each point of its boundary. By remarks in the example of Section 1, the smoothness means that $\|x\|$ has a differential at every point different from 0 ([17]).

Besides smoothness, we may require the norm to be strictly convex. In terms of the conjugate mapping, this property means that for $x \neq y$, $(x)^* \cap (y)^* = \emptyset$. An equivalent meaning is that the boundary of the unit ball $\partial B_1(0)$ does not contain any segment of straight line containing more than one point.

We can start now with the general case. The smallest distance $F_S(x)$ may be attained or not attained. In the first case we will define the projection $P_S(x)$ as the set

$$P_S(x) = \{y \in S:\ \|x - y\| = F_S(x)\}. \tag{2}$$

This projection P_S is in general multi-valued and the set of x's where it is not empty is the domain of P_S.

To explain the additional requirements we may impose on S we introduce the following notions: If τ is any locally convex Hausdorf topology on X weaker than the strong topology we will say that S is boundedly (bdd.) τ-closed or boundedly τ-compact if any intersection of S with a closed ball $\overline{B_R(x)}$ is τ-closed or τ-compact respectively. The topologies we will consider on X are the strong, the weak and the weak-* if X is a conjugate space. The last one, if it exists, is the weakest one. Hence, we have the implications: bdd. weak-*-closed $\Rightarrow$ bdd. weakly closed, bdd. strongly compact $\Rightarrow$ bdd. weakly compact $\Rightarrow$ bdd. weak-*-compact.

By using the fact that $\|x\|$ is lower semi-continuous for the weak and any weak-*-topology on X, we obtain immediately:

PROPOSITION 1. *If S is bdd. weakly compact, or, if X is a conjugate space and S is bdd. weak-*-closed, then the domain of P_S is X.*

([17]) This notion of smoothness of the norm is very weak.

THEOREM 1. *If $DF_S(x)$ is a differential, x being in the domain of P_S and $x \notin S$, then $\|DF_S(x)\| = 1$ and $DF_S(x) \in \left(\frac{x-y}{\|x-y\|}\right)^*$ for every $y \in P_S(x)$.*

Proof. Take an arbitrary $y \in P_S(x)$ then $F_S(x) = \|y-x\|$ and for $0<\varrho<1$,

$$F_S\big(x+\varrho(y-x)\big) = \inf_{y' \in S}\|x+\varrho(y-x)-y'\| \leqslant (1-\varrho)\|y-x\|.$$

On the other hand, for $y' \in S$

$$\|x+\varrho(y-x)-y'\| \geqslant \|x-y'\|-\varrho\|y-x\| \geqslant \|x-y\|-\varrho\|y-x\| = (1-\varrho)\|y-x\|.$$

Therefore, $F_S\big(x+\varrho(y-x)\big) = (1-\varrho)\|y-x\|$ and the differential quotient $\frac{1}{\varrho}[F_S\big(x+\varrho(y-x)\big)-F_S(x)] = -\|y-x\|$ which shows that $DF_S\left(x; \frac{x-y}{\|x-y\|}\right) = 1$ and proves our theorem.

COROLLARY 1. *Under the assumptions of Theorem 1, the hyperplane $-DF_S(x;u) = 1$ is a supporting hyperplane of the ball $B_1(0)$ the intersection of which with $\partial B_1(0)$ is a convex set containing all the points $\frac{y-x}{\|y-x\|}$ for $y \in P_S(x)$.*

The proof is immediate.

COROLLARY 2. *Under the assumptions of the theorem, if the norm in X is smooth, then the hyperplane $-DF_S(x;u) = 1$ is the unique supporting hyperplane for the unit ball at each of the points $\frac{y-x}{\|y-x\|}$ for $y \in P_S(x)$.*

This follows from the single-valuedness of the conjugate mapping.

COROLLARY 3. *If the norm in X is strictly convex, then, under the assumptions of Theorem 1, $P_S(x)$ is single-valued.*

This follows immediately from the properties of strict convexity mentioned above.

PROPOSITION 2. *If $S = \bigcup_1^n S_k$, S_k strongly closed, then $F_S(x) = \min_{1 \leqslant k \leqslant n} F_{S_k}(x)$.*

The proof is immediate.

PROPOSITION 3. *If S is convex, $P_S(x) = S \cap \partial B_{F_S(x)}(x)$ is a convex set. If in addition the norm in X is strictly convex, then $P_S(x)$ is single-valued for all x in its domain.*

Proof. Suppose that y_1 and $y_2 \in P_S(x)$. Then the closed segment $[y_1; y_2] \subset S$ and for each $y \in [y_1; y_2]$ we have $\|y-x\| \leqslant \max(\|y_1-x\|, \|y_2-x\|) = F_S(x)$, hence $[y_1; y_2] \subset P_S(x)$.

If the norm in X is strictly convex there exists no convex set on $\partial B_{F(x)}(x)$ containing more than a single point, hence the second part of Proposition 3.

PROPOSITION 4. *If S is convex, $F_S(x)$ is a convex function on X.*

Proof. Consider any segment $[x_1; x_2]$, any points y_1 and y_2 in S such that $F_S(x_k) > \|y_k - x_k\| - \varepsilon$, $k = 1, 2$, and $\varepsilon > 0$. The segment $[y_1; y_2] \subset S$ and we will consider the mapping of the interval $0 \leqslant t \leqslant 1$ into $\|(1-t)(y_1-x_1)+t(y_2-x_2)\|$. For any t we have

$$\|(1-t)(y_1-x_1)+t(y_2-x_2)\| \leqslant (1-t)\|y_1-x_1\|+t\|y_2-x_2\|.$$

It follows that

$$F_S\big((1-t)x_1+tx_2\big) \leqslant \|(1-t)(y_1-x_1)+t(y_2-x_2)\|$$
$$< (1-t)\big(F_S(x_1)+\varepsilon\big)+t\big(F_S(x_2)+\varepsilon\big) = (1-t)F_S(x_1)+tF_S(x_2)+\varepsilon.$$

Since ε is arbitrary, we obtain here the convexity inequality on the segment $[x_1; x_2]$. Hence, F_S is convex.

PROPOSITION 5. *If S is convex and if the norm in X is strictly convex and if the convex set is bdd. weakly compact (bdd. weak-*-closed when X is a conjugate space), then the projection P_S is defined on the whole of X and is single-valued.*

The proof follows from Proposition 1 and Proposition 3.

By Proposition 5 and Theorem 1 of Section 2 of the present chapter we obtain:

THEOREM 2. *If the norm in X is strictly convex and S is convex, then $DF_S(x)$ exists everywhere as a Gateaux differential and is a differential exc. $\mathfrak{A}^0$.*

EXAMPLES. I. Consider a hyperplane

$$S \equiv S_{v,a} = \{y \in X: \langle y, v\rangle = a \text{ for fixed } a \text{ and } v \in X^*, v \neq 0\}.$$

S is a linear variety of codimension 1. It divides the space X into two open half spaces X^- and X^+ of points x with $\langle x, v\rangle < a$ and $\langle x, v\rangle > a$ respectively. X^- will be called the left half space and X^+ the right half space. As a set of points S does not change when we replace v and a by βv and βa for any real $\beta \neq 0$. However, if $\beta > 0$, the left and right half spaces are not changed, whereas for $\beta < 0$ these half-spaces interchange. Hence, it will be convenient to consider that $v \in X^*$ defines an orientation in X which is not changed if we replace v by βv with $\beta > 0$ and changes to the opposite when we replace v by βv with $\beta < 0$. Consider now P_S and F_S. For every S, S lies obviously in the domain of P_S. It is checked immediately that for a hyperplane S if $P_S(x)$ is not empty for one $x \notin S$, then it is not empty for every $x \in X$. This case arises if and only if $v \in (z)^*$ for-some $z \in X$. Also, in this case if we denote by $*(v)$ the set of all z's such that $v \in (z)^*$ [18] we have $P_S(x) = x + \frac{F_S(x)}{\|v\|} *(v)$ or $x - \frac{F_S(x)}{\|v\|} *(v)$ for $x \in X^-$ or $\in X^+$ respectively.

We assume now that the norm in X is strictly convex. Then P_S is an affine mapping of X onto S. If S is a subspace (i.e. $a = 0$), then P_S is a linear projection.

Consider now F_S. By Theorem 2, $DF_S(x)$ exists for all x as a Gateaux differential. For $x \in X^-$, or $x \in X^+$, $DF_S(x)$ is a differential and $DF_S(x; u) = -\langle u, v/\|v\|\rangle$ or $+\langle u, v/\|v\|\rangle$ respectively. If $x \in S$, $DF_S(x)$ is actually a Gateaux differential such that $DF_S(x; u)$ is symmetric in u and vanishes for u such that $x + u \in S$.

Consider now $S = S_1 \cup S_2$ with S_1 and S_2 hyperplanes: $\langle y, v_1\rangle = a_1$ and $\langle y, v_2\rangle = a_2$ respectively, with v_1 and v_2 linearly independent.

Now $P_S(x)$ exists for every x if and only if $*(v_1)$ and $*(v_2)$ are both non-empty.

$X \setminus S$ decomposes in four connected open components X^{--}, X^{+-}, X^{-+}, and X^{++}, where X^{--} is the intersection of the left half spaces of S_1 and S_2 and similarly for the other components. If the norm in X is strictly convex and $P_S(x)$ exists everywhere, then $P_S(x)$ is composed of a single point or of two points.

In the open sets $G_1 = \{x \in X: F_{S_1}(x) < F_{S_2}(x)\}$ and $G_2 = \{x \in X: F_{S_1}(x) > F_{S_2}(x)\}$, $DF_S(x)$ is a differential $= DF_{S_1}(x)$ or $= DF_{S_2}(x)$ respectively. For x in the closed set where $F_{S_1}(x) = F_{S_2}(x)$, $DF_S(x)$ is a Gateaux differential with $DF_S(x; u) = 0$ for u parallel to $S_1 \cap S_2$.

II. Consider now $S = \partial B_1(0)$. It is obvious that $F_S(x) = |1 - \|x\||$.

We will assume now that the norm in X is strictly convex, then $P_S(x) = x/\|x\|$ for $x \neq 0$, and $P_S(0) = \partial B_1(0)$. If $x \in B_1(0)$, $F_S(x) = 1 - \|x\|$ and $DF_S(x) = -D\|x\|$ which is always a Gateaux differential. $DF_S(0; u) = -\|u\|$, and for $x \in B_1(0) \setminus (0)$, $DF_S(x)$ is a differential if and only if, at the point $x/\|x\|$, $B_1(0)$ has a unique supporting hyperplane (see the example in Sec. 2). For $x \in X \setminus B_1(0)$, $F_S(x) = \|x\| - 1$, $DF_S(x)$ is a Gateaux differential. Hence again $DF_S(x)$ is a differential if and only if at $x/\|x\|$, $B_1(0)$ has a unique supporting hyperplane. If $x \in \partial B_1(0)$ and there is only one supporting hyperplane at x, then $DF_S(x)$ is a Gateaux differential such that $DF_S(x; x) = DF_S(x; -x) = 1$, and $DF_S(x; u) = 0$ for all u parallel to the supporting hyperplane at x.

It seems that interesting results could be attained by introducing an adequate notion of supporting hyperplanes for an arbitrary closed set S and relating the differential (or Gateaux differential) $DF_S(x)$ with the supporting hyperplane at the points of $P_S(x)$. We do not introduce here the relevant notions since it would increase greatly the size of the present section and what is more, we are not able to answer, at this stage, many of the naturally arising questions connected with these notions.

[18] The set $*(v)$ is closed and convex.

CHAPTER IV

BOREL MEASURES ABSOLUTELY CONTINUOUS REL. $\mathfrak{A}$

1. Structure of measures absolutely continuous rel. $\mathfrak{A}$.

DEFINITION 1. Let $\mathfrak{B}$ be an arbitrary exceptional class in X and let μ be a σ-finite signed measure on X. We say that μ is *absolutely continuous* (*a.c.*) rel. $\mathfrak{B}$ if every set in $\mathfrak{B}$ is of μ-measure 0. If the whole measure μ is concentrated on a set in $\mathfrak{B}$ we will say that μ is *singular* rel. $\mathfrak{B}$.

PROPOSITION 1. *For any σ-finite signed measure μ on X and any exceptional class $\mathfrak{B}$ in X there exists a unique decomposition $\mu = \mu_{ac} + \mu_s$ into σ-finite signed measures such that μ_{ac} is a.c. rel. $\mathfrak{B}$ and μ_s is singular rel. $\mathfrak{B}$.*

The proof is completely similar to the classical proof of the decomposition of μ into absolutely continuous and singular parts relative to another measure μ_0 [19].

PROPOSITION 2. a) *In order that the σ-finite signed measure μ be a.c. or singular rel. $\mathfrak{B}$ it is necessary and sufficient that its positive and negative parts be a.c. or singular respectively.*

b) *In order that a non-negative σ-finite measure μ be a.c. or singular rel. $\mathfrak{B}$ it is necessary and sufficient that each finite part of it be a.c. or singular respectively.*

Again the proof follows the classical argument.

THEOREM 1. *For a measure μ in X to be a.c. rel. $\mathfrak{A}$ it is necessary and sufficient that μ be representable as a sum:*

$$\mu = \mu_0 + \sum_{k=1}^{\infty} \mu_k, \tag{1}$$

where all the measures in the right-hand member are mutually disjoint (i.e. they are concentrated on mutually disjoint sets), μ_k, $k = 1, 2, \ldots$, is a.c. rel. $\mathfrak{A}\{a_n^{(k)}\}$ for some sequence $(a_1^{(k)}, a_2^{(k)}, \ldots)$ complete in X, and μ_0, if it is not 0, is singular rel. all the classes $\mathfrak{A}\{a_n\}$, but a. c. rel. $\mathfrak{A}$.

Remark 1. The measures satisfying the properties of the measure μ_0 will be called *exceptional measures*. We do not know of any example of such a measure and we are not able to prove that they do not exist.

[19] The usual definition of absolute continuity and singularity rel. μ_0 coincides with our Definition 1 with $\mathfrak{B}$ replaced by the class of sets of μ_0-measure 0.

Proof of Theorem 1. That representation (1) is sufficient for μ to be a.c. rel. $\mathfrak{A}$ follows from the fact that μ_k being a.c. rel. $\mathfrak{A}\{a_n^{(k)}\}$ is a fortiori a.c. rel. $\mathfrak{A} \subset \mathfrak{A}\{a_n^{(k)}\}$.

To prove that the existence of representation (1) is necessary we use Proposition 2 to remark that it suffices to prove it for μ non-negative and finite; we will assume this in the remainder of the proof. We use transfinite induction. Let Γ be the first ordinal number of cardinality equal to the cardinality of the set of complete sequences $\{b_n\}$ in X. Hence, we can arrange all the sequences in a transfinite sequence $\{\{b_n^{(\gamma)}\}_n\}$ for $1 \leqslant \gamma \leqslant \Gamma$. For $\gamma = 1$, we put $\mu = \mu_1' + \nu_1$ where μ_1' is the absolutely continuous part of μ rel. $\mathfrak{A}\{b_n^{(1)}\}$ and ν_1 is the singular part. Suppose that for all $\gamma < \alpha \leqslant \Gamma$ we have already defined μ_γ' and ν_γ such that for each $\beta < \alpha$ we have $\mu = \nu_\beta + \sum_{\gamma \leqslant \beta} \mu_\gamma'$ with all the measures on the right-hand side mutually disjoint and non-negative so that μ_γ' be a.c. rel. $\mathfrak{A}\{b_n^{(\gamma)}\}$ and ν_β be singular rel. all $\mathfrak{A}\{b_n^{(\gamma)}\}$ for $\gamma \leqslant \beta$. It follows that only a countable number of measures μ_γ' are not identically zero and that the measures ν_γ form a decreasing sequence of measures. Hence

(2) $$\mu = \sum_{\gamma < \alpha} \mu_\gamma' + \nu_\alpha',$$

where ν_α' is the limit of the decreasing sequence of measures ν_γ, $\gamma < \alpha$. By our assumptions, all the measures in the right-hand member of (2) are mutually disjoint and ν_α' is singular rel. all $\mathfrak{A}\{b_n^{(\gamma)}\}$ with $\gamma < \alpha$. Hence, decomposing ν_α' into $\mu_\alpha' + \nu_\alpha$, the a.c. part and singular part of ν_α' rel. $\mathfrak{A}\{b_n^{(\alpha)}\}$, we achieve the inductive definition of all μ_γ' and ν_γ for $\gamma \leqslant \Gamma$. Since ν_Γ is singular rel. all $\mathfrak{A}\{b_n^{(\gamma)}\}$ and, on the other hand, as part of μ is a.c. rel. $\mathfrak{A}$, ν_Γ qualifies for the exceptional measure μ_0. The countable number of μ_γ''s which are not identically zero can be arranged in a simple sequence denoted by $\{\mu_k\}$ and the corresponding $\mathfrak{A}\{b_n^{(\gamma)}\}$ will be denoted $\mathfrak{A}\{a_n^{(k)}\}$. Thus representation (1) is achieved.

Remark 2. Representation (1) is certainly not unique.

Since we do not know about the existence of an exceptional measure μ_0 all the measures μ a.c. rel. $\mathfrak{A}$ which we will construct will be of the form (1) with $\mu_0 = 0$. Hence, the problem of constructing the measures μ will be reduced to constructing all the measures (or, at least, large classes of them) which are a.c. rel. $\mathfrak{A}\{a_n\}$ for some given sequence $\{a_n\}$ complete in X.

2. Cylindrical measures. For any complete sequence $\{a_n\} \subset X$ we can form an equivalent generalized basis $\{e_n\}$ in X such that $\mathfrak{A}\{a_n\} = \mathfrak{A}\{e_n\}$ and X is decomposable in direct sums $X = [e_1, \ldots, e_n] + \overline{[e_{n+1}, \ldots]}$.

The corresponding projections P_n of X onto $[e_1, \ldots, e_n]$ have the property that $P_n P_{n+1} = P_n$.

Consider now an arbitrary finite Borel non-negative measure μ on X (not necessarily a.c. rel. $\mathfrak{A}\{e_n\}$). It defines a series of Borel measures μ_n on $[e_1, \ldots, e_n]$ given by

(1) $$\mu_n(A_n) = \mu(A_n + \overline{[e_{n+1}, \ldots]}) \quad \text{for} \quad A_n \subset [e_1, \ldots, e_n].$$

The measures μ_n are called the *cylindrical* measures corresponding to the measure μ and the generalized basis. It is immediately checked that the measures μ_n must satisfy the compatibility condition:

(2) $$\mu_{n+1}(A_n + [e_{n+1}]) = \mu_n(A_n) \quad \text{for every } A_n \subset [e_1, \ldots, e_n], \quad n = 1, 2, \ldots$$

The total mass of a measure μ will be denoted by $|\mu|$. By (1) and (2) we obtain immediately

(3) $$|\mu_n| = |\mu|, \quad n = 1, 2, \ldots$$

(4) For every Borel set $A \subset X$, $\mu_n(P_n A) \geqslant \mu_{n+1}(P_{n+1} A) \geqslant \mu(A)$, $n = 1, 2, \ldots$

Remark 1. Most authors consider a somehow different kind of cylindrical measures[20]. In this other notion we don't consider cylindrical measures relative to a given generalized basis $\{e_n\}$. As cylindrical measures corresponding to μ are considered the measures $\mu_{X/F}$ defined on the quotient space X/F when F is a closed subspace of X of finite co-dimension. If $A \subset X/F$, then $\mu_{X/F}(A) = \mu(A + F)$. Almost all our developments in this section are based on ideas which were used in connection with the other notion of cylindrical measures. However, the slight difference in the notion warrants a brief account of the proofs. Our choice of the definition is justified by the fact that we will want to construct the measure μ by using the cylindrical measures μ_n defined on an increasing sequence of concrete finite dimensional spaces $[e_1, \ldots, e_n]$.

Proposition 1. *For any closed ball $\overline{B_R(y)}$ we have*

(5) $$\mu_n(P_n \overline{B_R(y)}) \searrow \mu(\overline{B_R(y)}).$$

Proof. Consider the basis $\{f_n\} \subset X^*$ dual to $\{e_n\}$. Then the subspace $[\{f_n\}]$ is weak-*-dense in X^*. Hence, the closed ball $\overline{B_R(y)}$ can be obtained as intersection of a countable number of closed half spaces

$$S_k = \{x \in X \colon \langle x, g_k \rangle \leqslant \alpha_k, \, g_k \in [\{f_n\}]\}.$$

[20] See Gelfand–Vilenkin [5] and L. Schwartz [10].

$\bigcap_1^m S_k$ is a cylinder with basis in $[e_1, \ldots, e_{n_m}]$, where n_m is the largest index of f_n figuring in the g_k, $k = 1, \ldots, m$, represented as finite linear combinations of the f_n's. Since

$$P_{n_m}\overline{B_R(y)} \subset P_{n_m}\bigcap_1^m S_k,$$

it follows that

$$\mu_{n_m}(P_{n_m}\overline{B_R(y)}) \leqslant \mu_{n_m}(P_{n_m}\bigcap_1^m S_k) = \mu(\bigcap_1^m S_k)$$

we get

$$\lim_{m\nearrow\infty}\mu_{n_m}(P_{n_m}\overline{B_R(y)}) \leqslant \lim \mu(\bigcap_1^m S_k) = \mu(\bigcap_1^\infty S_k) = \mu(\overline{B_R(y)}).$$

Comparing with (4), we obtain (5).

COROLLARY 1. *In order that the cylindrical measures $\{\mu_n\}$ correspond to some finite Borel measures it is necessary that for each $\varepsilon > 0$ there exists a closed ball $\overline{B_R(y)}$ such that*

$$\mu_n(P_n\overline{B_R(y)}) \geqslant |\mu_1| - \varepsilon. \tag{6}$$

We use here (3) and (5) and the fact that $\mu(B_R(y)) \nearrow |\mu| = |\mu_1|$ when $R \nearrow \infty$.

THEOREM 1 (Uniqueness Theorem). *If two Borel measures μ' and μ'' in X determine the same sequence of cylindrical measures $\{\mu_n\}$ in X, then $\mu' = \mu''$.*

Proof. By Proposition 1 we see that for any closed sphere $\overline{B_R(y)}$, $\mu'(\overline{B_R(y)}) = \mu''(\overline{B_R(y)})$. Furthermore, the same proof as in Proposition 1 gives us the extension of (5) to all finite intersections of closed spheres in X. It follows that for every set A which is a finite intersection of closed spheres $\mu'(A) = \mu''(A)$. This implies by a standard argument, since X is separable, that $\mu'(A) = \mu''(A)$ for all Borel sets A.

A sequence of measures μ_n defined on successive spaces $[e_1, \ldots, e_n]$ is called a *cylindrical sequence of measures* if the compatibility condition (2) is satisfied. From this condition already it follows that $|\mu_n| = |\mu_{n+1}|$ for $n = 1, 2, \ldots$ Our aim is to establish conditions under which such a compatible sequence of measures corresponds to a Borel measure μ. By Theorem 1, if it corresponds to some Borel measure this measure is unique.

For the existence of a Borel measure μ corresponding to a compatible sequence of cylindrical measures, we know that the condition of the Corollary 1 is necessary. In the next theorem we will show (by a simple application of a generalized Prokhorov's theorem([21]) that under certain restrictions on the space X and on the generalized basis $\{e_n\}$ the necessary condition of Corollary 1 is also sufficient.

THEOREM 2 (Existence Theorem). *If X is a conjugate space and the dual basis $\{f_n\}$ of the generalized basis $\{e_n\}$ lies in the preconjugate $^*X$, then any compatible sequence of cylindrical measures, satisfying the condition of Corollary 1, corresponds to a Borel measure μ in X.*

Proof. We notice first that if X is a conjugate space with a preconjugate $^*X$ the preconjugate must be separable, is canonically, isomorphically and isometrically embedded in X^*, that any complete sequence $\{v_n\}$ in $^*X$ is necessarily weak-*-complete in X^* so that by starting with any complete sequence $\{u_n\}$ in X and any complete $\{v_n\}$ in $^*X$ we will obtain, by Proposition 1 of Section 2, Chapter I, two dual bases $\{e_n\}$ and $\{f_n\}$ in X or $^*X$ respectively, $\{f_n\}$ being at the same time a basis for X^*. In this way if X is a conjugate space there is an infinite number of choices for the generalized basis $\{e_n\}$ to satisfy the requirement of our theorem. On X we consider the weak-*-topology (rel. $^*X$). By our requirements on $\{e_n\}$ the projections P_n are continuous in this topology and, furthermore, the closed balls $\overline{B_R(0)}$ are weak-*-compact. Therefore, by the generalized Prokhorov's theorem, the condition of Corollary 1 is sufficient for the existence of a Borel measure μ corresponding to the compatible sequence of cylindrical measures μ_n.

Remark 2. 1° The condition of Corollary 1 can be written equivalently in the form (for $y = 0$):

(7) For $R \nearrow \infty$, $\mu_n(P_n\overline{B_R(0)}) \nearrow |\mu_n| = |\mu_1|$ uniformly in n.

2° Since $P_n\overline{B_R(0)} \supset [e_1, \ldots, e_n] \cap \overline{B_R(0)}$, the following condition which is more easily checked (it does not require the knowledge of the projection P_n) is stronger than (7):

(7′) For $R \nearrow \infty$, $\mu_n([e_1, \ldots, e_n] \cap \overline{B_R(0)}) \nearrow |\mu_n| = |\mu_1|$ uniformly in n.

3° If $\{e_n\}$ is a weak Schauder basis, conditions (7) and (7′) are equivalent (since the projections P_n are then uniformly bounded).

4° A stronger condition than (7′) is

(7″) For some fixed R and every n, μ_n is concentrated on
$$[e_1, \ldots, e_n] \cap \overline{B_R(0)}.$$

This condition is often satisfied in concrete cases.

Remark 3. In many cases, even when X is not a conjugate space, we can obtain, by the generalized Prokhorov's theorem, the existence of

([21]) See L. Schwartz [10], Theorem 22, p. 81.

μ corresponding to the compatible sequence $\{\mu_n\}$ if we can find for every $\varepsilon>0$ a compact set $K_\varepsilon \subset X$ (in the weak topology) such that $\mu_n(P_n K_\varepsilon) > |\mu_1| - \varepsilon$. This is the case (already used) when $\mu_n = \nu_1 \times \ldots \times \nu_n$, where ν_k is a measure with total mass $|\nu_k| = 1$, ν_k being concentrated on a segment $[-\alpha_k e_k;\ \alpha_k e_k]$, $\alpha_k>0$ with $\sum_1^\infty \alpha_k \|e_k\| < \infty$. The corresponding infinite rectangle $\prod_1^\infty [-\alpha_k e_k;\ \alpha_k e_k]$ is then compact even in the strong topology of X.

3. Measures absolutely continuous rel. $\mathfrak{A}\{e_n\}$. In view of the developments of the preceding section we will accept throughout the present section that X is a conjugate space with preconjugate ${}^*X \subset X^*$ and that $\{e_n\}$ is a generalized basis with dual basis $\{f_n\} \subset {}^*X$. Therefore, each compatible sequence of cylindrical measures satisfying (7), Section 2, corresponds to a unique Borel measure in X. We want to investigate under which conditions the so determined measure in X is absolutely continuous rel. $\mathfrak{A}\{e_n\}$. We have first a necessary condition in the following:

PROPOSITION 1. *If the measure μ corresponding to a sequence of compatible measures μ_n is a.c. rel. $\mathfrak{A}\{e_n\}$, each μ_n is a.c. rel. to the Lebesgue measure in $[e_1, \ldots, e_n]$.*

Proof. If μ_n has a singular part ν_n rel. Lebesgue measure, not identically zero, concentrated on a set $A_n \subset [e_1, \ldots, e_n]$ of Lebesgue measure 0, then $\nu_n(A_n) = \mu_n(A_n) = \mu(A_n + \overline{[e_{n+1}, \ldots]})$. But the set $A_n + \overline{[e_{n+1}, \ldots]} \in \mathfrak{A}[e_1, \ldots, e_n] \subset \mathfrak{A}\{e_k\}$. Hence, μ wouldn't be a.c. rel. $\mathfrak{A}\{e_n\}$.

Remark 1. We do not know if the condition of Proposition 1 is sufficient for the measure μ to be a.c. rel. $\mathfrak{A}\{e_n\}$; we doubt that it is so. We can establish necessary and sufficient conditions for μ to be a.c. rel. $\mathfrak{A}\{e_n\}$ by using decomposition of measures (called by Bourbaki disintegration of measures). For each $k = 1, 2, \ldots$ put $X^{(k)} = \overline{[e_1, \ldots, e_{k-1}, e_{k+1}, \ldots]}$. Then we have a direct decomposition $X = X^{(k)} + \boldsymbol{R}^1 e_k$. To this decomposition corresponds the decomposition of the measure μ in the form $d\mu(x) = d\omega_k(x')\, d\theta_{x'}^{(k)}(\xi_k)$, where ω_k is a uniquely determined measure on $X^{(k)}$ of total mass $|\mu|$, $x' \in X^{(k)}$, $x = x' + \xi_k e_k$, $\theta_{x'}^{(k)}$ is a measure on the real line of variable ξ_k of total mass 1 which depends on x'. Two different determinations of the measures $\theta_{x'}^{(k)}$ differ only on a set of x''s of ω_k-measure 0. In terms of these decompositions the necessary and sufficient conditions for μ to be a.c. rel. $\mathfrak{A}\{e_n\}$ is that for each k the measures $\theta_{x'}^{(k)}$ be a.c. rel. the Lebesgue measure except perhaps on a set of x''s of ω_k-measure 0 [22].

[22] The sufficiency follows immediately by applying the Fubini's theorem to the integral $\int_X \chi_A(x)\, d\mu(x) = \int_{X^{(k)}} \int_{\boldsymbol{R}^1} \chi_A(x' + \xi_k e_k)\, d\theta_{x'}^{(k)}(\xi_k)\, d\omega_k(x')$, where χ_A is the characteristic function of a set $A \in \mathfrak{A}(e_k)$. The necessity is slightly more complicated.

However, this condition is not readily expressible in general in terms of the sequence of cylindrical measures $\{\mu_n\}$. In these conditions we will be satisfied with establishing large classes of sequences $\{\mu_n\}$, to which correspond measures μ a.c. rel. $\mathfrak{A}\{e_n\}$ without knowing if we exhaust all of them.

From now on we will consider only sequences μ_n of compatible cylindrical measures satisfying the uniformity condition (7) of Section 2 and the necessary condition of Proposition 1.

PROPOSITION 2. a) *If μ_n' and μ_n'' are non-negative measures satisfying our requirements, then the measures $\mu_n = \alpha\mu_n' + \beta\mu_n''$, $\alpha > 0$, $\beta > 0$, also satisfy the requirements and for the corresponding Borel measures on X we have $\mu = \alpha\mu' + \beta\mu''$.*

b) *If $\mu_n' \leqslant \mu_n$, $n = 1, 2, \ldots$, then for the corresponding Borel measures we have $\mu' \leqslant \mu$.*

Proof. a) That the requirements for $\{\mu_n\}$ are satisfied is clear. That the relation $\mu = \alpha\mu' + \beta\mu''$ holds is obtained by using the same idea as in the proof of Proposition 1, Section 2.

b) We can write $\mu_n = \mu_n' + \mu_n''$ and check immediately that our requirements are satisfied for μ_n'' if they are satisfied for μ_n and μ_n'. Hence, by a) $\mu = \mu' + \mu'' \geqslant \mu'$.

Let $n_1 = 1 < n_2 < n_3 < \ldots$ be an increasing sequence of integers. Consider on each subspace $[e_{n_k}, \ldots, e_{n_{k+1}-1}]$ a non-negative measure ν_k of total mass 1 a.c. rel. the Lebesgue measure. Such measures generate a sequence $\{\mu_n\}$ of compatible cylindrical measures where for $n_k \leqslant n \leqslant n_{k+1}$ and $A \subset [e_1, \ldots, e_n]$,

$$\mu_n(A) = (\nu_1 \times \nu_2 \times \ldots \times \nu_k)(A + [e_{n+1}, \ldots, e_{n_{k+1}-1}]). \tag{1}$$

This measure obviously satisfies our necessary condition from Proposition 1 but we will have to assume that the ν_k's are chosen so that the uniformity condition (7) of Section 2 is satisfied [23]. Under these conditions we have:

PROPOSITION 3. *The Borel measure μ corresponding to $\{\mu_n\}$ is a.c. rel. $\mathfrak{A}\{e_n\}$.*

Proof. In the present case we can achieve the decomposition of μ corresponding to the decomposition $X = X^{(n)} + \boldsymbol{R}^1 e_n$ (see Remark 1) by taking k with $n_k \leqslant n < n_{k+1}$ and decomposing the measure ν_k corresponding to the decomposition $[e_k, \ldots, e_{n_{k+1}-1}] = X^{(n)}[e_{n_k}, \ldots, e_{n_{k+1}-1}] + \boldsymbol{R}^1 e_n$. Since ν_k is a.c. rel. Lebesgue measure the decomposition of ν_k leads to measures $\theta_{x'}^{(n)}(\xi_n)$ which are all a.c. rel. Lebesgue measure. Hence we can use the Fubini's theorem.

[23] For instance, we can achieve it by assuming that all the product measures $\nu_1 \times \ldots \times \nu_k$ are concentrated in a fixed ball $B_R(0)$.

Starting with the class of cylindrical measures treated in the last proposition, we can define a whole convex cone of non-negative Borel measures a.c. rel. $\mathfrak{A}\{e_n\}$ (by using Proposition 2). If we place this cone in the space $\mathfrak{M}$ of all finite signed Borel measures on X with the norm $\|\mu\| = |\mu^+| + |\mu^-|$ (24), then its closure will still consist of measures of the same kind.

This is the class of measures a.c. rel. $\mathfrak{A}\{e_n\}$ which we can construct effectively by using the cylindrical measures.

(24) μ^+ and μ^- are the positive and negative parts of μ.

References

[1] E. Asplund, *Fréchet differentiability of convex functions*, Acta Math. 121 (1968), pp. 31–47.

[2] H. Brezis, *Opérateurs maximaux monotones*, etc., North-Holland–American Elsevier, Amsterdam 1973.

[3] J. P. R. Christensen, *Measure theoretic zero-sets in infinite dimensional spaces and applications to differentiability of Lipschitz mappings*, Preprint.

[4] I. M. Gelfand, *Abstrakte Funktionen und lineare Operatoren*, Math. Sborn. 4 (1938), pp. 235–286.

[5] — and N. Ya. Vilenkin, *Generalized functions*, Vol. 4, Academic Press, New York 1964.

[6] P. Mankiewicz, *On Lipschitz mappings between Fréchet spaces*, Studia Math. 41 (1972), pp. 225–241.

[7] — *On the differentiability of Lipschitz mappings in Fréchet spaces*, ibidem 45 (1973), pp. 15–29.

[8] H. B. Maynard, *A geometrical characterization of Banach spaces having the Radon–Nikodym property*, Trans. Amer. Math. Soc. 185 (1973), pp. 493–500.

[9] H. Rademacher, *Über partielle und totale Differenzierbarkeit von Funktionen mehrerer Variabeln and über die Transformation der Doppelintegrale*, Math. Ann. 79 (1919), pp. 340–359.

[10] L. Schwartz, *Radon measures on arbitrary topological spaces and cylindrical measures*, Oxford University Press, 1973.

[11] M. Sova, Czech. Math. J. 16 (1966), pp. 339–362.

[12] E. Zarantonello, *Dense single-valuedness of monotone operators*, Israel J. Math. 15 (1973).

Received February 12, 1975 (937)