SOME REMARKS ON VARIATIONS AND DIFFERENTIALS

M. Z. NASHED, Georgia Institute of Technology

1. Introduction. The treatment of differential calculus for functions of several variables is often dominated in undergraduate courses by computational formulas for differentials, gradients, directional derivatives, etc. The *intrinsic* nature of calculus and the *conceptual* meaning of these notions are seldom brought to light.

The calculus of functions of several variables is in many respects more subtel than the calculus of one real variable. For example, different notions of differentiability can be defined; the mean value theorem and Taylor's formula can be generalized in several ways and the theory of extrema is more involved.

It is generally recognized that the differential and integral calculus of several variables can be best studied in the setting of modern differential geometry, rather than in the traditional setting of real variable theory. For an interesting discussion of the merits of this approach, see [**12**]. On the other hand, an abstract formulation of some aspects of differential calculus can be given using vector spaces and linear operators. The undergraduate student in mathematics today is exposed to these notions in his study of linear algebra and analysis, and can climb up to such levels of formulations.

This approach sheds light on the ideas and arguments of multivariate calculus and the calculus of mapping on normed linear spaces, and unifies many notions and methods in analysis related to integral equations, the calculus of variations and numerical analysis. See, for instance, [**1**, **9**, **11**, **14**, **16**].

The purpose of this exposition is to discuss this approach for mappings whose domain and range are in normed linear spaces over the field of real numbers.

2. Linear and multilinear operators. To make the discussion in the following pages self-contained, we shall review in this section elementary properties of linear and multilinear operators. More details and numerous examples may be found, for instance, in the expository paper by Goffman in [**2**] and in [**3**, Chapter 5].

Throughout this paper, let E be a vector space over the real numbers and let θ denote the zero element in E. A norm on E is a mapping $\|\cdot\|$ which assigns to each element x a real number $\|x\|$ and which satisfies the following axioms:

(i) $\|x\| \geqq 0$; $\|x\| = 0$ only if $x = \theta$,

(ii) $\|\lambda x\| = |\lambda|\ \|x\|$ for any scalar λ,

(iii) $\|x+y\| \leqq \|x\| + \|y\|$.

A normed linear space is a vector space with a norm. Any normed linear space is also a metric space with the distance function $d(x, y) = \|x-y\|$. A complete normed linear space is called a Banach space.

Let E and Y be normed linear spaces. An operator L on E with range in Y is called

(i) additive if $L(x+y)=Lx+Ly$ for all x, y in E,
(ii) homogeneous if $L(\lambda x)=\lambda Lx$ for any scalar λ,
(iii) continuous at x in E if $\|Lx_n-Lx\|\to 0$ as $\|x_n-x\|\to 0$,
(iv) bounded if there exists a nonnegative number M such that $\|Lx\|\leqq M\|x\|$ for all x in E.

L is called *linear* if it is additive and homogeneous. It is well-known that a linear operator is continuous on the whole space if and only if it is continuous at $x=\theta$, and that it is bounded if and only if it is continuous.

If L is a bounded linear operator, then the *norm* of L, denoted by $\|L\|$, is defined as the greatest lower bound of all M satisfying the inequality in (iv). If L and S are bounded linear operators on E to E, then their sum $L+S$ and product LS are again linear operators and

$$\|L+S\| \leqq \|L\| + \|S\| \quad \text{and} \quad \|LS\| \leqq \|L\|\,\|S\|.$$

The set of all bounded linear operators on E to Y forms a Banach space if Y is complete [**3**, p. 102]. Denote this space by $\mathfrak{L}_1$. In particular, the Banach space of all bounded linear transformations from a Banach space E into its field of scalars is called the *dual* space of E and is denoted by E^*. These transformations are called linear functionals over E.

A bilinear operator L_2 on a normed linear space E is a bounded linear operator mapping E into $\mathfrak{L}_1$, i.e. L_2x is in $\mathfrak{L}_1$ and therefore

$$L_2x_1(x_2+x_3) = L_2x_1x_2 + L_2x_1x_3$$

$$L_2(x_1+x_2)x_3 = L_2x_1x_3 + L_2x_2x_3,$$

and

$$\|L_2x_1x_2\| \leqq M\|x_1\|\,\|x_2\|.$$

Similarly we may consider an m-linear operator L_m on a normed linear space E, mapping E into Y, as a bounded linear operator mapping E into $\mathfrak{L}_{m-1}$, where $\mathfrak{L}_{m-1}$ is the space of all $(m-1)$-linear operators in E. More generally, an m-linear operator may be defined on different normed linear spaces.

DEFINITION 1. *A mapping L_m defined on the product space $E_1\times\cdots\times E_m$ of m normed linear spaces E_i, $i=1,\cdots,m$, with range in a normed linear space Y, is called multilinear if it is additive and continuous (hence homogeneous and bounded) in each argument.*

A basic notion associated with multilinear operators is the notion of symmetry.

DEFINITION 2. *A multilinear operator is said to be symmetric if $E_1=E_2=\cdots=E_m$ and $L_m(x_1,\cdots,x_m)$ remains invariant under all permutations of the elements $x_1,\cdots,x_m$.*

3. Bibliographical comments on differential calculus in normed spaces. In building up necessary tools for studying a nonlinear functional in the neighborhood of a fixed element of a normed linear space, it is natural to seek a generalization of either the gradient or the differential as defined in the classical analysis of three-dimensional Euclidean spaces. To include mappings (as well as functionals) it turns out to be more expedient to first generalize the concepts of the differential and the directional derivative. These generalizations were first undertaken by Fréchet [**4**] and Gâteaux [**5**], respectively. Other definitions of differentials were given later by Michal, Zorn, Hyers (see [**8**] for references). Several generalizations of the differential to topological vector spaces, which are not necessarily normed, have been introduced by these and other authors. Such extensions have found useful applications in general differential geometry, dynamics and continuous group theory.

The most useful generalizations of the differential are those which preserve a fundamental idea in calculus, namely the "local" approximation of functions by linear functions. A differential $DF(x; h)$ of a mapping $F: X \to Y$, where X is an open subset of E, E and Y are normed linear spaces, is then a mapping $DF: X \times E \to Y$, where for each x in X, $DF(x; h)$ is assumed—in most definitions—to be linear and continuous in h. The basic difference in the various definitions of a differential is the sense in which $DF(x; h)$ approximates $F(x+h)-F(x)$.

Rolle's theorem does not hold for arbitrary sets in normed linear spaces of dimension greater than one. For such spaces, there are several forms of the mean-value theorem, Taylor's formula and theorem, and each form holds only in a certain sense. The validity of these extensions was established by Graves and Hildebrandt [**6, 7**], who also generalized the implicit function theorem, Kerner [**10**], Kantorovich [**9**], Vainberg [**16**] and others. The proofs used are different in most cases from the proof used in the case of a function of a real variable and as expected a slight strengthening of the hypotheses is required. Kerner considered integrability conditions of abstract vector fields and generalized Stoke's theorem. Rothe [**15**] studied topological properties of gradient mappings, which are generalizations of the idea of a conservative "force field" to abstract vector spaces.

4. Variations. We shall first discuss the Gâteaux variation, which is a generalization of the directional derivative in classical calculus and of the notion of the first variation arising in the calculus of variations.

Let Y be a Banach space and let (t_1, t_2) be an open interval of the real line. The first derivative $\Phi'(t_0)$ of $\Phi: (t_1, t_2) \to Y$, at t_0 in (t_1, t_2) is defined by

$$\Phi'(t_0) = \lim_{t \to t_0} \frac{\Phi(t) - \Phi(t_0)}{t - t_0}$$

if the limit exists, where the limit is taken in the sense of the norm of Y. We note that this derivative is unique and that if Φ has a first derivative at a point t_0, then Φ is continuous at t_0. Higher order derivatives are defined inductively as in classical analysis.

DEFINITION 3. *Let F be a mapping from an open subset X of E into Y, where E and Y are normed linear spaces. Let x_0 be a point in X and h an arbitrary nonzero fixed element in E. Then x_0+th is in X for $|t| \leqq \epsilon(x_0; h)$. Let*

$$\tau = \sup \{\epsilon: |t| \leqq \epsilon \Rightarrow x_0 + th \text{ in } X\}.$$

Then $F(x_0+th)$ is defined for $|t|<\tau$. If

$$\frac{d}{dt} F(x_0 + th) \Big|_{t=0}$$

exists, it is called the Gâteaux variation (or the weak differential) of F at x_0 with increment h and is denoted by $\delta F(x_0; h)$. If F has a Gâteaux variation, hereafter called G-variation, at every point x in X, then F is said to have a first variation on X.

Similarly, F has an nth variation $\delta^n F(x_0; h)$ at a point x_0 if the function $F(x_0+th)$ has an nth derivative with respect to t at $t=0$.

It follows from the definition, that the first variation is homogeneous in h of degree one, i.e. if $\delta F(x; h)$ exists, then for any scalar λ, $\delta F(x; \lambda h)$ exists and is equal to $\lambda \delta F(x; h)$. Similarly, $\delta^n F(x; \lambda h) = \lambda^n \delta F(x; h)$.

It should be emphasized, however, that the weak differential is not necessarily linear nor continuous in h, as may be seen from the following:

Example 1.

$$f(x_1, x_2) = \frac{x_1 x_2^2}{x_1^2 + x_2^2}, \qquad (x_1, x_2) \neq (0, 0); \qquad f(0, 0) = 0.$$

For each $h=(h_1, h_2)$, the G-variation exists and is equal to $h_1 h_2^2(h_1^2+h_2^2)^{-1}$, but the mapping $(h_1, h_2) \rightarrow h_1 h_2^2 (h_1^2+h_2^2)^{-1}$ is not linear in h.

The reader may contrast this remark and the next few results with properties of differentials and variations in *complex* normed spaces, where a different situation prevails [**8**, **11**].

Note that if F has a Gâteaux variation at x_0, then F is continuous in the direction h, i.e.

$$\lim_{t \to 0} \|F(x_0 + th) - F(x_0)\| = 0 \quad (h \text{ is fixed})$$

but is not necessarily continuous at x_0.

Example 2. Let E be the space of all functions $y=y(x)$ which have a continuous first derivative on $[a, b]$. Define a norm on E by

$$\|y\| = \max |y(x)| + \max |y'(x)|,$$

where the maximum is taken over $[a, b]$. Let $f(x, y, z)$ be a function which is defined and has continuous partial derivatives for all finite z and for $a \leqq x \leqq b$, $\Phi_1(x) \leqq y \leqq \Phi_2(x)$ for some prescribed functions Φ_1 and Φ_2. Let

$$J[y] = \int_a^b f(x, y(x), y'(x))\, dx.$$

Then a simple computation shows the G-variation of J at y, corresponding to the increment $h = h(x)$ is

$$\delta J[y; h] = \int_a^b [h(x) f_y(x, y, y') + h'(x) f_{y'}(x, y, y')]\, dx$$

which is the usual first variation.

5. Gâteaux differential. From the above remarks it is clear that the Gâteaux variation does not possess many of the important properties of total differentials for functions of several variables. This motivates the definitions of the Gâteaux and Fréchet differentials, hereafter called G- and F-differentials respectively, which will then enable us to study more effectively a functional or a nonlinear operator in the neighborhood of a fixed element in the space.

DEFINITION 4. *If* $\delta F(x_0; h)$ [Definition 3] *is linear and bounded in* h, *it is called the Gâteaux differential of* F *at* x_0 *with increment* h *and is denoted by* $DF(x_0; h)$.

The G-differential provides in some sense a local approximation property. More precisely, we have

THEOREM 1. *Let* X *be an open subset of* E *and let* F *be a nonlinear operator from* X *to* Y. *A necessary and sufficient condition for* F *to be* G-*differentiable at* x_0 *is that the following representation holds*:

$$F(x_0 + h) - F(x_0) = L(x_0; h) + R(x_0; h) \tag{5.1}$$

for every h *in* E *for which* $x_0 + h$ *is in* X, *where* $L(x_0; h)$ *is linear and continuous in* h *and*

$$\operatorname*{Lim}_{\tau \to 0} \frac{\|R(x_0; \tau h)\|}{\tau} = 0 \qquad \textit{for each } h. \tag{5.2}$$

Proof. We first remark that if such a representation exists, then it is unique. For if another representation exists with L' and R', then

$$\begin{aligned} L(x_0; h) - L'(x_0; h) &= \lim_{\tau \to 0} \tau^{-1}[L(x_0; \tau h) - L'(x_0; \tau h)] \\ &= \lim_{\tau \to 0} \tau^{-1}[R'(x_0; \tau h) - R(x_0; \tau h)] = 0. \end{aligned}$$

Now if the representation (5.1) holds, then

$$\begin{aligned} \left.\frac{dF(x_0 + \tau h)}{d\tau}\right|_{\tau=0} &= \lim_{\tau \to 0} \tau^{-1}[F(x_0 + \tau h) - F(x_0)] \\ &= L(x_0; h) + \lim_{\tau \to 0} \tau^{-1} R(x_0; \tau h) = L(x_0; h). \end{aligned}$$

Thus the G-variation exists and is linear and continuous in h. Conversely, if the G-differential exists, then

$$\tau^{-1}[F(x_0 + \tau k) - F(x_0)] = DF(x_0; k) + \epsilon(x_0; \tau k),$$

where $\epsilon(x_0; \tau k) \to 0$ as $\tau \to 0$. Letting $\tau k = h$, we get the representation (5.1), where $R(x_0; h) = \tau\epsilon(x_0; h)$ and thus (5.2) holds.

This theorem brings us closer to our objectives and a strengthening of condition (5.2) leads to the definition of the Fréchet differential in Section 6. In the rest of this section, we shall discuss other conditions for a G-variation to be a G-differential.

THEOREM 2. *A necessary and sufficient condition for $\delta F(x_0; h)$ to be linear and continuous in h is that F satisfies the following two conditions:*

(a) *To each h corresponds a $\delta(h)$ such that*

$$|t| \leqq \delta \text{ implies } \|F(x_0 + th) - F(x_0)\| \leqq M\|th\|,$$

where M does not depend on h.

(b) $\overset{2}{\Delta}_{th_1, th_2} F(x_0) = o(t)$ *where*

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

The proof of the theorem is straight-forward and is given for instance in [**16**, p. 39].

It was noted that $\delta F(x; h)$ is not necessarily linear nor continuous in h or x. It turns out, however, that if $\delta F(x; h)$ is continuous in x at x_0, then it is linear in h. More precisely we state:

THEOREM 3. (See, for instance [**16**, p. 37].) *If F has a G-variation in an open set U such that $\delta F(x; h)$ is continuous in x at some x_0 in U, then $\delta F(x; h)$ is additive in h, i.e. $\delta F(x_0; h+k)$ exists and is equal to $\delta F(x_0; h) + \delta F(x_0; k)$.*

Combining this result with the well-known property of linear operators stated in Section 2, we arrive at

THEOREM 4. *Let the G-variation of the operator F exist in some neighborhood of the point x_0 and let $\delta F(x; h)$ be continuous in x at x_0. Furthermore, assume that $\delta F(x_0; h)$ is continuous in h at $h = \theta$. Then $\delta F(x_0; h)$ is a G-differential.*

6. Fréchet differential.

DEFINITION 5. *The operator F is said to be Fréchet (strongly, totally) differentiable at x_0 if the representation* (5.1) *holds, where $L(x_0; h)$ is linear and continuous in h and moreover*

$$\lim_{h \to \theta} \frac{\|R(x_0; h)\|}{\|h\|} = 0. \tag{6.1}$$

The uniqueness of the Fréchet differential is a special case of the uniqueness of Gâteaux differential, in view of Theorem 5.

We write $L(x_0; h) = dF(x_0; h) = F'_{x_0}h$ and call it the F-differential of F at x_0 with increment h. The mapping $dF(x_0; \cdot) = F'_{x_0}(\cdot)$ which is a bounded linear operator is called the Fréchet *derivative* of F at x_0. It may be noted that the F-derivative is an element in the space $\mathfrak{L}_1$ (see Section 2), while the F-differential $dF(x_0; h)$ is an element in Y. This fact is obscured in the calculus of one real variable, where the derivative at a point is defined as a *number*, by the one-to-one correspondence that exists in this case between numbers and linear operators.

It is easy to show that if F is F-differentiable at x_0, then it is continuous at that point; this is not necessarily true, however, if F is G-differentiable. Furthermore, if F is continuous at x_0 then the requirement of continuity of $dF(x_0; h)$ in h in Definition 5 is redundant. This follows from the inequality

$$\|dF(x_0; h)\| \leqq \|F(x_0 + h) - F(x_0) - dF(x_0; h)\| + \|F(x_0 + h) - F(x_0)\|,$$

which shows that $dF(x_0; h)$ is continuous at $h = \theta$ and hence continuous everywhere.

REMARK 1. The norm $\|\cdot\|'$ is said to be equivalent to the norm $\|\cdot\|$ if there exist positive numbers m and M such that

$$m\|x\| \leqq \|x\|' \leqq M\|x\|$$

for all x in E. This is clearly an equivalence relation. The definitions of the differentials in the representations (5.1), (5.2), and (6.1) are given in terms of the norms on X and Y. However, it is easy to check that two equivalent norms lead to the same definitions of differentiability, i.e. (5.2) and (6.1) still hold if the norms are replaced by equivalent norms. In the case of a finite-dimensional space all norms are equivalent, so that the differentiability of a mapping F on E into Y is independent of the norms on E and Y, in addition to being independent of the coordinates.

Equivalent norms define the same "topology" so that differentiability depends only on the topologies of X and Y, in infinite-dimensional spaces.

Thus it is possible to extend the definitions of the F- and G-differentials to certain topological vector spaces. It may also be observed in view of (5.2) that G-differentiability is meaningful if E is a linear space and Y a topological linear space.

Example 3. Let $F: E^n \to E^m$, where E^n is the Euclidean n-space. That is, $y_i = F_i(x_1, \cdots, x_n)$, $i = 1, 2, \cdots, m$. Assume that F has an F-differential at $a = (a_1, \cdots, a_n)$. Then it is not hard to show that the partial derivatives $\partial F_i/\partial x_j$ exist at a, and that the F-derivative is the linear transformation whose matrix is $[\partial F_i/\partial x_j]$. Conversely, assume that there exists an $r > 0$ such that for x in $\|x - a\| \leqq r$, $\partial F_i/\partial x_j$ exist, and are continuous at $x = a$, then F has an F-differential at $x = a$. See, for instance, *Advanced Calculus*, by R. C. Buck.

Example 4. Let $K(s, t)$ be a continuous real function for $0 \leqq s, t \leqq 1$, and assume that K is symmetric, i.e. $K(s, t) = K(t, s)$. The functional

$$J[x] = \int_0^1 x^2(t)dt - \lambda \int_0^1 \int_0^1 K(s, t)x(s)x(t)dsdt$$

is defined on the space of all continuous real functions on $[0, 1]$ with norm $\|x\| = \max_{0 \leq t \leq 1} |x(t)|$. From a simple computation it follows that

$$\frac{d}{d\tau} J[x + \tau h] \Big|_{\tau=0} = 2\int_0^1 x(t)h(t)\, dt - 2\lambda \int_0^1 \int_0^1 K(s, t)x(s)h(t)\, dtds$$

which is linear and continuous in h. Condition (6.1) is also satisfied, so that the last expression is the F-differential of J.

7. Gradients. The definitions and properties given in the preceding sections are for mappings between normed linear spaces. Thus, they hold in particular for a functional f defined on a subset X of a normed linear space E and mapping each x in X into a real number $f(x)$. By Definitions 4 and 5, if f is differentiable on X then there exists a mapping df on $X \times E$ into the reals, which is linear and continuous in h in E. Another way of looking at df is to consider $df(x_0; h)$ as a continuous linear functional for each fixed x_0, and denote it by $f'_{x_0} h$. Then f'_{x_0} is an element of the dual space E^* of E. As x_0 varies over X, a mapping $f'_x(\cdot): X \rightarrow E^*$ is thus obtained, which Rothe [**15**] called the *gradient* mapping of f.

For example if f is a differentiable function of three real variables, then the differential of f at $\vec{x} = (x_1, x_2, x_3)$, with increment $\vec{h} = (h_1, h_2, h_3)$ is given by

$$(7.1) \qquad df(\vec{x}; \vec{h}) = \sum_{i=1}^{3} \frac{\partial f}{\partial x_i} h_i = \vec{\Gamma}(x) \cdot \vec{h}, \quad \text{where} \quad \vec{\Gamma}(x) = \operatorname{grad} f(x).$$

Thus $\vec{\Gamma}(x)$ assigns to each x in E a continuous linear functional df. In this example df is the inner product of grad f and h.

Similarly, if E is a complete inner product space H, ([**2**, p. 103] or [**3**, p. 111]) then the gradient mapping may be considered as a mapping from H into itself since the dual space may be indentified with H. Furthermore, $df(x_0; h)$, being a continuous linear functional in h, can be uniquely represented as an inner product [**3**, p. 117], i.e. there exists a unique $\Gamma(x_0)$ in H such that

$$(7.2) \qquad df(x_0; h) = (\Gamma(x_0), h),$$

where the parentheses denote inner product. $\Gamma(x_0)$, defined by (7.2) is called the gradient of the functional f at x_0 and is denoted by grad $f(x_0)$.

For instance it follows from Example 4, under the usual inner product $\int_a^b x(t)y(t)dt$, that

$$\tfrac{1}{2} \operatorname{grad} J[x] = x(t) - \lambda \int_0^1 K(s, t)x(s)\, ds.$$

The concept of gradient in abstract spaces was first introduced by M. Golomb in his study of nonlinear integral equations.

If in (7.2), we use the Gâteaux differential, then we obtain the definition of the *weak* gradient grad_w of f at x_0, i.e.

$$Df(x_0; h) = (\operatorname{grad}_w f(x_0), h). \tag{7.3}$$

REMARK 2. The gradient depends on the inner product. For instance if P is a positive operator, i.e. $(Px, x) > 0$ unless $x = \theta$, then we may define a *new* inner product $[x, y]$ by

$$[x, y] = (Px, y). \tag{7.4}$$

Let $\operatorname{grad} f$ and $\operatorname{grad}^\# f$ denote the gradient of f with respect to the original and new inner products respectively. Then by definition of gradient,

$$df(x; h) = [\operatorname{grad}^\# f, h] = (\operatorname{grad} f, h). \tag{7.5}$$

But from (7.4)

$$[\operatorname{grad}^\# f, h] = (P \operatorname{grad} f, h). \tag{7.6}$$

From (7.5) and (7.6) we get for all h, $(\operatorname{grad} f, h) = (P \operatorname{grad}^\# f, h)$. Hence,

$$\operatorname{grad} f = P \operatorname{grad}^\# f.$$

For example, if in E^3 we define "distance" by

$$d(x, y) = \left\{ \sum_{i,j=1}^{n} p_{ij}(y_j - x_j)(y_i - x_i) \right\}^{1/2} = (P(x - y), x - y)^{1/2},$$

where $P = [p_{ij}]$ is a positive definite matrix, then the gradient of f with respect to this metric is related to the gradient of f with respect to the usual Euclidean distance by

$$\operatorname{grad}^\# f = [p_{ij}]^{-1} \operatorname{grad} f.$$

8. Implication relationships between F- and G-differentiability. The only difference between the Fréchet and Gâteaux differentials is in the relations (5.2) and (6.1). We now show that (6.1) implies (5.2) but not conversely. This result is included in the following interesting characterization of the F-differential.

THEOREM 5. *The operator F is F-differentiable at x_0 if and only if the representation* (5.1) *holds, where $L(x_0; h)$ is continuous and linear in h and*

$$\lim_{\tau \to 0} \tau^{-1} \| R(x_0; \tau h) \| = 0 \tag{8.1}$$

uniformly with respect to h on each set $\|h\| = constant$.

Proof. Without any loss of generality, we may prove this for the set $\|h\| = 1$. If F is Fréchet differentiable at x_0, then

$$\lim_{\|h\|\to 0} \frac{\|R(x_0;\, h)\|}{\|h\|} = 0.$$

Letting $h = \tau k$, where k has a unit norm, we get $\lim_{\tau\to 0} \tau^{-1}\|R(x_0;\, \tau k)\| = 0$ uniformly on $\|k\| = 1$.

Conversely, if (8.1) holds uniformly on each bounded set, then, in view of Theorem 1, F has a G-differential $DF(x_0; h)$ at x_0. Thus for any given $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\|\tau^{-1}[F(x_0 + \tau h) - F(x_0)] - DF(x_0;\, h)\| < \epsilon,$$

whenever $|\tau| < \delta$. That is,

$$F(x_0 + \tau h) - F(x_0) = DF(x_0;\, \tau h) + R(x_0;\, \tau h),$$

where for $|\tau| < \delta$,

$$\frac{\|R(x_0;\, \tau h)\|}{\|\tau h\|} < \epsilon, \qquad (\|h\| = 1).$$

Letting $k = \tau h$, we get $F(x_0 + k) - F(x_0) = DF(x_0;\, k) + R(x_0;\, k)$, where

$$\lim_{k\to\theta} \frac{\|R(x_0;\, k)\|}{\|k\|} = 0.$$

Hence, $DF(x_0;\, k) = dF(x_0;\, k)$.

Thus if F is F-differentiable at x_0, then F is G-differentiable (and consequently it has a G-variation) at x_0. Furthermore

$$dF(x_0;\, h) = DF(x_0;\, h) = \delta F(x_0;\, h).$$

The converse holds if F is a function of one real variable, but does not necessarily hold in higher dimensions as may be seen from the following:

Example 5. Let $x = (x_1, x_2)$ where x_1 and x_2 are real variables and consider

$$f(x) = \frac{x_1}{x_2}(x_1^2 + x_2^2), \qquad x_2 \neq 0; \qquad f(x_1, 0) = 0,$$

and let $\|x\| = (|x_1|^2 + |x_2|^2)^{1/2}$. Then f has a G-variation at $x = (0, 0)$, which is (trivially) continuous and linear in h. In this case,

$$R(0;\, h) = \frac{h_1}{h_2}(h_1^2 + h_2^2), \quad h_2 \neq 0; \qquad R(0,\, h) = 0 \text{ if } h_2 = 0,$$

and hence (5.2) holds. However, the F-differential does not exist at (0, 0). For if we let $h_n = (n^{-1/2}, n^{-1})$ then $h_n \to \theta$ while

$$\frac{\|R(0;\, h_n)\|}{\|h_n\|} = \sqrt{1 + 1/n} \to 1, \qquad \text{as } n \to \infty.$$

REMARK 3. Neither the G-differential $Df(x_0; h)$ nor the F-differential $df(x_0; h)$ of a functional f is required to be continuous in x at x_0; and hence neither the gradient nor the weak gradient of f. It can be shown, however, that if $\text{grad}_w f(x)$ exists and is continuous in x on an open set Ω, then it coincides with grad $f(x)$. This then implies that f is F-differentiable and $df(x; h) = Df(x; h)$.

REMARK 4. The requirement of continuity of grad f is related to the notion of *uniform* Fréchet differentials. F is said to have a *locally uniform* F-differential $dF(x; h)$ on an open set Ω if F has an F-differential on Ω and the remainder $R(x_0; h)$ is locally uniformly bounded, i.e. for each $\epsilon > 0$ and an arbitrary x_0 in Ω, there exists a $\delta(x_0; \epsilon)$ and $\eta(x_0; \epsilon)$ such that

$$\|R(x; h)\| \leqq \epsilon\|h\| \quad \text{if} \quad \|h\| \leqq \delta \quad \text{and} \quad \|x - x_0\| \leqq \eta.$$

It turns out that a necessary and sufficient condition for grad $f(x)$ to be continuous in the sphere S: $\|x\| < a$ is that $df(x; h)$ have a locally uniform remainder and grad $f(x)$ be locally bounded. We leave the discussion of the interesting implications of these remarks and proofs in the case of a function of several real variables to the reader, (see Example 3).

9. Higher order differentials. The first order differential $dF(x; h)$ is a function of two variables. Thus, several notions for a second order differential may be defined. The most natural notion, which we will discuss here, is based on the observation that $F_x'(\cdot) = dF(x; \cdot)$ which is an element of the space $\mathfrak{L}_1$, is also an *operator* sending X into the space $\mathfrak{L}_1$. If this operator is F-differentiable, its derivative is called the second order Fréchet derivative of F and is denoted by $F_x''(\cdot, \cdot)$. Thus the second order derivative is an element of the space $\mathfrak{L}_2$ of all continuous linear operators from E into $\mathfrak{L}_1$, i.e. it is a bilinear operator from E to Y; it also has the representation

$$dF(x + k; h) - dF(x; h) = F_x'' hk + R(x; h, k),$$

where

$$\lim_{k \to 0} \frac{\|R(x; h, k)\|}{\|k\|} = 0.$$

$F_x'' hk = d^2F(x; h, k)$ is called the second order Fréchet differential of F.

The F-differential of the nth order may be defined inductively as follows:

DEFINITION 6. *Let E and Y be normed linear spaces over the field of real numbers and X be an open subset of E. Suppose that for some integer $m \geqq 2$, the m-th order F-differential $d^mF(x_0; h_1 \cdots h_m)$ of the mapping F: $X \to Y$, has been defined for all $(m+1)$-tuples $(x_0, h_1, \cdots, h_m)$ of elements of E such that $x_0 + \sum_{i=1}^m h_i$ is in X. Then F is said to have an F-differential of order $m+1$, if for all $(m+2)$-tuples of elements $(x_0, h_1, \cdots, h_{m+1})$ of E such that $x_0 + \sum_{i=1}^{m+1} h_i$ is in X, the following representation holds:*

$$d^mF(x_0 + h_{m+1}; h_1, \cdots, h_m) - d^mF(x_0; h_1, \cdots, h_m)$$
$$= d^{m+1}F(x_0; h_1, \cdots, h_m, h_{m+1}) + R(x_0; h_1, \cdots, h_{m+1}),$$

where the mapping $d^mF(x; h_1, \cdots, h_{m+1})$ *is linear and continuous in* h_{m+1} *and*

$$\lim_{\|h_{m+1}\|\to 0} \|h_{m+1}\|^{-1}\|R(x_0; h_1, \cdots, h_m, h_{m+1})\| = 0.$$

If such a representation exists, it is unique and $d^{m+1}F(x_0; h_1, \cdots, h_{m+1})$ is called the $(m+1)$-th F-differential of F at x_0. The operator $d^{m+1}F(x_0; \cdots)$ is called the $(m+1)$-th Fréchet derivative. An inductive argument shows that $d^{m+1}F(x_0; h_1, \cdots, h_{m+1})$ is $(m+1)$-linear in $h_1, \cdots, h_{m+1}$ (Definition 1). Symmetry of the mth F-differential in $h_1, \cdots, h_m$ may therefore be defined (Definition 2). An interesting result, which is a generalization of the sufficient condition that makes immaterial the order of mixed partial differentiation for real functions of several variables, may be stated as follows: a sufficient condition for $d^mF(x; h_1, \cdots, h_m)$ to be symmetric in $h_1, \cdots, h_m$ at $x=x_0$ is that $d^mF(x; h_1, \cdots, h_m)$ be continuous in x for all x in some neighborhood of x_0 (Theorem 8 in [**6**]).

The Gâteaux differential of order m may be defined similarly; implication relationships similar to those established in Section 8 may be stated. For example, if F has an F-differential of the mth order at x_0, then the mth order variation of F at that point exists and moreover

$$d^mF(x_0; h_1, \cdots, h_m) = \frac{\partial^m}{\partial t_1 \cdots \partial t_m} F\left(x_0 + \sum_{i=1}^{m} t_i h_i\right)\Bigg|_{t_1=\cdots=t_m=0}.$$

In particular, if $h_1= \cdots =h_m=h$, then

$$d^mF(x_0; h, \cdots, h) = \frac{d^m}{dt^m} F(x_0 + th)\Bigg|_{t=0}.$$

10. Remarks.

A. It should be noted that the notions presented in the preceding sections are *coordinate free* and that the derivative is defined in an invariant form as a linear transformation. This approach is useful in applied mathematics for formulating simultaneous algebraic equations, integral equations and boundary-value problems, etc., as operator equations, and for approximate and iterative methods for solving these equations. See for instance the interesting expository paper on Newton's method and variations by R. H. Moore in [**1**] and [**9, 14, 16**]. The compactness of notation that results from this approach and the abstractness of the notions are assets to the conceptual framework, unifying diverse situations in analysis and approximation theory.

B. Some of the recent books in advanced calculus and real variables, treat differentials of mapping from regions in E^n to E^m in the spirit of linear transformations. We refer specifically to the outstanding books on *Advanced Calculus*, by T. M. Apostol; R. C. Buck; W. Maak; H. K. Nickerson, D. C. Spencer, and N. E. Steenrod; and to *Principles of Mathematical Analysis* (2nd edition) by W. Rudin, *The Elements of Real Analysis* by R. G. Bartle and [**3**]. An elementary introduction is also given in *Calculus of Vector Functions* by R. H. Crowell and R. E. Williamson. See also the recent books by W. H. Fleming and C. Goffman on functions of several variables.

C. Differentiation rules, chain rules, mean value theorems, Taylor's formula, etc., can be developed for F- and G-differentials as in classical calculus. Partial F- and G-differentials can also be defined paralleling the classical theory.

11. Some open questions. One mathematician remarked that a colloquium lecture in mathematics should include at least one proof and one open problem. We assume, without further discussion, that this also holds for an expository article. We conclude therefore by mentioning some open questions which can be stated within the framework of this paper.

A. The derivative $dF(x;\ \cdot)$ is not necessarily continuous in x. The question then arises as when we can approximate a nonlinear operator by a continuous Fréchet derivative or, more generally, by another *nonlinear* map with a continuous Fréchet derivative, in the case of infinite-dimensional spaces.

B. It is known that a real function of a real variable which satisfies a Lipschitz condition is differentiable almost everywhere. This follows from the fact that if $|f(x)-f(y)| \leqq M|x-y|$ for all x and y in $[a, b]$, then

$$\sum_{i=1}^{n-1} |f(x_{i+1}) - f(x_i)| \leqq M \sum_{i=1}^{n-1} |x_{i+1} - x_i|,$$

for any partition $a \leqq x_1 < x_2 < \cdots < x_n \leqq b$. Thus f is of bounded variation, and hence being the difference of two monotonic functions is differentiable almost everywhere. The questions that arise are:

(i) Does an operator which is Lipschitz continuous, i.e.

$$\|F(x) - F(y)\| \leqq M\|x - y\| \quad \text{for all } x \text{ and } y,$$

have any G- or F-differentiability properties almost everywhere?

(ii) What additional hypotheses are sufficient to imply G- or F-differentiability for a Lipschitz continuous operator?

C. A functional defined on a linear space E (or on a convex subset of E) is said to be *convex* if for all x and y in the domain of f and for $0 \leqq a \leqq 1$, $f[ax+(1-a)y] \leqq af(x)+(1-a)f(y)$. A convex functional defined on an open subset has a one-sided G-variation, i.e.

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

exists, $t > 0$. Furthermore, if f is continuous and convex on the real interval $[a, b]$, then f has a right-hand and left-hand derivative at every point and the subset on which f' does not exist is countable. See, for instance, pages 195–196 in *Analysis* by E. Hille.

We ask the same questions as in **B** for convex functionals, i.e., what hypotheses imply G- or F-differentiability of a convex functional, which would not imply the same for arbitrary functionals?

A similar question can be posed for monotone operators (see the paper by Dolph and Minty in [1]).

References

1. P. M. Anselone, editor, Nonlinear integral equations, Univ. of Wisconsin Press, Madison, 1964.

2. R. C. Buck, editor, Studies in modern analysis, Vol. 1, MAA Studies, 1962.

3. J. Dieudonné, Foundations of modern analysis, Academic Press, New York, 1960.

4. M. Fréchet, La notion de différentielle dans l'analyse générale, Ann. Sci. École Norm. Sup., 42 (1925) 293–323.

5. R. Gâteaux, Sur les fonctionnelles continues et les fonctionnelles analytiques, Bull. Soc. Math. France, 50 (1922) 1–21.

6. L. M. Graves, Riemann integration and Taylor's theorem in general analysis, Trans. Amer. Math. Soc., 29 (1927) 163–177.

7. T. H. Hildebrandt and L. M. Graves, Implicit functions and their differentials in general analysis, Trans Amer. Math. Soc., 29 (1927) 127–153.

8. E. Hille and R. S. Phillips, Functional analysis and semigroups, Amer. Math. Soc. Coll. Publ., Providence, 1957, (see Chapter 26).

9. L. V. Kantorovich and G. P. Akhilov, Functional Analysis in Normed Spaces, Pergamon Press, New York, 1964.

10. M. Kerner, Die Differentiale in der allgemeinen Analysis, Ann. of Math., 34 (1933) 546–572.

11. A. D. Michal, Le calcul différentiel dans les espaces de Banach, vol. 1, Gauthier-Villars, Paris, 1958.

12. M. E. Munroe, Bringing calculus up-to-date, this MONTHLY, 65 (1958) 81–90.

13. M. E. Munroe, Modern Multidimensional Calculus, Addison-Wesley, Reading, Mass., 1963.

14. M. Z. Nashed, On general iterative methods for the solutions of a class of nonlinear operator equations, Math. Comp., 19 (1965) 14–24.

15. E. H. Rothe, Gradient mappings, Bull. Amer. Math. Soc., 59 (1953) 5–19.

16. M. M. Vainberg, Variational Methods for the Study of Nonlinear Operators, Holden-Day, San Francisco, 1964.

17. A. Bronsted and R. T. Rochafellar, On the subdifferentiability of convex functions, Proc. Amer. Math. Soc., 16 (1965) 605–611.

18. E. B. Leach, Differentiability of sub-convex functionals on Banach spaces, Amer. Math. Soc. Notices, 625–164, 1965.

19. T. Rado and P. V. Reichelderfer, Continuous Transformations in Analysis, Springer Verlag, Berlin, 1955.