A UNIFIED APPROACH TO MATHEMATICAL OPTIMIZATION AND LAGRANGE MULTIPLIER THEORY FOR SCIENTISTS AND ENGINEERS

RICHARD A. TAPIA

Appendix E: Differentiation in Abstract Spaces

It should be no surprise that the differentiation of functionals (real valued functions) defined on abstract spaces plays a fundamental role in continuous optimization theory and applications. Differential tools were the major part of the early calculus of variations and actually it was their use that motivated the terminology calculus of variations. The tools that we develop that allow us to work in the full generality of vector spaces are the so-called one-sided directional variation, the directional variation, and the directional derivative. Their development will be our first task. In some theory and applications we need more powerful differentiation tools. Towards this end our second task will be the presentation of the Gâteaux and Fréchet derivatives. Since these derivatives offer more powerful tools, their definition and use will require stronger structure than just vector space structure. The price that we have to pay is that we will require a normed linear space structure. While normed linear spaces are not as general as vector spaces, the notion of the Fréchet derivative in a normed linear space is a useful theoretical tool and carries with it a very satisfying theory. Hence, it is the preferred differentiation notion in theoretical mathematics. Moreover, at times in our optimization applications we will have to turn to this elegant theory. However, we stress that the less elegant differentiation notions will often lead us to surprisingly general and useful applications; therefore, we promote and embrace the study of all these theories.

We stress that the various differentiation notions that we present are not made one bit simpler, or proofs shorter, if we restrict our attention to finite dimensional spaces, indeed to $I\!R^n$. Dimension of spaces is not an inherent part of differentiation when properly presented. However, we quickly add that we will often turn to $I\!R^n$ for examples because of its familiarity and rich base of examples. The reader interested in consulting references on differentiation will find a host of treatments in elementary analysis books in the literature. However, we have been guided mostly by Ortega and Rheinboldt [2]. Our treatment is not the same as theirs but it is similar. In consideration of the many students that will read our presentation, we have included numerous examples throughout our presentation and have given them in unusually complete detail.

1. Introduction. The notion of the derivative of a function from the reals to the reals has been known and understood for centuries. It can be traced back to Leibnitz and Newton in the late 1600's and, at least in understanding, to Fermat and Descartes in the early 1600's. However, the extension from $f : I\!R \to I\!R$ to $f : I\!R^n \to I\!R$. and then to $f : I\!R^n \to I\!R^m$ was not handled well. A brief presentation of this historical development follows. We motivate this discussion with the leading question: What comes first, the derivative or the gradient? Keep in mind that the derivative should be a linear form, and not a number, as was believed incorrectly for so many years. It should be even more embarrassing that this belief in the number interpretation persisted even in the presence of the notions of the directional derivative, the Gâteaux derivative, and the Fréchet derivative, all linear forms that we study in this appendix. It must have been that these latter notions were all viewed as different animals, whose natural habitat was the abstract worlds of vector spaces and not $I\!R^n$.

Consider a function $f : I\!R \to I\!R$. From elementary calculus we have the following notion of the derivative of f at x:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}. \tag{1.1}$$

If instead we consider $f : I\!R^n \to I\!R$, then (1.1) no longer makes sense, since we cannot form $x + \triangle x$, i.e., we cannot add scalars and vectors and we cannot divide by vectors. So, historically as is so often done in mathematics the problem at hand is reduced to a problem with known solution, i.e., to the known one dimensional case, by considering the partial derivatives of f at x

$$\partial_i f(x) = \lim_{\Delta x \to 0} \frac{f(x_1, \ldots, x_i + \Delta x, \ldots, x_n) - f(x_1, \ldots, x_n)}{\Delta x}, \quad i = 1, \ldots, n. \tag{1.2}$$

The partial derivatives are then used to form $\nabla f(x)$, the so-called gradient of f at x, i.e.,

$$\nabla f(x) = \begin{bmatrix} \partial_1 f(x) \\ \partial_2 f(x) \\ \vdots \\ \partial_n f(x) \end{bmatrix}. \tag{1.3}$$

Then the gradient is used to build the so-called total derivative or differential of f at x as

$$f'(x)(\eta) = \langle \nabla f(x), \eta \rangle. \tag{1.4}$$

These ideas were readily extended to $F : \mathbb{R}^n \to \mathbb{R}^m$ by writing

$$F(x) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))^T$$

and defining

$$F'(x)(\eta) = \begin{bmatrix} \langle \nabla f_1(x), \eta \rangle \\ \vdots \\ \langle \nabla f_m(x), \eta \rangle \end{bmatrix}. \tag{1.5}$$

If we introduce the so-called Jacobian matrix,

$$JF(x) = \begin{bmatrix} \partial_1 f_1(x) & \ldots & \partial_n f_1(x) \\ & \ldots & \\ \partial_1 f_m(x) & \ldots & \partial_n f_m(x) \end{bmatrix}, \tag{1.6}$$

then we can write

$$F'(x)(\eta) = JF(x)\eta. \tag{1.7}$$

This standard approach is actually backwards and is flawed. It is flawed because it depends on the notion of the partial derivatives which in turn depends on the notion of a natural basis for $\mathbb{R}^n$, and the dependence becomes rather problematic once we leave $\mathbb{R}^n$. The backwardness follows from the realization that the gradient and hence the Jacobian matrix follow from the derivative and not the other way around. This will become readily apparent in §2 and §3.

Historically, the large obstacle to producing satisfactory understanding in the area of differentiation came from the failure to view $f'(x)$ in (1.1) as the linear form $f'(x) \cdot \eta$ evaluated at $\eta = 1$. In $\mathbb{R}$ the number a and the linear form $a \cdot \eta$ are easily identified, or at least not separated. It followed that historically a derivative was always seen as a number and not a linear form. This led to the notion of partial derivative, and in turn to the linear form $f'(x)$ given in (1.3). This linear form was not seen as a number, hence it was not seen as a derivative. This motivated the terminology differential and at times total derivative, for the linear form defined in (1.3). Much confusion resulted in terms of the usage derivative, differential, and total derivative. We reinforce our own critical statements by quoting from the introductory page to Chapter **??**, the differential calculus chapter, in Dieudonne [].

"*That presentation, which throughout adheres strictly to our general geometric outlook on analysis, aims at keeping as close as possible to the fundamental idea of calculus, namely the local approximation of functions by **linear** functions. In the classical teaching of Calculus, this idea is immediately obscured by the accidental fact that, on a one-dimensional vector space, there is a one-to-one correspondence between linear forms and numbers, and therefore the derivative at a point is defined as a **number** instead of a **linear form**. This slavish subservience to the Shibboleth of numerical interpretation at any cost becomes worse when dealing with functions of several variables.*"

2. Basic Differential Notions. We begin with a formal definition of our differential notions except for the notion of the Fréchet derivative, it will be presented in §4.

Definition 2.1. *Consider $f : X \to Y$ where X is a vector space and Y is a topological linear space. Given $x, \eta \in X$, if the (one-sided) limit*

$$f'_+(x)(\eta) = \lim_{t \downarrow 0} \frac{f(x+t\eta) - f(x)}{t} \tag{2.1}$$

exists, then we call $f'_+(x)(\eta)$ the right or forward directional variation of f at x in the direction η.

If the (two-sided) limit

$$f'(x)(\eta) = \lim_{t \to 0} \frac{f(x+t\eta) - f(x)}{t} \tag{2.2}$$

exists, then we call $f'(x)(\eta)$ the directional variation of f at x in the direction η. The terminology forward directional variation of f at x, and the notation $f'_+(x)$, imply that $f'_+(x)(\eta)$ exists for all $\eta \in X$ with an analogous statement for the directional variation.

If $f'(x)$, the directional variation of f at x, is a linear form, i.e., $f'(x)(\eta)$ is linear in η, then we call $f'(x)$ the directional derivative of f at x.

If, in addition, X and Y are normed linear spaces and $f'(x)$, the directional derivative of f at x, exists as a bounded linear form ($f'(x) \in L[X, Y]$), then we call $f'(x)$ the Gâteaux derivative of f at x.

Furthermore, we say that f is directionally, respectively Gâteaux, differentiable in $D \subset X$ to mean that f has a directional, respectively Gâteaux, derivative at all points $x \in D$.

Remark 2.2. *Clearly the existence of the directional variation means that the forward directional variation $f'_+(x)(\eta)$, and the backward directional variation*

$$f'_-(x)(\eta) = \lim_{t \uparrow 0} \frac{f(x+t\eta) - f(x)}{t} \tag{2.3}$$

exist and are equal.

Remark 2.3. *In the case where $f : I\!R^n \to I\!R^m$ the notions of directional derivative and Gâteaux derivative coincide, since in this setting all linear operators are bounded. See Theorem C.5.4.*

Remark 2.4. *The literature abounds with uses of the terms derivative, differential, and variation with qualifiers directional, Gâteaux, or total. In this context, we have strongly avoided the use of the confusing term differential and the even more confusing qualifier total. Moreover, we have spent considerable time and effort in selecting terminology that we believe is consistent, at least in flavor, with that found in most of the literature.*

Remark 2.5. *It should be clear that we have used the term variation to mean that the limit in question exists, and have reserved the use of the term derivative to mean that the entity under consideration is linear in the direction argument. Hence derivatives are always linear forms, variations are not necessarily linear forms.*

Remark 2.6. *The term variation comes from the early calculus of variations and is usually associated with the name Lagrange.*

In all our applications the topological vector space Y will be the reals $I\!R$. We presented our definitions in the more general setting to emphasize that the definition of variation only requires a notion of convergence in the range space.

In terms of homogeneity of our variations, we have the following properties.

Proposition 2.7. *Consider $f : X \to Y$ where X is a vector space and Y is a topological vector space. Also, consider $x, \eta \in X$. Then*

(i) The forward directional variation of f at x in the direction η is non-negatively homogeneous in η, i.e.,

$$f'_+(x)(\alpha\eta) = \alpha f'_+(x)(\eta) \quad \textit{for} \quad \alpha \geq 0.$$

(ii) If $f'_+(x)(\eta)$ is homogeneous in η, then $f'_-(x)(\eta)$ exists and the two are equal. Hence, if $f'_+(x)(\eta)$ is homogeneous in η, then $f'_+(x)(\eta)$ is the directional variation of f at x.

(iii) The directional variation of f at x in the direction η is homogeneous in η, i.e,

$$f'(x)(\alpha\eta) = \alpha f'(x)(\eta) \quad \textit{for all} \quad \alpha \in I\!R.$$

Proof. The proofs follow directly from the definitions and the fact that

$$f'_-(x)(\eta) = -f'_+(x)(-\eta).$$

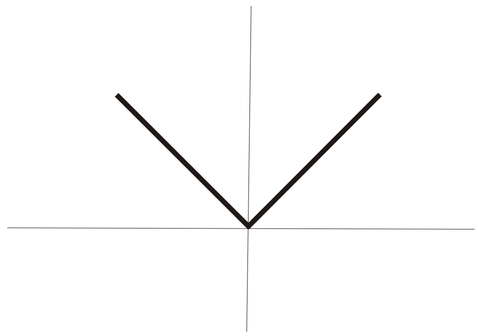

FIG. 2.1. *Absolute value function*

We now consider several examples. Figure E.1 shows the graph of the function $f(x) = |x|$. For this function $f'_+(0)(1) = 1$ and $f'_-(0)(1) = -1$. More generally, if X is a normed linear space and we define $f(x) = \|x\|$, then $f'_+(0)(\eta) = \|\eta\|$ and $f'_-(0)(\eta) = -\|\eta\|$. This shows the obvious fact that a forward (backward) variation need not be a directional variation.

The directional variation of f at x in the direction η, $f'(x)(\eta)$, will usually be linear in η; however, the following example shows that this is not always the case.

Example 2.8. Let $f : R^2 \to R$ be given by $f(x) = \dfrac{x_1 x_2^2}{x_1^2 + x_2^2}$, $x = (x_1, x_2) \neq 0$ and $f(0) = 0$. Then $f'(0)(\eta) = \dfrac{\eta_1 \eta_2^2}{\eta_1^2 + \eta_2^2}$.

The following example shows that the existnce of the partial derivatives is not a sufficient condition for the existence of the directional variation.

Example 2.9. Let $f : R^2 \to R$ be given by

$$f(x) = \frac{x_1 x_2}{x_1^2 + x_2^2}, \quad x = (x_1, x_2)^T \neq 0 \quad \text{and} \quad f(0) = 0.$$

Then $f'(0)(\eta) = \lim_{t \to 0} \dfrac{1}{t} \dfrac{\eta_1 \eta_2}{(\eta_1^2 + \eta_2^2)}$ exists if and only if $\eta = (\eta_1, 0)$ or $\eta = (0, \eta_2)$.

In §5 we will show that the existence of continuous partial derivatives is sufficient for the existence of the Gâteaux derivative, and more.

The following observation, which we state formally as a proposition, is extremely useful when one is actually calculating variations.

Proposition 2.10. *Consider $f : X \to I\!R$, where X is a real vector space. Given $x, \eta \in I\!R$, let $\phi(t) = f(x+t\eta)$. Then*

$$\phi'_+(0) = f'_+(x)(\eta) \quad \textit{and} \quad \phi'(0) = f'(x)(\eta).$$

Proof. The proof follows directly once we note that

$$\frac{\phi(t) - \phi(0)}{t} = \frac{f(x + t\eta) - f(x)}{t}.$$

Example 2.11. Given $f(x) = x^T A\, x$, with $A \in I\!R^{n \times n}$ and $x, \eta \in I\!R^n$, find $f'(x)(\eta)$. First, we define

$$\phi(t) = f(x + t\,\eta)\& = x^T\ A\, x + t\, x^T\ A\, \eta + t\, \eta^T\ A\, x + t^2 \eta^T A\, \eta. \tag{2.4}$$

Then, we have

$$\phi'(t) = x^T\ A\eta + \eta^T\ Ax + 2t\eta^T\ A\eta, \tag{2.5}$$

and by the proposition

$$\phi'(0) = x^T\ A\eta + \eta^T\ Ax \qquad\qquad = f'(x)(\eta). \tag{2.6}$$

Recall that the ordinary mean-value theorem from differential calculus says that given $a, b \in I\!R$ if the function $f : I\!R \to I\!R$ is differentiable on the open interval (a, b) and continuous at $x = a$ and $x = b$, then there exists $\theta \in (0, 1)$ so that

$$f(b) - f(a) = f'(a + \theta(b - a))(b - a). \tag{2.7}$$

We now consider an analog of the mean-value theorem for functions which have a directional variation. We will need this tool in proofs in §5.

Proposition 2.12. *Let X be a vector space and Y a normed linear space. Consider an operator $f : X \to Y$. Given $x, y \in X$ assume f has a directional variation at each point of the line $\{x + t(y - x) : 0 \le t \le 1\}$ in the direction $y - x$. Then for each bounded linear functional $\delta \in Y^*$*

(i) $\delta(f(y) - f(x)) = \delta[f'(x + \theta(y - x))(y - x)]$, *for some* $0 < \theta < 1$.

Moreover,

(ii) $\|f(y) - f(x)\| \le \sup\limits_{0<\theta<1} \|f'(x + \theta(y - x))(y - x)\|$.

(iii) If in addition X is a normed linear space and f is Gâteaux differentiable on the given domain, then

$$\|f(y) - f(x)\| \le \sup_{0<\theta<1} \|f'(x + \theta(y - x))\|\|y - x\|.$$

(iv) If in addition f has a Gâteaux derivative at an additional point $x_0 \in X$, then

$$\|f(y) - f(x) - f'(x_0)(y - x)\| \le \sup_{0<\theta<1} \|f'(x + \theta(y - x)) - f'(x_0)\|\|y - x\|.$$

Proof. Consider $g(t) = \delta(f(x + t(y - x)))$ for $t \in [0, 1]$. Clearly $g'(t) = \delta(f'(x + t(y - x))(y - x))$. So the conditions of the ordinary mean-value theorem (2.7) are satisfied for g on $[0, 1]$ and $g(1) - g(0) = g'(\theta)$ for some $0 < \theta < 1$. This is equivalent to *(i)*. Using Corollary C.7.3 of the Hahn-Banach Theorem C.7.2 choose $\delta \in Y^*$ such that $\|\delta\| = 1$ and $\delta(f(y) - f(x)) = \|f(y) - f(x)\|$. Clearly *(ii)* now follows from *(i)* and *(iii)* follows from *(ii)*. Inequality *(iv)* follows from *(iii)* by first observing that $g(x) = f(x) - f'(x_0)(x)$ satisfies the conditions of *(iii)* and then replacing f in *(iii)* with g.

While the Gâteaux derivative notion is quite useful; it has its shortcomings. For example, the Gâteaux notion is deficient in the sense that a Gâteaux differentiable function need not be continuous. The following example demonstrates this phenomenon.

Example 2.13. Let $f : R^2 \to R$ be given by $f(x) = \frac{x_1^3}{x_2}$, $x = (x_1, x_2) \ne 0$, and $f(0) = 0$. Then $f'(0)(\eta) = 0$ for all $\eta \in X$. Hence $f'(0)$ exists and is a continuous linear operator, but f is not continuous at 0.

We accept this deficiency, but quickly add that for our more general notions of differentiation we have the following limited, but actually quite useful, form of continuity. Ortega and Rheinboldt [2] call this hemicontinuity.

Proposition 2.14. *Consider the function $f : X \to I\!R$ where X is a vector space. Also consider $x, \eta \in X$. If $f'(x)(\eta)$ (respectively $f'_+(x)(\eta)$) exists, then*

$$\phi(t) = f(x + t\eta)$$

as a function from $I\!R$ to $I\!R$ is continuous (respectively continuous from the right) at $t = 0$.

Proof. The proof follows by writing

$$f(x + t\eta) - f(x) = \frac{(f(x + t\eta) - f(x))t}{t}$$

and then letting t approach 0 in the appropriate manner.

Example 2.15. *Consider $f : I\!R^n \to I\!R$. It should be clear that the directional variation of f at x in the coordinate directions $e^1 = (1, 0, \ldots, 0)^T, \ldots, e^n = (0, \ldots, 0, 1)^T$ are the partial derivatives $\partial_i f(x)$, $i = 1, \ldots, n$ given in* (1.2).

Next consider the case where $f'(x)$ is linear. Then from Theorem **??** *we know that $f'(x)$ must have a representation in terms of the inner product, i.e., there exists $a(x) \in I\!R^n$ such that*

$$f'(x)(\eta) = \langle a(x), \eta\rangle \quad \forall \eta \in I\!R^n. \tag{2.8}$$

Now by the linearity of $f'(x)(\eta)$ in $\eta = (\eta_1, \ldots, \eta_n)^T$ we have

$$\begin{aligned} f'(x)(\eta) &= f'(x)(\eta_1 e_1 + \ldots + \eta_n e_n) = \eta_1 \partial_1 f(x) + \ldots + \eta_n \partial_n f(x) \\ &= \langle \nabla f(x), \eta \rangle \end{aligned} \tag{2.9}$$

where $\nabla f(x)$ is the gradient vector of f at x given in (1.3)*. Hence, $a(x)$, the representer of the directional derivative at x, is the gradient vector at x and we have*

$$f'(x)(\eta) = \langle \nabla f(x), \eta \rangle \quad \forall \eta \in I\!R^n. \tag{2.10}$$

Example 2.16. *Consider $F : I\!R^n \to I\!R^m$ which is directionally differentiable at x. We write*

$$F(x) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))^T.$$

By Theorem C 5.4 we know that there exists an $m \times n$ matrix $A(x)$ so that

$$F'(x)(\eta) = A(x)\eta \quad \forall \eta \in I\!R^n.$$

The argument used in Example 2.15 can be used to show that

$$F'(x)(\eta) = \begin{bmatrix} \langle \nabla f_1(x), \eta \rangle \\ \vdots \\ \langle \nabla f_m(x), \eta \rangle \end{bmatrix}.$$

It now readily follows that the matrix representer of $F'(x)$ is the so-called Jacobian matrix, $JF(x) =$ given in (1.5)*, i.e., $F'(x)(\eta) = JF(x)\eta$.*

Remark 2.17. *We stress that these representation results are straightforward because we have made the strong assumption that the derivatives exist. The challenge is to show that the derivatives exist from assumptions on the partial derivatives. We do exactly this in §5.*

3. The Gradient Vector. While (1.4) and (2.8) end up at the same place, it is important to observe that in (1.4) the derivative was defined by first obtaining the gradient, while in (2.8) the gradient followed from the derivative as its representer. This distinction is very important if we want to extend the notion of the gradient to infinite dimensional spaces. We reinforce these points by extending the notion of gradient to inner product spaces.

Definition 3.1. *Consider $f : H \to I\!R$ where H is an inner product space with inner product $\langle \cdot, \cdot \rangle$. Assume that f is Gâteaux differentiable at x. Then if there exists $\nabla f(x) \in H$ such that*

$$f'(x)(\eta) = \langle \nabla f(x), \eta \rangle \quad \textit{for all } \eta \in H,$$

we call this (necessarily unique) representer of $f'(x)(\cdot)$ the gradient of f at x.

Remark 3.2. *If our inner product space H is complete, i.e. a Hilbert space, then by the Riesz representation theorem (Theorem C.6.1) the gradient $\nabla f(x)$ always exists. However, while the calculation of $f'(x)$ is almost always a straightforward matter, the determination of $\nabla f(x)$ may be quite challenging.*

Remark 3.3. *Observe that while $f' : H \to H^*$ we have that $\nabla f : H \to H$.*

It is well known that in this generality $\nabla f(x)$ is the unique direction of steepest ascent of f at x in the sense that $\nabla f(x)$ is the unique solution of the optimization problem

$$\begin{array}{ll} \text{maximize} & f'(x)(\eta) \\ \text{subject to} & \|\eta\| \le \|\nabla f(x)\|. \end{array}$$

The proof of this fact is a direct application of the Cauchy-Schwarz inequality. The well-known method of steepest descent considers iterates of the form $x - \alpha \nabla f(x)$ for $\alpha > 0$. This statement is made to emphasize that an expression of the form $x - \alpha f'(x)$ would be meaningless. This notion of gradient in Hilbert space is due to Golomb [ ] in 1934. Extension of the notion of the gradient to normed linear spaces is an interesting topic and has been considered by several authors, see Golomb and Tapia [ ] 1972 for example. This notion is discussed in Chapter 14.

4. The Fréchet Derivative. The directional notion of differentiation is quite general. It allows us to work in the full generality of real-valued functionals defined on a vector space. In many problems, including problems from the calculus of variations this generality is exactly what we need and serves us well. However, for certain applications we need a stronger notion of differentiation. Moreover, many believe that a satisfactory notion of differentiation should possess the property that differentiable functions are continuous. This leads us to the concept of the Fréchet derivative.

Recall that by $L[X,Y]$ we mean the vector space or bounded linear operators from the normed linear space X to the normed linear space Y. We now introduce the notion of the Fréchet derivative.

Definition 4.1. *Consider $f : X \to Y$ where both X and Y are normed linear spaces. Given $x \in X$, if a linear operator $f'(x) \in L[X,Y]$ exists such that*

$$\lim_{\triangle x \to 0} \frac{\|f(x+\triangle x) - f(x) - f'(x)(\triangle x)\|}{\|\triangle x\|} = 0, \quad \triangle x \in X, \tag{4.1}$$

then $f'(x)$ is called the Fréchet derivative of f at x. The operator $f' : X \to L[X,Y]$ which assigns $f'(x)$ to x is called the Fréchet derivative of f.

Remark 4.2. *Contrary to the directional variation at a point, the Fréchet derivative at a point is by definition a continuous linear operator.*

Remark 4.3. *By* (4.1) *we mean, given $\epsilon > 0$ there exists $\delta > 0$ such that*

$$\|f(x+\triangle x) - f(x) - f'(x)(\triangle x)\| \le \epsilon \|\triangle x\| \tag{4.2}$$

for every $\Delta x \in X$ such that $\|\Delta x\| \le \delta$.

Proposition 4.4. *Let X and Y be normed linear spaces and consider $f : X \to Y$. If f is Fréchet differentiable at x, then f is Gâteaux differentiable at x and the two derivatives coincide.*

Proof. If the Fréchet derivative $f'(x)$ exists, then by replacing Δx with $t\Delta x$ in (4.1) we have

$$\lim_{t\to 0} \left\| \frac{f(x+t\Delta x) - f(x)}{t} - f'(x)(\Delta x) \right\| = 0.$$

Hence $f'(x)$ is also the Gâteaux derivative of f at x.

Corollary 4.5. *The Fréchet derivative, when it exists, is unique.*

Proof. The Gâteaux derivative defined as a limit must be unique, hence the Fréchet derivative must be unique.

Proposition 4.6. *If $f : X \to Y$ is Fréchet differentiable at x, then f is continuous at x.*

Proof. Using the left-hand side of the triangle inequality on (4.2) we obtain

$$\|f(x+\triangle x) - f(x)\| \le (\epsilon + \|f'(x)\|)\|\triangle x\|.$$

The proposition follows.

Example 4.7. *Consider a linear operator $L : X \to Y$ where X and Y are vector spaces. For $x, \eta \in X$ we have*

$$\frac{L(x+t\eta) - L(x)}{t} = L(\eta).$$

Hence it is clear that for a linear operator L the directional derivative at each point of X exists and coincides with L.[1] Moreover, if X and Y are normed linear spaces and L is bounded then $L'(x) = L$ is both a Gâteaux and a Fréchet derivative.

Example 4.8. *Let X be the infinite-dimensional normed linear space defined in Example C.5.1. Moreover, let δ be the unbounded linear functional defined on X given in this example. Since we saw above that $\delta'(x) = \delta$, we see that for all x, δ has a directional derivative which is never a Gâteaux or a Fréchet derivative, i.e., it is not bounded.*

Example 2.9 describes a situation where $f'(0)(\eta)$ is a Gâteaux derivative. But it is not a Fréchet derivative since f is not continuous at $x = 0$. For the sake of completeness, we should identify a situation where the function is continuous and we have a Gâteaux derivative at a point which is not a Fréchet derivative. Such a function is given by (6.4).

[1] There is a slight technicality here, for if X is not a topological linear space, then the limiting process in the definition of the directional derivative is not defined, even thought it is redundant in this case. Hence, formally we accept the statement.

5. Conditions Implying the Existence of the Fréchet Derivative. In this section, we investigate conditions implying that a directional variation is actually a Fréchet derivative and others implying the existence of the Gâteaux or Fréchet derivative. Our first condition is fundamental.

Proposition 5.1. *Consider $f : X \to Y$ where X and Y are normed linear spaces. Suppose that for some $x \in X$, $f'(x)$, is a Gâteaux derivative. Then the following two statement are equivalent:*

(i) $f'(x)$ is a Fréchet derivative.

(ii) The convergence in the directional variation defining relation (2.2) is uniform with respect to all η such that $\|\eta\| = 1$.

Proof. Guided by (4.2) we consider the statement

$$\frac{\|f(x+\Delta x) - f(x) - f'(x)(\Delta x)\|}{\|\Delta x\|} \leq \epsilon \quad \text{whenever} \quad \|\Delta x\| \leq \delta. \tag{5.1}$$

Consider the change of variables $\Delta x = t\eta$ in (5.1) to obtain the statement

$$\left\| \frac{f(x+t\eta) - f(x)}{t} - f'(x)(\eta) \right\| \leq \epsilon \quad \text{whenever} \quad \eta \in X, \|\eta\| = 1 \tag{5.2}$$
$$\text{and} \quad 0 < |t| \leq \delta.$$

Our steps can be reversed so (5.1) and (5.2) are equivalent statements, and correspond to *(i)* and *(ii)* of the proposition.

There is value in identifying conditions which imply that a directional variation is actually a Gâteaux derivative as we assumed in the previous proposition. The following proposition gives such conditions.

Proposition 5.2. *Let X and Y be normed linear spaces. Assume $f : X \to Y$ has a directional variation in X. Assume further that*

(i) for fixed x, $f'(x)(\eta)$ is continuous in η at $\eta = 0$, and

(ii) for fixed η, $f'(x)(\eta)$ is continuous in x for all $x \in X$.

Then $f'(x) \in L[X,Y]$, i.e., $f'(x)$ is a bounded linear operator.

Proof. By Proposition **??** $f'(x)$ is a homogeneous operator; hence $f'(x)(0) = 0$. By *(i)* above there exists $r > 0$ such that $\|f'(x)(\eta)\| \leq 1$ whenever $\|\eta\| \leq r$. It follows that

$$\|f'(x)(\eta)\| = \left\| \frac{\|\eta\|}{r} f'(x)\left(\frac{r\eta}{\|\eta\|}\right) \right\| \leq \frac{1}{r}\|\eta\|;$$

hence $f'(x)$ will be a continuous linear operator once we show it is additive. Consider $\eta_1, \eta_2 \in X$. Given $\epsilon > 0$ from Definition 2.1 there exists $\tau > 0$ such that

$$\begin{aligned}
&\|f'(x)(\eta_1+\eta_2) - f'(x)(\eta_1) - f'(x)(\eta_2) \\
&- \left[\frac{f(x+t\eta_1+t\eta_2) - f(x)}{t}\right] - \left[\frac{f(x+t\eta_1) - f(x)}{t}\right] - \left[\frac{f(x+t\eta_2) - f(x)}{t}\right]\Bigg\| \\
&\leq 3\epsilon,
\end{aligned}$$

whenever $|t| \leq \tau$. Hence

$$\begin{aligned}
&\|f'(x)(\eta_1+\eta_2) - f'(x)(\eta_1) - f'(x)(\eta_2)\| \\
&\leq \frac{1}{|t|}\|f(t+t\eta_1+t\eta_2) - f(x+t\eta_1) - f(x+t\eta_2) - f(x)\| + 3\epsilon.
\end{aligned}$$

By *(i)* of Proposition 2.12 and Corollary C.7.3 of the Hahn-Banach theorem there exists $\delta \in Y^*$ such that $\|\delta\| = 1$ and

$$\begin{aligned}
&\|f(x+t\eta_1+t\eta_2)-f(x+t\eta_1)-f(x+t\eta_2)-f(x)\| \\
&=\delta(f(x+t\eta_1+t\eta_2)-f(x+t\eta_1))-\delta(f(x+t\eta_2)-f(x)) \\
&=t\delta(f'(x+t\eta_1+\theta_1 t\eta_2)(\eta_2))-t\delta(f'(x+\theta_2 t\eta_2)(\eta_2)) \\
&=t\delta[f'(x+t\eta_1+\theta_1 t\eta_2)(\eta_2)-f'(x)(\eta_2) \\
&\qquad +f'(x)(\eta_2)-f'(x+\theta_2 t\eta_2)(\eta_2)] \\
&\quad \le |t|\ \|f'(x+t\eta_1+\theta_1 t\eta_2)(\eta_2)-f'(x)(\eta_2)\| \\
&\quad + |t|\ \|f'(x)(\eta_2)-f'(x+\theta_2 t\eta_2)\|, \quad 0<\theta_1,\theta_2<1, \\
&\quad \le 2\ |t|\ \epsilon \quad \text{if } t \text{ is sufficiently small.}
\end{aligned}$$

It follows that

$$\|f'(x)(\eta_1+\eta_2)-f'(x)(\eta_1)-f'(x)\eta_2\| \le 5\epsilon$$

since $\epsilon > 0$ was arbitrary $f'(x)$ is additive. This proves the proposition.

We next establish the well-known result that a continuous Gâteaux derivative is a Fréchet derivative.

Proposition 5.3. *Consider $f : X \to Y$ where X and Y are normed linear spaces. Suppose that f is Gâteaux differentiable in an open set $D \subset X$. If f' is continuous at $x \in D$, then $f'(x)$ is a Fréchet derivative.*

Proof. Consider $x \in D$ a continuity point of f'. Since D is open there exists $\hat{r} > 0$ satisfying for $\eta \in X$ and $\|\eta\| < \hat{r}$, $x + t\eta \in D$ for $t \in (0, 1)$. Choose such an η. Then by part *(i)* of Proposition 2.12 for any $\delta \in Y^*$

$$\delta[f(x+\eta)-f(x)-f'(x)(\eta)] = \delta[f'(x+\theta\eta)(\eta)-f'(x)(\eta)], \quad \text{for some} \quad \theta \in (0,1).$$

By Corollary C.7.3 of the Hahn-Banach theorem we can choose δ so that

$$\|f(x+\eta)-f(x)-f'(x)(\eta)\| \le \|f'(x+\theta\eta)-f'(x)\|\|\eta\|.$$

Since f' is continuous at x, given $\epsilon > 0$ there exists $r, 0 < r < \hat{r}$, such that

$$\|f(x+\eta)-f(x)-f'(x)(\eta)\| \le \epsilon\|\eta\|$$

whenever $\|\eta\| \le r$. This proves the proposition.

Consider $f : I\!R^n \to I\!R$. Example 2.9 shows that the existence of the partial derivatives does not imply the existence of the directional variation, let alone the Gâteaux or Fréchet derivatives. However, it is a well-known fact that the existence of continuous partial derivatives does imply the existence of the Fréchet derivative. Proofs of this result can be found readily in advanced calculus and introductory analysis texts. We found several such proofs, these proofs are reasonably challenging and not at all straightforward. For the sake of completeness of the current appendix we include an elegant and comparatively short proof that we found in the book by Fleming [1].

Proposition 5.4. *Consider $f : D \subset I\!R^n \to I\!R$ where D is an open set. Then the following are equivalent.*

(i) f has continuous first-order partial derivatives in D.

(ii) f is continuously Fréchet differentiable in D.

Proof.

[$(i) \Rightarrow (ii)$] We proceed by induction on the dimension n. For $n = 1$ the result is clear. Assume that the result is true in dimension $n - 1$. Our task is to establish the result in dimension n. Hats will be used to denote $(n-1)$-vectors. Hence $\hat{x} = (x_1, \ldots, x_{n-1})^T$. Let $x^0 = (x^0, \ldots, x^0_n)^T$ be a point in D and choose $\delta_0 > 0$ so that $B(x^0, \delta_0) \subset D$. Write

$$\phi(\hat{x}) = f(x_1, \ldots, x_{n-1}, x^0_n) = f(\hat{x}, x^n_0) \quad \text{for} \quad \hat{x} \quad \text{such that} \quad (\hat{x}, x^0_n) \in D.$$

Clearly,

$$\partial_i\phi(\hat{x}) = \partial_i f(\hat{x}, x^0_n), \quad i = 1, \ldots, n-1 \tag{5.3}$$

where ∂_i denotes the partial derivative with respect to the ith variable. Since $\partial_i f$ is continuous $\partial_i \phi$ is continuous. By the induction hypothesis ϕ is Fréchet differentiable at $\hat{x}^0$. Therefore, given $\epsilon > 0$ there exists δ_1, satisfying $0 < \delta_1 < \delta_0$ so that

$$|\phi(\hat{x}^0 + \hat{h}) - \phi(\hat{x}^0) - \langle \nabla\phi(\hat{x}^0), \hat{h}\rangle| < \epsilon\|\hat{h}\| \quad \text{whenever} \quad \|\hat{h}\| < \delta_1. \tag{5.4}$$

Since $\partial_n f$ is continuous there exists δ_2 satisfying $0 < \delta_2 < \delta_1$ such that

$$|\partial_n f(y) - \partial_n f(x^0)| < \epsilon \quad \text{whenever} \quad \|y - x^0\| < \delta_2. \tag{5.5}$$

Consider $h \in I\!R^n$ such that $\|h\| < \delta_2$. By the ordinary mean-value theorem, see (2.7),

$$\begin{aligned} f(x_0 + h) - \phi(\hat{x}^0 + \hat{h}) &= f(\hat{x} + \hat{h}, x_n^0 + h_n) - f(\hat{x}^0 + \hat{h}, x_n^0) \\ &= \partial_n f(\hat{x}^0 + \hat{h}, x_n^0 + \theta h_n)h_n \end{aligned} \tag{5.6}$$

for some $\theta \in (0, 1)$. Setting

$$y = (\hat{x}^0 + \hat{h}, x_n^0 + \theta h_n) \tag{5.7}$$

gives

$$\|y - x_0\| = \theta|h_n| < \|h\| < \delta_2. \tag{5.8}$$

Since $f(x^0) = \phi(\hat{x}^0)$ we can write

$$f(x^0 + h) - f(x_0) = [f(x^0 + h) - \phi(\hat{x}^0 + \hat{h}) + \phi(\hat{x}^0 + \hat{h}) - \phi(\hat{x}^0). \tag{5.9}$$

Now, $\langle \nabla f(x^0), h\rangle$ is a bounded linear functional in h. Our goal is to show that it is the Fréchet derivative of f at x^0. Towards this end we first use (5.6) and (5.7) to rewrite (5.9) as

$$f(x^0 + h) - f(x^0) = \partial_n f(y)h_n + \phi(\hat{x}^0 + \hat{h}). \tag{5.10}$$

Now we subtract $\nabla f(x^0, h)$ from both sides of (5.10) and recall (5.3) to obtain

$$\begin{aligned} f(x^0 + h) - f(x^0) - \langle \nabla f(x^0), h\rangle &= [\partial_n f(y)h_n - \partial_n f(x^0)h_n] \\ &\quad + [\phi(\hat{x}^0 + \hat{h}) - \phi\hat{x}^0 - \langle \nabla\phi(\hat{x}^0), \hat{h}\rangle]. \end{aligned} \tag{5.11}$$

Observing that $|h_n| \leq \|h\|$ and $\|\hat{h}\| \leq \|h\|$ and calling on (5.4) and (5.5) for bounds we see that (5.11) readily leads to

$$|f(x^0 + h) - f(x^0) - \langle \nabla f(x^0), h\rangle| < \epsilon|h_n| + \epsilon\|\hat{h}\| \leq 2\epsilon\|h\| \tag{5.12}$$

whenever $\|h\| < \delta_2$. This shows that f is Fréchet differentiable at x^0 and

$$f'(x^0)(h) = \langle \nabla f(x^0), h\rangle. \tag{5.13}$$

Since the partials $\partial_i f$ are continuous at x^0, the gradient operator ∇f is continuous at x^0 and it follows that the Fréchet derivative f' is continuous at x^0.

$[(ii) \Rightarrow (i)]$ Consider $x^0 \in D$. Since f' is a Fréchet derivative we have the representation (5.13) (see (2.10)). Moreover, continuous partials implies a continuous gradient and in turn a continuous derivative. This proves the proposition.

Remark 5.5. *The challenge in the above proof was establishing linearity. The proof would have been immediate if we had assumed the existence of a directional derivative.*

Corollary 5.6. *Consider $F : D \subset I\!R^n \to I\!R^m$ where D is an open set. Then the following are equivalent.*

(i) F has continuous first-order partial derivatives in D.

(ii) F is continuously Fréchet differentiable in D.

Proof. Write $F(x) = (f_1(x), \dots, f_m(x))^T$. The proof follows from the proposition by employing the relationship

$$F'(x)(\eta) = \begin{pmatrix} \langle \nabla f_1(x), \eta \rangle \\ \vdots \\ \langle \nabla f_m(x), \eta \rangle \end{pmatrix} = JF(x)\eta$$

where $JF(x)$ represents the Jacobian matrix of F at x. See Example 2.16. This proof activity requires familiarity with the material discussed in §**??** of Appendix C.

6. The Chain Rule. The chain rule is a very powerful and useful tool in analysis. We now investigate conditions which guarantee a chain rule.

Proposition 6.1. *(Chain Rule). Let X be a vector space and let Y and Z be normed linear spaces. Assume:*

(i) $h : X \to Y$ has a forward directional variation, respectively directional variation, directional derivative, at $x \in X$, and

(ii) $g : Y \to Z$ is Fréchet differentiable at $h(x) \in Y$.

Then $f = g \circ h : X \to Z$ has a forward directional variation, respectively directional variation, directional derivative, at $x \in X$ and

$$f'_+(x) = g'(h(x))h'_+(x). \tag{6.1}$$

If X is also a normed linear space and $h'(x)$ is a Gâteaux, respectively Fréchet, derivative, then $f'(x)$ is also a Gâteaux, respectively Fréchet, derivative.

Proof. Given x consider $\eta \in X$. Let $y = h(x)$ and $\triangle y = h(x + t\eta) - h(x)$. Then

$$\begin{aligned} \frac{f(x+t\eta) - f(x)}{t} &= \frac{g(y + \triangle y) - g(y)}{t} \\ &= \frac{g'(y)(\triangle y) + g(y + \triangle y) - g(y) - g'(y)(\triangle y)}{t} \\ &= g'(y)\frac{(h(x+t\eta) - h(x))}{t} \\ &\quad + \frac{g(y + \triangle y) - g(y) - g'(y)(\triangle y)}{\|\triangle y\|} \frac{\|h(x+t\eta) - h(x)\|}{t}. \end{aligned}$$

Hence

$$\begin{aligned} &\left\| g'(h(x))h'(x)(\eta) - \left[\frac{f(x+t\eta) - f(x)}{t}\right] \right\| \\ &\le \|g'(h(x))\| \left\| h'(x)(\eta) - \left[\frac{h(x+t\eta) - h(x)}{t}\right] \right\| \\ &+ \frac{\|g(y + \triangle y) - g(y) - g'(y)(\triangle y)\|}{\|\triangle y\|} \left\| \frac{h(x+t\eta) - h(x)}{t} \right\|. \end{aligned} \tag{6.2}$$

Observe that if h has a forward directional variation at x, then although h may not be continuous at x, we do have that h is continuous at x in each direction (see Proposition 2.14), i.e., $\|\triangle y\| \to 0$ as $t \downarrow 0$. By letting $t \downarrow 0$ in (6.2) we see that f has a forward directional variation at x. Clearly, the same holds for two-sided limits. To see that variation can be replaced by derivative we need only recall that the composition of two linear forms is a linear form.

Suppose that X is also a normed linear space. From (6.1) it follows that $f'(x)$ is a bounded linear form if $g'(h(x))$ and $h'(x)$ are bounded linear forms. Hence, $f'(x)$ is a Gâteaux derivative whenever $h'(x)$ is a Gâteaux derivative. Now suppose that $h'(x)$ is a Fréchet derivative. Our task is to show that $f'(x)$ is Fréchet. We do this by establishing uniform convergence in (6.2) for η such that $\|\eta\| = 1$. From Proposition 4.4 h is continuous at x. Hence given $\epsilon > 0 \; \exists \; \delta > 0$ so that

$$\|\Delta y\| = \|h(x + t\eta) - h(x)\| < \epsilon$$

whenever $\|t\eta\| = |t| < \delta$. So as $t \to 0$, $\Delta y \to 0$ uniformly in η satisfying $\|\eta\| = 1$.

Using the left-hand side of the triangle inequality we can write

$$\left\| \frac{h(x+t\eta) - h(x)}{t} \right\| \leq \|h'(x))\| \|\eta\| + \left\| \frac{h(x+t\eta) - h(x)}{t} - h'(x)(\eta) \right\|. \tag{6.3}$$

Since h is Fréchet differentiable at x the second term on the right-hand side of the inequality in (6.3) goes to zero as $t \to 0$ uniformly for η satisfying $\|\eta\| = 1$. Hence the term on the left-hand side of the inequality (6.3) is bounded for small t uniformly in η satisfying $\|\eta\| = 1$. It now follows that as $t \to 0$ the expression on the left-hand side of the inequality (6.3) goes to zero uniformly in η satisfying $\|\eta\| = 1$. This shows by Proposition 5.1 that $f'(x)$ is a Fréchet derivative.

We now give an example where in Propositions 6.1 the function h is Fréchet differentiable, the function g has a Gâteaux derivative and yet the chain rule (6.1) does not hold. The requirement that g be Fréchet differentiable cannot be weakened.

Example 6.2. *Consider $g : I\!R^2 \to I\!R$ defined by $g(0) = 0$ and for nonzero $x = (x_1, x_2)^T$*

$$g(x) = \frac{x_2(x_1^2 + x_2^2)^{3/2}}{[(x_1^2 + x_2^2)^2 + x_2^2]}. \tag{6.4}$$

Also consider $h : I\!R^2 \to I\!R$ defined by

$$h(x) = (x_1, x_2^2)^T. \tag{6.5}$$

We are interested in the composite function $f : I\!R^2 \to I\!R$ obtained as $f(x) = g(h(x))$. We can write

$$f(x) = \frac{x_2(x_1^2 + x_4^2)^{3/2}}{[(x_1^2 + x_4^2)^2 + x_2^4]} \quad x \neq 0 \tag{6.6}$$

and $f(0) = g(h(0)) = g(0) = 0$. To begin with (see Example 2.16)

$$h'(x)(\eta) = \begin{pmatrix} 1 & 0 \\ 0 & 2x_2 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \tag{6.7}$$

so

$$h'(0)(\eta) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \eta_1 \\ 0 \end{pmatrix}. \tag{6.8}$$

Moreover, $h'(0)$ in (6.8) is clearly Fréchet since $h'(x)$ in (6.7) gives a bounded linear operator which is continuous in x (see Proposition 5.3). Direct substitution shows that

$$g'(0)(\eta) = \lim_{t \to 0} \frac{(g(t\eta) - g(0))}{t} = \lim_{t \to 0} \frac{g(t\eta)}{t} = 0. \tag{6.9}$$

Hence $g'(0)$ is a Gâteaux derivative. It follows that

$$g'(h(0))(h'(0)) = 0. \tag{6.10}$$

Again direct substitution shows that

$$f'(0)(\eta) = \frac{\eta_1^3 \eta_2^2}{\eta_1^4 + \eta_2^4}. \tag{6.11}$$

Since (6.10) and (6.11) do not give the same right-hand side our chain rule fails. It must be that $g'(0)$ is not a Fréchet derivative. We now demonstrate this fact directly. At this juncture there is value in investigating the continuity of g at $x = 0$. If g is not continuous at $x = 0$, we will have demonstrated that $g'(0)$ cannot be a Fréchet derivative. On the other hand, if we show that g is continuous at $x = 0$, then our example of chain rule failure is even more compelling.

We consider showing that $g(x) \to 0$ as $x \to 0$. From the definition of g in (6.4) we write

$$g(x) = \frac{x_2 \delta^3}{\delta^4 + x_2^2} \tag{6.12}$$

where $x_1^2 + x_2^2 = \delta^2$. Since δ is the Euclidean norm of $(x_1, x_2)^T$ a restatement of our task is to show that $g(x) \to 0$ as $\delta \to 0$. We therefore study $\phi : I\!R \to I\!R$ defined by

$$\phi(x) = \frac{x\delta^3}{\delta^4 + x^2} \quad \text{for} \quad x \geq 0. \tag{6.13}$$

Direct calculations show that $\phi(0) = 0$, $\phi(x) > 0$ for $x > 0$, $\phi'(x) > 0$ for $x < \delta^2$, $\phi'(x) = 0$ for $x = \delta^2$, and $\phi'(x) < 0$ for $x > \delta^2$. It follows that ϕ has a unique minimum at $x = \delta^2$ and $\phi(\delta^2) = \frac{\delta}{2}$. This means that

$$|g(x)| \leq \frac{\delta}{2} \quad \forall\ x.$$

Hence, $g(x) \to 0$ as $\delta \to 0$ and g is continuous at $x = 0$.

We encounter this result with slightly mixed emotions. While it adds credibility to the example, i.e., we have an example where the function is continuous and the Gâteaux derivative is not a Fréchet derivative, we now have to do more work to show that $g'(0)$ is not a Fréchet derivative. Our only remaining option is to demonstrate that the convergence in (6.9) is not uniform with respect to η satisfying $\|\eta\| = 1$. Toward this end consider $\eta = (\eta_1, \eta_2)^T$ such that $\|\eta\| = 1$, i.e., $\eta_1^2 + \eta_2^2 = 1$. Then

$$\frac{g(t\eta)}{t} = \frac{t\eta_2}{t^2 + \eta_2^2}. \tag{6.14}$$

If we choose $\eta_2 = t$, then

$$\frac{g(t\eta)}{t} = \frac{1}{2}. \tag{6.15}$$

So for small t we can always find an η with norm one of the form $\eta = (\eta_1, t)$. Hence the convergence (6.9) is not uniform with respect to η satisfying $\|\eta\| = 1$.

The choice $\eta_2 = t$ should follow somewhat directly by reflecting on (6.14); however, this is not the way we found it. Being so pleased with our optimization analysis of the function ϕ in (6.13) we did the same thing with the function given in (6.14). It turns out that it is maximized in η_2 with the choice $\eta_2 = t$. If any choice should demonstrate the lack of uniform convergence, the maximizer should certainly do it. The student should explore our approach.

7. A Schematic Overview and an Example. We begin by presenting our hierarchy of notions of differentiation schematically in Figure 7.1. For this purpose it suffices to work with a functional $f : X \to I\!R$ where X is a vector space. For a given $x \in X$ we will consider arbitrary $\eta \in X$.

We now illustrate these notions and implications with an important example.

Example 7.1. *Let $C^1[a, b]$ be the vector space of all real-valued functions which are continuously differentiable on the interval $[a, b]$. Suppose $f : R^3 \to R$ has continuous partial derivatives with respect to the second and third variables. Consider the functional $J\colon C^1[a, b] \to R$ defined by*

$$J(y) = \int_a^b f(x, y(x), y'(x))\, dx. \tag{7.1}$$

We denote the partial derivatives of f by f_1, f_2, and f_3 respectively. Using the technique described in Proposition 2.10 for $\eta \in C^1[a, b]$, we define

$$\phi(t) = J(y + t\eta) = \int_a^b f(x, y(x) + t\eta(x), y' + t\eta'(x))dx.$$

Then,

$$\begin{aligned}\phi'(t) = \int_a^b &[f_2(x, y(x) + t\eta(x), y'(x) + t\eta'(x))\eta(x) \\ &+ f_3(x, y(x) + t\eta(x), y'(x) + t\eta'(x))\eta'(x)]dx,\end{aligned}$$

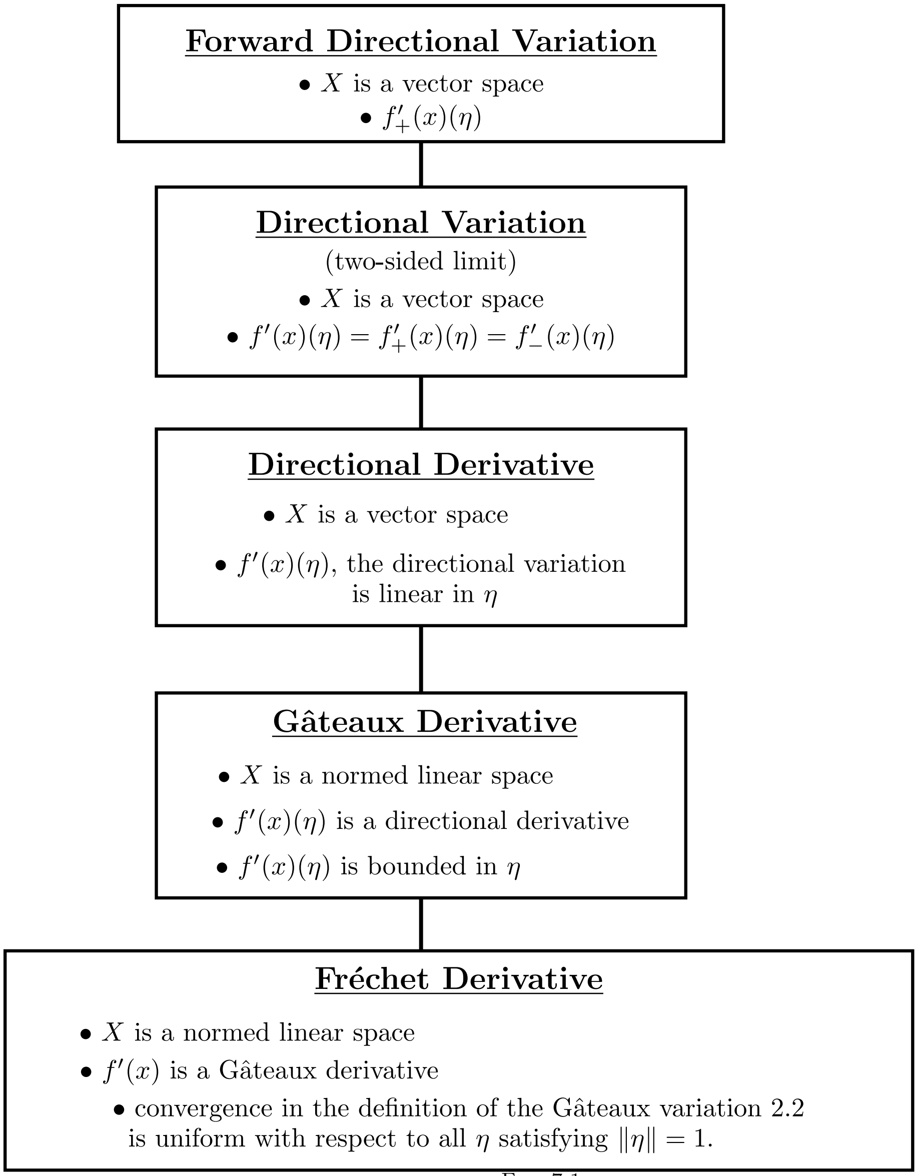

FIG. 7.1.

so that

$$
\begin{aligned}
J'(y)(\eta) &= \phi'(0) \\
&= \int_a^b [f_2(x, y(x), y'(x))\eta(x) + f_3(x, y(x), y'(x))\eta'(x)]dx. \qquad (7.2)
\end{aligned}
$$

It is readily apparent that the directional variation $J'(y)(\eta)$ is actually a directional derivative, i.e., $J'(y)(\eta)$ is linear in η.

Before we can discuss Gâteaux or Fréchet derivatives of J we must introduce a norm on $C^1[a, b]$. Toward this end let

$$\|\eta\| = \max_{a \le x \le b} |\eta(x)| + \max_{a \le x \le b} |\eta'(x)|$$

for each $\eta \in C^1[a, b]$. Recall that we have assumed f has continuous partial derivatives with respect to the second and third variables. Also y and y' are continuous on $[a, b]$. Hence, the composite functions

$f_i(x, y(x), y'(x))$ $i = 2, 3$ are continuous on the compact set $[a, b]$ and therefore attain their maxima on $[a, b]$. It follows that

$$|J'(y)(\eta)| \leq (b-a)\left[\max_{a\leq x\leq b}|f_2(x, y(x), y'(x))| + \max_{a\leq x\leq b}|f_3(x, y(x), y'(x))|\right]\|\eta\|;$$

hence $J'(y)$ is a bounded linear operator, and is therefore a Gâteaux derivative. We now show that J' is the Fréchet derivative. Using the ordinary mean-value theorem for real valued functions of several real variables allows us to write

$$\begin{aligned} J(y+\eta) - J(y) &= \int_a^b [f(x, y(x)+\eta(x), y'(x)+\eta'(x)) - f(x, y(x), y'(x))]dx \\ &= \int_a^b [f_2(x, y(x)+\theta(x)\eta(x), y'(x)+\theta(x)\eta'(x)) \\ &\quad + f_3(x, y(x)+\theta(x)\eta(x), y'(x)+\theta(x)\eta'(x), y'(x))]dx \end{aligned} \tag{7.3}$$

where $0 < \theta(x) < 1$. Using (7.2) and (7.3) we write

$$\begin{aligned} &J(y+\eta) - J(y) - J'(y)(\eta) = \\ &\quad\int_a^b \{f_2(x, y(x)+\theta(x)\eta(x), y'(x)+\theta(x)\eta(x)) - f_2(x, y(x), y'(x))]\eta(x) \\ &\quad + [f_3(x, y(x)+\theta(x)\eta(x), y'(x)+\theta(x)\eta(x)) - f_3(x, y(x), y'(x))\eta'(x)]\}dx. \end{aligned} \tag{7.4}$$

For our given $y \in C^1[a, b]$ choose a δ_0 and consider the set $D \subset I\!R^3$ defined by

$$D = [a, b] \times [-\|y\| - \delta_0, \|y\| + \delta_0] \times [-\|y\| - \delta_0, \|y\| + \delta_0]. \tag{7.5}$$

Clearly D is compact and since f_2 and f_3 are continuous, they are uniformly continuous on D. So, given $\epsilon > 0$, by uniform continuity of f_2 and f_3 on D there exists $\delta > 0$ such that if $\|\eta\| < \delta$ then all the arguments of the function f_2 and f_3 in (7.4) are contained in D and

$$|J(y+\eta) - J(y) - J'(y)(\eta)| \leq 2\epsilon(b-a)\|\eta\|. \tag{7.6}$$

Hence, $J'(y)$ is the Fréchet derivative of J at y according to the definition of the Fréchet derivative, see Definition 4.1 and Remark 4.3.

Alternatively, we can show that $J'(y)$ is the Fréchet derivative by demonstrating that the map J' defined from $C^1[a, b]$ into its topological dual space (see §C.7) is continuous at y and then appealing to Proposition 5.3. Since we are attempting to learn as much as possible from this example we pursue this path and then compare the two approaches. Familiarity with §C.5 will facilitate the presentation.

From the definitions and (7.2) we obtain for $y, h \in C^1[a, b]$

$$\begin{aligned} \|J'(y+h) - J'(y)\| &= \sup_{\|\eta\|=1}|J'(y+h)(\eta) - J'(y)(\eta)| \\ &\leq (b-a)\left[\max_{a\leq x\leq b}|f_2(x, y(x)+h(x), y'(x)+h'(x)) - f_2(x, y(x), y'(x))|\right. \\ &\quad\left. + \max_{a\leq x\leq b}|f_3(x, y(x)+h(x), y'(x)+h'(x)) - f_3'(x, y(x), y'(x))|\right]. \end{aligned} \tag{7.7}$$

Observe that the uniform continuity argument used to obtain (7.4) and (7.6) can be applied to (7.7) to obtain

$$\|J'(y+h) - J'(y)\| \leq 2\epsilon(b-a) \quad \text{whenever} \quad \|h\| \leq \delta.$$

Hence J' is continuous at y and $J'(y)$ is a Fréchet derivative.

We now comment on what we have learned from our two approaches. Clearly, the second approach was more direct. Let's explore why this was so. For the purpose of adding understanding to the role that uniform continuity played in our application we offer the following observation. Consider the function

$$g(y) = f_i(x, y(x), y'(x))$$

(for $i = 1$ or $i = 2$) viewed as a function of $y \in C^1[a, b]$ into $C^0[a, b]$, the continuous functions on $[a, b]$ with the standard max norm. It is exactly the uniform continuity of f_i on the compact set $D \subset I\!R^3$ that enables the continuity of g at y.

Observe that in our first approach we worked with the definition of the Fréchet derivative which entailed working with the expression

$$J(y+\eta) - J(y) - J'(y)(\eta). \tag{7.8}$$

The expression (7.8) involves functions of differing differential orders; hence can be problematic. However, in our second approach we worked with the expression

$$J'(y+h) - J'(y) \tag{7.9}$$

which involves only functions of the same differential order. This should be less problematic and more direct.

Now, observe that it was the use of the ordinary mean-value theorem that allowed us to replace (7.8) with

$$(J'(y+\theta\eta) - J'(y))(\eta). \tag{7.10}$$

Clearly (7.10) has the same flavor as (7.9), So it should not be a surprise that the second half of our first approach and our second approach were essentially the same. However, the second approach gave us more for our money; in the process we also established the continuity of the derivative.

Our summarizing point is that if one expects J' to also be continuous, as we should have here in our example, it is probably wise to consider working with Proposition 5.3 instead of working with the definition and Proposition 5.1 when attempting to establish that a directional derivative is a Fréchet derivative.

8. The Second Directional Variation. Again consider $f : X \to Y$ where X is a vector space and Y is a topological vector space. As we have shown this allows us sufficient structure to define the directional variation. We now turn our attention towards the task of defining a second variation. It is mathematically satisfying to be able to define a notion of a second derivative as the derivative of the derivative. If f is directionally differentiable in X, then we know that

$$f' : X \to [X, Y],$$

where $[X, Y]$ denotes the vector space of linear operators from X into Y. Since in this generality, $[X, Y]$ is not a topological vector space we can not consider the directional variation of the directional derivative as the second directional variation. If we require the additional structure that X and Y are normed linear spaces and f is Gâteaux differentiable, we have

$$f' : X \to L[X, Y]$$

where $L[X, Y]$ is the normed linear space of bounded linear operators from X into Y described in Appendix **??**. We can now consider the directional variation, directional derivative, and Gâteaux derivative. However, requiring such structure is excessively restrictive for our vector space applications. Hence we will take the approach used successfully by workers in the early calculus of variations. Moreover, for our purposes we do not need the generality of Y being a general topological vector space and it suffices to choose $Y = I\!R$. Furthermore, because of *(ii)* of Proposition 2.7 we will not consider the generality of one-sided second variations.

Definition 8.1. *Consider $f : X \to I\!R$, where X is a real vector space and $x \in X$. For $\eta_1, \eta_2 \in X$ by the second directional variation of f at x in the directions η_1 and η_2, we mean*

$$f''(x)(\eta_1, \eta_2) \;=\; \lim_{t \to 0} \frac{f'(x+t\eta_1)(\eta_2) - f'(x)(\eta_2)}{t}, \tag{8.1}$$

whenever these directional derivatives and the limit exist. When the second directional variation of f at x is defined for all $\eta_1, \eta_2 \in X$ we say that f has a second directional variation at x. Moreover, if $f''(x)(\eta_1, \eta_2)$ is bilinear, i.e., linear in η_1 and η_2, we call $f''(x)$ the second directional derivative of f at x and say that f is twice directionally differentiable at x.

We now extend our previous observation to include the second directional variation.

Proposition 8.2. *Consider a functional $f : X \to I\!R$. Given $x, \eta \in X$, let*

$$\phi(t) = f(x + t\eta).$$

Then,

$$\phi'(0) = f'(x)(\eta) \tag{8.2}$$

and

$$\phi''(0) = f''(x)(\eta, \eta) \tag{8.3}$$

Moreover, given $x, \eta_1, \eta_2 \in X$, let

$$\omega(t) = f'(x + t\eta_1)(\eta_2).$$

Then,

$$\omega'(0) = f''(x)(\eta_1, \eta_2) \tag{8.4}$$

Proof. The proof follows directly from the definitions.

Example 8.3. Given $f(x) = x^T A x$, with $A \in I\!R^{n\times n}$ and $x, \eta \in I\!R^n$, find $f''(x)(\eta, \eta)$.
From Example 2.11, we have

$$\phi'(t) \;=\; x^T A\eta + \eta^T Ax + 2t\eta^T A\eta.$$

Thus,

$$\phi''(t) \;=\; 2\,\eta^T A\eta,$$

and

$$\phi''(0) \;=\; 2\,\eta^T A\eta \;=\; f''(x)(\eta, \eta).$$

Moreover, letting

$$\omega(t) = (x + t\eta_1)^T A\eta_2 + \eta_2^T A(x + t\eta_1),$$

we have

$$\omega'(0) = \eta_1^T A\eta_2 + \eta_2^T A\eta_1 = f''(x)(\eta_1, \eta_2).$$

In essentially all our applications we will be working with the second directional variation only in the case where the directions η_1, and η_2 are the same. The terminology second directional variation at x in the direction η will be used to describe this situation.

Some authors only consider the second variation when the directions η_1 and η_2 are the same. They then use (8.4) as the definition of $f''(x)(\eta, \eta)$. While this is rather convenient and would actually suffice in most of our applications, we are concerned that it masks the fact that the second derivative at a point is inherently a function of two independent variables with the expectation that it be a symmetric bilinear form under reasonable conditions. This will become apparent when we study the second Fréchet derivative in later sections. The definition (7.10) retains the flavor of two independent arguments.

The second directional variation is usually linear in each of η_1 and η_2, and it is usually symmetric, in the sense that

$$f''(x)(\eta_1, \eta_2) \;=\; f''(x)(\eta_2, \eta_1).$$

This is the case in Example 8.3 above even when A is not symmetric. However, this is not always the case.

Example 8.4. *Consider the following example. Define $f : I\!R^2 \to I\!R^3$ by $f(0) = 0$ and by*

$$f(x) = \frac{x_1 x_2 (x_1^2 - x_2^2)}{x_1^2 + x_2^2}$$

for $x \neq 0$.

Clearly, $f'(0)(\eta) = 0 \;\forall \eta \in I\!R^2$. Now, for $x \neq 0$, the partial derivatives of f are clearly continuous. Hence f is Fréchet differentiable for $x \neq 0$ by Proposition 5.3. It follows that we can write

$$f'(x)(\eta) = \langle \nabla f(x), \eta \rangle. \tag{8.5}$$

A direct calculation of the two partial derivatives allows us to show that

$$f'(tx)(\eta) = \langle \nabla f(tx), \eta \rangle = t\langle \nabla f(x), \eta \rangle = tf'(x)(\eta). \tag{8.6}$$

Using (8.6) we have that

$$f''(0)(u, v) = \lim_{t\to 0} \left[\frac{f'(tu)(v) - f'(0)(v)}{t} \right] = f'(u)(v).$$

Hence, we can write

$$f''(0)(u, v) = \langle \nabla f(u), v \rangle. \tag{8.7}$$

Again, appealing to the calculated partial derivatives we see that $f''(0)$ as given by (8.7) is not linear in u and is not symmetric. We conclude that $f''(0)$ as a second directional variation is neither symmetric nor a second directional derivative (bilinear). However, as we shall demonstrate in §10 the second Fréchet derivative at a point is always a symmetric bilinear form.

In the following section of this appendix we will define the second Fréchet derivative. A certain amount of satisfaction will be gained from the fact that, unlike the second directional variation (derivative), it is defined as the derivative of the derivative. However, in order to make our presentation complete, we first define the second Gâteaux derivative. It too will be defined as the derivative of the derivative, and perhaps represents the most general situation in which this can be done. We then prove a proposition relating the second Gâteaux derivative to Fréchet differentiability. After several examples concerning Gâteaux differentiability in $\mathbb{R}^n$, we close the section with a proposition giving several basic results in $\mathbb{R}^n$.

Definition 8.5. *Let X and Y be normed linear spaces. Consider $f : X \to Y$. Suppose that f is Gâteaux differentiable in an open set $D \subset X$. Then by the second Gâteaux derivative of f we mean $f'' : D \subset X \to L[X, L[X, Y]]$, the Gâteaux derivative of the Gâteaux derivative $f' : D \subset X \to L[X, Y]$.*

As we show in detail in the next section, $L[X, L[X, Y]]$ is isometrically isomorphic to $[X^2, Y]$ the bounded bilinear forms defined on X that map into Y. Hence the second Gâteaux derivative can be viewed naturally as a bounded bilinear form.

The following proposition is essentially Proposition 5.2 restated to accommodate the second Gâteaux derivative.

Proposition 8.6. *Let X and Y be normed linear spaces. Suppose that $f : X \to Y$ has a first and second Gâteaux derivative in an open set $D \subset X$ and f'' is continuous at $x \in D$. Then both $f'(x)$ and $f''(x)$ are Fréchet derivatives.*

Proof. By continuity $f''(x)$ is a Fréchet derivative (Proposition 5.1). Hence f' is continuous at x (Proposition 5.2) and is therefore the Fréchet derivative at x (Proposition 5.1).

Some concern immediately arises. According to our definition given in the next section, the second Fréchet derivative is the Fréchet derivative of the Fréchet derivative. So it seems as if we have identified a slightly more general situation here where the second Fréchet derivative could be defined as the Fréchet derivative of the Gâteaux derivative $f' : X \to X^*$. However, this situation is not more general. For if f'' is the Fréchet derivative of f', then by Proposition 4.4 f' is continuous and, in turn by Proposition 5.2 f' is a Fréchet derivative. So, a second Gâteaux derivative which is a Fréchet derivative is a second Fréchet derivative according to the definition that will be given.

Recall that if $f : \mathbb{R}^n \to \mathbb{R}$, then the notions of directional derivative and Gâteaux derivative coincide since in theis setting all linear operators are bounded, see §**??**.

Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose that f is Gâteaux differentiable. Then we showed in the Example 2.15 that the gradient vector is the representer of the Gâteaux derivative, i.e., for $x, \eta \in \mathbb{R}^n$

$$f'(x)(\eta) = \langle \nabla f(x), \eta \rangle.$$

Now suppose that f is twice Gâteaux differentiable. The first thing that we observe, via Example 2.16, is that the Hessian matrix $\nabla^2 f(x)$ is the Jacobian matrix of the gradient vector; specifically

$$J\nabla f(x) = \begin{bmatrix} \partial_1\partial_1 f(x) & \dots & \partial_n\partial_1 f(x) \\ & \dots & \\ \partial_1\partial_n f(x) & \dots & \partial_n\partial_n(x) \end{bmatrix}. \tag{8.8}$$

Next, observe that

$$\begin{aligned} f''(x)(\eta_1, \eta_2) &= \lim_{t\to 0} \frac{f'(x + t\eta_1)(\eta_2) - f'(x)(\eta_2)}{t} \\ &= \lim_{t\to 0} \frac{\langle \nabla f(x + t\eta_1), \eta_2\rangle - \langle \nabla f(x), \eta_2\rangle}{t} \\ &= \langle \lim_{t\to 0} \frac{\nabla f(x + t\eta_1) - \nabla f(x)}{t}, \eta_2 \rangle \\ &= \langle \nabla^2 f(x)\eta_1, \eta_2\rangle \\ &= \eta_2^T \nabla^2 f(x)\eta_1. \end{aligned}$$

Following in this direction consider twice Gâteaux differentiable $f : \mathbb{R}^n \to \mathbb{R}^m$. Turning to the component functions we write $f(x) = (f_1(x) \dots, f_m(x))^T$. Then,

$$f'(x)(\eta) = Jf(x)\eta = (\langle \nabla f_1(x), \eta\rangle, \dots, \langle \nabla f_m(x), \eta\rangle)^T,$$

and

$$\begin{aligned} f''(x)(\eta_1, \eta_2) &= (\langle \nabla^2 f_1(x)\eta_1, \eta_2\rangle, \dots, \langle \nabla^2 f_m(x)\eta_1, \eta_2\rangle)^T \\ &= (\eta_1^T \nabla^2 f_1(x)\eta_2, \dots, \eta_1^T \nabla^2 f_m(x)\eta_2)^T. \end{aligned} \tag{8.9}$$

This discussion leads naturally to the following proposition.

Proposition 8.7. *Consider $f : D \subset \mathbb{R}^n \to \mathbb{R}^m$ where D is an open set in $\mathbb{R}^n$. Suppose that f is twice Gâteaux differentiable in D. Then for $x \in D$*

- *(i) $f''(x)$ is symmetric if and only if the Hessian matrices of the component functions at x, $\nabla^2 f_1(x), \dots, \nabla^2 f_m(x)$, are symmetric.*

(ii) f'' *is continuous at* x *if and only if all second-order partial derivatives of the component functions* $f_1, \ldots, f_m$ *are continuous at* x.

(iii) If f'' *is continuous at* x*, then* $f'(x)$ *and* $f''(x)$ *are Fréchet derivatives and* $f''(x)$ *is symmetric.*

Proof. The proof of *(i)* follows directly from (8.9) since second-order partial derivatives are first-order partial derivatives of the first-order partial derivatives. Part *(ii)* follows from Corollary 5.6. Part *(iii)* follows Proposition 5.3, Proposition 4.6, Proposition 5.3 again, and the yet to be proved Proposition 9.4, all in that order.

Hence the Hessian matrix is the representer, via the inner product, of the bilinear form $f''(x)$. So, the gradient vector is the representer of the first derivative and the Hessian matrix is the representer of the second derivative. This is satisfying.

At this point we pause to collect these observations in proposition form, so they can be easily referenced.

Proposition 8.8. *Consider* $f : I\!R^n \to I\!R$.

(i) If f *is directionally differentiable at* x*, then*

$$f'(x) = \langle \nabla f(x), \eta \rangle \quad \forall \eta \in I\!R^n$$

where $\nabla f(x)$ *is the gradient vector of* f *at* x *defined in* (1.3).

(ii) If f *is directionally differentiable in a neighborhood of* x *and twice directionally differentiable at* x*, then*

$$f''(x)(\eta_1, \eta_2) = \langle \nabla^2 f(x)\eta_1, \eta_2 \rangle \quad \forall \eta_1, \eta_2 \in X$$

where $\nabla^2 f(x)$ *is the Hessian matrix of* f *at* x*, described in* (8.8).

Remark 8.9. *It is worth noting that the Hessian matrix can be viewed as the Jacobian matrix of the gradient vector function. Moreover, in this context our directional derivatives are actually Gâteaux derivatives, since we are working in* $I\!R^n$*. Finally, in this generality we can not guarantee that the Hessian matrix is symmetric. This would follow if* $f''(x)$ *was actually a Fréchet derivative, see Proposition 8.10.*

The following proposition plays a strong role in Chapter **??** where we develop our fundamental principles for second-order necessity and sufficiency. Our presentation follows that of Ortega and Rheinboldt [].

Proposition 8.10. *Assume that* $F : D \subset X \to Y$*, where* X *and* Y *are normed linear spaces and* D *is an open set in* X*, is Gâteaux differentiable in* D *and has a second Gâteaux derivative at* $x \in D$.
Then,

(i) $\lim_{t\to 0} \frac{1}{t^2}\left[F(x+th) - F(x) - F'(x)(th) - \frac{1}{2}F''(x)(th, th)\right] = 0$

for an $h \in X$.
Moreover, if $f''(x)$ *is a Fréchet derivative, then*

(ii) $\lim_{h\to 0} \left(\frac{1}{\|h\|^2}\right)\left[F(x+h) - F(x) - F'(x)(h) - \frac{1}{2}F''(x)(h, h)\right]$.

Proof. For given $h \in X$ and t sufficiency small we have that $x + th \in D$ and

$$G(t) = F(x+th) - F(x) - F'(x)(th) - \frac{1}{2}F''(x)(th, th)$$

is well-defined. Clearly,

$$G'(t) = F'(x+th)(h) - F'(x)(h) - tF''(x)(h, h).$$

From the definition of $f''(x)(h, h)$ we have that given $\varepsilon > 0$ $\exists$ $\delta > 0$ so that

$$\|G'(t)\| \le \varepsilon|t|$$

when $|t| < \delta$. Now, from (iii) of Proposition 2.12

$$\|G(t)\| = \|G(t) - G(0)\| \le \sup_{0\le\theta\le 1} \|G'(\theta t)\||t| \le \varepsilon t^2$$

whenever $|t| < \delta$. Since $\varepsilon > 0$ was arbitrary we have established (i). We therefore turn our attention to (ii) by letting

$$R(h) = F(x+h) - F(x) - F'(x)(h) - \frac{1}{2}F''(x)(h, h).$$

Notice that R is well-defined for h sufficiently small and is Gâteaux differentiable in a neighborhood of $h = 0$, since F is defined in a neighborhood of x. It is actually Fréchet differentiable in this neighborhood, but we only need Gâteaux differentiability. Since $F''(x)$ is a Fréchet derivative, by definition, given $\varepsilon > 0$ $\exists$ $\delta > 0$ so that

$$\|R'(h)\| = \|F'(x+h) - F'(x) - F''(x)(h, \cdot) \le \varepsilon\|h\|$$

provided $\|h\| < \delta$.
As before, (iii) of Proposition 2.12 gives

$$\|R(h)\| = \|R(h) - R(0)\| \le \sup_{0\le\theta\le 1} \|R'(\theta h)\|\|h\| \le \varepsilon\|h\|^2$$

provided $\|h\| < \delta$. Since ε was arbitrary we have established (ii) and the proposition.

9. Higher Order Fréchet Derivatives. Let X and Y be normed linear spaces. Suppose $f : X \to Y$ is Fréchet differentiable in D an open subset of X. Let $L_1[X,Y]$ denote $L[X,Y]$ the normed linear space of bounded linear operators from X into Y. Recall that $f' : D \subset X \to L_1[X,Y]$. Consider f'', the Fréchet derivative of the Fréchet derivative of f. Clearly

$$f'' : D \subset X \to L_1\left[X, L_1[X,Y]\right]. \tag{9.1}$$

In general let $L_n[X,Y]$ denote $L_1[X, L_{n-1}[X,Y]], n = 2,3,\cdots$. Then $f^{(n)}, n = 2,3,\ldots$, the n-th Fréchet derivative of f is by definition the Fréchet derivative of $f^{(n-1)}$, the $(n-1)$-st Fréchet derivative of f. Clearly

$$f^{(n)} : D \subset X \to L_n[X,Y].$$

It is not immediately obvious how to interpret the elements of $L_n[X,Y]$. The following interpretation is very helpful. Recall that the Cartesian product of two sets U and V is by definition $U \times V = \{(u,v) : u \in U, v \in V\}$. Also by U^1 we mean U and by U^n we mean $U \times U^{n-1}, n = 2,3,\cdots$. Clearly U^n is a vector space in the obvious manner whenever U is a vector space.

An operator $K : X^n \to Y$ is said to be an n-linear operator from X into Y if it is linear in each of the n variables, i.e., for real α and β

$$\begin{aligned} K(x_1,\cdots,\alpha x_i' + \beta x_i'',\cdots,x_n) &= \alpha K(x_1,\cdots,x_i',\cdots,x_n) \\ &\quad + \ \beta K(x_1,\cdots,x_i'',\cdots,x_n), \quad i = 1,\cdots,n. \end{aligned}$$

The n-linear operator K is said to be bounded if there exists $M > 0$ such that

$$\|K(x_1,\cdots,x_n)\| \ \le M\|x_1\|\,\|x_2\|\ \cdots\|x_n\|, \text{ for all } (x_1,\cdots,x_n) \in X^n. \tag{9.2}$$

The vector space of all bounded n-linear operators from X into Y becomes a normed linear space if we define $\|K\|$ to be the infimum of all M satisfying (9.2). This normed linear space we denote by $[X^n, Y]$.

Clearly by a 1-linear operator we mean a linear operator. Also a 2-linear operator is usually called a bilinear operator or form.

The following proposition shows that the spaces $L_n[X,Y]$ and $[X^n,Y]$ are essentially the same except for notation.

Proposition 9.1. *The normed linear spaces $L_n[X,Y]$ and $[X^n,Y]$ are isometrically isomorphic.*

Proof. These two spaces are isomorphic if there exists

$$T_n : L_n[X,Y] \to [X^n,Y]$$

which is linear, one-one, and onto. The isomorphism is an isometry if it is also norm preserving. Clearly T_n^{-1} will have the same properties. Since $L_1[X,Y] = [X^1,Y]$, let T_1 be the identity operator. Assume we have constructed T_{n-1} with the desired properties. Define T_n as follows. For $W \in L_n[X,Y]$ let $T_n(W)$ be the n-linear operator from X into Y defined by

$$T_n(W)(x_1,\cdots,x_n) = T_{n-1}(W(x_1))(x_2,\cdots,x_n)$$

for $(x_1,\cdots,x_n) \in X^n$. Clearly $\|T_n(W)\| \le \|W\|$; hence $T_n(W) \in [X^n,Y]$. Also T_n is linear. If $U \in [X^n,Y]$, then for each $x \in X$ let $W(x) = T_{n-1}^{-1}(U(x,\cdot,\cdots,\cdot))$. It follows that $W : X \to L_{n-1}[X,Y]$ is linear and $\|W\| \le \|U\|$; hence $W \in L_n[X,Y]$ and $T_n(W) = U$.This shows that T_n is onto and norm preserving. Clearly a linear norm preserving operator must be one-one. This proves the proposition.

Remark 9.2. *The Proposition impacts the Fréchet derivative in the following manner. The n-th Fréchet derivative of $f : X \to Y$ at a point can be viewed as a bounded n-linear operator from X^n into Y.*

It is quite satisfying, that without asking for continuity of the n-th Fréchet derivative, $f^n : X \to [X^n,Y]$ we actually have symmetry as is described in the following proposition.

Proposition 9.3. *If $f : X \to Y$ is n times Fréchet differentiable in an open set $D \subset X$, then the n-linear operator $f^n(x)$ is symmetric for each $x \in D$.*

A proof of this symmetry proposition can be found in Chapter 8 of Dieudonné []. There the result is proved by induction on n. The induction process is initiated by first establishing the result for $n = 2$. We will include that proof because it is instructive, ingenious, and elegant. However, the version of the proof presented in Dieudonné is difficult to follow. Hence, we present Ortega and Rheinboldt's [] adaptation of that proof.

Proposition 9.4. *Suppose that $f : X \to Y$ is Fréchet differentiable in an open set $D \subset X$ and has a second Fréchet derivative at $x \in D$. Then the bilinear operator $f''(x)$ is symmetric.*

Proof. The second Fréchet derivative is by definition the Fréchet derivative of the Fréchet derivative. Hence applying (4.2) for second Fréchet derivatives we see that for given $\epsilon > 0\ \exists\ \delta > 0$ so that $y \in D$ and

$$\|f'(y) - f'(x) - f''(x)(y-x,\cdot)\| \le \epsilon\|x-y\| \tag{9.3}$$

whenever $\|x-y\| < \delta$. Consider $u, v \in B(0,\frac{\delta}{2})$. The function $G : [0,1] \to Y$ defined by

$$G(t) = f(x+tu+v) - f(x+tu)$$

is Fréchet differentiable on $[0,1]$. Moreover by the chain rule

$$G'(t) = f'(x+tu+v)(u) - f'(x+tu)(u). \tag{9.4}$$

We can write

$$\begin{aligned} G'(t) - f''(x)(v,u) = {} & [f'(x+tu+v)(u) - f'(x)(u) - f''(x)(tu+v,u)] \\ & - [f'(x+tu)(u) - f'(x)(u) - f''(x)(tu,u)]. \end{aligned} \tag{9.5}$$

Now, observe that since $t \in [0,1]$, $\|u\| \le \frac{\delta}{2}$, and $\|v\| \le \frac{\delta}{2}$. It follows that $\|y - x\| < \delta$ where y denotes either $x+tu+v$ or $x+tu$. Hence (9.3) can be used in (9.5) to obtain

$$\begin{aligned} \|G'(t) - f''(x)(v,u)\| &\le \epsilon\|tu+v\|\|u\| + \epsilon\|tu\|\|u\| \\ &\le 2\epsilon\|u\|\|u+v\|. \end{aligned} \tag{9.6}$$

Therefore,

$$\begin{aligned} \|G'(t) - G'(0)\| &\le \|G'(t) - f''(x)(v,u)\| + \|G'(0) - f''(x)v,y\| \\ &\le 4\epsilon\|u\|(\|u\| + \|v\|). \end{aligned} \tag{9.7}$$

An application of the mean-value theorem, *(iv)* of Proposition 2.12, (9.5) and (9.6) leads to

$$\begin{aligned} \|G(1) - G(0) - f''(x)(v,u)\| &\le \|G(1) - G(0) - G'(0)\| + \|G'(0) - f''(x)(v,u)\| \\ &\le \sup_{0 \le t \le 1} \|G'(t) - G'(0)\| + 2\epsilon\|u\|\|u+v\| \\ &\le 6\epsilon\|u\|\|u+v\|. \end{aligned} \tag{9.8}$$

The same argument with u and v interchanged may be applied to

$$\overline{G}(t) = f(x+u+tv) - f(x+tv)$$

to obtain

$$\|\overline{G}(1) - \overline{G}(0) - f''(x)(u,v)\| \le 6\epsilon\|v\|\|u+v\|. \tag{9.9}$$

However, $\overline{G}(1) - \overline{G}(0) = G(1) - G(0)$. So (9.8) and (9.9) give

$$\|f''(x)(u,v) - f''(x)(v,u)\| \le 6\epsilon\|u+v\|^2 \tag{9.10}$$

for $u, v \in B(0, \frac{\delta}{2})$. Now, for arbitrary $u, v \in X$ choose $t > 0$ so that $\|tu\| < \frac{\delta}{2}$ and $\|tv\| < \frac{\delta}{2}$. Then (9.10) gives

$$\|f''(x)(tu,tv) - f''(x)(tv,tu)\| \le 6\epsilon\|tu+tv\|^2. \tag{9.11}$$

Clearly, (9.11) simplifies to

$$t^2\|f''(x)(u,v) - f''(x)(v,u)\| \le 6\epsilon t^2\|u+v\|^2, \tag{9.12}$$

which in turn gives (9.10). Hence (9.10) holds for all $u, v \in X$. Since $\epsilon > 0$ was arbitrary we must have

$$f''(x)(u,v) = f''(x)(v,u).$$

10. Calculating the Second Derivative and an Example. In §7 we illustrated our hierarchy of first-order differential notions. In this section we do a similar activity for our second-order differential notions. For this purpose, as we did in §7, we restrict our attention to functionals.

In all applications the basic first step is the calculation of $f''(x)(\eta_1, \eta_2)$, the second directional variation given in Definition 8.1. Our objectives are to first derive conditions that guarantee that $f''(x)$ is a second Fréchet derivative, then apply these conditions to an important example from the calculus of variations literature. Our first objective requires us to consider the various convergence properties inherent in the expression (8.1) written conveniently as

$$\left| f''(x)(\eta_1,\eta_2) - \frac{f'(x+t\eta_1)(\eta_2) - f'(x)(\eta_2)}{t} \right| \to 0 \quad \text{as} \quad t \to 0. \tag{10.1}$$

We first observe that the convergence described in (10.1) is not convergence in an operator space, but what we might call pointwise convergence in $I\!R$. This pointwise approach is employed so we can define the second variation of a functional in the full generality of a vector space, i.e., no norms needed. As such our second directional variation at x is not defined as a differential notion of a differential notion, e.g., derivative of a derivative. However, such structure is necessary in order to utilize the tools, theory, and definitions presented in our previous sections. Hence we cannot move forward in our quest for understanding without requiring more structure on the vector space X. Toward this end we require X to be a normed linear space. Recall that the dual space of X, the normed linear space of bounded linear functionals on X, is denoted by X^*, see §C.7. Also recall that from §9 we know that the second Fréchet derivative f'' can be viewed as either

$$f'' : X \to [X^2, I\!R]$$

or

$$f'' : X \to L[X, X^*].$$

There is value in keeping both interpretations in mind.

Consider $x \in X$. Suppose it has been demonstrated that f is Gâteaux differentiable in D, an open neighborhood of x, i.e., $f' : D \subset X \to X^*$, and also that for $\eta_1 \in X$, $f''(x)(\eta_1, \cdot) \in X^*$. If the convergence described in (10.1) is uniform with respect to all $\eta_2 \in X$ satisfying $\|\eta_2\| = 1$, then $f''(x)(\eta_1, \cdot)$ is the directional variation at x in the direction η_1 of the Gâteaux derivative f'. To see this write (10.1) as given $\epsilon > 0 \; \exists \; \delta > 0$ such that

$$\left| f''(x)(\eta_1, \eta_2) - \frac{f'(x + t\eta_1)(\eta_2) - f'(x)(\eta_2)}{t} \right| \le \epsilon \quad \text{whenever} \quad |t| < \delta \quad \text{and} \quad \|\eta_2\| = 1. \tag{10.2}$$

Now take supremum of the left-hand side of (10.2) over all $\eta_2 \in X$ satisfying $\|\eta_2\| = 1$ to obtain

$$\left\| f''(x)(\eta_1, \cdot) - \frac{f'(x + t\eta_1) - f'(x)}{t} \right\| \le \epsilon \quad \text{whenever} \quad |t| < \delta. \tag{10.3}$$

Since $\epsilon > 0$ was arbitrary (10.3) implies convergence in operator norm, and $f''(x)(\eta_1, \cdot)$ is the directional variation of the Gâteaux derivative. If in addition $f''(x)(\eta_1, \cdot)$ is linear in η_1, then $f''(x)(\cdot, \cdot)$ is bilinear and is not only the second directional derivative of f at x in the sense of Definition 4.1, but is also the directional derivative of the Gâteaux derivative. Continuing on, suppose that in addition $f''(x)(\cdot, \cdot)$ is a bounded bilinear form. Then it is the Gâteaux derivative of the Gâteaux derivative. If the convergence described by (10.1) is uniform with respect to both η_1 and η_2 satisfying $\|\eta_1\| = \|\eta_2\| = 1$, then $f''(x)$ is the Fréchet derivative of the Gâteaux derivative $f' : X \to X^*$. This follows from writing the counterpart of (10.2) and (10.3) for the case $\|\eta_1\| = \|\eta_2\| = 1$, and then appealing to Proposition 6.1. Recall that if f'' is the Fréchet derivative of f' the Fâteaux derivative, then by Proposition 5.4 f' is continuous and, in turn by Proposition 5.2 f' is the Fréchet derivative. So, a second Gâteaux derivative which is a Fréchet derivative is a second Fréchet derivative according to our definition.

We now summarize what we have learned. While the schematic form (Figure 7.1 in §7) worked well for our first-order differentiation understanding, we prefer a proposition format to illustrate our second-order differentiation understanding. Our primary objective is to give conditions which allow one to conclude that a second directional variation is actually a second Fréchet derivative. However, there is value in first identifying conditions that imply that our second directional variation is a second Gâteaux derivative in the sense that it is the Gâteaux derivative of the Gâteaux derivative. If this second Gâteaux derivative is continuous, then it will be a second Fréchet derivative. Appreciate the fact that continuity of the second Gâteaux derivative makes it a Fréchet derivative, which guarantees continuity of the first Gâteaux derivative, and this in turn makes it a Fréchet derivative.

Proposition 10.1. *Consider $f : X \to I\!R$ where X is a normed linear space. Suppose that for a given $x \in X$ it has been demonstrated that f is Gâteaux differentiable in D, an open neighborhood of x, and $f''(x)$ is a bounded bilinear form. Then the following statements are equivalent:*

(i)

$$\left| f''(x)(\eta_1, \eta_2) - \frac{f'(x + t\eta_1)(\eta_2) - f'(x)(\eta_2)}{t} \right| \to 0 \quad \text{as} \quad t \to 0. \tag{10.4}$$

uniformly for all η_2 satisfying $\|\eta_2\| = 1$.

(ii)

$$\left\| f''(x)(\eta_1, \cdot) - \frac{f'(x + t\eta_1) - f'(x)}{t} \right\| \to 0 \quad \text{as} \quad t \to 0. \tag{10.5}$$

(iii) $f''(x)$ is the Gâteaux derivative at x of f', the Gâteaux derivative of f.

Moreover, if any one of conditions (i) - (iii) holds and

(iv) $f'' : D \subset X \to [X^2, I\!R]$ is continuous at x

then $f''(x)$ is the second Fréchet derivative of f at x.

Proof. The proof follows from arguments not unlike those previously presented.

A version of Proposition 10.1 that concerns the second Fréchet derivative follows.

Proposition 10.2. *Consider $f : X \to I\!R$ where X is a normed linear space. Suppose that for a given $x \in X$ it has been demonstrated that f is Gâteaux differentiable in D an open neighborhood of x and $f''(x)$ is a bounded bilinear form. Then the following statements are equivalent:*

(i)

$$\left| f''(x)(\eta_1, \eta_2) - \frac{f'(x + t\eta_1)(\eta_2) - f'(x)(\eta_2)}{t} \right| \to 0 \quad as \quad t \to 0. \tag{10.6}$$

uniformly for all η_1, η_2 satisfying $\|\eta_1\| = \|\eta_2\| = 1$.

(ii)

$$\left\| f''(x)(\eta_1, \cdot) - \frac{f'(x + t\eta_1) - f'(x)}{t} \right\| \to 0 \quad as \quad t \to 0. \tag{10.7}$$

uniformly for all $\eta_1 \in X$ satisfying $\|\eta_1\| = 1$.

(iii) $f''(x)$ is the second Fréchet derivative of f' at x.

Proof. Again, the proof follows from arguments not unlike those previously presented.

We make several direct observations. If it is known that f has a second Fréchet derivative at x, then it must be $f''(x)$ as given by the directional variation. Hence, if we calculate $f''(x)$ as the second directional variation and it is not a symmetric bounded bilinear form, then f does not have a second Fréchet derivative at x. The symmetry issue is fascinating and subtle (see the remarkable proof of Proposition 9.4) and on the surface seems to not be reflected in statements *(i)* and *(ii)* of the above proposition; but it is there, well-hidden perhaps.

Suppose that we have calculated $f''(x)$ as the second directional variation of f at x and have observed not only that it is a symmetric bounded bilinear form, but $f'' : X \to [X^2, I\!R]$ is continuous at x. Can we conclude that $f''(x)$ is the second Fréchet derivative of f at x? The somewhat surprising answer is no, as the following example demonstrates.

Example 10.3. *As in Example 4.8 we consider the unbounded linear functional $f : X \to I\!R$ defined in Example C 5.1. Since f is linear*

$$f'(x) = f$$

and

$$f''(x)(\eta_1, \eta_2) = 0 \quad for\ all \quad x, \eta_1, \eta_2 \in X.$$

Clearly, $f''(x)$ is a symmetric bounded bilinear form. Also, $f'' : X \to [X^2, I\!R]$ is continuous, indeed it is the zero functional for all x. However, $f''(x)$ cannot be a second Fréchet derivative, since $f'(x)$ as an unbounded linear operator is neither a Gâteaux derivative nor a Fréchet derivative. It is interesting to observe that the functional in this example satisfies the convergence requirements in conditions *(i)* and *(ii)* of Proposition 10.1. However, the assumption that is violated is that f is not Gâteaux differentiable.

Immediately we ask if we could conclude that $f''(x)$ was a second Fréchet derivative if we knew that in addition f was Gâteaux differentiable in a neighborhood of x. This question is answered affirmatively in the following proposition. In fact, we do not have to postulate the symmetry of $f''(x)$. It follows from the uniform convergence that follows from the continuity of f'' at x.

Proposition 10.4. *Consider $f : X \to I\!R$ where X is a normed linear space. Let x be a point in X and let D be an open neighborhood of x. Suppose that*

(i) f is Gâteaux differentiable in D.

(ii) The second directional variation $f''(y)$ exists as a bounded bilinear form for $y \in D$ and $f'' : D \subset X \to [X^2, I\!R]$ is continuous at x.

Then $f''(x)$ is the second Fréchet derivative of f at x.

Proof. Guided by Proposition 8.2, for $x, \eta_1, \eta_2 \in X$ consider $\omega : I\!R \to I\!R$ defined by

$$\omega(t) = f'(x + t\eta_1)(\eta_2).$$

Since we are assuming the existence of $f''(y)$ for $y \in D$ it follows that ϕ is differentiable for sufficiently small t. Hence, by the mean-value theorem

$$\omega(t) - \omega(0) = \omega'(\theta t)t = f''(x + \theta t\eta_1)(t\eta_1, \eta_2)$$

for some $\theta \in (0, 1)$. Therefore,

$$f'(x + t\eta_1)(\eta_2) - f'(x)(\eta_2) = f''(x + \theta t\eta_1)(t\eta_1, \eta_2). \tag{10.8}$$

By continuity of f'' at x given $\epsilon > 0 \; \exists \; \delta > 0$ such that

$$\|f''(x) - f''(x+h)\| < \epsilon \quad \text{whenever} \quad \|h\| < \delta. \tag{10.9}$$

Calling on (10.8) and (10.9) we have

$$\begin{aligned} \left| f''(x)(\eta_1, \eta_2) - \frac{f'(x+t\eta_1)(\eta_2) - f'(x)(\eta_2)}{t} \right| &= |f''(x)(\eta_1, \eta_2) - f''(x+\theta t\eta_1)(\eta_1, \eta_2)| \\ &\leq \|f''(x) - f''(x+\theta t\eta_1)\| \|\eta_1\| \|\eta_2\| \\ &\leq \epsilon \end{aligned}$$

whenever $|t| < \delta$ and $\|\eta_1\| = \|\eta_2\| = 1$. This demonstrates that condition *(i)* of Proposition 10.2 holds. Hence, $f''(x)$ is the second Fréchet derivative of f at x.

In order to better appreciate Propositions 10.1, 10.2 and 10.4 we present the following example.

Example 10.5. *Consider the normed linear space X and the functional $J : X \to I\!R$ defined in Example 7.1. In this application we assume familiarity with the details including notation and the arguments used in Example 7.1. There we assumed that $f : I\!R^3 \to I\!R$ appearing in the definition of J in (6.15) had continuous partial derivatives with respect to the second and third variables. Here we will need continuous second-order partials with respect to the second and third variables.*

It was demonstrated (see (7.2)) that

$$J'(y)(\eta) = \int_a^b [f_2(x, y, y')\eta + f_3(x, y, y')\eta'] dx. \tag{10.10}$$

For convenience we have suppressed the argument x in y, y, η and η' and will continue to do so in these quantities and analogous quantities. For given $y, \eta_1, \eta_2 \in X$ let

$$\phi(t) = J'(y + t\eta_1)(\eta_2).$$

Then

$$\phi'(0) = J''(y)(\eta_1, \eta_2).$$

So from (10.10) we obtain

$$\begin{aligned} J''(y)(\eta_1, \eta_2) = \int_a^b [& f_{22}(x, y, y')\eta_1\eta_2 + f_{23}(x, y, y')\eta_1'\eta_2 \\ & + f_{32}(x, y, y')\eta_1\eta_2' + f_{33}(x, y, y')\eta_1'\eta_2'] dx. \end{aligned} \tag{10.11}$$

Observe that since we have continuous second-order partials with respect to the second and third variables it follows that $f_{23} = f_{32}$. Hence, $J''(y)(\cdot, \cdot)$ is a symmetric bilinear form. Moreover,

$$|J''(y)(\eta_1, \eta_2| \leq (b-a)\|\eta_1\|\|\eta_2\| \sum_{i,j=2}^{3} \max_{a \leq x \leq b} |f_{ij}(x, y, y')|. \tag{10.12}$$

Arguing as we did in Example 7.1, because of continuity of the second-order partials, we can show that the maxima in (10.12) are finite; hence $J''(y)$ is a bounded symmetric bilinear form. Furthermore, for $y, z, \eta_1, \eta_2 \in X$ we can write

$$|J''(y)(\eta_1, \eta_2) - J''(z)(\eta_1, \eta_2)| \leq (b-a)\|\eta_1\|\|\eta_2\| \sum_{i,j=2}^{3} \max_{a \leq x \leq b} |f_{ij}(x, y, y') - f_{ij}(x, z, z')|.$$

It follows by taking the supremum over η_1 and η_2 satisfying $\|\eta_1\| = \|\eta_2\| = 1$ that

$$\|J''(y) - J''(z)\| \leq (b-a) \sum_{i,j=2}^{3} \max_{a \leq x \leq b} |f_{ij}(x, y, y') - f_{ij}(x, z, z')|. \tag{10.13}$$

Now recall the comments surrounding (7.7). They were made with this application in mind. They say that it follows from continuity that the function

$$f_{ij}(x, y(x), y'(x))$$

viewed as a function of $y \in X = C^1[a, b]$ into $C^0[a, b]$ with the max norm is continuous at y. Hence given $\epsilon > 0 \; \exists \; \delta_{ij} > 0$ such that

$$\max_{a \leq x \leq b} |f_{ij}(x, y, y') - f_{ij}(x, z, z')| < \epsilon$$

whenever $\|y - z\| < \delta_{ij}$. *Letting* $\delta = \min_{ij} \delta_{ij}$; *it follows from* (10.13) *that*

$$\|J''(y) - J''(z)\| \leq 4\epsilon(b-a) \tag{10.14}$$

whenever $\|y - z\| < \delta$. *Hence* $J'' : X \to [X^2, I\!R]$ *is continuous. In Example 7.1 it was demonstrated that* $J' : X \to X^*$ *was a Fréchet derivative. We have just demonstrated that the second directional variation has the property that* $J'' : X \to [X^2, I\!R]$ *and is continuous. Hence by Proposition 10.4 it must be that* J'' *is the second- Fréchet derivative of* J.

If instead we chose to establish *(i)* of Proposition 10.2 directly, instead of turning to the continuity of J'' and Proposition 10.4, the proof would not be much different once we used the mean-value theorem to write

$$J'(y + t\eta_1)(\eta_2) - J'(y)(\eta_2) = J''(y + \theta t\eta_1)(t\eta_1, \eta_2).$$

The previous argument generalizes in the obvious manner to show that if f in (6.15) has continuous partial derivatives of order n with respect to the second and third arguments, then $J^{(n)}$ exists and is continuous.

11. Closing Comment. Directional variations and derivatives guarantee that we have good behavior along lines, for example hemicontinuity as described in Proposition 2.8. However, good behavior along lines is not sufficient for good behavior in general, for example, continuity, see Example 2.13. The implication of general good behavior is the strength of the Fréchet theory. However, it is exactly behavior along lines that is reflected in our optimization necessity theory described and developed in Chapter **??**. For this reason, and others, we feel that optimization texts that require Fréchet differentiability produce theory that is limited in its effectiveness, it requires excessive structure.

REFERENCES

[1] Wendell H. Fleming. *Functions of Several Variables.* Addison-Wesley Publishing Co.,Inc. Reading, Mass.-London, 1965.

[2] J.M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Society for Industrial and Applied Mathematics (SIAM), 2000. Reprint of the 1970 original.