Fréchet & Gâteaux Derivatives[1]and the Chain Rule

by

Francis J. Narcowich

January, 2021

There are two types of derivatives that get used in connection with nonlinear functions (and functionals), the Gâteaux (weak) derivative and the Fréchet (strong) derivative.

Let V and W be Banach spaces, Ω an open set in V, and F a function that maps Ω into W. Fix $\mathbf{y}$ in Ω.

Definition 0.1 (Gâteaux Derivative)**.** *If there is a bounded linear map*[2] $\Delta_{\mathbf{y}}F : V \to W$ *given by*[3]

$$\Delta_{\mathbf{y}}F(\mathbf{v}) := \lim_{\varepsilon\to 0} \frac{F(\mathbf{y}+\varepsilon\mathbf{v}) - F(\mathbf{y})}{\varepsilon}, \text{ for all } v \in V, \tag{0.1}$$

then we say that F is Gâteaux differentiable at $\mathbf{y}$ and that $\Delta_y F$ is the Gáteaux derivative of F at $\mathbf{y}$.

It is easy to show that, if the derivative exists, it is unique.

As a simple example of the Gâueax derivative, let $V = W = \mathbb{R}^2$ and define F by

$$F(x,y) := \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix}, \tag{0.2}$$

where f_1 and f_2 are differentiable functions in a neighborhood of $\mathbf{y} = (x_0, y_0)$. Let $\mathbf{v} = (u, v)$. Using the chain rule, we see that

$$\Delta_{\mathbf{y}}F(\mathbf{v}) := \begin{pmatrix} \frac{\partial f_1}{\partial x}\frac{\partial x}{\partial \varepsilon} + \frac{\partial f_1}{\partial y}\frac{\partial y}{\partial \varepsilon} \\ \frac{\partial f_2}{\partial x}\frac{\partial x}{\partial \varepsilon} + \frac{\partial f_2}{\partial y}\frac{\partial y}{\partial \varepsilon} \end{pmatrix}\Bigg|_{\varepsilon=0} = \begin{pmatrix} \frac{\partial f_1}{\partial x}u + \frac{\partial f_1}{\partial y}v \\ \frac{\partial f_2}{\partial x}u + \frac{\partial f_2}{\partial y}v \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}$$

The matrix on the far right is called the Jacobian matrix, and it's usually denoted by J or $J_{\mathbf{y}}$. The calculation above works for any system of equations of the form $F(x_1, \ldots, x_n) = [f_1(x_1, \ldots, x_n) \cdots f_m(x_1, \ldots, x_n)]^T$. The Jacobian

[1]The main reference for these notes is Jordan Bell, Fr'echet and Gâteaux Derivatives, `https://individual.utoronto.ca/jordanbell/`.

[2]This is Gâteaux's definition. Other authors use a directional derivative and even allow nonlinear derivatives.

[3]Note that ε is *not* required to be positive.

matrix for the larger system is $m \times n$ and has entries $J_{ij} = \frac{\partial f_i}{\partial x_j}$. The same calculation as above works for the more general case. Thus, $\Delta_{\mathbf{y}} F = J_{\mathbf{y}}$.

Another example, which differs from the one above, is the following: Let $H : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$H(x,y) = \begin{cases} \frac{|x|^3 y}{x^4+y^2} & (x,y) \neq (0,0) \\ 0 & x=0, y=0. \end{cases} \tag{0.3}$$

$H(x,y))$ is continuous at $(0,0)$. At $(x,y) \neq (0,0)$,

$$|H(x,y| = \frac{|x|x^2|y|}{x^4+y^2} \leq \frac{1}{2}|x|\frac{x^4+y^2}{x^4+y^2} = \frac{1}{2}|x|,$$

Hence, as $(x,y) \to (0,0)$, $H(x,y) \to 0$, so H is continuous at $(0,0)$.

Let's compute the Gâteaux derivative. Let $\mathbf{v} \neq (0,0)$ and $\mathbf{y} = (0,0)$, then the difference quotient is

$$\frac{H(\varepsilon(u,v)) - H(0,0)}{\varepsilon} = \frac{|\varepsilon|^3 \varepsilon |u|^3 v}{\varepsilon(\varepsilon^4 u^4 + \varepsilon^2 v^2)} = \frac{|\varepsilon||u|^3 v}{\varepsilon^2 u^4 + v^2}.$$

Letting $\varepsilon \to 0$ results in the Gâteaux derivative being $\Delta_{(0,0)} H(\mathbf{v}) = 0$. It is important to note that $\Delta_{(0,0)} H$ is *linear* in $\mathbf{v}$. If it were not, the Gâteaux would not exist. For example, the function

$$K(x,y) := \begin{cases} \frac{x^2 y}{x^2+y^2} & xy > 0 \\ 0 & xy \leq 0 \end{cases}$$

satisfies

$$\frac{K(\varepsilon(u,v)) - G(0,0)}{\varepsilon} = K(u,v),$$

which is obviously nonlinear in $\mathbf{v}$. Consequently it's Gâteaux doesn't exist.

One last remark. The existence of the Gâteaux derivative doesn't necessarily imply continuity of a function. For instance, let

$$M(x,y) := \begin{cases} \frac{x^4 y}{x^6+y^3} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

A computation nearly identical to the one above shows that $\Delta_{(0,0)} M(\mathbf{v}) = 0$. However, $M(x,x^2) = \frac{1}{2} \neq 0$. Thus, M is discontinuous at $(0,0)$, even though the Gâteaux derivative exists.

The Fréchet derivative is defined in a way that is somewhat different than the Gâteaux derivative. Let V, W, Ω and F be as defined earlier. Again, fix $\mathbf{y} \in \Omega$.

Definition 0.2 (Fréchet Derivative)**.** *If there is a bounded linear map $\Delta_{\mathbf{y}}F : V \to W$ that satisfies*

$$\lim_{\|\mathbf{v}\|\to 0} \frac{\|F(\mathbf{y}+\mathbf{v}) - F(\mathbf{y}) - \Delta_{\mathbf{y}}F(v)\|_W}{\|\mathbf{v}\|_V} = 0, \tag{0.4}$$

then we say that F is Fréchet differentiable at $\mathbf{y}$ and we call $\Delta_{\mathbf{y}}F$ the the Fréchet derivative of F at $\mathbf{y}$.

Like the Gâteaux, one can show that the Fréchet derivative is unique, if it exists. However, unlike the definition of the Gâteaux derivative, where the Gâteaux derivative is given *directly* via the formula in (0.1), the Fréchet derivative is defined *indirectly*, without a formula, as a linear operator satisfying (0.4). This would seem to make it more difficult to compute the Fréchet derivative. However, if the Fréchet derivative does exist, it can always be computed via the Gâteaux derivative:

Proposition 0.3. *If the Fréchet derivative exists, then the Gâteaux derivative also exists, and the two are equal.*

Proof. Suppose that the Fréchet derivative exists at $\mathbf{y}$. Fix $\mathbf{v}$ and note that the linearity of $\Delta_{\mathbf{y}}F$ implies that $\Delta_{\mathbf{y}}F(\varepsilon\mathbf{v}) = \varepsilon\Delta_{\mathbf{y}}F(\mathbf{v})$, so that

$$\left|\frac{F(\mathbf{y}+\varepsilon\mathbf{v}) - F(\mathbf{y})}{\varepsilon} - \Delta_{\mathbf{y}}F(\mathbf{v})\right| = \|v\|\left|\frac{F(\mathbf{y}+\varepsilon\mathbf{v}) - F(\mathbf{y}) - \Delta_{\mathbf{y}}F(\varepsilon\mathbf{v})}{\|\varepsilon\mathbf{v}\|}\right|.$$

By (0.4) and the fact that any approach to zero for $\|v\|$ is allowed in the Fréchet case, the limit on the right is zero. Hence, the limit on the left is also zero. Thus, the Gâteaux derivative exists and is equal to the Fréchet derivative, whenever the latter exists. □

Consider a function $F(x_1, \ldots, x_n) = [f_1(x_1, \ldots, x_n) \cdots f_m(x_1, \ldots, x_n)]^T$. When is F guaranteed to be Fréchet diffentiable? We leave it as an exercise to show that F is Fréchet differentiable at a given point if and only if all of the first partials $\frac{\partial f_j}{\partial k}$ are continuous there.

The Gâteaux derivative may exist, but the Fréchet derivative may not. Earlier, we showed that the Gâteaux derivative for the function defined in

(0.3) was $\Delta_{(0,0)}H(\mathbf{v}) = 0$. If the Fréchet derivative exists, it is equal to the Gâteaux derivative. Thus, by (0.4), we have that

$$\lim_{\|\mathbf{v}\|\to 0} \frac{\|H(u,v) - H(0,0) - \Delta_{(0,0)}H(\mathbf{v})\|}{\|\mathbf{v}\|} = \lim_{\|\mathbf{v}\|\to 0} \frac{|u|^3 v}{(u^2+v^2)^{\frac{1}{2}}(u^4+v^2)} = 0.$$

Choose $v = u^2$. Then, as $u \to 0$ we have

$$\frac{|u|^3 v}{(u^2+v^2)^{\frac{1}{2}}(u^4+v^2)} = \frac{|u|^5}{(1+u^2)^{\frac{1}{2}}|u|(2u^4)} = \frac{1}{2(1+u^2)^{\frac{1}{2}}} \longrightarrow 1/2,$$

which is a contradiction. Consequently, the Fréchet derivative doesn't exist.

There are other important properties of these derivatives, such as the sum rule, product rule and the chain rule. To state and prove these, we note that the Fréchet derivative may also be defined in terms of "little oh" notation: The Fréchet derivative exists at $\mathbf{y}$ if and only if $F(\mathbf{y}+\mathbf{v}) = F(\mathbf{y}) + \Delta_y F(\mathbf{v}) + o(\|\mathbf{v}\|)$ for all $\mathbf{v}$. The same is true for the Gáteaux derivative; it will exist at $\mathbf{y}$ if and only if for every fixed $\mathbf{v}$ we have $F(\mathbf{y}+\varepsilon\mathbf{v}) = F(\mathbf{y}) + \varepsilon\Delta_y F(\mathbf{v}) + o(\varepsilon)$. Our aim is to prove the chain rule.

Theorem 0.4 (Chain Rule)**.** *Suppose that U, V and W are normed spaces, that Ω is an open subset of U, $\widetilde{\Omega}$ is an open subset of V, and that $F : U \to V$, with $F(\Omega) \subseteq \widetilde{\Omega}$. In addition, assume that F is Fréchet differentiable at $\mathbf{x} \in \Omega$ and the $F(\mathbf{x}) = \mathbf{y} \in \widetilde{\Omega}$. If $G : \widetilde{\Omega} \to W$ is Fréchet differentiable at $\mathbf{y}$. Then*

$$\Delta_{\mathbf{x}}(G \circ F) = \Delta_{\mathbf{y}} G \circ \Delta_{\mathbf{x}} F,$$

which is the chain rule, holds.

Proof. Let $\mathbf{v}$ be in Ω. Since F is Fréchet differentiable at $\mathbf{x}$, $F(\mathbf{x}+\mathbf{v}) = \underbrace{F(\mathbf{x})}_{\mathbf{y}} + \Delta_{\mathbf{x}} F(\mathbf{v}) + o(\|\mathbf{v}\|)$. In addition,

$$\begin{aligned} G \circ F(\mathbf{x}+\mathbf{v}) &= G(\mathbf{y} + \Delta_{\mathbf{x}} F(\mathbf{v}) + o(\|\mathbf{v}\|)) \\ &= G(\mathbf{y}) + \Delta_{\mathbf{y}} G(\Delta_{\mathbf{x}} F(\mathbf{v}) + o(\|\mathbf{v}\|)) \\ &= G(\mathbf{y}) + \Delta_{\mathbf{y}} G \circ \Delta_{\mathbf{x}} F(\mathbf{v}) + \Delta_{\mathbf{y}} G(o(\|\mathbf{v}\|)) \end{aligned}$$

Because $\Delta_{\mathbf{y}} G$ is a bounded linear operator, $\|\Delta_{\mathbf{y}} G(o(\|\mathbf{v}\|))\| \le \|\Delta_{\mathbf{y}} G\| \, \|o(\|\mathbf{v}\|)\| = o(\|\mathbf{v}\|)$. From this, we have that

$$G \circ F(\mathbf{x}+\mathbf{v}) = G(\mathbf{y}) + \Delta_{\mathbf{y}} G \circ \Delta_{\mathbf{x}} F(\mathbf{v}) + o(\|\mathbf{v}\|),$$

so $\Delta_{\mathbf{x}}(G \circ F)(\mathbf{v}) = \Delta_{\mathbf{x}} G \circ \Delta_{\mathbf{x}} F(\mathbf{v})$. Thus, $\Delta_{\mathbf{x}}(G \circ F) = \Delta_{\mathbf{y}} G \circ \Delta_{\mathbf{x}} F$ $\square$