Differentiation in Banach spaces

John D. Cook

May 24, 1994

Throughout these notes, X and Y will be Banach spaces. $\mathcal{L}(X,Y)$ will denote the continuous linear operators from X to Y.

The derivative of a path $u : (0,1) \to X$ is given by

$$u'(t) = \lim_{h \to 0} \frac{u(t+h) - u(t)}{h}.$$

Suppose $x, h \in X$ and let U be an open neighborhood of x. Let $F : U \to Y$ and suppose the function $\varphi : \lambda \mapsto F(x + \lambda h)$ is differentiable at 0. Then $\varphi'(0)$ is called the **Gâteaux variation** of F at x in the direction h, denoted $\delta F(x; h)$; the Gâteaux variation is a generalization of the idea of a directional derivative.

If there exists an operator $A \in \mathcal{L}(X,Y)$ such that

$$\delta F(x; h) = Ah$$

then A is called the **Gâteaux** derivative of F at x, denoted $DF(x)$.

If the stronger condition

$$\lim_{\|h\| \to 0} \frac{\|F(x+h) - F(x) - Ah\|}{\|h\|} = 0$$

holds then $DF(x)$ is called the **Fréchet** derivative of F at x. The Fréchet derivative is the closer analog of the derivative of calculus.

If the Gâteaux derivative of F exists and is continuous at x, then F is Fréchet differentiable at x.

Mean Value Theorem: Let $[x_1, x_2]$ denote the line segment joining two points x_1, x_2 in an open set $U \subseteq X$. If F is Gâteaux differentiable in U, then

$$\|F(x_1) - F(x_2)\| \le \|x_1 - x_2\| \sup_{x \in [x_1, x_2]} \|DF(x)\|.$$

Proof outline: Define $\varphi(t) = F((1-t)x_1 + tx_2)$. By Hahn-Banach there exists a $y^* \in Y^*$ of unit norm such that

$$y^*(\varphi(0) - \varphi(1)) = \|\varphi(0) - \varphi(1)\|.$$

Apply the ordinary mean value theorem to $y^*\varphi$.

Suppose $X = X_1 \oplus \cdots \oplus X_m$. Then one can define **partial Fréchet derivatives** $D_j F(x_1 \ldots x_m)$ by analogy with partial derivatives from ordinary calculus. If $h = (h_1, \ldots h_m)$ and all the partial Fréchet derivatives exist and are continuous at x then

$$DF(x)h = \sum_{j=1}^{m} D_j F(x) h_j.$$

Chain Rule: If $F : X \to Y$ and $G : Y \to Z$ are Fréchet differentiable then $G \circ F$ is Fréchet differentiable and

$$D(G \circ F)(x) = DG(F(x))DF(x).$$

Product Rule: Let $X = X_1 \oplus \cdots X_n$ and $Y = Y_1 \oplus \cdots Y_n$. Suppose we are given a product on Y, *i.e.,* a continuous n-linear function from Y to a space Z. Let $F_i : X_i \to Y_i$ be Fréchet differentiable for $1 \leq i \leq n$. Then the map $F : X \to Y$ given by

$$F(x) = P(F_1(x_1), \ldots, F_n(x_n))$$

is Fréchet differentiable and its derivative is given by

$$DF(x)h = \sum_{i=1}^{n} P(F(x_1), \ldots, F_{i-1}(x_{i-1}), DF_i(x_i)h, F_{i+1}(x_{i+1}), \ldots, F_n(x_n)).$$

Define $\mathcal{L}^n(X, Y)$ to be the space of continuous n-linear maps from $X \oplus \cdots \oplus X$ (n copies) to Y. This space is a Banach space under the norm

$$\|F\| = \sup_{\|x_i\| \leq 1,\ i=1\ldots n} \|F(x_1, \ldots, x_n)\|.$$

$\mathcal{L}^n(X, Y)$ is isomorphic to $\mathcal{L}(x, \mathcal{L}^{n-1}(X, Y))$.

If $F : X \to Y$ is n times Fréchet differentiable, then $DF : X \to \mathcal{L}(X, Y)$ and under the above identification, $D^n F : X \to \mathcal{L}^n(X, Y)$. For each $x \in X$, $D^n F(x)$ is a symmetric n-linear map.

We define $C^n(X, Y)$ as might be expected: the set of maps whose first n Fréchet derivatives exist and are continuous.

Inverse Function Theorem: Suppose $F \in C^1(X, Y)$ and that for some $x_0 \in X$, $DF(x_0)$ is invertible. Then there exists a neighborhood of $y_0 \equiv F(x_0)$ on which F is invertible and

$$D(F^{-1})(y_0) = (DF(x_0))^{-1}.$$

In other words, the best linear approximation to the inverse function is the inverse of the best linear approximation to the function. If $X = Y = \mathbb{R}$,

$$\frac{dx}{dy} = \left(\frac{dy}{dx}\right)^{-1}.$$

Implicit Function Theorem: Let $F \in C^1(X \oplus Y, Z)$ and suppose that the partial Fréchet derivative $D_2F(x_0, y_0) \in \mathcal{L}(Y, Z)$ is invertible for some (x_0, y_0). Then there exists a neighborhood U of x_0, a neighborhood V of $F(x_0, y_0)$, and a unique map $G \in C^1(U \times V, Y)$ such that

$$F(x, G(x, z)) = z.$$

In other words, for each $z \in V$, the relation $F(x, y) = z$ implicitly defines y as a C^1 function of x for x's in some neighborhood of x_0 and the dependence on z is also C^1.

The implicit function theorem follows from applying the inverse function theorem to $G(x, y) : (x, y) \mapsto (x, F(x, y))$.

Lagrange Multipliers: Let U be an open set of a Hilbert space H. Let an objective function F_0 and and constraint functions $F_1, \ldots, F_n$ be in $C^1(U, \mathbb{R})$. Define

$$M \equiv \cap_{i=1}^n F_i^{-1}(0).$$

Then if the restriction of F to M has a local maximum at $x_0 \in M$, then there exist constants $\lambda_0 \ldots \lambda_n$, not all zero, such that

$$\sum_{i=0}^{n} \lambda_i DF_i(x_0) = 0.$$

In the case $n = 1$ and $X = \mathbb{R}^m$, an intuitive proof can be given as follows. Consider the manifold $F_1^{-1}(0)$ as stationary and consider $F_0^{-1}(y)$ for y bigger that the maximum of F_0 on $F_1^{-1}(0)$. Gradually decrease y until the two level sets are first tangent. At that point, the gradients of F_0 and F_1 must be parallel.