Basic Differentiable Calculus Review

Jimmie Lawson
Department of Mathematics
Louisiana State University

Spring, 2003

1 Introduction

Basic facts about the multivariable differentiable calculus are needed as background for differentiable geometry and its applications. The purpose of these notes is to recall some of these facts. We state the results in the general context of Banach spaces, although we are specifically concerned with the finite-dimensional setting, specifically that of $\mathbb{R}^n$.

Let U be an open set of $\mathbb{R}^n$. A function $f : U \to \mathbb{R}$ is said to be a C^r *map* for $0 \leq r \leq \infty$ if all partial derivatives up through order r exist for all points of U and are continuous. In the extreme cases C^0 means that f is continuous and C^∞ means that all partials of all orders exists and are continuous on U. A function $f : U \to \mathbb{R}^m$ is a C^r *map* if $f_i := \pi_i f$ is C^r for $i = 1, \ldots, m$, where $\pi_i : \mathbb{R}^m \to \mathbb{R}$ is the i^{th} projection map defined by $\pi_i(x_1, \ldots, x_m) = x_i$. It is a standard result that mixed partials of degree less than or equal to r and of the same type up to interchanges of order are equal for a C^r-function (sometimes called Clairaut's Theorem).

We can consider a category with objects nonempty open subsets of $\mathbb{R}^n$ for various n and morphisms C^r-maps. This is indeed a category, since the composition of C^r maps is again a C^r map.

2 Normed Spaces and Bounded Linear Operators

At the heart of the differential calculus is the notion of a differentiable function. The notion of a differentiable function is that of a function which can be appropriately approximated locally by a linear function. To develop and study this concept in a general setting, one needs both a notion of linearity (a vector space notion) and a notion of nearness or approximation (a metric or topological notion). These two notions are happily wed in the setting of a normed vector space.

Definition 2.1 *A normed space is a vector space E over $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ endowed with a norm $\|\cdot\|\colon E \to \mathbb{R}^+$ satisfying*

(i) $\|v\| = 0$ if and only if $v = 0$.

(ii) $\|\alpha v\| = |\alpha|\|v\|$ for each $v \in E$ and $\alpha \in \mathbb{F}$.

(iii) $\|v + w\| \leq \|v\| + \|w\|$, for all $v, w \in E$.
Then E is a metric space with respect to the metric defined from the norm by $d(x, y) := \|x - y\|$. A normed space for which this metric is a complete metric is called a Banach space

The metric on a normed space provides us with a notion of nearness and limits.

Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space V are *equivalent* if there exist constants $b, c > 0$ such that

$$b\|x\|_1 \leq \|x\|_2 \leq c\|x\|_1$$

for all $x \in V$. *It is a standard fact that all norms on the real vector space $\mathbb{R}^n$ are equivalent, and hence induce the same topology, the usual product topology of* $\mathbb{R}^n$. Furthermore, $\mathbb{R}^n$ equipped with any of these norms is a Banach space. In particular, this is true for the *Euclidean norm* $\|x\| = (x \cdot x)^{1/2}$, where $x \cdot y$ is the usual *Euclidean inner product* on $\mathbb{R}^n$ given by

$$x \cdot y = \sum_{i=1}^{n} x_i y_i \quad \text{where} \quad x = (x_1, \ldots, x_n),\ y = (y_1, \ldots, y_n).$$

We turn now to linear maps.

Definition 2.2 *Let E, F be normed vector spaces. A function $L \colon E \to F$ is called a linear operator if (i) for all $x, y \in E$, $L(x + y) = L(x) + L(y)$, and (ii) for all $\alpha \in \mathbb{F}$ and $x \in E$, $L(\alpha x) = \alpha L(x)$.*

Proposition 2.3 *Suppose that L is a linear operator from the normed vector space E to the normed vector space F. The following are equivalent:*

(i) L is continuous;

(ii) L is continuous at 0;

(iii) there exists a real number $\beta > 0$ such that $\|L(x)\| \leq \beta\|x\|$ for all $x \in E$.

Linear operators which satisfy (iii) are called *bounded linear operators*. (Note that there are really two norms involved in this notion). According to the proposition these are precisely the continuous ones.

Any linear operator $L : \mathbb{R}^n \to \mathbb{R}^m$ is continuous (hence bounded) and has a unique matrix representation, once bases in $\mathbb{R}^n$ and $\mathbb{R}^m$ are specified.

The *operator norm* $\|L\|$ of a bounded linear operator L is the greatest lower bound of all positive real numbers β such that $\|L(x)\| \leq \beta\|x\|$ for all $x \in E$. It is thus the smallest non-negative real number $\|L\|$ satisfying $\|L(x)\| \leq \|L\| \, \|x\|$ for all $x \in E$. The space of all bounded linear operators from E to F is denoted $\mathcal{L}(E, F)$.

Proposition 2.4 *Under pointwise addition and scalar multiplication and the operator norm, the space $\mathcal{L}(E, F)$ is a normed linear space. It is a Banach space if F is.*

Exercises

Exercise 2.1 *Show that the metric of Definition 2.1 is indeed a metric. Show that vector space addition and scalar multiplication are continuous.*

Exercise 2.2 *Show that the usual euclidean norm on $\mathbb{R}^n$ yields a Banach space.*

Exercise 2.3 *Prove Proposition 2.3.*

Exercise 2.4 *Prove Proposition 2.4.*

Exercise 2.5 *Show that a linear operator $T : \mathbb{R}^n \to \mathbb{R}^m$ is continuous.*

3 Fréchet Derivatives

Let E and F be two normed linear spaces and let $f : U \to F$, where $U \neq \emptyset$ is an open subset of E.

Definition 3.1 *The function f is* differentiable *at $x \in U$ if there is a bounded linear operator $L : E \to F$ that approximates f at x in the sense that for all $h \in E$ such that $x + h \in U$, we have*

$$f(x+h) = f(x) + L(h) + \|h\|\eta(h) \quad \textit{where } \lim_{h\to 0} \eta(h) = 0. \tag{1}$$

The function f is differentiable on U if it is differentiable at each point of U.

If L exists at x, then it is uniquely determined.

Lemma 3.2 *If L and T are bounded linear operators satisfying Equation (1), then $L = T$.*

Proof. Let $h \neq 0 \in E$ (clearly $L(0) = 0 = T(0)$). Then for small enough $t > 0$, $x + th \in U$, so

$$\lim_{t\to 0^+} \frac{1}{t}\big(f(x+th) - f(x)\big) = \lim_{t\to 0^+} \frac{1}{t} L(th) + \frac{1}{t}\|th\|\eta(th) = L(h)$$

and similarly $T(h) = \lim_{t\to 0^+}(1/t)\big(f(x+th) - f(x)\big)$; thus $T = L$.

Definition 3.3 *Let E and F be two normed linear spaces and let $f : U \to F$, where $U \neq \emptyset$ is an open subset of E. If f is differentiable at $x \in U$, then (by Lemma 3.2) there exists a unique bounded linear operator $Df(x)\colon E \to F$ satisfying*

$$f(x+h) = f(x) + Df(x)(h) + \|h\|\eta(h) \quad \textit{where } \lim_{h\to 0} \eta(h) = 0, \tag{2}$$

called the (Frechét) derivative *or the* total derivative *of f at x.*

We list some standard basic properties of deriviatives.

Proposition 3.4 *If $f\colon U \to F$ is differentiable at x, then it is continuous at x. If f is continuous at x and if there exists a linear operator L satisfying Equation (1), then L is continuous and equal to $Df(x)$.*

The Frechét derivative in multivariable calculus is given by the Jacobian.

Proposition 3.5 *Let U be a nonempty open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}^m$ be a C^1 map. Then f is differentiable at all $x \in U$ and*

$$Df = Jf := \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

Thus $Df(x) = Jf(x) : \mathbb{R}^n \to \mathbb{R}^m$, where the entries of the Jacobian matrix are evaluated at $x = (x_1, \ldots, x_n)$. Hence $Df(x)(u)$ is given by multiplying the matrix $Df(x)$ times the $n \times 1$ column vector u.

A weaker notion of differentiability, directional differentiability, is given by the limit (if it exists) in the next proposition. In the presence of differentiability, it always exists and can be computed from the Frechét derivative.

Proposition 3.6 *If $f : U \to F$ is differentiable at $x \in U$, then*

$$df(x, h) := \lim_{t \to 0} \frac{f(x + th) - f(x)}{t} = Df(x)(h).$$

If the normed space E is the real line $\mathbb{R}$ (with norm absolute value), then a linear operator from $\mathbb{R}$ to a normed space Y has the form $t \mapsto ty : \mathbb{R} \to Y$ for some $y \in Y$. If $\alpha : (a, b) \to Y$ is differentiable at $x \in (a, b)$, an open interval in $\mathbb{R}$, then we typically identify the linear map $t \mapsto ty = D\alpha(x)$ with the vector y and call it the *tangent vector* to the curve α at $\alpha(x)$.

Exercises

Exercise 3.1 *Let E, F be normed spaces, U an open nonempty subset of E, and $f : U \to F$.*

(i) f is differentiable at $x_0 \in U$ iff there exists a bounded linear operator $L : E \to F$ such that

$$f(x) = f(x_0) + L(x - x_0) + \|x - x_0\|\phi(x), \text{ where } \lim_{x \to x_0} \phi(x) = 0.$$

In this case $L = Df(x_0)$.

(ii) *f is differentiable at $x \in U$ iff there exists a bounded linear operator $L : E \to F$ such that*

$$f(x+h) = f(x) + L(h) + r(h), \text{ where } \lim_{h\to 0} \frac{r(h)}{\|h\|} = 0.$$

The latter holds iff

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - L(h)\|}{\|h\|} = 0.$$

In either case $L = Df(x)$.

Exercise 3.2 *Prove Proposition 3.4.*

Exercise 3.3 *Let $f : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^n$. Interpret the derivative $Df(x)$ as the gradient vector at x identified with a member of the dual space of $\mathbb{R}^n$ by means of the inner product.*

Exercise 3.4 *Show that $t \mapsto ty : \mathbb{R} \to Y$ for some $y \in Y$, a normed space, is a bounded linear operator and every linear operator from $\mathbb{R}$ to Y has this form.*

4 Differentiation Rules

Proposition 4.1 *Let E, F be normed spaces, U open in E, $f, g : U \to F$. If f and g are differentiable at $x \in U$, then*

(i) *$f + g$ is differentiable at x and $D(f+g)(x) = \lambda D(f)(x) + D(g)(x)$;*

(ii) *λf is differentiable at x and $D(\lambda f)(x) = \lambda Df(x)$.*

Proposition 4.2 (The Chain Rule) *Suppose that E_1, E_2, and E_3 are normed linear spaces, $U \subseteq E_1$ and $V \subseteq E_2$ are non-empty open sets, and $f\colon U \to V$ and $g\colon V \to E_3$ are differentiable at x and $f(x)$ resp., then $g \circ f$ is differentiable at x with derivative $D(g \circ f)(x) = Dg(f(x)) \circ Df(x)$.*

The next proposition is a general version of the mean-value theorem.

Proposition 4.3 *Let E, F be normed spaces, U open in E, $f : U \to F$. Suppose that U contains the interval S with endpoints p and q and suppose f is differentiable at each point of S. Then*

$$\|f(q) - f(p)\| \leq \|q - p\| \sup_{x \in S} \|Df(x)\|.$$

Proposition 4.4 *If $f \colon U \to F$ is a constant function, then it is differentiable on U and the derivative is the 0-operator at every point.*

Proposition 4.5 *If $L \colon E \to F$ is a bounded linear operator, then L is differentiable on E and $DL(x) = L$ for all $x \in E$.*

Exercises

Exercise 4.1 *Prove Propositions 4.1.*

Exercise 4.2 *Prove Propositions 4.3 through 4.5.*

Exercise 4.3 *Let $(V, \|\cdot\|)$ be a normed vector space. Show that the addition map $(x, y) \mapsto x + y : V \times V \to V$ is a bounded linear map, and hence its derivative at every point is the addition function.*

5 Partial Derivatives

Let E_1, E_2, F be Banach spaces, and let $f : W \to F$ be a function, where W is a nonempy open subset of $E_1 \times E_2$. We define the *partial derivative* with respect to the first variable at $(x_0, y_0) \in U$ to be the derivative at x_0 of the map $x \mapsto f(x, y_0)$ defined on some neighborhood U of x_0 such that $U \times V$ is a neighborhood of (x_0, y_0) contained in W. The partial derivative is denoted $D_1 f(x_0, y_0)$. The partial with respect to the second variable, $D_2 f(x_0, y_0)$ is defined analogously.

Proposition 5.1 *If in the previous setting f is differentiable at (x_0, y_0), then the partial derivatives of f at (x_0, y_0) exist and $Df(x_0, y_0)(u, v) = D_1 f(x_0, y_0)u + D_2 f(x_0, y_0)v$. If $(x, y) \mapsto Df_1(x, y)$ and $(x, y) \mapsto Df_2(x, y)$ are continuous on W, then f is differentialble on W and the last formula of the preceding sentence continues to hold.*

Exercises

Exercise 5.1 *Formulate a matrix version of the preceding proposition for differentiable* $f : E_1 \times E_2 \to F_1 \times F_2$*. (Hint: consider the derivative of the composite* $x \mapsto (x, y_0) \mapsto f(x, y_0) \mapsto \pi_i f(x, y_0) = f_i(x, y_0)$*, where* π_i *is projection onto the* i*-th coordinate for* $i = 1, 2$*; the derivative is a linear map from* E_1 *to* F_i*. The derivative is the "Jacobian" with the four partials for entries.)*

Exercise 5.2 *Let* $F : E_1 \times E_2 \to F$ *be a continuous bilinear map for normed spaces* E_1, E_2, F*. Show that* F *is differentiable and* $dF_{(x,y)}(u, v) = F(x, v) + F(u, y)$*.*

Exercise 5.3 *Apply the preceding exercise to the scalar multiplication function* $(\lambda, x) \mapsto \lambda x : \mathbb{R} \times E \to E$ *of a Banach space* E*.*

Exercise 5.4 *Show that for* $p, q \in Y$*, a Banach space, the map* $t \mapsto tp + (1 - t)q : \mathbb{R} \to Y$ *is differentiable and find its derivative.*

6 Higher Derivatives and Smooth Functions

Let $f : U \to F$ be a differentiable function, where $U \subseteq E$ is a nonempty open subset and E, F are Banach spaces. Then $Df(x)$ exists for every $x \in U$. Thus we may define $Df : U \to \mathcal{L}(E, F)$ by assigning to $x \in U$ the bounded linear operator $Df(x)$. With respect to the operator norm $\mathcal{L}(E, F)$ is again a Banach space. Thus we can apply our theory to talk about the differentiability or non-differentiability of $Df : U \to \mathcal{L}(E, F)$. If Df is differentiable, we call its derivative the *second derivative* of f, denoted $D^2 f : U \to \mathcal{L}(E, \mathcal{L}(E, F))$. Inductively we can define *higher order derivatives* of all orders. The function f is said to be of *class* C^r if the r^{th} derivative exists and is continuous, and is *smooth* or of *class* C^∞ if it is of class C^r for every nonnegative integer r. A function is a C^r*-diffeomorphism* if it is a C^r-bijection with a C^r-inverse.

If E and F are finite dimensional, the identification of the Fréchet derivative with the Jacobian allows one to see directly that $f : U \to F$ is C^1 in the above sense if and only if all partials exist and are continuous. It follows that we may alternatively define class C^r and smoothness in terms of the existence and continuity of the appropriate partials, as we did at the beginning of these notes.

Exercises

Exercise 6.1 *For E, F Banach spaces, let $\mathcal{B}(E\times E, F)$ denote all continuous bilinear maps. Then with respect to pointwise addition and scalar multiplication $\mathcal{B}(E \times E, F)$ is a vector space and is a Banach space with respect to the norm*

$$\|\beta\| := \inf\{K \geq 0 : \forall x, y \in E,\ \|\beta(x, y)\| \leq K\|x\|\|y\|\}.$$

Exercise 6.2 *For E, F Banach spaces, $\mathcal{B}(E \times E, F)$ is linearly isometric to $\mathcal{L}(E, \mathcal{L}(E, F))$ with respect to the currying map $\beta \mapsto \hat{\beta}$, where $\hat{\beta}(x)(y) = \beta(x, y)$. Thus with respect to this identification we may identify the second derivatives with continuous bilinear maps. In an analogous way, higher derivatives may be identified with continuous multilinear maps.*

7 Key Theorems

One of the basic tools of the theory is the Inverse Function Theorem.

Theorem 7.1 (Inverse Function Theorem) *Suppose E, F are Banach spaces, $U \neq \emptyset$ is open in U, and $f : U \to F$ is C^r for $1 \leq r \leq \infty$ or $r = \omega$. If for $x \in U$, $Df(x) : E \to F$ is a continuous vector space isomorphism, then there exists V open, $x \in V \subseteq U$ and W open in F such that $f(x) \in W$ and $f|_V : V \to W$ is a C^r-diffeomorphism.*

Theorem 7.2 (Constant-Rank Theorem) *Let M and N be smooth finite-dimensional manifolds and let $F : M \to N$ be a smooth mapping whose tangent mappings $DF(x)$ have constant rank k at all points of M. Then*

(i) for each y in the range of F, the inverse image $F^{-1}(y)$ is an embedded submanifold of M of dimension $m - k$, where $m = \dim M$;

(ii) for each $x \in M$, there exists a neighborhood U_x of x in M such that $F(U_x)$ is an embedded submanifold of N of dimension k.

Proof. (i) Let y be in the range of F. Let $F(p) = y$. Let V be an open set around y which is the domain of a chart ψ for which $\psi(y) = 0$. Let ϕ be a chart of M such that the domain U of ϕ contains p, $\phi(p) = 0$, and $F(U) \subseteq V$. Since $D\psi(\phi(p))\circ F\circ\phi^{-1}$ has rank k, there exists bases $\{e_1, \ldots, e_m\}$ of $\mathbb{R}^m$ and $\{f_1, \ldots, f_n\}$ of $\mathbb{R}^n$ such that $d_{\phi(p)}\psi\circ F\circ\phi^{-1}$ restricted to the subspace spanned by $e_1, \ldots, e_k$ is a vector space isomorphism onto the subspace spanned by $f_1, \ldots, f_k$. We define a function G from U to $\mathbb{R}^m$ by projecting $\psi \circ F(x)$ into the first k-coordinates with respect to the basis $\{f_1, \ldots, f_n\}$ and $\phi(x)$ into the last $m-k$ coordinates with respect to the basis $\{e_1, \ldots, e_m\}$. Since a change of coordinates between bases is a linear mapping, we conclude that the two blocks of G of dimension k and $m-k$ resp. are smooth, and hence that G is smooth.

Consider the mapping $G \circ \phi^{-1}$ on $\mathbb{R}^m$. With respect to the coordinate system given by the basis $\{e_1, \ldots e_m\}$, this mapping has its Jacobian matrix at 0 lower block triangular with the lower $m-k \times m-k$ diagonal block being the identity and the upper $k \times k$ block being invertible (from the rank condition). Thus the Jacobian matrix is invertible, the Inverse Function Theorem applies, and we conclude that $G \circ \phi^{-1}$ is smooth.