

A Note on Differential Calculus in \mathbf{R}^n

by James Hebda
August 2010

I. Partial Derivatives of Functions

Let $f : U \rightarrow \mathbf{R}$ be a real valued function defined in an open neighborhood U of the point $a = (a_1, \dots, a_n)$ in the n -dimensional Euclidean space \mathbf{R}^n . For each $i = 1, \dots, n$, one defines the i -th partial derivative of f at a to be the limit

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \rightarrow 0} \frac{f(a_1, \dots, a_{i-1}, a_i + h, a_{i+1}, \dots, a_n) - f(a)}{h},$$

provided the limit exists.

If the i -th partial derivatives at a of the two real valued functions f and g exist, then so does the i -partial derivative of their sum and product. In this case the sum and product rules hold:

$$\frac{\partial(f + g)}{\partial x_i}(a) = \frac{\partial f}{\partial x_i}(a) + \frac{\partial g}{\partial x_i}(a)$$

and

$$\frac{\partial(fg)}{\partial x_i}(a) = \frac{\partial f}{\partial x_i}(a)g(a) + f(a)\frac{\partial g}{\partial x_i}(a).$$

Moreover, if $g(a) \neq 0$, the i -th partial derivative of the quotient f/g exists and is given by the quotient rule:

$$\frac{\partial(f/g)}{\partial x_i}(a) = \frac{\frac{\partial f}{\partial x_i}(a)g(a) - f(a)\frac{\partial g}{\partial x_i}(a)}{g(a)^2}.$$

These differentiation formulas are consequences of the corresponding formulas for functions of one variable because the i -th partial derivative of a function f is the derivative of the one variable function $\tilde{f}(x) = f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$ obtained by holding all but the i -th coordinate fixed.

Recall that the proofs of the product and quotient rules in the single variable case depend upon the fact that differentiability implies the continuity of the function at the point. In contrast, the existence of all the partial derivatives of a function of several variables at a point does not imply the continuity of the function. For a simple example let $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ be defined by

$$f(x, y) = \begin{cases} 0 & \text{if } xy = 0 \\ 1 & \text{if } xy \neq 0. \end{cases}$$

Clearly, f is discontinuous at $(0, 0)$ although $\frac{\partial f}{\partial x}(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0$.

II. Differentiability

Let $f = (f_1, \dots, f_m) : U \rightarrow \mathbf{R}^m$ be an \mathbf{R}^m valued function defined in an open neighborhood U of the point $a = (a_1, \dots, a_n)$ in the n -dimensional Euclidean space \mathbf{R}^n . The function f is defined to be differentiable at a if there exists a linear transformation $T : \mathbf{R}^n \rightarrow \mathbf{R}^m$ such that

$$\lim_{x \rightarrow a} \frac{|f(x) - f(a) - T(x - a)|}{|x - a|} = 0,$$

where $|-|$ denotes the norms on \mathbf{R}^n and \mathbf{R}^m .¹

If f is differentiable at a , then the linear transformation T is unique. In fact T is given by the formula:

$$T(v) = \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t}. \tag{*}$$

¹ Since all norms on finite dimensional vector spaces are equivalent, the differentiability of a function is independent of which norm is used.

This unique linear transformation is called the derivative of f at a and is denoted $Df(a)$.² When $m = 1$, it may also be called the total derivative or total differential.

It is important to observe that the mapping f is differentiable at a if and only if each of the component functions f_i is differentiable at a . Indeed, if this is the case, $Df_i(a)$ is just the projection of $Df(a)$ on the i -th coordinate of \mathbf{R}^m . This is proved from the definition by using the norm inequalities $|y_i| \leq |y| \leq \sum_{i=1}^m |y_i|$ for every $y = (y_1, \dots, y_m) \in \mathbf{R}^m$.

Equation (*) shows that if f is differentiable at a then the “directional derivatives” of the component functions f_i of f exist along all lines through the point a . In particular, the partial derivatives of the component functions f_i exist at a , and the $m \times n$ -matrix for $Df(a)$ with respect to the standard bases of Euclidean space takes the form:

$$[Df(a)]_{ij} = \frac{\partial f_i}{\partial x_j}(a).$$

This matrix is called the Jacobian matrix of f at a .

The existence of all the partial derivatives $\frac{\partial f_i}{\partial x_j}(a)$ is not sufficient for differentiability. What's more, even the existence of a linear transformation T satisfying equation (*) is not sufficient for differentiability. For example, define

$$f(x, y) = \begin{cases} 1 & \text{if } y = x^2 \text{ and } (x, y) \neq (0, 0) \\ 0 & \text{otherwise} \end{cases}.$$

Then $T(v) = 0$ satisfies equation (*) at $a = (0, 0)$ for all $v \in \mathbf{R}^2$. On the other hand f is not continuous at $(0, 0)$, and thus is not differentiable there because:

Differentiability implies continuity. *If f is differentiable at a then f is continuous at a .*

Proof: We may write $f(x) = f(a) + Df(a)(x - a) + \phi(x)$ where ϕ is a \mathbf{R}^m -valued function satisfying $\lim_{x \rightarrow a} \frac{|\phi(x)|}{|x - a|} = 0$. This limit clearly implies that $\lim_{x \rightarrow a} \phi(x) = 0$. Because $\lim_{x \rightarrow a} Df(a)(x - a) = 0$ as well, we conclude $\lim_{x \rightarrow a} f(x) = f(a)$.

Examples: If $f = T$ is a linear transformation, then f is differentiable at every point $a \in \mathbf{R}^n$, and $Df(a) = T$ because $|f(x) - f(a) - T(x - a)| = 0$ for all x .

Similarly, if $f = B : \mathbf{R}^{n_1} \times \mathbf{R}^{n_2} \rightarrow \mathbf{R}^m$ is bilinear, then f is differentiable at every point $(a, b) \in \mathbf{R}^{n_1} \times \mathbf{R}^{n_2}$, and $Df(a, b)(v, w) = B(a, w) + B(v, b)$ for every $(v, w) \in \mathbf{R}^{n_1} \times \mathbf{R}^{n_2}$. This can be proved by first observing $f(x, y) - f(a, b) - B(a, y - b) - B(x - a, b) = B(x - a, y - b)$ and next seeing there is a constant C such that $|B(v, w)| \leq C|v||w| \leq C|(v, w)|^2$ for all $(v, w) \in \mathbf{R}^{n_1} \times \mathbf{R}^{n_2}$. Thus $\lim_{(v, w) \rightarrow (0, 0)} \frac{|B(v, w)|}{|(v, w)|} = 0$ which implies the differentiability of f .³ Similar results hold for multilinear mappings.

The Chain Rule. *Let g be differentiable at a , and let f be differentiable at $g(a)$, then the composition $f \circ g$ is differentiable at a , and*

$$D(f \circ g)(a) = Df(g(a)) \circ Dg(a).$$

Proof: Write

$$\begin{aligned} g(x) &= g(a) + Dg(a)(x - a) + \phi(x) \\ f(y) &= f(g(a)) + Df(g(a))(y - g(a)) + \psi(y) \end{aligned}$$

where $\lim_{x \rightarrow a} \frac{|\phi(x)|}{|x - a|} = 0$ and $\lim_{y \rightarrow g(a)} \frac{|\psi(y)|}{|y - g(a)|} = 0$. Thus on substituting $g(x)$ for y ,

$$f(g(x)) = f(g(a)) + Df(g(a))(Dg(a)(x - a)) + Df(g(a))(\phi(x)) + \psi(g(x)).$$

We need to show (1) $\lim_{x \rightarrow a} \frac{|Df(g(a))(\phi(x))|}{|x - a|} = 0$ and (2) $\lim_{x \rightarrow a} \frac{|\psi(g(x))|}{|x - a|} = 0$. Because $Df(g(a))$ is a linear transformation, there is a constant C such that $|Df(g(a))(y)| \leq C|y|$ for all y . Consequently,

$$\frac{|Df(g(a))(\phi(x))|}{|x - a|} \leq C \frac{|\phi(x)|}{|x - a|}$$

² This is called the Fréchet derivative of f . A weaker notion of differentiability is the Gateau derivative defined by equation (*).

³ In the case $m = 1$, one may take C equal to $n_1 n_2$ times the largest of the numbers $|B(e_i, e'_j)|$ as e_i and e'_j run through the standard bases elements of \mathbf{R}^{n_1} and \mathbf{R}^{n_2} respectively.

from which limit (1) follows. Given $\epsilon > 0$, there exists a $\delta > 0$ such that $|\psi(y)| < \epsilon|y - g(a)|$ whenever $0 < |y - g(a)| < \delta$. Furthermore, since g is continuous at a , there exists an $\eta > 0$ such that $|g(x) - g(a)| < \delta$ whenever $|x - a| < \eta$. Thus if $0 < |x - a| < \eta$,

$$|\psi(g(x))| \leq \epsilon|g(x) - g(a)| \leq \epsilon|Dg(a)(x - a)| + \epsilon|\phi(x)|.$$

Therefore, for some constant C' depending only on the linear transformation $Dg(a)$,

$$\frac{|\psi(g(x))|}{|x - a|} \leq \epsilon C' + \epsilon \frac{|\phi(x)|}{|x - a|},$$

whenever $0 < |x - a| < \eta$. Since ϵ is arbitrary, limit (2) follows.

The next result applies to all varieties of products, including dot and cross products.

The Omnibus Product Rule. *Let $B : \mathbf{R}^{m_1} \times \mathbf{R}^{m_2} \rightarrow \mathbf{R}^k$ be bilinear. Suppose f and g are maps of a neighborhood of $a \in \mathbf{R}^n$ into \mathbf{R}^{m_i} , $i = 1$ and 2 respectively, which are differentiable at a . Then $B(f, g)$ is differentiable at a and*

$$D(B(f, g))(a) = B(f(a), Dg(a)) + B(Df(a), g(a)).$$

Proof: Apply the chain rule to the composition of the maps (f, g) and B .

We end this section with:

A sufficient criterion for differentiability. *Assume that $\frac{\partial f_i}{\partial x_j}$ are defined throughout a neighborhood of a , and are continuous at a . Then $f = (f_1, \dots, f_m)$ is differentiable at a .*

Proof: It suffices to consider the case $m = 1$. Then the mean value theorem implies

$$\begin{aligned} f(x) - f(a) &= f(x_1, x_2, \dots, x_n) - f(a_1, x_2, \dots, x_n) + f(a_1, x_2, \dots, x_n) - f(a_1, a_2, x_3, \dots, x_n) + \dots \\ &\quad \dots + f(a_1, \dots, a_{n-1}, x_n) - f(a_1, \dots, a_n) \\ &= \frac{\partial f}{\partial x_1}(\xi_1, x_2, \dots, x_n)(x_1 - a_1) + \frac{\partial f}{\partial x_2}(a_1, \xi_2, x_3, \dots, x_n)(x_2 - a_2) + \dots \\ &\quad \dots + \frac{\partial f}{\partial x_n}(a_1, \dots, a_{n-1}, \xi_n)(x_n - a_n) \end{aligned}$$

where ξ_i lies between x_i and a_i . Set $\hat{\xi}_i = (a_1, \dots, a_{i-1}, \xi_i, x_{i+1}, \dots, x_n)$. Then

$$\begin{aligned} \left| f(x) - f(a) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a)(x_i - a_i) \right| &= \left| \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(\hat{\xi}_i) - \frac{\partial f}{\partial x_i}(a) \right) (x_i - a_i) \right| \\ &\leq \sum_{i=1}^n \epsilon |x_i - a_i| \\ &\leq n\epsilon |x - a| \end{aligned}$$

whenever $|x - a| < \delta$ for $\delta > 0$ chosen so that $|x - a| < \delta$ implies that

$$\left| \frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a) \right| < \epsilon$$

for all i . This completes the proof.

On the other hand the continuity of the partial derivatives is not a necessary requirement for differentiability. For example, $x^2 \sin(1/x)$ is differentiable at 0, but the derivative is not continuous at 0. One can make a function of several variables out of this: $(x^2 + y^2) \sin(1/\sqrt{x^2 + y^2})$. (In both these examples the function is assumed to have the value 0 at the origin.)

III. Higher Derivatives

Let f be a real valued function defined in an open set $U \subset \mathbf{R}^n$. f is said to be C^0 if f is continuous. For an integer $k > 0$, we say f is of differentiability class C^k , or simply C^k , if the partial derivatives $\frac{\partial f}{\partial x_i}$ are C^{k-1} functions on U for $i = 1, \dots, n$. We will let $C^k(U)$ denote the collection of all C^k functions on U . Using the sum and product rules for partial derivatives and the differentiability criterion, we find by induction that the $C^k(U)$ form a nested decreasing sequence of real algebras

$$C^0(U) \supset C^1(U) \supset C^2(U) \supset C^3(U) \dots$$

under the operations of pointwise addition and multiplication of functions. f is said to be infinitely differentiable, C^∞ , or smooth if f is C^k for all k . Thus the collection of C^∞ functions on U is the intersection

$$C^\infty(U) = \cap_{k=0}^\infty C^k(U)$$

which forms a real algebra under pointwise addition and multiplication.

The k -order partials of a function are defined by successive differentiation:

$$\frac{\partial^k f}{\partial x_{\alpha_k} \cdots \partial x_{\alpha_1}} = \frac{\partial}{\partial x_{\alpha_k}} \cdots \frac{\partial}{\partial x_{\alpha_1}} f.$$

In general the order of differentiation is important. For example $\frac{\partial^2 f}{\partial x \partial y} \neq \frac{\partial^2 f}{\partial y \partial x}$ at $(0, 0)$ for the function

$$f(x, y) = \begin{cases} xy \frac{x^2 - y^2}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

Symmetry of Mixed Partials. If f is C^k , then the k -order partial derivatives of f are independent of the order of differentiation. In particular, for C^∞ functions, their higher order partial derivatives are independent of the order of differentiation.

Using induction on k , the proof of the symmetry of mixed partials reduces to showing the following result in 2 dimensions:

Lemma. If $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ are continuous on U , then $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$ on U .

Proof: On one hand, if $\frac{\partial^2 f}{\partial x \partial y}(a) > \frac{\partial^2 f}{\partial y \partial x}(a)$ at some point a in U , then by continuity there is a rectangle R contained in U for which $\frac{\partial^2 f}{\partial x \partial y}(x, y) > \frac{\partial^2 f}{\partial y \partial x}(x, y)$ for all $(x, y) \in R$. Thus on integrating over R ,

$$\iint_R \frac{\partial^2 f}{\partial x \partial y} dA > \iint_R \frac{\partial^2 f}{\partial y \partial x} dA.$$

On the other hand, application of Fubini's theorem with $R = [x_0, x_1] \times [y_0, y_1]$ leads to the contradiction

$$\iint_R \frac{\partial^2 f}{\partial x \partial y} dA = \int_{y_0}^{y_1} \int_{x_0}^{x_1} \frac{\partial^2 f}{\partial x \partial y} dx dy = \int_{x_0}^{x_1} \int_{y_0}^{y_1} \frac{\partial^2 f}{\partial y \partial x} dy dx = \iint_R \frac{\partial^2 f}{\partial y \partial x} dA.$$

For an infinitely differentiable function, Taylor's Theorem takes the following form.

Taylor's Formula. If U is an open set in \mathbf{R}^n which is star-shaped about the point a , and f is a C^∞ real-valued function defined in U then for each $k \geq 0$ and $x \in U$

$$\begin{aligned} f(x) &= f(a) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a)(x_i - a_i) + \frac{1}{2!} \sum_{i_1, i_2=1}^n \frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2}}(a)(x_{i_1} - a_{i_1})(x_{i_2} - a_{i_2}) + \dots \\ &\quad \dots + \frac{1}{k!} \sum_{i_1, \dots, i_k=1}^n \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(a)(x_{i_1} - a_{i_1}) \cdots (x_{i_k} - a_{i_k}) + \\ &\quad \frac{1}{(k+1)!} \sum_{i_1, \dots, i_{k+1}=1}^n h_{i_1 \dots i_{k+1}}(x)(x_{i_1} - a_{i_1}) \cdots (x_{i_{k+1}} - a_{i_{k+1}}), \end{aligned}$$

where $h_{i_1 \dots i_{k+1}}$ is C^∞ and satisfies $h_{i_1 \dots i_{k+1}}(a) = \frac{\partial^{k+1} f}{\partial x_{i_1} \dots \partial x_{i_{k+1}}}(a)$.

Proof: A set $U \subset \mathbf{R}^n$ is star-shaped about $a \in U$ if for every $x \in U$, the line segment joining a to x is contained in U . Hence by the Fundamental Theorem of Calculus and the chain rule, if $x \in U$, then

$$f(x) - f(a) = \int_0^1 \frac{d}{dt} f(a + t(x - a)) dt = \sum_{i=1}^n \int_0^1 \frac{\partial f}{\partial x_i}(a + t(x - a))(x_i - a_i) dt \quad (\dagger)$$

which is Taylor's formula when $k = 0$ setting $h_i(x) = \int_0^1 \frac{\partial f}{\partial x_i}(a + t(x - a)) dt$. (The smoothness of the h_i follow from the theorems about differentiating under the integral sign.) Having established Taylor's formula for k , the formula for $k + 1$ is obtained by replacing the coefficient functions $h_{i_1 \dots i_{k+1}}$ by their expansions as in (\dagger) and symmetrizing the resulting coefficient functions of the monomials of degree $k + 2$.

We say that a mapping $f = (f_1, \dots, f_m) : U \rightarrow \mathbf{R}^m$ is C^k , if each of the m component functions f_i are C^k . Using induction over k , the chain rule shows that the composition of C^k mappings are C^k . Likewise, the omnibus product formula shows that "products" of C^k mappings are C^k .

Here is an application of Taylor's formula.

C^∞ mappings are locally Lipschitz. Suppose $f : U \rightarrow \mathbf{R}^m$ is a C^∞ mapping of the open set $U \subset \mathbf{R}^n$, then for any compact convex subset $K \subset U$, there is a constant L_K , such that

$$|f(x) - f(y)| \leq L_K |x - y|$$

for all $x, y \in K$.

Proof: Write $f = (f_1, \dots, f_m)$ and let M be the maximum of the supremums $\sup\{|\frac{\partial f_i}{\partial x_j}(x)| : x \in K\}$. Then applying equation (\dagger) to each component function shows that

$$\begin{aligned} |f(x) - f(y)| &\leq \sum_{i=1}^m |f_i(x) - f_i(y)| \leq \sum_{i=1}^m \sum_{j=1}^n \left(\int_0^1 \left| \frac{\partial f_i}{\partial x_j}(y + t(x - y)) \right| dt \right) |x_j - y_j| \\ &\leq \sum_{i=1}^m \sum_{j=1}^n M |x_j - y_j| \leq mnM |x - y| \end{aligned}$$

if $x, y \in K$. Therefore $L_K = mnM$.

IV. Three Theorems About Smooth Mappings.

We state without proof:

The Inverse Function Theorem. Let f be a C^∞ mapping from an open set $U \subset \mathbf{R}^n$ into \mathbf{R}^n . Suppose that the derivative $Df(a)$ is an invertible linear transformation at the point a in \mathbf{R}^n . Then there exists an open neighborhood V of a with $V \subset U$ such that the restriction $f|V$ is a one to one mapping onto an open neighborhood W of $f(a)$ such the the inverse mapping $f|V^{-1} : W \rightarrow V$ is C^∞ . Moreover $Df^{-1}(f(a)) = Df(a)^{-1}$.

An smooth mapping that has an smooth inverse is called a *diffeomorphism*. We remark that if $f : V \rightarrow W$ is a diffeomorphism with inverse mapping $g : W \rightarrow V$ where $V \subset \mathbf{R}^n$ and $W \subset \mathbf{R}^m$, then $m = n$ since by the chain rule the derivatives $Df(a)$ and $Dg(f(a))$ would be inverse linear transformations of each other for every $a \in V$.

Using this terminology we restate the inverse function theorem: If the derivative a smooth mapping f is invertible at a point a , then f restricted to a sufficiently small neighborhood of the point a is a diffeomorphism of that neighborhood onto a neighborhood of the point $f(a)$.

The Rank Theorem. Let $f : U \rightarrow \mathbf{R}^m$ be a smooth mapping defined on an open set U in \mathbf{R}^n . Suppose that the rank of $Df(x)$ is a constant r for all x in a neighborhood of the point $a \in U$. Then there exist neighborhoods V of a and W of $f(a)$, together with diffeomorphisms $\phi : V \rightarrow \phi(V) \subset \mathbf{R}^n$, and $\psi : W \rightarrow \psi(W) \subset \mathbf{R}^m$ such that the mapping

$$\psi \circ f \circ \phi^{-1} : \phi(V) \rightarrow \psi(W)$$

takes the form

$$\psi \circ f \circ \phi^{-1}(t_1, \dots, t_n) = (t_1, \dots, t_r, 0, \dots, 0).$$

The Implicit Function Theorem. Let f be a C^∞ mapping from an open set $U \subset \mathbf{R}^n \times \mathbf{R}^m$ into \mathbf{R}^m . Suppose $f(a, b) = 0$. Writing the coordinates $(x, y) = (x_1, \dots, x_n, y_1, \dots, y_m)$ in $\mathbf{R}^n \times \mathbf{R}^m$, assume that the $m \times m$ square matrix

$$\left[\frac{\partial f_i}{\partial y_j}(a, b) \right]$$

is nonsingular. Then there exists a C^∞ map $\phi : V \rightarrow \mathbf{R}^m$ defined in an open neighborhood V of a in \mathbf{R}^n such that $\phi(a) = b$ and $f(x, \phi(x)) = 0$ for all x in V . The function ϕ is unique in the sense that any two such functions agree in a neighborhood of a .