Higher Derivatives and Taylor's Formula Via Multilinear Maps

August 26, 2021

1 Differentiability

Let V and W be finite-dimensional vector spaces over $\mathbb{R}$ of dimensions m and n respectively and let U be an open set of V. A map $f\colon U \to W$ is **differentiable** if and only if for each $u \in U$ there exists a (necessarily unique) linear map $\mathrm{D}f(u)\colon V \to W$ such that

$$\frac{\|f(u+h) - f(u) - (\mathrm{D}f(u))(h)\|}{\|h\|} \to 0$$

as $h \to 0$ in V (where the norms on the top and bottom are on V and W, and the choices do not impact the definition since any two norms on a finite-dimensional $\mathbb{R}$-vector space are bounded by a constant positive multiple of each other). If we fixed ordered bases for V and W, then we can identity V with $\mathbb{R}^m$, W with $\mathbb{R}^n$, U with an open subset of $\mathbb{R}^m$, and $\mathrm{D}f(u)$ with an $n \times m$ matrix over $\mathbb{R}$. In this case, the map $f\colon U \to W$ has the form

$$\begin{aligned} f\colon U &\longrightarrow W \\ u = (u_1, \dots, u_m)^\top &\mapsto (f_1(u), \dots, f_n(u))^\top = f(u), \end{aligned}$$

where the $f_j\colon U \to \mathbb{R}$ for $1 \leq j \leq n$ are called the **component functions** of f. We claim that for each $u \in U$, the linear map $\mathrm{D}f(u)$ is none other than the **Jacobian matrix** of f at u:

$$\mathrm{D}f(u) = \mathrm{J}_f(u) := \begin{pmatrix} (\partial_{x_1} f_1)(u) & \cdots & (\partial_{x_m} f_1)(u) \\ \vdots & \ddots & \vdots \\ (\partial_{x_1} f_n)(u) & \cdots & (\partial_{x_m} f_n)(u) \end{pmatrix}.$$

Indeed, suppose $h = (0, \dots, h_i, \dots, 0)^\top$ and let $\mathrm{D}f(u)_i$ (respectively $\mathrm{D}f(u)_i^j$) denote the ith column vector (respectively the (j,i) entry) of the matrix $\mathrm{D}f(u)$. Then as $h_i \to 0$, we see that

$$\frac{\|f(u+h) - f(u) - (\mathrm{D}f(u))(h)\|}{\|h\|} = \left\| \frac{f(u_1, \dots, u_i + h_i, \dots, u_m)^\top - f(u_1, \dots, u_m)^\top - h_i \mathrm{D}f(u)_i}{h_i} \right\| \to 0$$

as $h_i \to 0$. In particular, this implies the jth component of the vector

$$\frac{f(u_1, \dots, u_i + h_i, \dots, u_m)^\top - f(u_1, \dots, u_m)^\top - h_i \mathrm{D}f(u)_i}{h_i} \in \mathbb{R}^n$$

tends to zero as $h_i \to 0$. The jth component of this vector is

$$\frac{f_j(u_1, \dots, u_i + h_i, \dots, u_n) - f_j(u_1, \dots, u_n) - h_i \mathrm{D}f(u)_i^j}{h_i}.$$

In particular, this means $\mathrm{D}f(u)_i^j = \partial_{x_i} f_j(u)$. Thus our claim is proved.

Example 1.1. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x) = x_1^2 + x_1 x_2 + x_2$ for all $x = (x_1, x_2)$ in $\mathbb{R}^2$. Then f is differentiable with its derivative defined by

$$\mathrm{D}f(x) = \begin{pmatrix} 2x_1 + x_2 & x_1 + 1 \end{pmatrix}$$

for all $x = (x_1, x_2)$ in $\mathbb{R}^2$.

Example 1.2. Let U be the open subset of $\mathbb{R}^2$ given by

$$U = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 x_2 \in \mathbb{R}\backslash\{\pi/2 + \pi\mathbb{Z}\} \text{ and } x_1 + x_2 \in (0, \infty)\}$$

and let $f\colon U \to \mathbb{R}^2$ be defined by

$$f(u) = (\tan(u_1 u_2), \ln(u_1 + u_2))$$

for all $u = (u_1, u_2)$ in U. Then f is differentiable with its derivative defined by

$$\mathrm{D}f(u) = \begin{pmatrix} u_2 \sec^2(u_1 u_2) & u_1 \sec^2(u_1 u_2) \\ \frac{1}{u_1+u_2} & \frac{1}{u_1+u_2} \end{pmatrix}$$

for all $u = (u_1, u_2)$ in U.

1.1 Chain Rule

Proposition 1.1. *Let V, V',and V'' be finite-dimensional $\mathbb{R}$-vector spaces of dimensions n, n', and n'' respectively, let U be an open subset of V, let U' be an open subset of V', and let $f\colon U \to V'$ and $g\colon U' \to V''$ be differentiable functions. Then the map $g \circ f\colon U \cap f^{-1}(U') \to V''$ is differentiable and for all $u \in U \cap f^{-1}(U')$, we have*

$$\mathrm{D}(g \circ f)(u) = \mathrm{D}g(f(u)) \circ \mathrm{D}f(u). \tag{1}$$

Proof. After choosing bases for V, V', and V'', we may identify $\mathrm{D}(g \circ f)(u)$ with the $n'' \times n$ matrix whose (i'', i) entry is $\partial_{x_i}(g \circ f)_{i''}(u)$, we may identify $\mathrm{D}g(f(u))$ with the $n'' \times n'$ matrix whose (i'', i') entry is $\partial_{f_{i'}} g_{i''}(f(u))$, and we may identify $\mathrm{D}f(u)$ with the $n' \times n$ matrix whose (i', i) entry is $\partial_{x_i} f_{i'}(u)$. Then (1) turns into the matrix equation

$$(\partial_{x_i}(g \circ f)_{i''}(u)) = (\partial_{f_{i'}} g_{i''}(f(u))) \cdot (\partial_{x_i} f_{i'}(u))$$

which gives us a system of equations

$$\partial_{x_i}(g \circ f)_{i''}(u) = \sum_{i'=1}^{n'} \partial_{f_{i'}} g_{i''}(f(u)) \partial_{x_i} f_{i'}(u) \tag{2}$$

for each $1 \le i \le n$ and $1 \le i'' \le n''$. Therefore (1) holds if and only if (2) holds for all i, i'', and (2) holds since this is just the chain rule in the classical case. □

1.2 C^1 maps

Let $f\colon U \to W$ be differentiable as above. Then for each $u \in U$ we get a linear map $\mathrm{D}f(u)\colon V \to W$. Hence, we get a linear map

$$\mathrm{D}f\colon U \to \mathrm{Hom}(V, W)$$

into a *new* target vector space, namely $\mathrm{Hom}(V, W)$. In terms of linear coordinates, this is a "matrix-valued" function on U, but we want to consider this target space of matrices as a vector space in its own right and hence on par with the initial target W.

What does it mean to say that $\mathrm{D}f\colon U \to \mathrm{Hom}(V, W)$ is continuous? Upon fixing linear coordinates on V and W, such continuity amounts to continuity for each of the component functions $\partial_{x_i} f_j\colon U \to \mathbb{R}$ of the matrix-valued $\mathrm{D}f$, and so the concrete definition of f being C^1 is equivalent to the coordinate-free property that $f\colon U \to W$ is differentiable and that the associated total derivative map $\mathrm{D}f\colon U \to \mathrm{Hom}(V, W)$ from U to a new vector space $\mathrm{Hom}(V, W)$ is continuous. With this latter point of view, wherein $\mathrm{D}f$ is a map from the open set $U \subseteq V$ into a finite-dimensional vector space $\mathrm{Hom}(V, W)$, a very natural question is: what does it mean to say that Df is differentiable, or even continuously so?

Lemma 1.1. *Suppose $f\colon U \to W$ is a C^1 map, and let $\mathrm{D}f\colon U \to \mathrm{Hom}(V, W)$ be the associated total derivative map. As a map from an open set in V to a finite-dimensional vector space, $\mathrm{D}f$ is C^1 if and only if (relative to a choice of linear coordinates on V and W) all second-order partials $\partial_{x_{i_1}} \partial_{x_{i_2}} f_j\colon U \to \mathbb{R}$ exist and are continuous.*

Proof. Fixing linear coordinates identifies $\mathrm{D}f$ with a map from an open set $U \subseteq \mathbb{R}^m$ to a Euclidean space of $n \times m$ matrices, with component functions $\partial_{x_i} f_j$ for $1 \le i \le m$ and $1 \le j \le n$. Hence, this map is C^1 if and only if these components admit all first-order partials that are moreover continuous, and this is exactly the statement that the f_j's admit all second-order partials and that such partials are continuous. □

Let us say that $f\colon U \to W$ is C^2 when it is differentiable and $\mathrm{D}f\colon U \to \mathrm{Hom}(V,W)$ is C^1. By the lemma, this is just a fancy way to encode the concrete definition that all component functions of f (relative to linear coordinizations of V and W) admit continuous second-order partials. Next let us consider the total derivative of $\mathrm{D}f$, that is,

$$\mathrm{D}^2 f = \mathrm{D}(\mathrm{D}f)\colon U \to \mathrm{Hom}(V, \mathrm{Hom}(V,W)).$$

More to the point, how do we work with the vector space $\mathrm{Hom}(V,\mathrm{Hom}(V,W))$? I claim that it is not nearly as complicated as it may seem, and that once we understand how to think about this iterated Hom-space we will see that the theory of higher-order partials admits a very pleasing reformulation in the language of multilinear mappings. The underlying mechanism is a certain isomorphism in linear algebra, so we now digress to discuss the algebraic preliminaries in a purely algebraic setting over any field.

Definition 1.1. In general, for an integer $p \geq 1$ we say that $f\colon U \to W$ is a C^p **map**, or is p **times continuously differentiable**, if it is differentiable and $\mathrm{D}f\colon U \to \mathrm{Hom}(V,W)$ is a C^{p-1} map. If f is a C^p map for every p, we shall say that f is a C^∞ **map**, or is **infinitely differentiable**.

2 Higher Derivatives as Symmetric Multilinear Maps

Let V and W be finite-dimensional vector spaces over $\mathbb{R}$, and let U be open in V. Let $f : U \to W$ be a map of sets. We say f is a C^0 **map** if it is continuous. We have seen above that f is differentiable with

$$Df : U \to \mathrm{Hom}(V,W)$$

continuous if and only if, with respect to a choice of linear coordinates, the components f_j of f admit continuous first-order partial derivatives across all of U with respect to the coordinates on V. This property of f is called being a C^1 **map**, and we may rephrase it as the property that f is differentiable and Df is continuous. We now make a recursive definition:

Definition 2.1. In general, for an integer $p \geq 1$ we say that $f : U \to W$ is a C^p **map**, or is p **times continuously differentiable**, if it is differentiable and

$$Df : U \to \mathrm{Hom}(V,W)$$

is a C^{p-1} map. If f is a C^p map for every p, we shall say that f is a C^∞ **map**, or is **infinitely differentiable**.

Assuming f is C^2, we write $D^2 f(u)$ to denote $D(Df)(u)$, and by definition since $Df : U \to \mathrm{Hom}(V,W)$ is a differentiable map from an open in V to the vector space $\mathrm{Hom}(V,W)$, we see that $D^2 f(u)$ is a linear map from V to $\mathrm{Hom}(V,W)$. That is, we have

$$D^2 f : U \to \mathrm{Hom}(V, \mathrm{Hom}(V,W))$$

and this is continuous (as f is C^2). More generally, if f is C^p, then for $i \leq p$ we write $D^i f = D(D^{i-1} f)$, and arguing recursively we see that $D^p f(u)$ is a linear map from V to $\mathrm{Hom}(V, \mathrm{Hom}(V, \ldots, \mathrm{Hom}(V,W)\ldots))$ where there are $p-1$ iterated Hom's. That is, we have

$$D^p f : U \to \mathrm{Hom}(V, \mathrm{Hom}(V, \ldots, \mathrm{Hom}(V,W)\ldots)) \simeq \mathrm{Mult}(V^p, W).$$

Theorem 2.1. *Suppose $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$. Let $U \subseteq V$ be open an let $f_i : U \to \mathbb{R}$ denote the ith component of f, so f is described as a map $f = (f_1, \ldots, f_m) : U \to \mathbb{R}^m = W$. Let $p \geq 0$ be a non-negative integer. Then f is a C^p map if and only if all p-fold iterated partial derivatives of the f_i's exist and are continuous on U. Likewise, f is C^∞ if and only if all f_i's admit all iterated partials of all orders.*

Proof. We induct on p, the case $p = 0$ being the old result that a map f to a product space is continuous if and only if its component maps f_i are continuous. For $p = 1$, the theorem is our earlier observation that f is differentiable with $Df : U \to \mathrm{Hom}(V,W)$ continuous if and only if the component functions f_i of f admit continuous first-order partials.

Now we assume $p > 1$, so in either direction of implication in the theorem we know (from the C^1 case which has been established) that f admits a continuous derivative map Df and that all partials $\partial_{x_j} f_i$ exist as continous functions on U. Also, we know that the map

$$Df : U \to \mathrm{Hom}(V,W) \simeq \mathrm{Mat}_{m \times n}(\mathbb{R})$$

to the vector space of $m \times n$ matrices has as its component functions (i.e. "matrix entries") precisely the first-order partials $\partial_{x_j} f_i : U \to \mathbb{R}$.

By definition, f is C^p if and only if Df is C^{p-1}, but since this latter map has the $\partial_{x_j} f_i$'s as its component functions, by the inductive hypothesis applied to Df (with the target vector space now $\mathrm{Hom}(V,W)$ rather than

W, and linear coordinates given by matrix entries), it follows that Df is C^{p-1} if and only if all $\partial_{x_j} f_i$'s admit all $(p-1)$-fold iterated partial derivatives in the linear coordinates on V and that these are continuous. Since an arbitrary $(p-1)$-fold partial of an arbitrary first order partial $\partial_{x_j} f_i$ is nothing more or less than an arbitrary p-fold partial of f_i with respect to the linear coordinates on V, we conclude that f is C^p if and only if all p-fold partials of all f_i's with respect to the linear coordinates on V exist and are continuous. $\square$

Let $f : U \to W$ be a C^p mapping with $p \geq 1$, and consider the continuous pth derivative mapping

$$D^p f : U \to \mathrm{Mult}(V^p, W).$$

We want to describe this in terms of partial derivatives using linear coordinates on V and W. That is, we fix ordered bases $\{e_1, \ldots, e_n\}$ of V and $\{w_1, \ldots, w_m\}$ of W, so for each $u \in U$ the multilinear mapping

$$D^p f(u) : V^p \to W = \mathbb{R}^m$$

is uniquely determined by the m-tuples

$$D^p f(u)(e_{j_1}, \ldots, e_{j_p}) \in W = \mathbb{R}^m$$

for $1 \leq j_1, \ldots, j_p \leq n$. What are the m components of this vector in $\mathbb{R}^m$? The answer is very nice:

Theorem 2.2. *With notation as above, let $x_1, \ldots, x_n \in V^\vee$ be the dual basis to the basis $\{e_1, \ldots, e_n\}$ of V. Let ∂_j denote ∂_{x_j}, and let $f_1, \ldots, f_m$ be the component functions of $f : U \to W$ with respect to the basis of w_i's of W. For $1 \leq j_1, \ldots, j_p \leq n$,*

$$D^p f(u)(e_{j_1}, \ldots, e_{j_p}) = ((\partial_{j_p} \cdots \partial_{j_1} f_1)(u), \ldots, (\partial_{j_p} \cdots \partial_{j_1} f_m)(u)) \in \mathbb{R}^m.$$

Proof. We induct on p, the base case $p = 1$ being the old theorem on the determination of the matrix for the derivative map $Df(u) : V \to W$ in terms of first-order partials of the component functions for f (using linear coordinates on W to define these component functions, and using linear coordinates on V to define the relevant partial derivative operators on these functions). Now we assume $p \geq 2$.

By definition of the isomorphism in Corollary (**??**), we have

$$D^p f(u)(v_1, \ldots, v_p) = (\cdots((D^p f(u)(v_1))(v_2)\cdots)(v_p) \in W$$

for any ordered p-tuple $v_1, \ldots, v_p \in V$. Let $F = Df : U \to \mathrm{Hom}(V, W)$. Using the given linear coordinates on V and W, the associated "matrix entries" are taken as the linear coordinates on $\mathrm{Hom}(V, W)$ to get component functions F_{ij} for F (with $1 \leq i \leq m$ and $1 \leq j \leq n$). Considering $v_2, \ldots, v_p$ as fixed but v_1 as varying, we have

$$D^p f(u)(\cdot, v_2, \ldots, v_p) = (\cdots((D^{p-1}F)(u)(v_2))\cdots)(v_p) = D^{p-1}F(u)(v_2, \ldots, v_p) \in \mathrm{Hom}(V, W)$$

where $\mathrm{Hom}(V, W)$ is the target vector space for F. Setting $v_k = e_{j_k}$ for $2 \leq k \leq p$, the inductive hypothesis applied to $F : U \to \mathrm{Hom}(V, W) = \mathrm{Mat}_{m\times n}(\mathbb{R})$ gives

$$D^{p-1}F(u)(e_{j_2}, \ldots, e_{j_p}) = (\partial_{j_p} \cdots \partial_{j_2} F_{ij}(u)) \in \mathrm{Mat}_{m\times n}(\mathbb{R}).$$

In view of how the matrix coordinatization of $\mathrm{Hom}(V, W)$ was *defined* using the chosen ordered bases on V and W, evaluating e_{j_1} in $\mathrm{Hom}(V, W) \simeq \mathrm{Mat}_{m\times n}(\mathbb{R})$ corresponds to pass to the j_1th column of a matrix. Hence taking $v_1 = e_{j_1}$ gives

$$D^p f(u)(e_{j_1}, e_{j_2}, \ldots, e_{j_p}) = (\partial_{j_p} \cdots \partial_{j_2} F_{1j_1}(u), \ldots, \partial_{j_p} \cdots \partial_{j_2} F_{mj_1}(u)) \in \mathbb{R}^m = W.$$

By the C^1 case, $F = Df : U \to \mathrm{Hom}(V, W) = \mathrm{Mat}_{m\times n}(\mathbb{R})$ has ij-component function $F_{ij} = \partial_j f_i$, so $F_{ij_1} = \partial_{j_1} f_i$. Thus, we get the desired formula. $\square$

Example 2.1. Suppose $W = \mathbb{R}$ and let $x_1, \ldots, x_m$ be linear coordinates on V relative to some ordered basis $\boldsymbol{e} = (e_1, \ldots, e_m)$ on V. Then $\mathrm{D}^2 f(u)$ is identified with the **Hessian** of f at u:

$$\mathrm{D}^2 f(u) = \mathrm{H}_f(u) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(u)\right).$$

Hence, the Hessian that appears in the second derivative test in several variables is *not* a linear map (as might be suggested by its traditional presentation as a matrix) but rather is intrinsically seen to be a symmetric bilinear form.

3 Higher-Dimensional Taylor's Formula: Motivation and Preparations

As an application of the formalism of higher derivatives as multilinear mappings, we wish to state and prove Taylor's formula (with an integral remainder term) for C^α maps $f : U \to W$ on any open $U \subseteq V$. In the special case $V = W = \mathbb{R}$ and U a non-empty open interval, this will recover the usual Taylor formula from calculus. There is also a more traditional version of the multivariable Taylor formula given with loads of mixed partials and factorials, and we will show that this traditional version is equivalent to the version we will prove in the language of higher derivatives as multilinear mappings. The power of our point of view is that it permits one to give a proof of Taylor's formula that is virtually identical in appearance to the proof in the classical case (with $V = W = \mathbb{R}$); proofs of Taylor's formula in the classical language of mixed partials tend to become a big mess with factorials, and the integral formula and error bound for the remainder term are unpleasant to formulate in the classical language.

Before we state the general case, let us first recall the 1-variable Taylor formula for a C^p function $f\colon I \to \mathbb{R}$ on an interval $I \subseteq \mathbb{R}$ with $a \in I$ an interior point: for $|h|$ sufficiently small so that $(a-h, a+h) \in I$ we have

$$f(a+h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \cdots + \frac{f^{(p)}(a)}{p!}h^p + R_{p,a}(h)$$

with error term $R_{p,a}$ is given by

$$R_{p,a}(h) = \int_0^1 \frac{f^{(p)}(a+th) - f^{(p)}(a)}{(p-1)!} \cdot (1-t)^{p-1} h^p \mathrm{d}t = h^p \psi_{p,a}(h),$$

where $|\psi_{p,a}(h)|$ can be made below any desired ε for h near 0 (uniformly for a in a compact subinterval of I) since the continuous $f^{(p)}$ is uniformly continuous on compacts in I. In particular, as $h \to 0$ we have $|R_{p,a}(h)|/|h|^p \to 0$ *uniformly* for a in a compact subinterval of I.

4 Taylor's Formula: Statement and Proof

Let V and W be finite-dimensional $\mathbb{R}$-vector spaces, U and open subset in V, and $f : U \to W$ a C^p map with $p \geq 1$. We choose $a \in U$ and $r > 0$ such that $B_r(a) \subseteq U$ (relative to an arbitrary but fixed choice of norm on V). Thus, $f(a+h)$ makes sense for $h \in V$ satisfying $\|h\| < r$. Now we can state and prove Taylor's formula by essentially just copying the proof from calculus!

Theorem 4.1. *With the notation as above,*

$$f(a+h) = \sum_{j=0}^{p} \frac{(\mathrm{D}^j f)(a)}{j!}(h^{(j)}) + R_{p,a}(h) \tag{3}$$

in W, where

$$R_{p,a}(h) = \int_0^a \frac{(1-t)^{p-1}}{(p-1)!} \left((\mathrm{D}^p f)(a+th) - (\mathrm{D}^p f)(a)\right)(h^{(p)})\mathrm{d}t$$

satisfies

$$\|R_{p,a}(h)\| \leq C_{p,h,a}\|h\|^p \text{ and } \lim_{h\to 0} C_{p,h,a} = 0 \tag{4}$$

with

$$C_{p,h,a} = \sup_{t\in[0,1]} \frac{\|(\mathrm{D}^p f)(a+th) - (\mathrm{D}^p f)(a)\|}{p!}.$$

The convergence $C_{p,h,a} \to 0$ as $h \to 0$ is uniform for a supported in a compact subset of U.

One important consequence of the error estimate (4) is that it shows the error $R_{p,a}(h)$ in the "degree p" expansion (3) of $f(a+h)$ about a dies off more rapidly than $\|h\|^p$ as $h \to 0$, that is $\|R_{p,a}(h)\|/\|h\|^p \to 0$ as $h \to 0$ with the rate of such decay actually *uniform* for a supported in a fixed compact subset of U. This is tremendously important for some applications.

A particular important case is $p = 2$: the approximation

$$f(a+h) = f(a) + (\mathrm{D}f(a))(h) + (\mathrm{D}^2 f(a))(h,h) + (\dots)$$

has an error which dies more rapidly than $\|h\|^2$. This is what underlies the reason why the symmetric bilinear Hessian $\mathrm{H}_f(a) = (\mathrm{D}^2 f)(a)$ governs the structure of f near critical points (that is those with $\mathrm{D}f(a) = 0$, such as local extrema) in the case when $W = \mathbb{R}$. That is, the signature of the quadratic form associated to $\mathrm{H}_f(a)$ encodes much of the local geometry for f near a when $\mathrm{D}f(a) = 0$.