Math 396. Higher derivatives and Taylor's formula via multilinear maps

Let V and W be finite-dimensional vector space over $\mathbf{R}$, and $U \subseteq V$ an open subset. Recall that a map $f : U \to W$ is *differentiable* if and only if for each $u \in U$ there exists a (necessarily unique) linear map $Df(u) : V \to W$ such that

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

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 $\mathbf{R}$-vector space are bounded by a constant positive multiple of each other). We want to interpret the theory of higher derivatives in the language of multilinear mappings. From this vantage point we will acquire an insight into the true meaning of the equality of mixed higher partial derivatives, and moreover Taylor's formula will take on its most natural formulation in a manner that looks virtually identical to the classical version in calculus. (In contrast, the usual coordinatized presentation of Taylor's formula that involves hoardes of factorials and indices; we will derive the classical presentation from the "clean" version that we prove.)

1. Some motivation

If U is open in $\mathbf{R}^n$ and $f : U \to \mathbf{R}^m$ is a map, there's a down-to-earth way to define what it means to say that f is p-times continuously differentiable: for each component function f_i of f $(1 \leq i \leq m)$ all p-fold iterated partial derivatives

$$\partial_{x_{j_1}} \dots \partial_{x_{j_p}} f_i : U \to \mathbf{R}$$

should exist and be continuous on U, for arbitrary collections (with repetition allowed) of p indices $1 \leq j_1, \dots, j_p \leq n$. If this holds for all $p \geq 1$, we say f is a C^∞ map. We want to give an alternative definition that does not require coordinates and is better-suited to giving a clean statement and proof of the multivariable Taylor formula.

To get started, let V and W be finite-dimensional vector spaces over $\mathbf{R}$ and let $f : U \to W$ be a map on an open subset $U \subseteq V$. If f is differentiable, then for each $u \in U$ we get a linear map $Df(u) : V \to W$. Hence, we get a map

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

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. By avoiding coordinates it will be easier to focus on the underlying linear structure of the target and to not put too much emphasis on whether the target is a space of column vectors, matrices, and so on.

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

Lemma 1.1. *Suppose $f : U \to W$ is a C^1 map, and let $Df : U \to \mathrm{Hom}(V, W)$ be the associated total derivative map. As a map from an open set in V to a finite-dimensional vector space, Df is C^1 if and only if (relative to a choice of linear coordinates on V and W) all second-order partials $\partial_{x_{j_1}} \partial_{x_{j_2}} f_i : U \to \mathbf{R}$ exist and are continuous.*

Proof. Fixing linear coordinates identifies Df with a map from an open set $U \subseteq \mathbf{R}^n$ to a Euclidean space of $m \times n$ matrices, with component functions $\partial_{x_j} f_i$ 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_i's admit all second-order partials and that such partials are continuous. ■

Let us say that $f : U \to W$ is C^2 when it is differentiable and $Df : U \to \mathrm{Hom}(V, W)$ is C^1. By the lemma, this is just a fancy way to encode the concrete condition that all component functions of f (relative to linear coordinatizations of V and W) admit continuous second-order partials. What sort of structure is the total derivative of Df? That is, what sense can we make of the map

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

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.

2. Hom-spaces and multilinear mappings

Let V, V', W be arbitrary vector spaces over a field F. (We will only need the finite-dimensional case, but such a restriction is not relevant for what we are about to do on the algebraic side.) We want to establish some general relationships between spaces of bilinear (and more generally, multilinear) mappings. This will help us to think about the higher-dimensional "total second derivative" map as a symmetric bilinear pairing (the symmetry encoding the equality of mixed second-order partials for a C^2 function), and similarly for higher-order derivatives using multilinear mappings.

Let $\mathrm{Bil}_F(V \times V', W)$ denote the set of F-bilinear maps $V \times V' \to W$, endowed with its natural F-vector space structure (i.e., using the linear structure on W we can add such pairings, and multiply them by scalars, in the usual pointwise manner on $V \times V'$). Suppose we're given a bilinear map

$$\psi : V \times V' \to W.$$

Thus, for each fixed $v \in V$ we get a linear map

$$\psi(v, \cdot) : V' \to W$$

given by $v' \mapsto \psi(v, v')$. In other words, we have a set-theoretic map

$$V \to \mathrm{Hom}_F(V', W)$$

which assigns to any $v \in V$ the linear map $\psi(v, \cdot)$.

We wish to show that this set map $V \to \mathrm{Hom}_F(V', W)$ is also *linear*, and hence gives rise to a general construction assigning to each bilinear $\psi : V \times V' \to W$ an element in

$$\mathrm{Hom}_F(V, \mathrm{Hom}_F(V', W)).$$

That is, we'll have constructed a map of sets

$$\mathrm{Bil}_F(V \times V', W) \to \mathrm{Hom}_F(V, \mathrm{Hom}_F(V', W)).$$

In fact, we'll even show that this latter map is linear. And we'll go one step further and show it is a linear isomorphism.

Who cares? Why would anyone want to deal with such a bewildering array of complexity? The point is this. In the higher-dimensional theory of the second derivative, we have seen the intervention of

$$\mathrm{Hom}_F(V, \mathrm{Hom}(V, W))$$

(with $F = \mathbf{R}$). But this may seem to be a hard thing to internalize. What does it mean? The construction we outlined above will essentially lead us to the fact that this iterated Hom-space can be naturally (and linearly) identified with the space

$$\mathrm{Bil}_F(V \times V, W)$$

of bilinear pairings on V with values in W. This latter point of view is much more tractable and geometric, as we shall see, and leads to the higher-dimensional second derivative test. Thus, the ability to repackage $\mathrm{Hom}_F(V, \mathrm{Hom}_F(V', W))$ into a space of bilinear maps is quite convenient in various contexts.

We claim that the map

$$\begin{array}{rcl} \xi : \mathrm{Hom}_F(V, \mathrm{Hom}_F(V', W)) & \to & \mathrm{Bil}_F(V \times V', W) \\ \varphi & \mapsto & ((v, v') \mapsto (\varphi(v))(v')) \end{array}$$

is linear and the map

$$\begin{array}{rcl} \eta : \mathrm{Bil}_F(V \times V', W) & \to & \mathrm{Hom}_F(V, \mathrm{Hom}_F(V', W)) \\ \psi & \mapsto & (v \mapsto (v' \mapsto \psi(v, v'))) \end{array}$$

is also linear, and that these are inverse to each other (and hence are linear isomorphisms). This provides the explicit mechanism for translating into the language of bilinear pairings. Part of the work necessary to verify these claims is to see that things live where they should (e.g., in the first map ξ, we must check that the expression $\varphi(v)(v')$ is actually *bilinear* in v and v').

We first check that $\xi(\varphi) : V \times V' \to W$ is actually bilinear. That is, we want

$$(v, v') \mapsto (\varphi(v))(v')$$

to be bilinear, which is to say that it should be linear in v for fixed v' and it should be linear in v' for fixed v. Since $\varphi(v) \in \mathrm{Hom}_F(V', W)$ is a *linear* map, this gives exactly the linearity in v' for fixed v. Meanwhile, if v' is fixed that since $v \mapsto \varphi(v)$ is linear (by the very definition of the Hom-space in which φ lives!) we have

$$\varphi(c_1 v_1 + c_2 v_2) = c_1 \varphi(v_1) + c_2 \varphi(v_2)$$

in $\mathrm{Hom}_F(V', W)$. Now evaluating both sides on $v' \in V'$ and recalling what it *means* to add and scalar multiply in $\mathrm{Hom}_F(V', W)$ yields

$$\varphi(c_1 v_1 + c_2 v_2)(v') = c_1 \varphi(v_1)(v') + c_2 \varphi(v_2)(v').$$

This is exactly the linearity in v for fixed v'. This completes the verification that $\xi(\varphi)$ lives in the asserted space of bilinear maps.

Now that ξ makes sense, let's check it is linear. We want $\xi(c\varphi + c'\varphi') = c\xi(\varphi) + c'\xi(\varphi')$ in $\mathrm{Bil}_F(V \times V', W)$. We simply evaluate on a pair $(v, v') \in V \times V'$:

$$\begin{aligned}
\xi(c\varphi + c'\varphi')(v, v') &= ((c\varphi + c'\varphi')(v))(v') \\
&= (c(\varphi(v)) + c'(\varphi'(v)))(v') \\
&= c(\varphi(v))(v') + c'(\varphi'(v))(v') \\
&= c\xi(\varphi)(v, v') + c'\xi(\varphi')(v, v') \\
&= (c\xi(\varphi) + c'\xi(\varphi'))(v, v')
\end{aligned}$$

so the linearity of ξ follows.

Now we turn to study η. Pick $\psi \in \mathrm{Bil}_F(V \times V', W)$. For $v \in V$, the map $v' \mapsto \psi(v, v')$ is linear in v' (as ψ is bilinear), so in other words $f_v : v' \mapsto \psi(v, v')$ is an element in $\mathrm{Hom}_F(V', W)$. Now we check that the map $V \to \mathrm{Hom}_F(V', W)$ defined by $v \mapsto f_v$ is actually linear, so indeed $\eta(\psi)$ will make sense as an element in $\mathrm{Hom}_F(V, \mathrm{Hom}_F(V', W))$. In order to verify the linearity of $v \mapsto f_v$, we must show

$$f_{c_1 v_1 + c_2 v_2} = c_1 f_{v_1} + c_2 f_{v_2}$$

inside of $\mathrm{Hom}_F(V', W)$. That is, upon evaluating both sides at an arbitrary $v' \in V'$, we want (using the definition of f_v)

$$\psi(c_1 v_1 + c_2 v_2, v') \stackrel{?}{=} c_1 f_{v_1}(v') + c_2 f_{v_2}(v') = c_1 \psi(v_1, v') + c_2 \psi(v_2, v').$$

But this follows from the bilinearity of ψ.

With η at least now meaningful, we check it is actually *linear*. That is, we want

$$\eta(c_1 \psi_1 + c_2 \psi_2) = c_1 \eta(\psi_1) + c_2 \eta(\psi_2).$$

In order to check an equality among elements in $\mathrm{Hom}_F(V, \mathrm{Hom}_F(V', W))$, we evaluate on an arbitrary $v \in V$ and hope to get the same result in $\mathrm{Hom}_F(V', W)$ on each side. That is, we want an equality

$$\eta(c_1 \psi_1 + c_2 \psi_2)(v) = c_1 \eta(\psi_1)(v) + c_2 \eta(\psi_2)(v)$$

in $\mathrm{Hom}_F(V', W)$. To check an equality in here, we evaluate each side on an arbitrary $v' \in V'$ and hope to get the same result on each side. That is, we want an equality

$$(c_1 \psi_1 + c_2 \psi_2)(v, v') = c_1 \psi_1(v, v') + c_2 \psi_2(v, v').$$

But this is how the linear structure on $\mathrm{Bil}_F(V \times V', W)$ is *defined*!

Having done the exhaustive work to check that ξ and η are meaningful and linear, the fact that they're inverses is actually a lot easier, since it is just a set-theoretic problem of computing the composites. First we consider $\xi \circ \eta$. Evaluating on a bilinear map $\psi : V \times V' \to W$, we want $\xi(\eta(\psi)) = \psi$, or in other words

$$\xi(\eta(\psi))(v, v') = \psi(v, v')$$

for all $v \in V$, $v' \in V'$. By unwinding the definitions of ξ and η, we compute

$$\xi(\eta(\psi))(v, v') = (\eta(\psi)(v))(v') = \psi(v, v'),$$

as desired. To compute the composite the other way around, we choose $\varphi \in \mathrm{Hom}_F(V, \mathrm{Hom}_F(V', W))$ and we want $\eta(\xi(\varphi)) = \varphi$. Evaluating both sides on $v \in V$, we want

$$(\xi(\varphi))(v, \cdot) = \varphi(v)$$

in $\mathrm{Hom}_F(V', W)$. Evaluating on an arbitrary $v' \in W$, we want

$$\xi(\varphi)(v, v') = (\varphi(v))(v').$$

But this is true by the very definition of the bilinear form $\xi(\varphi)$!

Having succeeded in identifying $\mathrm{Hom}_F(V, \mathrm{Hom}_F(V', W))$ with $\mathrm{Bil}(V \times V', W)$, we want to push this further. For any positive integer d and vector spaces $V_1, \dots, V_d$, and W over F, we will now define a natural F-*linear* isomorphism

$$\mathrm{Hom}_F(V_1, \mathrm{Hom}_F(V_2, \mathrm{Hom}_F(V_3, \dots, \mathrm{Hom}_F(V_d, W) \dots))) \simeq \mathrm{Mult}(V_1 \times \cdots \times V_d, , W)$$

to the space of multilinear mappings

$$\mu : V_1 \times \cdots \times V_d \to W$$

(made into an F-vector space just as in the case of $\mathrm{Mult}(V_1 \times V_2, W) = \mathrm{Bil}(V_1 \times V_2, W)$ for $d = 2$, namely via pointwise operations on $V_1 \times \cdots \times V_d$ and the linear structure on W). In the case $d = 2$ we will recover the preceding considerations in the bilinear setting. Much of this is quite similar to what was just done, so we won't write out the complete details of a mechanical argument which was worked out fully above when it essentially carries over *verbatim* in the new situation.

Theorem 2.1. *With notation as introduced above, there is a natural linear isomorphism*

$$\xi : \mathrm{Hom}_F(V_1, \mathrm{Hom}_F(V_2, \mathrm{Hom}_F(V_3, \dots, \mathrm{Hom}_F(V_d, W) \dots))) \simeq \mathrm{Mult}(V_1 \times \cdots \times V_d, , W).$$

Proof. We define ξ as follows: for $\varphi \in \mathrm{Hom}_F(V_1, \mathrm{Hom}_F(V_2, \mathrm{Hom}_F(V_3, \dots, \mathrm{Hom}_F(V_d, W) \dots)))$, $\xi(\varphi) \in \mathrm{Mult}(V_1 \times \cdots \times V_d, , W)$ is the multilinear mapping

$$(\xi(\varphi))(v_1, \dots, v_d) = (\dots((\varphi(v_1))(v_2))(\dots))(v_d) \in W.$$

If $d = 2$, this recovers our construction in the bilinear case: $(\xi(\varphi))(v_1, v_2) = (\varphi(v_1))(v_2)$. If $d = 3$ then

$$(\xi(\varphi))(v_1, v_2, v_3) = ((\varphi(v_1))(v_2))(v_3) \in W.$$

Make sure you understand the definition of ξ in these special cases, and then the general case, before reading further.

The definition of $\xi(\varphi)$ is multilinear in any one of $v_1, \dots, v_d$ with all other v_j's held fixed; the argument goes just as in the previous section. Work out the case $d = 3$ (and review the case $d = 2$) to see what is going on, and then you'll see the general pattern. Thus, $\xi(\varphi)$ makes sense as a multilinear mapping from V^d into W. This defines ξ as a map of sets. To see that ξ is actually linear in φ amounts to exactly the same style of calculation we did earlier, once one sees that the collection of vectors $v_1, \dots, v_d$ essentially moves through the calculation rather formally. We leave this to the reader (and again suggest both reviewing the case $d = 2$ and working out $d = 3$ by hand to see the pattern in the argument).

Next we define an inverse

$$\eta : \mathrm{Mult}(V_1 \times \cdots \times V_d, , W) \to \mathrm{Hom}_F(V_1, \mathrm{Hom}_F(V_2, \mathrm{Hom}_F(V_3, \dots, \mathrm{Hom}_F(V_d, W) \dots))).$$

For a multilinear ψ, we define $\eta(\psi)$ by the rule

$$(\dots(((\eta(\psi))(v_1))(v_2))\dots)(v_d) = \psi(v_1, v_2, \dots, v_d).$$

To see that this makes sense, let us focus on the case $d = 3$. In this case, for a fixed $v_1 \in V_1$, the linear mapping $(\eta(\psi))(v_1) \in \mathrm{Hom}_F(V_2, \mathrm{Hom}_F(V_3, W))$ is the map sending a fixed $v_2 \in V_2$ to the element of $\mathrm{Hom}_F(V_3, W)$ given by $v_3 \mapsto \psi(v_1, v_2, v_3)$. When v_1 and v_2 are fixed, this construction is indeed linear in v_3 because ψ is linear in v_3 when the other variables are fixed. Moreover, for a fixed v_1 the association $v_2 \mapsto \psi(v_1, v_2, \cdot) \in \mathrm{Hom}_F(V_3, W)$ really is linear in v_2 because ψ is linear in v_2 when the other variables are fixed, and because the linear structure on the target is defined by pointwise operations. (This goes as in our earlier direct treatment of the bilinear case $d = 2$.) Finally, it follows from linearity of ψ in the first variable when the others are fixed that

$$v_1 \mapsto (v_2 \mapsto \psi(v_1, v_2, \cdot)) \in \mathrm{Hom}_F(V_2, \mathrm{Hom}_F(V_3, W))$$

is itself linear in v_1. That is, $\eta(\psi)$ really is a linear map from V_1 to $\mathrm{Hom}_F(V_2, \mathrm{Hom}_F(V_3, W))$, so for $d = 3$ the map η makes sense as a set-theoretic map from a space of multilinear maps to an interated Hom-space. Once you understand the case $d = 3$, you should be able to see that the general case works in exactly the same manner.

With η at least defined now as a map of sets, we have to check it is *linear*. Here again the argument is basically identical to what we did in the case $d = 2$ because of how the linear structure on spaces of multilinear mappings is *defined*. The fact that ξ and η are inverse to each other is exactly the same calculation that we did in the case $d = 2$. The details are left to the reader. ■

Setting $V_1 = \cdots = V_d$, we get:

Corollary 2.2. *Let V and W be vector spaces over a field. For any positive integer d, there is a natural linear isomorphism*

$$\mathrm{Hom}_F(V, \mathrm{Hom}_F(V, \dots, \mathrm{Hom}_F(V, W) \dots)) \simeq \mathrm{Mult}_F(V^d, W)$$

(*with d iterated* Hom*'s*) *to the space of multilinear mappings $V \times \cdots \times V \to W$ on a d-fold product of copies of V.*

3. Higher derivatives as symmetric multilinear mappings

This section is partly a review of the theory of higher derivatives from Math 296, and also a reformulation of that theory in terms of multilinear mappings. Let V and W be finite-dimensional vector spaces over $\mathbf{R}$, 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_i of f admit continuous first-order partials 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 3.1. 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.*

Even if we are ultimately most interested in C^∞ mappings, the merit of first working in the context of C^p maps with finite p is that it opens the door to the possibility of doing proofs by induction on $p \geq 0$ (something we can't do if we insist on working from the start with only the concept of an infinitely differentiable map).

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 $i \le p$ we write $D^i f = D(D^{i-1}f)$, and arguing recursively (check low-degree cases by hand) we see that $D^p f(u)$ is a linear map from V to $\mathrm{Hom}(V, \mathrm{Hom}(V, \dots, \mathrm{Hom}(V, W) \dots))$ where there are $p-1$ iterated Hom's. That is, we have

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

where there are p iterated Hom's. By Corollary 2.2, we may rewrite this:

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

It behooves us to now relate our abstract definition of C^p maps with a more down-to-earth one in terms of iterated partials. We begin with:

Theorem 3.2. *Suppose $V = \mathbf{R}^n$ and $W = \mathbf{R}^m$. Let $U \subseteq V$ be open and let $f_i : U \to \mathbf{R}$ denote the ith component of f, so f is described as a map $f = (f_1, \dots, f_m) : U \to \mathbf{R}^m = W$. Let $p \ge 0$ 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.*

The criterion in this theorem is exactly the traditional definition of a C^p map for $1 \le p \le \infty$.

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 continuous functions on U. Also, we know that the map

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

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

By definition, f is C^p if and only if the map 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 if 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 all 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. ■

Let $f : U \to W$ be a C^p mapping with $p \ge 1$, 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 fixed ordered bases $\{e_1, \dots, e_n\}$ of V and $\{w_1, \dots, w_m\}$ of W, so for each $u \in U$ the multilinear mapping

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

is uniquely determined by the m-tuples

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

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

Theorem 3.3. *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 $1 \le j_1, \ldots, j_p \le n$,*

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

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 $p \ge 2$.

By definition of the isomorphism in Corollary 2.2, which is how we identify $D^p f(u)$ with a multilinear map $V^p \to W$, we have

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

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 $1 \le i \le m$ and $1 \le j \le n$). Considering $v_2, \ldots, v_p$ as fixed but v_1 as varying, we have

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

where $\mathrm{Hom}(V, W)$ is the target vector space for F. (Make sure you understand this displayed equation.) Setting $v_k = e_{j_k}$ for $2 \le k \le p$, the inductive hypothesis applied to $F : U \to \mathrm{Hom}(V, W) = \mathrm{Mat}_{m \times n}(\mathbf{R})$ gives

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

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

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

By the C^1 case, $F = Df : U \to \mathrm{Hom}(V, W) = \mathrm{Mat}_{m \times n}(\mathbf{R})$ 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. ■

By induction on the positive integer p, the theorem on equality of mixed partials for C^2 functions gives the general equality of mixed p-fold partials for C^p functions f on U: for an ordered p-tuple $(j_1, \ldots, j_p)$ of integers between 1 and n the function $\partial_{x_{j_1}} \ldots \partial_{x_{j_p}} f : U \to \mathbf{R}$ is insensitive to permutation of the j_k's. This has the following beautiful coordinate-free interpretation:

Corollary 3.4. *For any positive integer p and any C^p map $f : U \to W$ on an open subset $U \subseteq V$, the pth total derivative map $D^p f : U \to \mathrm{Mult}(V^p, W)$ takes values in the subspace of symmetric multilinear maps. That is, for $u \in U$ and $v_1, \ldots, v_p \in V$, $D^p f(u)(v_1, \ldots, v_p) \in W$ is invariant under permutation of the v_j's.*

In particular, if $f : U \to W$ is C^∞ then for all positive integers p, the multilinear mapping $D^p f(u) : V^p \to W$ is symmetric for all $u \in U$.

Proof. The symmetry assertion is that for any $\sigma \in \mathfrak{S}_p$ and $u \in U$,

$$D^p f(u)(v_1, \ldots, v_p) = D^p f(u)(v_{\sigma(1)}, \ldots, v_{\sigma(p)})$$

in W for any $v_1, \ldots, v_p \in V$. Both sides of the equality are multilinear in $(v_1, \ldots, v_p) \in V^p$, and so to check the equality it suffices to check when the v_j's are taken from a fixed basis $\{e_1, \ldots, e_n\}$ of

V. Taking $v_k = e_{j_k}$ for $1 \leq j_1, \dots, j_p \leq n$ and inspecting the proposed equality in each component of the vector values in $W \simeq \mathbf{R}^m$ (using a basis of W), Theorem 3.3 reduces the proposed equality to the known symmetry of p-fold mixed partials on C^p functions. ■

Example 3.5. When $p = 2$, $W = \mathbf{R}$, and $\dim V = n$ with linear coordinates $x_1, \dots, x_n$ on V relative to some ordered basis $\mathbf{e} = \{e_1, \dots, e_n\}$, then the *Hessian* $H_f(u) = (D^2 f)(u)$ is the symmetric bilinear form on V (or equivalently, self-dual map $V \to V^\vee$) represented by the matrix

$$_{\mathbf{e}^*}[H_f(a)]_{\mathbf{e}} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right).$$

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 (and the traditional matrix is thereby seen to be an instance of the description of symmetric bilinear forms in terms of matrices for self-dual mappings $V \to V^\vee$).

We now apply the Hessian to give the higher-dimensional "second derivative" test:

Theorem 3.6. *Let $U \subseteq V$ be an open set and let $f : U \to \mathbf{R}$ be a C^2 function. Suppose that a is a critical point for f in the sense that $Df(a) = 0$ for some $a \in U$. Let $H_f(a) : V \times V \to \mathbf{R}$ be the symmetric bilinear Hessian $D^2 f(a)$, and let $q_{f,a} : V \to \mathbf{R}$ be the associated quadratic form. If $H_f(a)$ is non-degenerate then f has an isolated local minimum at a when $q_{f,a}$ is positive-definite, an isolated local maxiumum at a when $q_{f,a}$ is negative-definite, and neither a local minimum nor maximum in the indefinite case.*

In the 1-dimensional case this recovers the usual second derivative test from calculus at critical points: the non-degeneracy condition specializes to the 1-variable condition $f''(a) \neq 0$, and that leaves only the positive-definite and negative-definite cases. The remaining possibities are so-called "saddle points" and are a strictly higher-dimensional phenomenon. As with the second derivative test in calculus, when $H_f(a)$ is not non-degenerate, no inference can be made. (Consider $f(x, y) = \pm(x^4 + y^4)$, for which $Df(0, 0) = 0$ and $D^2 f(0, 0) = 0$.)

Proof. Replacing f with $f - f(a)$, we may assume $f(a) = 0$. By Taylor's formula in the higher-dimensional formulation given in Theorem 5.1, for small h we have $f(a + h)/\|h\|^2 = H_f(a)(\widehat{h}, \widehat{h}) + R_a(h) = q_{f,a}(\widehat{h}) + R_a(h)$ where $R_a(h) \to 0$ as $h \to 0$ and $\widehat{h} = h/\|h\|$ is a unit vector (with respect to the norm) pointing in the same direction as h. (Here we have fixed an arbitrary norm on V). Thus, $f(a + h)/\|h\|^2$ is approximated by $q_{f,a}(\widehat{h})$ up to an error that tends to 0 locally uniformly in a as $h \to 0$. Provided that $q_{f,a}$ is *non-degenerate*, in the positive-definite case it is bounded below by some $c > 0$ on the unit sphere for the chosen norm, and hence (depending on c) by taking h sufficiently small we get $f(a+h)/\|h\|^2 \geq c/2 > 0$. This shows that f has an isolated local minimum at a, and a similar argument gives an isolated local maximum at a if $q_{f,a}$ is negative-definite.

Now suppose that $q_{f,a}$ is indefinite. By the spectral theorem, if we choose the norm on V to come from an inner product then the pairing $H_f(a)$ is given by the inner product against an orthogonal linear map. Hence, in such cases we can find an orthnormal basis with respect to which $q_{f,a}$ is diagonalized, and so in the indefinite case there are lines on which the restriction of $q_{f,a}$ is positive-definite and there are lines on which the restriction of $q_{f,a}$ is negative-definite. Approaching a along such directions gives different types of behavior for f at a (isolated local minimum when approaching through the positive light cone for $q_{f,a}$, and an isolated local maximum when approaching through the negative light cone for $q_{f,a}$, provided the approach is not tangential to the null cone of vectors $v \in V$ for which $q_{f,a}(v) = 0$). This gives the familiar "saddle point" picture for the behavior of

f, with the shape of the saddle governed by the eigenspace decomposition for the orthogonal map arising from the Hessian $H_f(a)$ and the choice of inner product on V. ■

4. Higher-dimensional Taylor's formula: motivation and preparations

Fix an integer $p \geq 1$. As an application of the formalism of higher derivatives as multilinear mappings, we wish to state and proof Taylor's formula (with an integral remainder term) for C^p maps $f : U \to W$ on any open $U \subseteq V$. In the special case $V = W = \mathbf{R}$ 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 $V = W = \mathbf{R}$); 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 $f : I \to \mathbf{R}$ on an interval $I \subseteq \mathbf{R}$ with $a \in I$ an interior point: for $|h|$ sufficiently small so that $(a - r, a + r) \in I$ we have

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

with error term given by

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

where $|(\ldots)|$ 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.

The remainder formula (1) is related to the version on p. 392 of the Spivak text by means of the change of variable $t \leftrightarrow (t-a)/h$ to convert $\int_a^{a+h}$ into $\int_0^1$. The reason we want to get parameters out of the integral bounds is that we're going to want to be in a situation where a and h will become vectors. Strictly speaking, our integral formula for the remainder is a mild (but useful!) modification on the integral formula in Spivak's book. The point is that Spivak merely gives a Taylor expansion to degree $p-1$ and integrates with $f^{(p)}(a+th)$, whereas we give a Taylor expansion to degree p (which seems fitting for a C^p function!) and in the integral we are subtracting the constant $f^{(p)}(a)$. However, if one directly integrates our subtraction term $f^{(p)}(a)(1-t)^{p-1}h^p/(p-1)!$ one gets $f^{(p)}(a)h^p/p!$, and so this exactly cancels the degree-p Taylor polynomial term that we have inserted. Hence, the formula above really is equivalent to what is in Spivak's *Calculus*.

In order to state a higher-dimensional Taylor formula that resembles the classical case, we need some convenient notation:

Definition 4.1. For symmetric $T \in \mathrm{Mult}(V^p, W)$ and $v \in V$, $T(v^p)$ means $T(v, \ldots, v)$. For $v, v' \in V$ and $0 \leq i \leq p$, $T(v^i, v'^{p-i})$ means $T(v, \ldots, v, v', \ldots, v')$ with v appearing i times and v' appearing $p-i$ times.

In the second piece of notation introduced here, it actually does not matter where the v's and v''s are placed, as T is assumed to be symmetric. The utility of this notation is due to the "binomial formula":

$$T((v+v')^p) = \sum_{i=0}^{p} \binom{p}{i} T(v^i, v'^{p-i}).$$

The proof is the same as for the binomial formula. Indeed, since

$$T((v+v')^p) = T(v+v', \dots, v+v')$$

with p appearances of $v+v'$, if we expand out in each slot via multilinearity of T, we get 2^p terms, each consisting of T evaluated on some set of v's and v''s. The symmetry of T ensures that the value of each term depends only on the number of v's and v''s, and not on their specific positions (this plays the role of the commutative law of multiplication in the usual proof of the binomial theorem), and so $T((v+v')^p)$ is given as a sum of terms $T(v^i, v'^{p-i})$ with $0 \le i \le p$, and the term for a given i shows up as many times as there are i-element subsets of a set of size p. This gives the binomial multiplier coefficients.

In general, the higher-dimensional analogue of the term

$$\frac{f^{(i)}(a)}{i!} \cdot h^i \in \mathbf{R}$$

in Taylor's formula is

$$\frac{((D^i f)(a))}{i!}(h^{(i)}) \in W \tag{2}$$

where

$$h^{(i)} = (h, \dots, h) \in V^i.$$

For example, when $i = 2$, $W = \mathbf{R}$, and V has linear coordinate functions $x_1, \dots, x_n$ relative to some ordered basis $\mathbf{e}$, then

$$\frac{((D^2 f)(a))}{2!}(h,h) = \frac{1}{2!}[h]^t_{\mathbf{e}} \cdot {}_{\mathbf{e}^*}[H_f(a)]_{\mathbf{e}} \cdot [h]_{\mathbf{e}} = \frac{1}{2!}[h]^t_{\mathbf{e}} \cdot \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\right) \cdot [h]_{\mathbf{e}}, \tag{3}$$

where $[h]_{\mathbf{e}}$ is the column vector encoding the expansion of $h \in V$ in the standard $\mathbf{e}$-coordinates (and $[h]^t_{\mathbf{e}}$ is the corresponding transposed row vector). Expanding out the bilinear expression on the right side of (3), the cross-terms

$$h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a)$$

for $i \ne j$ appear twice, cancelling out the 2! in the denominator, so we get

$$\frac{((D^2 f)(a))}{2!}(h,h) = \sum_{i<j} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \cdot h_i h_j + \sum_{i=1}^{m} \frac{1}{2!} \frac{\partial^2 f}{\partial x_i^2}(a) \cdot h_i^2. \tag{4}$$

The general shape of (2) in terms of partial derivatives, generalizing (4), is given by the following theorem.

Theorem 4.2. *Let* $\dim V = n < \infty$ *and choose an ordered basis, with associated linear coordinate functions* $x_1, \dots, x_n$*. Let* $f : U \to W$ *be a* C^p *map, with* $U \subseteq V$ *open. Choose* $a \in U$ *and* $r > 0$ *such that* $B_r(a) \subseteq U$ *for a choice of norm on* V*. Choose* $h = \sum h_j e_j \in V$ *with* $\|h\| < r$*. For non-negative integers* $k \le p$ *we have an equality*

$$\frac{(D^k f)(a)}{k!}(h^{(k)}) = \sum_{i_1 + \dots + i_n = k} \frac{1}{i_1! \cdots i_n!} h_1^{i_1} \cdots h_n^{i_n} \cdot \frac{\partial^k f}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}}(a), \tag{5}$$

in W, where $h^{(k)} = (h, \dots, h) \in V^k$ and the sum is taken over all ordered n-tuples $(i_1, \dots, i_n)$ of non-negative integers whose sum is k.

In this theorem we are taking partial derivatives of the W-valued (rather than $\mathbf{R}$-valued) mapping f, but such partials only require the target to be a finite-dimensional vector space rather than to be $\mathbf{R}$. (Alternatively, one can choose a basis to identify W with some $\mathbf{R}^m$ and then such W-valued partials are componentwise computed as iterated partials of the $\mathbf{R}$-valued C^p component functions of f.) It is easy to check that when $k = 2$ this recovers the explicit formula (4). It is also instructive to write out the right side of this general formula says for $k = 3$:

$$\sum_{i=1}^{n} \frac{h_i^3}{3!} \cdot \frac{\partial^3 f}{\partial x_i^3}(a) + \sum_{i \neq j} \frac{h_i^2 h_j}{2! \cdot 1!} \cdot \frac{\partial^2 f}{\partial x_i^2 \partial x_j}(a) + \sum_{i \neq j \neq k \neq i} h_i h_j h_k \cdot \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(a).$$

Now let's proof Theorem 4.2.

Proof. If we multiply through both sides of (5) by $k!$, then on the left side we get

$$((D^k f)(a))(h, \dots, h)$$

and on the right side we have coefficients

$$\frac{k!}{i_1! \cdots i_n!}.$$

This ratio is just a binomial coefficient when $n = 2$ and in general is a positive integer which counts exactly the number of ways to decompose a set S of k things into an ordered collection of disjoint subsets $S_1, \dots, S_n$ with S_j of size i_j. For $n = 2$ this is just the combinatorial interpretation of the coefficients in the binomial theorem, and for $n > 2$ it analogously corresponds to the "multinomial theorem" (which is readily verified by induction).

In other words, if we consider all k-fold partial derivative operators

$$\frac{\partial^k}{\partial x_{j_1} \cdots \partial x_{j_k}} \tag{6}$$

with *ordered* k-tuples of indices $(j_1, \dots, j_k)$, then the ordering among the k partial derivative operators $\partial/\partial x_{j_s}$ does not affect the value of (6) and if we let i_t denote the number of times t appears in a fixed $(j_1, \dots, j_k)$, then (6) can be rewritten as

$$\frac{\partial^k}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}},$$

and this particular ordered n-tuple of "exponents" $(i_1, \dots, i_n)$ arises from exactly

$$\frac{k!}{i_1! \cdots i_n!}$$

of the expressions (6) as we run over all k-tuples $(j_1, \dots, j_k)$ with $1 \leq j_1, \dots, j_k \leq n$. Thus, $k!$ times (5) can be written in the form

$$((D^k f)(a))(h, \dots, h) \stackrel{?}{=} \sum_{(j_1, \dots, j_k)} h_{j_1} \cdots h_{j_k} \cdot \frac{\partial^k f}{\partial x_{j_1} \cdots \partial x_{j_k}}(a) \tag{7}$$

in W where the sum is taken over *all* n^k ordered k-tuples $(j_1, \dots, j_k)$ of integers between 1 and n.

A more general assertion we can try to prove is

$$((D^k f)(a))(v_1, \dots, v_k) = \sum_{(j_1,\dots,j_k)} c_{j_1,1} \cdots c_{j_k,k} \cdot \frac{\partial^k f}{\partial x_{j_1} \cdots \partial x_{j_k}}(a), \tag{8}$$

where $v_s = \sum c_{is} e_i \in V$ are k arbitrary vectors (setting all v_s's to be equal to a common $h \in V$ then recovers (7)). We will prove the more general (8). By multilinearity of both sides in $(v_1, \dots, v_k) \in V^k$, it suffices to treat the case when the v_s's come from the initially chosen basis of V. This special case is exactly the identity in Theorem 3.3. ■

5. Taylor's formula: statement and proof

Let V and W be finite-dimensional over $\mathbf{R}$, U an open 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 5.1. *With notation as above,*

$$f(a+h) = \sum_{j=0}^{p} \frac{(D^j f)(a)}{j!}(h^{(j)}) + R_{p,a}(h) \tag{9}$$

in W, where

$$R_{p,a}(h) = \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!}((D^p f)(a+th) - (D^p f)(a))(h^{(p)})\, \mathrm{d}t \tag{10}$$

satisfies

$$\|R_{p,a}(h)\| \leq C_{p,h,a}\|h\|^p, \quad \lim_{h \to 0} C_{p,h,a} = 0 \tag{11}$$

with

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

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

Remark 5.2. The norm on $\mathrm{Mult}(V^p, W)$ that is implicit in the numerator defining $C_{p,h,a}$ is defined in terms of arbitrary but fixed choices of norms on V and W: for any multilinear $\mu : V^p \to W$ there exists a constant $B \geq 0$ such that $\|\mu(v_1, \dots, v_p)\| \leq B \prod_{j=1}^{p} \|v_j\|$, by elementary arguments exactly as in the simplest case $p = 1$, and the infimum of all such B's also works and is called $\|\mu\|$. More concretely, $\|\mu\|$ is the minimum of $\|\mu(v_1, \dots, v_p)\|$ for the compact set of points $(v_1, \dots, v_p) \in V^p$ satisfying $\|v_j\| = 1$ for all j. It is easy to check that $\mu \mapsto \|\mu\|$ is a norm on the finite-dimensional vector space $\mathrm{Mult}^p(V, W)$, and in particular it is a continuous $\mathbf{R}$-valued function on this space of multilinear mappings.

The supremum defining $C_{p,h,a}$ is finite because $[0,1]$ is compact and $t \mapsto (D^p f)(a+th)$ is continuous (since f is a C^p mapping).

Before giving the proof of the general Taylor formula modelled on the one from calculus, we stress that Theorem 4.2 makes (9) more concrete when we choose linear coordinates on V, say with $n = \dim V$. Here is the "explicit" version of (9) with the remainder (10) in such coordinates: $f(a+h) \in W$ is equal to

$$\sum_{i_1+\cdots+i_n \leq p} \frac{h_1^{i_1} \cdots h_n^{i_n}}{i_1! \cdots i_n!} \frac{\partial^{i_1+\cdots+i_n} f}{\partial x_i^{i_1} \cdots \partial x_n^{i_n}}(a)$$

$$+\sum_{\sum i_j=p}\frac{(1-t)^{p-1}}{(p-1)!}\int_0^1\frac{h_1^{i_1}\cdots h_n^{i_n}}{i_1!\cdots i_n!}\left(\frac{\partial^p f}{\partial x_i^{i_1}\cdots\partial x_n^{i_n}}(a+th)-\frac{\partial^p f}{\partial x_i^{i_1}\cdots\partial x_n^{i_n}}(a)\right)\mathrm{d}t.$$

There is no doubt that (9) and (10) are far simpler expressions than this coordinatized version. Describing $C_{p,h,a}$ (or a reasonable substitute) in terms of such coordinates is a total mess.

One important consequence of the error estimate (11) is that it shows the error $R_{p,a}(h)$ in the "degree p" expansion (9) of $f(a+h)$ about a dies off more rapidly than $\|h\|^p$ as $h \to 0$ (i.e., $\|R_{p,a}(h)\|/\|h\|^p \to 0$ 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 particularly important case is $p = 2$: the approximation

$$f(a+h) = f(a) + ((Df)(a))(h) + ((D^2 f)(a))(h,h) + (\dots)$$

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

Now we are ready to present the proof of Taylor's formula. We emphasize that it is really Theorem 4.2 that encapsulates all of the gritty work with bases and coordinates. In the general coordinate-free setup in which we have presently placed ourselves, the use of total derivatives will make arguments from 1-variable calculus adapt almost without change; part of the purpose of the multilinear formalism is to enable us to make clean multivariable arguments "as if" we were in the 1-variable case. Trying to prove Theorem 5.1 by using the classical explicit formula with all of the factorials and iterated partials looks unpleasant.

Proof. (of Theorem 5.1). Granting the formula for a moment, let us briefly explain the error estimates. The uniform continuity of the continuous $\|D^p f\|$ on compacts is the reason for the uniformity (with respect to a in a compact) for the rate of decay $C_{p,h,a} \to 0$ as $h \to 0$. As for the bound $C_{p,h,a}\|h\|^p$ on the norm of the integral formula for $R_{p,a}(h)$, since $\int_0^1 (1-t)^{p-1}/(p-1)! = 1/p!$ we can use the general inequality

$$\|\mu(v_1,\dots,v_p)\| \le \|\mu\| \cdot \prod_{j=1}^p \|v_j\|$$

for $\mu \in \mathrm{Mult}(V^p, W)$ to reduce the problem to proving $\|\int_0^1 g(t)\,\mathrm{d}t\| \le \int_0^1 \|g(t)\|\,\mathrm{d}t$ for any continuous map $g : [0,1] \to W$. This inequality for integrals is an analogue of the estimate $|\int_a^b \phi| \le \int_a^b |\phi|$ in calculus, and since we require our W-valued g to be continuous (and hence it is uniformly continuous, as $[0,1]$ is compact), we may express g as a *uniform* limit of W-valued step functions. Thus, by chopping up the interval $[0,1]$ into pieces (depending on the discontinuous step-function approximation) it is easy to thereby reduce the general inequality $\|\int_0^1 g\| \le \int_0^1 \|g\|$ to the trivial special case when the W-valued function g is constant. This establishes the error estimate (11).

Now we prove the formulas (9) and (10) for all p. By the second Fundamental Theorem of Calculus (applied componentwise using a basis of W, say), we have

$$f(a+h) = f(a) + \int_0^1 ((Df)(a+th))(h)\,\mathrm{d}t = f(a) + (Df)(a)(h) + \int_0^1 ((Df)(a+th) - (Df)(a))(h)\,\mathrm{d}t$$

in W. This takes care of the case $p = 1$.

Now we assume $p > 1$ and we use induction. Since f is also of class C^{p-1}, we have

$$f(a+h) = \sum_{j=0}^{p-1} \frac{(D^j f)(a)}{j!}(h^{(j)}) + \int_0^1 \frac{(1-t)^{p-2}}{(p-2)!}((D^{p-1}f)(a+th) - (D^{p-1}f)(a))(h^{(p-1)})\,\mathrm{d}t$$

in W. Thus, we just have to show

$$\begin{aligned}&\int_0^1 \frac{(1-t)^{p-2}}{(p-2)!}((D^{p-1}f)(a+th) - (D^{p-1}f)(a))(h^{(p-1)})\,\mathrm{d}t \\ \stackrel{?}{=}\ & \frac{(D^p f)(a)}{p!}(h^{(p)}) + \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!}((D^p f)(a+th) - (D^p f)(a))(h^{(p)})\,\mathrm{d}t\end{aligned}$$

in W. Since the first term on the right exactly cancels what is being subtracted within the integral on the right, by using the identification $\mathrm{Mult}(V^p, W) \simeq \mathrm{Hom}(V, \mathrm{Mult}(V^{p-1}, W))$ it suffices to prove that in $\mathrm{Mult}(V^{p-1}, W)$ we have an equality

$$\int_0^1 (p-1)(1-t)^{p-2}((D^{p-1}f)(a+th) - (D^{p-1}f)(a))\,\mathrm{d}t \stackrel{?}{=} \int_0^1 (1-t)^{p-1}((D^p f)(a+th))(h)\,\mathrm{d}t,$$

(where the evaluation at $h \in V$ on the right is really evaluation in the first slot of a symmetric multilinear mapping on V^p), since then evaluation on $(h, \dots, h) \in V^{p-1}$ (which can be moved inside a definite integral!) and division by $(p-1)!$ will yield what we want.

Let $g = D^{p-1}f : U \to \mathrm{Mult}(V^{p-1}, W)$, so g is a C^1 map and we want

$$\int_0^1 (p-1)(1-t)^{p-2}(g(a+th) - g(a))\,\mathrm{d}t \stackrel{?}{=} \int_0^1 (1-t)^{p-1}((Dg)(a+th))(h)\,\mathrm{d}t.$$

This essentially comes down to integration by parts. We can rewrite our desired equation as

$$g(a) \stackrel{?}{=} \int_0^1 ((p-1)(1-t)^{p-2}g(a+th) - (1-t)^{p-1}((Dg)(a+th))(h))\,\mathrm{d}t. \tag{12}$$

Consider the map

$$\phi : (-1, 1+\varepsilon) \mapsto \mathrm{Mult}(V^{p-1}, W)$$

defined by

$$\phi(t) = -(1-t)^{p-1}g(a+th),$$

where $\varepsilon > 0$ is small enough so that $(1+\varepsilon)\|h\| < r$ and hence $a + th \in B_r(a)$ for $t \in (-1, 1+\varepsilon)$.

Since g is C^1, a straightfoward application of the Chain Rule yields that ϕ is C^1 with

$$\phi'(t) \stackrel{\text{def}}{=} ((D\phi)(t))(1) = (p-1)(1-t)^{p-2}g(a+th) - (1-t)^{p-1}((Dg)(a+th))(h)$$

in $\mathrm{Mult}(V^{p-1}, W)$. This is exactly the integrand in (12), so we are reduced to proving $g(a) = \int_0^1 \phi'$. But by the second Fundamental Theorem of Calculus (applied componentwise with respect to a basis of the vector space $\mathrm{Mult}(V^{p-1}, W)$, say), this latter integral is equal to $\phi(1) - \phi(0)$, and from the definition of ϕ this is exactly $g(a)$ as desired! It is an instructive exercise to check that when $V = W = \mathbf{R}$, this *is* the proof in Spivak's *Calculus* (up to linear change of variable on the integrals). ■