Higher order derivatives

Let $E \subset \mathbb{R}^n$, and let $f : E \to \mathbb{R}^m$ be differentiable. Then $f'$ is a function from $E$ to $L(\mathbb{R}^n, \mathbb{R}^m)$, which is a finite-dimensional normed space. Everything we’ve done with derivatives carries over to functions between more general normed spaces than Euclidean spaces. So, it makes sense to differentiate $f'$. The derivative of $f'$ at $a \in E$ would be a linear map from $\mathbb{R}^n$ to $L(\mathbb{R}^n, \mathbb{R}^m)$. A moment’s thought reveals that this might have limited value beyond being the logical consequence of the definitions. As it happens, there is a much more practical interpretation of higher-order derivatives, namely as multilinear maps. But we will not need this, so we pursue it no further. Instead, we work completely in the context of higher-order partial derivatives, and moreover we consider only real-valued functions. These are precisely the higher-order derivatives you learned about in calculus, so we only sketch the basics of the theory. In fact, we only need to agree on our notational conventions and recall Clairaut’s Theorem (see below), which you probably learned in calculus.

Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}$, and $a \in U$. Suppose that the partial derivative $\frac{\partial f}{\partial x_i}$ exists on $U$. Then $\frac{\partial f}{\partial x_i}$ is a real-valued function on $U$, and we can consider its partial derivatives. Recall the notation: for $i, j \in \{1, \ldots, n\}$ we write
\[
\frac{\partial^2 f}{\partial x_j \partial x_i}(a) = \frac{\partial}{\partial x_j} \left( \frac{\partial f}{\partial x_i} \right)(a).
\]

If $i = j$ we write
\[
\frac{\partial^2 f}{\partial x_i^2}(a) = \frac{\partial^2 f}{\partial x_i \partial x_i}(a).
\]

These are second-order partial derivatives. Inductively, for any $n > 1$ an $(n+1)$th-order partial derivative is a partial derivative of an $n$th-order partial derivative, and we use an obvious extension of the above notational scheme. We say $f$ is $n$-times continuously differentiable, or $C^n$, if all its $n$th-order partial derivatives exist and are continuous.

A downside to our decision to stay within the context of higher-order partial derivatives is that we cannot use the phrase “$n$-times differentiable” (although we could, if we wanted to, say “$n$-times partially differentiable”). We’ll have to be satisfied with “$n$-times continuously differentiable”.

A little experimentation with the notation of higher-order partials leads to the conclusion that things quickly get out of hand. For example, if $f$ is a function of two variables $x, y$ and we partially differentiate $f$ with respect to $y, x, y, x, x, x$ (in that order!) then we get
\[
\frac{\partial^7 f}{\partial x \partial x \partial x \partial y \partial y \partial x \partial y}.
\]
However, if we knew that the order of the partial derivatives didn’t matter then we could rearrange the above as

\[ \frac{\partial^2 f}{\partial x^4 \partial y^3} \]

which is certainly an improvement. Fortunately, this is ok as long as the partials are continuous:

**Theorem 1** (Clairaut’s Theorem). Let \( U \subset \mathbb{R}^n \) be open and \( f : U \to \mathbb{R} \) be \( C^k \). Then

\[ \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}} = \frac{\partial^k f}{\partial x_{j_1} \cdots \partial x_{j_k}} \]

for all \((i_1, \ldots, i_k) \in \{1, \ldots, n\}^k\) and every rearrangement \((j_1, \ldots, j_k)\) of \((i_1, \ldots, i_k)\).

**Proof.** The idea of this proof is fairly simple, basically involving a couple of applications of the single-variable Mean Value Theorem. But to make the argument precise involves a mildly surprising amount of fussiness. First we simplify: since every rearrangement of coordinates can be obtained by switching pairs of coordinates finitely many times, by induction it suffices to show that if \( \frac{\partial^2 f}{\partial x_i \partial x_j} \) is continuous for all \( i, j = 1, \ldots, n \) then \( \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \) for all \( i, j \).

Due to the manner in which partial derivatives are computed, it suffices to consider \( f \) as a function of two variables \( x, y \), and show that if \( \frac{\partial^2 f}{\partial x \partial y} \) and \( \frac{\partial^2 f}{\partial y \partial x} \) are both continuous on an open set \( U \subset \mathbb{R}^2 \) then they coincide at any point \((b, c) \in U\).

Fix \((b, c) \in U\) and let \( \varepsilon > 0 \), and choose \( \delta > 0 \) such that if \( \| (s, t) \| < \delta \) then \((b + s, c + t) \in U\) and both

\[ \left| \frac{\partial^2 f}{\partial x \partial y} (b + s, c + t) - \frac{\partial^2 f}{\partial x \partial y} (b, c) \right| < \varepsilon \]

\[ \left| \frac{\partial^2 f}{\partial y \partial x} (b + s, c + t) - \frac{\partial^2 f}{\partial y \partial x} (b, c) \right| < \varepsilon. \]

Define \( \Delta : B_\delta(0, 0) \to \mathbb{R} \) by

\[ \Delta(s, t) = f(b + s, c + t) - f(b + s, c) - f(b, c + t) + f(b, c). \]

Temporarily fix \((s, t) \in B_\delta(0, 0)\), and define a real-valued function \( g \) in the closed interval with endpoints \( 0, s \) by

\[ g(r) = f(b + r, c + t) - f(b + r, c). \]

Then by the Chain Rule \( g \) is differentiable and

\[ g'(r) = \frac{\partial f}{\partial x} (b + r, c + t) - \frac{\partial f}{\partial x} (b + r, c). \]

Thus by the single-variable Mean Value Theorem there exists \( r \) between \( 0 \) and \( s \) such that

\[ f(b + s, c + t) - f(b + s, c) - f(b, c + t) + f(b, c) = g(s) - g(0) = g'(r)s = \left( \frac{\partial f}{\partial x} (b + r, c + t) - \frac{\partial f}{\partial x} (b + r, c) \right)s, \]
and then by the Mean Value Theorem for the partial derivative $\partial f/\partial y$ there exists $u$ between 0 and $t$ such that the above can be contused as

$$\frac{\partial}{\partial y} \left( \frac{\partial f}{\partial x} \right) (b + r, c + u)ts$$

$$= \frac{\partial^2 f}{\partial y \partial x}(b + r, c + u)st.$$ 

Thus for all $(s, t) \in B_\delta(0,0)$ there exists $(r, u) \in B_\delta(0,0)$ such that

$$\left| \frac{\Delta(s, t)}{st} - \frac{\partial^2 f}{\partial y \partial x}(b, c) \right| = \left| \frac{\partial^2 f}{\partial y \partial x}(b + r, c + u) - \frac{\partial^2 f}{\partial y \partial x}(b, c) \right| < \varepsilon.$$

We conclude that

$$\lim_{(s, t) \to (0, 0)} \frac{\Delta(s, t)}{st} = \frac{\partial^2 f}{\partial y \partial x}(b, c).$$

Now, clearly the above construction could have been done with the roles of $x$ and $y$ (and $s$ and $t$) reversed, giving

$$\lim_{(s, t) \to (0, 0)} \frac{\Delta(s, t)}{st} = \frac{\partial^2 f}{\partial x \partial y}(b, c),$$

and we are done. \hfill \square

Even better, we can further clean up the notation for higher-order partials using multi-index notation:

**Definition 2.** A multi-index is an $n$-tuple $\alpha = (\alpha_1, \ldots, \alpha_n)$ of nonnegative integers. We write

- $|\alpha| = \alpha_1 + \cdots + \alpha_n$
- $\alpha! = \alpha_1! \cdots \alpha_n!$
- $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$
- $\frac{\partial^{|\alpha|} f}{\partial x^\alpha} = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$.

It helps to keep in mind the Multinomial Theorem:

$$(x_1 + \cdots + x_n)^k = \sum_{i_1, \ldots, i_k=1}^n x_{i_1} \cdots x_{i_k}$$

$$= \sum_{\alpha_1 + \cdots + \alpha_n = k} \frac{k!}{\alpha_1! \cdots \alpha_n!} x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$
**Theorem 3** (Taylor’s Theorem). Let \( a \in \mathbb{R}^n \) and \( r > 0 \), and let \( f : B_r(a) \to \mathbb{R} \) be \( C^{m+1} \). Then for all \( x \in B_r(0) \) there exists \( t \in (0, 1) \) such that

\[
 f(a + x) = \sum_{|\alpha| \leq m} \frac{1}{\alpha!} \frac{\partial^{|\alpha|} f(a)}{\partial x^\alpha} x^\alpha + \sum_{|\alpha| = m+1} \frac{1}{\alpha!} \frac{\partial^{|\alpha|} f(a + tx)}{\partial x^\alpha} x^\alpha
\]

**Proof.** Define \( g : [0, 1] \to \mathbb{R}^m \) by \( g(t) = a + tx \). Then \( f \circ g : [0, 1] \to \mathbb{R} \) is \( C^{m+1} \), so by the one-variable Taylor Theorem there exists \( t \in (0, 1) \) such that

\[
 f(a + x) = f \circ g(1) = \sum_{k \leq m} \frac{(f \circ g)^{(k)}(0)}{k!} + \frac{(f \circ g)^{(m+1)}(t)}{(m + 1)!}.
\]

Thus it suffices to observe that if \( 0 \leq k \leq m + 1 \) and \( t \in [0, 1] \) then

\[
 (f \circ g)^{(k)}(t) = \sum_{i_1, \ldots, i_k = 1}^n \frac{\partial}{\partial x_{i_1}} \cdots \frac{\partial}{\partial x_{i_k}} f(a + tx)x_{i_1} \cdots x_{i_k}
\]

\[
 = \sum_{\alpha_1 + \cdots + \alpha_n = k} \frac{k!}{\alpha_1! \cdots \alpha_n!} \frac{\partial^k f}{\partial x_{i_1}^{\alpha_1} \cdots \partial x_{i_k}^{\alpha_k}} (a + tx)x_{i_1}^{\alpha_1} \cdots x_{i_k}^{\alpha_k}
\]

\[
 = \sum_{|\alpha| = k} \frac{k!}{\alpha!} \frac{\partial^{|\alpha|} f(a + tx)}{\partial x^\alpha} x^\alpha. \quad \square
\]

When \( m = 0 \) the above version Taylor’s Theorem is a slightly weakened version of the Mean Value Theorem (because the latter doesn’t require continuity of the derivative). Actually, we could have weakened our hypotheses in Taylor’s Theorem so that it would include the Mean Value Theorem, but it would have meant significantly more work, and the extra generality is typically not useful. The exercises explore how the case \( m = 1 \) leads to a Second Derivative Test for functions of several variables.