Derivatives

Here our purpose is to give a rigorous foundation of the principles of differentiation in $\mathbb{R}^n$. Much of what we’ll do generalizes easily to more general normed spaces (even infinite-dimensional ones), but we only need the theory in $\mathbb{R}^n$, so it will be more efficient for us to restrict to that context.

As you know, I’m assuming that you have a familiarity with linear algebra, and a thorough knowledge of calculus, including multivariable calculus. In particular, I assume you are thoroughly familiar with partial differentiation, at least for functions of two or three variables — the general theory for functions of $n$ variables poses nothing new, except for messier notation involving general subscripts.

My main goals here are to present the general theory efficiently and clearly; these are in some sense competing objectives, and hence will require some compromises.

The general context here is functions defined on a subset of $\mathbb{R}^n$ and taking values in $\mathbb{R}^m$, for some positive integers $n, m$. But for a little while here at the beginning let’s focus on real-valued functions. As you learned in calculus, for a function of several variables, the basic idea of a partial derivative is to hold all the variables fixed except one, and differentiate with respect to that single variable. Thus, all the theory of single-variable differentiation carries over verbatim to partial derivatives, e.g., linearity, the power rule, the product and quotient rules, the chain rule, the Mean Value Theorem, and Taylor’s Theorem.

First let’s recall the definition of partial derivative. If $E \subset \mathbb{R}^n$ and $f : E \to \mathbb{R}$, we will only allow ourselves to consider (partial) derivatives at interior points; when $n = 1$ it’s easy to make sense out of derivatives on closed intervals, but in higher dimensions the boundary points are a much bigger problem, so we just exclude them from consideration as places at which to differentiate. So, let $a \in E^o$, and recall that the *partial derivative of $f$ at $a$ with respect to $x_i$ is* 

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t \to 0} \frac{f(a + te_i) - f(a)}{t},$$

where $e_1, \ldots, e_n$ denotes the standard basis for $\mathbb{R}^n$. Also, notice that

$$\frac{\partial f}{\partial x_i}(a) = \frac{d}{dx_i} f(a_1, \ldots, x_i, \ldots, a_n) \bigg|_{x_i=a_i} = \frac{d}{dt} f(a + te_i) \bigg|_{t=0}.$$

If $f$ takes values in a higher-dimensional space $\mathbb{R}^m$, we can take partial derivatives of the component functions $f_1, \ldots, f_m$, and we organize them into a matrix:
**Definition 1.** If $E \subset \mathbb{R}^n$ and $f : E \to \mathbb{R}^m$, the Jacobian matrix of $f$ is

\[
    Jf = \begin{pmatrix}
        \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\
        \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\
        \vdots & \vdots & \ddots & \vdots \\
        \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n}
    \end{pmatrix}.
\]

If $n = m$, then $Jf$ is a square matrix, and its determinant is also sometimes (confusingly) called the Jacobian of $f$.

Unfortunately, the Jacobian matrix doesn’t necessarily do what we want from a derivative (unless the partial derivatives are continuous, as we’ll see soon). You’ll recall from calculus that derivatives (and partial derivatives for a function of several variables) are used to give linear approximations to the function. This is exactly what we want the derivative to do, and the general situation requires a different approach:

**Definition 2.** Let $E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}^m$, and $a \in E^0$. We say $f$ is differentiable at $a$ if there exists a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

\[
    \lim_{x \to 0} \frac{f(a + x) - f(a) - Tx}{\|x\|} = 0.
\]

The linear map $T$ is unique if it exists, and is called the derivative of $f$ at $a$, denoted by $f'(a)$.

Thus, just as for the single-variable case, when $f$ is differentiable at $a$ we have

\[
    f(x + h) = f(a) + f'(a)x + q(x)\|x\|,
\]

where

\[
    \lim_{x \to 0} q(x) = 0.
\]

However, in general we can’t compute the derivative as the limit of a “difference quotient”, because we can’t divide by a vector $x$.

Notice again that we only allow ourselves to talk of differentiability at an interior point of the domain $E$. We say $f$ is differentiable if it is differentiable at each point of $E$, which in particular requires that $E$ be an open set. When $f$ is differentiable, we have the derivative function $f' : E \to L(\mathbb{R}^n, \mathbb{R}^m)$. Since $L(\mathbb{R}^n, \mathbb{R}^m)$ is a normed space, it makes sense to consider continuity:

**Definition 3.** $f$ is continuously differentiable, or $C^1$, if it is differentiable and the derivative function $f' : E \to L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous.

**Proposition 4.** Let $E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}^m$, and $a \in E^0$. If $f$ is differentiable at $a$, then all the partial derivatives $\frac{\partial f_i}{\partial x_j}(a)$ exist, and we have

\[
    f'(a) = \left( \frac{\partial f_i}{\partial x_j}(a) \right).
\]

Moreover, $f$ is $C^1$ if and only if every $\frac{\partial f_i}{\partial x_j}$ is continuous on $E$. 
As usual for linear maps between coordinate spaces, we mean that the matrix \( \frac{\partial f_i}{\partial x_j}(a) \) represents the linear map \( f'(a) \) relative to the standard bases.

**Proof.** Fix \( i \in \{1, \ldots, n\} \) and \( j \in \{1, \ldots, m\} \), and let \( m_{ij} \) be the \( ij \)-entry of the matrix \( f'(a) \). Let \( \{e_1, \ldots, e_n\} \) and \( \{v_1, \ldots, v_m\} \) be the standard bases of \( \mathbb{R}^n \) and \( \mathbb{R}^m \), respectively. Then for \( t > 0 \) we have
\[
v_i \cdot \left( \frac{f(a + te_j) - f(a) - f'(a)(te_j)}{\|te_j\|} \right) = \frac{v_i \cdot f(a + te_j) - v_i \cdot f(a) - tv_i \cdot f'(a)e_j}{t} = \frac{f_i(a + te_j) - f_i(a)}{t} - v_i \cdot f'(a)e_j = \frac{f_i(a + te_j) - f_i(a)}{t} - m_{ij} \to \left. \frac{\partial f_i}{\partial x_j}(a) \right|_{t=0} = m_{ij}.
\]
But the limit is 0 by differentiability, so \( m_{ij} = \left. \frac{\partial f_i}{\partial x_j}(a) \right|_{t=0} \).

Finally, the other part now follows since limits of matrices can be computed entry-wise. \( \square \)

Thus, the derivative can be computed using the partial derivatives. In fact, some of the elementary theory of derivatives can be easily derived from the properties of partial derivatives, but it is more instructive to see the development in a “coordinate-free” context whenever possible. Actually, and unsurprisingly, some of the basic properties of derivatives follow in a completely straightforward way from elementary properties of limits, and constitute worthwhile exercises in the definitions. For example:

**Proposition 5.** If \( f : \mathbb{R}^n \to \mathbb{R}^m \) is linear, then it is \( C^1 \), and
\[
f'(a) = f \quad \text{for all } a \in \mathbb{R}^n.
\]

**Proposition 6.** If \( f \) is differentiable at \( a \), then it is continuous there.

**Proposition 7.** If \( f \) and \( g \) are differentiable at \( a \), then so are \( f + g \) and \( cf \) for \( c \in \mathbb{R} \), and
\[
\begin{align*}
(1) \quad (f + g)'(a) &= f'(a) + g'(a); \\
(2) \quad (cf)'(a) &= cf'(a).
\end{align*}
\]

The product rule(s) are a little harder than the single-variable case. Actually, in multivariable analysis there are various versions of the product rule, which can be proved by a common method. Here are two versions, and we only include the proof of the first:

**Proposition 8.** Let \( E \subseteq \mathbb{R}^n \), \( a \in E^0 \), and \( x \in \mathbb{R}^n \).
\[
\begin{align*}
(1) \quad \text{If } f : E \to \mathbb{R} \text{ and } g : E \to \mathbb{R}^m \text{ are differentiable at } a, \text{ then so is the pointwise product } fg, \text{ and } \\
(fg)'(a) &= f(a)g'(a)x + (f'(a)x)g(a). \\
(2) \quad \text{If } f, g : E \to \mathbb{R}^m \text{ are differentiable at } a, \text{ then so is the pointwise dot product } f \cdot g, \text{ and } \\
(f \cdot g)'(a)x &= f(a) \cdot (g'(a)x) + (f'(a)x) \cdot g(a).
\end{align*}
\]
You should pause to ponder the meaning if each of the above product rules, and in particular why they are written the way they are.

**Proof.** We only prove the first; the second is similar. Let
\[
q(x) = \frac{f(a + x) - f(a) - f'(a)x}{\|x\|},
\]
\[
r(x) = \frac{g(a + x) - g(a) - g'(a)x}{\|x\|},
\]
so that by differentiability both \( q(x) \) and \( r(x) \) go to 0 as \( x \to 0 \). Let’s compute:
\[
fg(a + x) - fg(a) - f'(a)yg(a) - f(a)g'(a)x
\]
\[
= f(a + x)g(a + x) - f(a)g(a) - f'(a)xg(a) - f(a)g'(a)x \pm f(a)g(a + x)
\]
\[
= (f(a + x) - f(a))g(a + x) + f(a)(g(a + x) - g(a) - g'(a)x) - f'(a)xg(a)
\]
\[
= (f(a + x) - f(a))g(a + x) \pm f'(a)xg(a + x) + f(a)r(x)\|x\| - f'(a)xg(a)
\]
\[
= (f(a + x) - f(a) - f'(a)x)g(a + x) + f'(a)x(g(a + x) - g(a)) + f(a)r(x)\|x\|
\]
\[
= q(x)\|x\|g(a + x) + f'(a)x(g(a + x) - g(a)) + f(a)r(x)\|x\|
\]
and dividing by \( \|x\| \) gives
\[
\frac{fg(a + x) - fg(a) - f'(a)yg(a) - f(a)g'(a)x}{\|x\|} = q(x)g(a + x) + \left(f'(a)\frac{x}{\|x\|}\right)(g(a + x) - g(a)) + f(a)r(x)
\]
\[
\begin{array}{c}
x \to 0, \\
because:
\end{array}
\]
\begin{itemize}
\item \( g(a + x) \to g(a) \) by continuity;
\item \( x/\|x\| \) is bounded and the linear map \( f'(a) \) is also bounded;
\item \( g(a + x) - g(a) \to 0 \) by continuity (again).
\end{itemize}

Here is the multivariable version of the chain rule:

**Proposition 9.** Let \( E \subseteq \mathbb{R}^n \), \( D \subseteq \mathbb{R}^m \), \( f : E \to D \), and \( g : D \to \mathbb{R}^k \). Suppose that \( a \in E^0 \) and \( f(a) \in D^0 \). If \( f \) is differentiable at \( a \) and \( g \) is differentiable at \( f(a) \), then \( g \circ f \) is differentiable at \( a \) and
\[
(g \circ f)'(a) = g'(f(a))f'(a).
\]

Again, you should pause to contemplate the meaning of the chain rule: the right-hand side is a composition of linear maps.

**Proof.** Put \( b = f(a) \), and let
\[
q(x) = \frac{f(a + x) - f(a) - f'(a)x}{\|x\|}
\]
\[
r(y) = \frac{g(b + y) - g(b) - g'(b)y}{\|y\|},
\]
so that \( \lim_{x \to 0} q(x) = 0 \) and \( \lim_{y \to 0} r(y) = 0 \). Letting \( y = f(a + x) - f(a) \), a little algebra shows that
\[
g \circ f(a + x) - g \circ f(a) = \frac{g'(f(a))f'(a)x}{\|x\|} - g'(b)q(x) + \frac{r(y)\|y\|}{\|x\|}.
\]
Both terms go to 0 as \( x \to 0 \); this is clear for the first term since \( g'(b) \) is continuous at 0. For the other term, we have \( \lim_{x \to 0} r(y) = 0 \) since \( \lim_{x \to 0} y = 0 \) by continuity of \( f \) at \( a \), and
\[
\|y\| = \|f'(a)x + q(x)\| \leq \|f'(a)x\| + \|q(x)\|\|x\| \leq \|f'(a)\|\|x\| + \|q(x)\|\|x\|,
\]
so \( \|y\|/\|x\| \) is bounded as \( x \to 0 \).

\[\square\]

**Lemma 10.** Let \( E \subset \mathbb{R}^n \), \( f = (f_1, \ldots, f_m) : E \to \mathbb{R}^m \), and \( a \in E^0 \). Then \( f \) is differentiable at \( a \) if and only if every component function \( f_i \) is.

The single-variable Mean Value Theorem generalizes satisfactorily to real-valued functions of several variables:

**Definition 11.** Let \( x, y \in \mathbb{R}^n \). The line segment joining \( x \) and \( y \) is the subset
\[
[x, y] := \{(1 - t)x + ty : 0 \leq t \leq 1\}.
\]

A subset \( A \subset \mathbb{R}^n \) is convex if \( [x, y] \subset A \) for all \( x, y \in A \).

**Theorem 12** (Mean Value Theorem). Let \( U \subset \mathbb{R}^n \) be open and convex, and let \( f : U \to \mathbb{R} \) be differentiable. Then for all \( x, y \in U \) there exists \( z \in [x, y] \) such that
\[
f(x) - f(y) = f'(z)(x - y).
\]

Actually, we could slightly strengthen the conclusion to say that \( z \) can be found “strictly between” \( x \) and \( y \), but we will not need this.

**Proof.** Fix \( x, y \in U \), and define \( g : [0, 1] \to U \) by
\[
g(t) = tx + (1 - t)y,
\]
so that the range of \( g \) is the line segment \([x, y]\). Then \( f \circ g \) is differentiable\(^1\) on \([0, 1]\), so by the one-variable Mean Value Theorem there exists \( c \in [0, 1] \) such that
\[
f(x) - f(y) = f \circ g(1) - f \circ g(0) = (f \circ g)'(c)(1 - 0)
\]
\[
= f'(g(c))g'(c)
\]
\[
= f'(cx + (1 - c)y)(x - y)
\]
so we can take \( z = cx + (1 - c)y \).

\[\square\]

In general, however, we have to be satisfied with an inequality:

**Corollary 13** (Mean Value Theorem). Let \( U \subset \mathbb{R}^n \) be open and convex, and let \( f : U \to \mathbb{R}^n \) be differentiable. Suppose \( \|f'(x)\| \leq M \) for all \( x \in U \). Then
\[
\|f(x) - f(y)\| \leq M\|x - y\| \quad \text{for all } x, y \in U.
\]

\(^1\) and note that here we allow the closed interval for a single-variable function.
Proof. Fix \( x, y \in U \), put \( u = f(x) - f(y) \), and define a linear map \( T : \mathbb{R}^m \to \mathbb{R} \) by \( T(v) = u \cdot v \). Then \( T \circ f \) is differentiable, so by the above Mean Value Theorem (Theorem 12) there exists \( z \in [x, y] \) such that

\[
T \circ f(x) - T \circ f(y) = (T \circ f)'(z)(x - y)
\]

(2)

On the other hand,

\[
T \circ f(x) - T \circ f(y) = T(f(x) - f(y)) = u \cdot u = \|u\|^2,
\]

so combining with (2) we get

\[
\|u\|^2 = u \cdot (f'(z)(x - y))
\]

\[
\leq \|u\|\|f'(z)(x - y)\|
\]

\[
\leq \|u\|\|f'(z)\|\|x - y\|
\]

\[
\leq \|u\|M\|x - y\|
\]

and (1) follows. \( \square \)

The Mean Value Theorem can be used to derive a practical test for (continuous) differentiability, which can also be regarded as a complement to Proposition 4:

**Theorem 14.** Let \( U \subset \mathbb{R}^n \) be open and \( f : U \to \mathbb{R}^m \). If all the partial derivatives \( \frac{\partial f_i}{\partial x_j} \) exist and are continuous on \( U \), then \( f \) is \( C^1 \).

**Proof.** By Proposition 4 it suffices to show that \( f \) is differentiable. Moreover, by Lemma 10 without loss of generality \( f \) is real-valued.

Let \( a \in U \) and \( \varepsilon > 0 \). Choose \( \delta > 0 \) such that for all \( z \in U \), if \( \|z - a\| < \delta \) then

\[
\left| \frac{\partial f}{\partial x_j}(z) - \frac{\partial f}{\partial x_j}(a) \right| < \frac{\varepsilon}{n} \quad \text{for } j = 1, \ldots, n.
\]

Choose \( 0 < r < \delta \) such that \( B_r(a) \subset U \), and let \( x \in \mathbb{R}^n \) with \( \|x\| < r \). Define points \( a_0, \ldots, a_n \in B_r(a) \) by

\[
a_j = \begin{cases} 
a & \text{if } j = 0 \\
a_{j-1} + xe_j & \text{if } j = 1, \ldots, n.
\end{cases}
\]

Then we have a telescoping sum:

\[
f(a + x) - f(a) = \sum_{1}^{n}(f(a_j) - f(a_{j-1})).
\]

\( ^2 \)draw a picture!
Now, for each $j$ we have $[a_{j-1}, a_j] \subset B_r(a)$, so we can apply the Mean Value Theorem for partial derivatives to find $z_j \in [a_{j-1}, a_j]$ such that

$$f(a_j) - f(a_{j-1}) = \frac{\partial f}{\partial x_j}(z_j)x_j.$$  

Thus

$$\left| \frac{f(a + x) - f(a) - \sum_{j=1}^n \frac{\partial f}{\partial x_j}(a)x_j}{\|x\|} \right| = \left| \sum_{i=1}^n \left( \frac{\partial f}{\partial x_j}(z_i) - \frac{\partial f}{\partial x_j}(a) \right) x_i \right|$$

$$\leq \sum_{i=1}^n \left| \frac{\partial f}{\partial x_j}(z_i) - \frac{\partial f}{\partial x_j}(a) \right| \frac{|x_j|}{\|x\|}$$

$$\leq \sum_{i=1}^n \left| \frac{\partial f}{\partial x_j}(z_i) - \frac{\partial f}{\partial x_j}(a) \right|$$

$$< \varepsilon,$$  

showing that $f'(a) = \left( \frac{\partial f}{\partial x_1}(a) \cdots \frac{\partial f}{\partial x_n}(a) \right)$ (as expected).  $\square$