Differentiability

For a function of one variable, differentiability is synonymous with the existence of the derivative. However, the notion of differentiability is much more subtle for functions of more than one variable.

Recall that a function $f(x)$ is *differentiable* at $x = a$ if $f'(a) = \lim\limits_{h\to 0} \dfrac{f(a+h)-f(a)}{h}$ exists. Let's recast this definition is little. First, set $x = a+h$, then $h = x-a$. Now $h \to 0$ becomes $x \to a$. This means that being differentiable at $x = a$ is equivalent to the existence of the limit $\lim\limits_{x\to a} \dfrac{f(x)-f(a)}{x-a}$.

Let's manipulate our current definition a little more. We have $f'(a) = \lim\limits_{x\to a} \dfrac{f(x)-f(a)}{x-a}$ so…

$$0 = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} - f'(a) = \lim_{x\to a} \frac{f(x)-f(a)-f'(a)(x-a)}{x-a} = \lim_{x\to a} \frac{f(x)-[f(a)+f'(a)(x-a)]}{x-a}$$

This means that $f(x)$ is differentiable at $x = a$ if $\dfrac{f(x) - [\text{linearization of } f \text{ at } x=a]}{x-a} \to 0$ as $x \to a$. We have recast differentiability into a statement about a comparison between f and its linearization. We have arrived at a working definition of differentiability in general.

Definition: A function is **differentiable** at a point if it can be *well-approximated* by a linearization at that point.

Let's make the above definition more concrete. Consider a function of n variables, $f(\mathbf{x}) = f(x_1,\dots,x_n)$, and fix a point $\mathbf{a} = \langle a_1,\dots,a_n\rangle$. We will use $\mathbf{a}$ as our base point. Let $\mathbf{v} = \langle v_1,\dots,v_n\rangle$ be some fixed vector. We can create a linear function based at $\mathbf{x} = \mathbf{a}$ as follows: $L(\mathbf{x}) = f(\mathbf{a}) + \mathbf{v} \bullet (\mathbf{x}-\mathbf{a})$. In fact, this would be the linearization of $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{a}$ if $\mathbf{v} = \nabla f(\mathbf{a})$. However, for now let's leave $\mathbf{v}$ ambiguous.

Consider $f(\mathbf{x}) - L(\mathbf{x})$. We cannot "divide by $\mathbf{x}-\mathbf{a}$" since in the same way we divided by $x - a$ when we had a function of a single variable since now our objects are vectors. However, dividing magnitudes *does* make sense. Now we have a more concrete definition of differentiability…

Definition: $f(\mathbf{x})$ is **differentiable** at $\mathbf{x} = \mathbf{a}$ if there is some vector $\mathbf{v}$ such that…

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{|f(\mathbf{x}) - L(\mathbf{x})|}{\|\mathbf{x}-\mathbf{a}\|} = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\Big|f(\mathbf{x}) - [f(\mathbf{a}) + \mathbf{v}\bullet(\mathbf{x}-\mathbf{a})]\Big|}{\|\mathbf{x}-\mathbf{a}\|} = 0$$

When this is the case, we *define* $\nabla f(\mathbf{a}) = \mathbf{v}$.

Notice that in the special case $n = 1$ (i.e. a single variable function), we have…

$$\lim_{x\to a} \frac{|f(x) - [f(a)+v(x-a)]|}{|x-a|} = 0 \quad \Longleftrightarrow \quad \lim_{x\to a} \frac{f(x) - [f(a)+v(x-a)]}{x-a} = 0 \quad \Longleftrightarrow \quad \lim_{x\to a} \frac{f(x)-f(a)}{x-a} = v$$

In other words, for single variable functions, differentiability is the same as the derivative existing (in this case $f'(a) = v$).

Next, recall that if a limit exists, it exists and matches if we approach along any continuous curve. Let's see what happens when we approach along coordinate curves (approach parallel to coordinate axes), say $\mathbf{r}(t) = \langle a_1,\dots,a_{i-1},t,a_{i+1},\dots,a_n\rangle$ (so as $t \to a_i$ we get $\mathbf{r}(t) \to \mathbf{a}$). We must have that…

$$\begin{aligned} 0 &= \lim_{t\to a_i} \frac{|f(\mathbf{r}(t)) - f(\mathbf{a}) - \mathbf{v}\bullet(\mathbf{r}(t)-\mathbf{a})|}{\|\mathbf{r}(t)-\mathbf{a}\|} \\ &= \lim_{t\to a_i} \frac{|f(a_1,\dots,a_{i-1},t,a_{i+1},\dots,a_n) - f(a_1,\dots,a_n) - \mathbf{v}\bullet\langle 0,\dots,0,t-a_i,0,\dots,0\rangle|}{\|\langle 0,\dots,0,t-a_i,0,\dots0\rangle\|} \\ &= \lim_{t\to a_i} \frac{|f(a_1,\dots,a_{i-1},t,a_{i+1},\dots,a_n) - f(a_1,\dots,a_n) - v_i(t-a_i)|}{|t-a_i|} \end{aligned}$$

This is equivalent to saying that…

$$\lim_{t\to a_i} \frac{f(a_1,\dots,a_{i-1},t,a_{i+1},\dots,a_n) - f(a_1,\dots,a_n)}{t-a_i} = v_i$$

which is precisely the same as stating that $f_{x_i}(\mathbf{a}) = v_i$. In other words, *if* our function is differentiable at $\mathbf{x} = \mathbf{a}$, then $\nabla f(\mathbf{a}) = \langle f_{x_1}(\mathbf{a}),\dots,f_{x_n}(\mathbf{a})\rangle$. This means that our "new" definition of the gradient matches our old definition (as it should). Also, we just learned that…

Theorem: Differentiability implies the existence of partials.
In particular, if $f(\mathbf{x})$ is differentiable at $\mathbf{x} = \mathbf{a}$, then $f_{x_i}(\mathbf{a})$ exists for $i = 1,\dots,n$.

The converse of this theorem does not hold! This should make pretty good sense at this point. We know that just because a limit exists along several lines, does not mean that the full multivariate limit exists. So it should not be surprising to learn that: *existence of partials does* **not** *imply differentiability!* I will forgo giving an actual counterexample. We will soon see why we never run into this problem in practice.

We learn in Calculus I that differentiable functions are always continuous functions. This is still true for functions of more than one variable.

Theorem: Differentiability implies continuity.

Proof: Suppose that $f(\mathbf{x})$ is differentiable at $\mathbf{x} = \mathbf{a}$. Then $\displaystyle\lim_{\mathbf{x}\to\mathbf{a}} \frac{|f(\mathbf{x}) - L(\mathbf{x})|}{\|\mathbf{x}-\mathbf{a}\|} = 0$. For this limit to be 0, the numerator must limit to 0. This means that $\displaystyle\lim_{\mathbf{x}\to\mathbf{a}} \Big| f(\mathbf{x}) - [f(\mathbf{a}) + \mathbf{v} \bullet (\mathbf{x}-\mathbf{a})] \Big| = 0$ so that $\displaystyle\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) - f(\mathbf{a}) - \mathbf{v} \bullet (\mathbf{x}-\mathbf{a}) = 0$. Now $\mathbf{v} \bullet (\mathbf{x}-\mathbf{a}) \to 0$ as $\mathbf{x} \to \mathbf{a}$. Thus $\displaystyle\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) - f(\mathbf{a}) = 0$ and so $\displaystyle\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$ which is means $f(\mathbf{x})$ is continuous at $\mathbf{x} = \mathbf{a}$.

It should come as no surprise that there are non-differentiable continuous functions (i.e. the converse of this theorem does not hold). In fact, we knew this in Calculus I. It is easy to come up with continuous functions which have "sharp corners" where they cannot be differentiated. A two variable example would be something like $f(x,y) = |x-y|$. This function isn't differentiable at any point where $x = y$ (the graph of this function looks like a creased piece of paper with the fold along the line $y = x$).

Now for a final theorem which lays all concerns to rest.

Theorem: Continuous partials implies differentiability.

I will not provide a proof of this theorem. Its proof is more technical than the last two results. Also, just as with the other theorems, the converse of this theorem does not hold. There are differentiable functions which have discontinuous partials. Again, I will forgo giving a concrete counterexample – such an example is tricky to cook up. Every function we run into in this class will have continuous partials (where they are defined). This means that for us, computing partials (and calling on this theorem) will *prove* differentiability.

Example: $f(x,y) = \sin(xy^2)$ has partial derivatives $f_x = \cos(xy^2) \cdot y^2$ and $f_y = \cos(xy^2) \cdot 2xy$. Since f_x and f_y are continuous everywhere, we have (by this last theorem) that $f(x,y) = \sin(xy^2)$ is differentiable everywhere.

$$\begin{tikzcd}
 & \text{Continuous Partials} \arrow[d] \arrow[ddl, bend right] \arrow[ddr, bend left] & \\
 & \text{Differentiable} \arrow[dl] \arrow[dr] & \\
\text{Partials Exist} & & \text{Continuous}
\end{tikzcd}$$

The figure above summarizes our "big" theorems. Keep in mind that none of the arrows go backwards in general. Well, unless we have single variable functions, then "partials exist" (meaning the derivative exists) is the *same* as "differentiable". But again, that's only for functions of one variable.

Note:

Our textbook (like many Calculus III texts) defines differentiability as follows: $f(x,y)$ is differentiable at $(x,y) = (a,b)$ if the partial derivatives $f_x(a,b)$ and $f_y(a,b)$ exist as well as locally defined functions $\epsilon_1(x,y)$ and $\epsilon_2(x,y)$ such that

$$f(x,y) = \underbrace{f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)}_{\text{linearization}} + \underbrace{\epsilon_1(x,y)(x-a) + \epsilon_2(x,y)(y-b)}_{\text{error terms}}$$

and in addition $\displaystyle\lim_{(x,y)\to(a,b)} \epsilon_1(x,y) = 0$ and $\displaystyle\lim_{(x,y)\to(a,b)} \epsilon_2(x,y) = 0$.

It turns out that this definition and our definition (in the case of two variables) are equivalent. You can see that the book's definition says that $f(x,y)$ is differentiable at $(x,y) = (a,b)$ if $f(x,y)$ is equal to its linearization at $(x,y) = (a,b)$ plus some suitably structured error terms. If the term $f(a,b)$ is brought over to the other side of the defining equation, we get something like $dz = f_x\,dx + f_y\,dy + \epsilon_1\,dx + \epsilon_2\,dy$ so $dz = (f_x + \epsilon_1)\,dx + (f_y + \epsilon_2)\,dy$. In other words, f_x and f_y don't perfectly capture the change in f, but come close (up to some error terms).

It is my opinion that the definition presented in this handout is more conceptually clear. It also has the advantage of being immediately generalizable to functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ and even to functions on arbitrary Banach spaces (whatever those are). Adapting our definition to functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ leads to the definition of the *Jacobian matrix*. More on that later.