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 $f'(a) = \lim_{h \to 0} \dfrac{f(a+h) - f(a)}{h}$ 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 $\lim_{x \to a} \dfrac{f(x) - f(a)}{x - a}$.

Let's manipulate our current definition a little more. We have $f'(a) = \lim_{x \to a} \dfrac{f(x) - f(a)}{x - a}$ 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 $f : \mathbb{R}^n \to \mathbb{R}^m$ and a fixed point $\mathbf{a} = (a_1, \ldots, a_n)$ [actually, we only need f to be defined on an open ball containing $\mathbf{a}$]. Fix notation

$$f(\mathbf{x}) = f(x_1, \ldots, x_n) = (f^1(x_1, \ldots, x_n), \ldots, f^m(x_1, \ldots, x_n))$$

so that $f^j(\mathbf{x})$ is a scalar valued function on $\mathbb{R}^n$ (i.e. $f^j : \mathbb{R}^n \to \mathbb{R}$).

Definition: $f(\mathbf{x})$ is **differentiable** at $\mathbf{x} = \mathbf{a}$ if there is some linear map $J_\mathbf{a} : \mathbb{R}^n \to \mathbb{R}^m$ such that...

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

Here we have $L(\mathbf{x}) = f(\mathbf{a}) + J_\mathbf{a}(\mathbf{x} - \mathbf{a})$ is the *linearization* of $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{a}$ and $J_\mathbf{a}$ is the *Jacobian* of $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{a}$.

Theorem: If $f(\mathbf{x})$ is differentiable at $\mathbf{x} = \mathbf{a}$, then the Jacobian at $\mathbf{x} = \mathbf{a}$ is unique.

Proof: *(Sketch)* Let J and K both be Jacobians at $\mathbf{x} = \mathbf{a}$. Then $\|J(\mathbf{x} - \mathbf{a}) - K(\mathbf{x} - \mathbf{a})\| = \|f(\mathbf{x}) - [f(\mathbf{a}) + K(\mathbf{x} - \mathbf{a})] - (f(\mathbf{x}) - [f(\mathbf{a}) + J(\mathbf{x} - \mathbf{a})])\| \leq \|f(\mathbf{x}) - [f(\mathbf{a}) + K(\mathbf{x} - \mathbf{a})]\| + \|f(\mathbf{x}) - [f(\mathbf{a}) + J(\mathbf{x} - \mathbf{a})]\|$. Thus

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{\|J(\mathbf{x} - \mathbf{a}) - K(\mathbf{x} - \mathbf{a})\|}{\|\mathbf{x} - \mathbf{a}\|} \leq \lim_{\mathbf{x} \to \mathbf{a}} \frac{\|f(\mathbf{x}) - [f(\mathbf{a}) + K(\mathbf{x} - \mathbf{a})]\|}{\|\mathbf{x} - \mathbf{a}\|} + \frac{\|f(\mathbf{x}) - [f(\mathbf{a}) + J(\mathbf{x} - \mathbf{a})]\|}{\|\mathbf{x} - \mathbf{a}\|} = 0 + 0 = 0$$

Sketchy part: This means that the norm of the operator $J - K$ is $\|J - K\| = 0$ and thus $J - K = 0$ so that $J = K$. ■

Often we prefer to represent linear operators as (coordinate) matrices. First, consider projecting onto the i-th output coordinate: f projects to f^i. Say $J_\mathbf{a}$ projects to $J^i_\mathbf{a}$ (in fact, let's identify this linear map its coordinate matrix/vector). Thus $f^i(\mathbf{x}) \approx L^i(\mathbf{x}) = f^i(\mathbf{a}) + J^i_\mathbf{a} \bullet (\mathbf{x} - \mathbf{a})$. In particular, we have that $\lim_{\mathbf{x} \to \mathbf{a}} \dfrac{f^i(\mathbf{x}) - [f^i(\mathbf{a}) + J^i_\mathbf{a} \bullet (\mathbf{x} - \mathbf{a})]}{\|\mathbf{x} - \mathbf{a}\|} = 0$. Since this limit exists along any continuous curve through $\mathbf{x} = \mathbf{a}$, we have (approaching along $x_k = a_k$, $k \neq j$ and $x_j = t$), $\lim_{t \to a_j} \dfrac{f^i(a_1, \ldots, a_{j-1}, t, a_{j+1}, \ldots, a_n) - [f^i(\mathbf{a}) + J^i_\mathbf{a} \bullet (0, \ldots, 0, t - a_j, 0, \ldots, 0)]}{|t - a_j|} = 0$. Denote the j-th entry of the coordinate matrix/vector $J^i_\mathbf{a}$ by $J_{ij}(\mathbf{a})$. Then we have $\lim_{t \to a_j} \dfrac{f^i(a_1, \ldots, a_{j-1}, t, a_{j+1}, \ldots, a_n) - [f^i(\mathbf{a}) + J_{ij}(\mathbf{a})(t - a_j)]}{t - a_j} = 0$. After some algebraic manipulations, we get $\lim_{t \to a_j} \dfrac{f^i(a_1, \ldots, a_{j-1}, t, a_{j+1}, \ldots, a_n) - f^i(a_1, \ldots, a_n)}{t - a_j} = J_{ij}(\mathbf{a})$. We call the $m \times n$ matrix with (i,j)-entries $J_{ij}(\mathbf{a})$ the *Jacobian matrix* of $f(\mathbf{x})$ based at $\mathbf{x} = \mathbf{a}$ (this is the coordinate matrix for the operator $J_\mathbf{a}$).

Theorem: If $f(\mathbf{x})$ is differentiable at $\mathbf{x} = \mathbf{a}$, then its component functions' partial derivatives exist (at $\mathbf{x} = \mathbf{a}$). Moreover, the (i,j)-entry of the Jacobian matrix of $f(\mathbf{x})$ is $J_{ij}(\mathbf{a}) = (f^i)_{x_j}(\mathbf{a})$ (i.e. the partial derivative of the i-th component function with respect to the j-th input variable). In particular, the partials of the component functions of f exist (at $\mathbf{x} = \mathbf{a}$).

The converse of this theorem does not hold! From Calculus III, 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 don't tend to run into this problem in practice.

We learn in Calculus I that differentiable functions are always continuous functions. This is still true.

Theorem: [Differentiability implies continuity] Let $f(\mathbf{x})$ be differentiable at $\mathbf{x} = \mathbf{a}$. Then $f(\mathbf{x})$ is continuous at $\mathbf{x} = \mathbf{a}$.

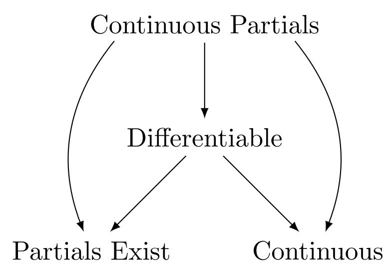
Proof: Suppose that $f(\mathbf{x})$ is differentiable at $\mathbf{x} = \mathbf{a}$. Then $\displaystyle\lim_{\mathbf{x}\to\mathbf{a}} \frac{\|f(\mathbf{x}) - L(\mathbf{x})\|}{\|\mathbf{x} - \mathbf{a}\|} = 0$. For this limit to be 0, the numerator must limit to 0. This means that $\displaystyle\lim_{\mathbf{x}\to\mathbf{a}}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

Theorem: [Continuous partials implies differentiability] Let the component functions of $f(\mathbf{x})$ have continuous (first) partials at $\mathbf{x} = \mathbf{a}$. Then $f(\mathbf{x})$ is differentiable at $\mathbf{x} = \mathbf{a}$.

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 typically run into has continuous partials (where they are defined). This means that for us, computing partials (and calling on this theorem) will *prove* differentiability.



The figure above summarizes our main differentiability 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: Most Calculus III texts define 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 scalar valued functions of two variables) are equivalent. You can see that this 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).