Differentiable functions of several variables

Sean Fitzpatrick

June 13, 2018

Since I'm using a different defintion of differentiability than the one found in most textbooks, I've prepared this handout with the details. I'll give the general result for n variables, using vector notation. Feel free to translate back into coordinate notation for the cases of two or three variables.

1   Real-valued functions

Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a given function of n variables (you can assume $n = 1, 2$ or 3 if you prefer). Let us denote a point $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ using the vector $\mathbf{x} = \langle x_1, x_2, \ldots, x_n \rangle$, so that $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$. Let $\mathbf{a} = \langle a_1, a_2, \ldots, a_n \rangle$ denote a fixed point $(a_1, a_2, \ldots, a_n) \in D$.

Definition 1. *The **gradient** of f at $\mathbf{a} \in D$ is the vector $\nabla f(\mathbf{a})$ defined by*

$$\nabla f(\mathbf{a}) = \left\langle \frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right\rangle.$$

*The **linearization** of f at $\mathbf{a}$ is the function $L_{\mathbf{a}}(\mathbf{x})$ defined by*

$$L_{\mathbf{a}}(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}).$$

Note that when $n = 1$, we get the linearization $L_a(x) = f(a) + f'(a)(x - a)$, which is the definition of the linearization given in a single-variable calculus course. (You might also notice that $L_a(x)$ is the first-degree Taylor polynomial of f about $x = a$. The same is true of the linearization of f for more than one variable, although we will not be considering Taylor polynomials in several variables.) If $n = 2$, we get the linearization given in class:

$$\begin{aligned} L_{(a,b)}(x, y) &= f(a, b) + \langle f_x(a, b), f_y(a, b) \rangle \cdot \langle x - a, y - b \rangle \\ &= f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b). \end{aligned}$$

Remark 2. If the point $\mathbf{a} \in D$ at which we are considering the linearization is fixed/clear in a given problem, we can drop the subscript in the notation, and simply write $L(\mathbf{x})$ instead of $L_{\mathbf{a}}(\mathbf{x})$.

Definition 3. *We say that f is* **differentiable** *at* $\mathbf{a} \in D$ *if* $\nabla f(\mathbf{a})$ *exists, and* $f(\mathbf{x})$ *and* $L_{\mathbf{a}}(\mathbf{x})$ *agree to first order at* $\mathbf{a}$; *that is, if*

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - L_{\mathbf{a}}(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} = \lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \langle \mathbf{x} - \mathbf{a} \rangle}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

What this definition says is that the linearization $L_{\mathbf{a}}(\mathbf{x})$ is a good linear approximation to f at $\mathbf{a}$. In fact, it's the *only* (and hence, best) linear approximation: if a linear approximation exists, it has to be $L_{\mathbf{a}}(\mathbf{x})$. If you want to see why this has to be true, recall that since the above limit exists, we have to be able to evaluate it along any path we like. In particular, we can choose a path where we let $x_i \to a_i$ for some $i \in \{1, \dots, n\}$, while setting $x_j = a_j$ for all $j \neq i$. (What do you get if you evaluate the limit along such a path?)

Remark 4. Note that $\|\mathbf{x} - \mathbf{a}\| = \sqrt{(x_1 - a_1)^2 + \cdots + (x_n - a_n)^2}$ is the distance from $\mathbf{x}$ to $\mathbf{a}$. In general, we would say that two functions $f(\mathbf{x})$ and $g(\mathbf{x})$ "agree up to order k" at $\mathbf{a}$ if

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - g(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|^k} = 0.$$

Exercise: Check that, for $n = 1$, two functions f and g agree up to order k at a if and only if their degree k Taylor polynomials are equal. (A similar statement is true in more than one variable.)

Remark 5. When $n = 1$, we have that $\nabla f(a) = f'(a)$ (a vector with one only component is just a number), and existence of $f'(a)$ immediately implies differentiability: since by definition we have

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a},$$

it follows from the limit laws that

$$0 = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} - f'(a) = \lim_{x \to a} \left(\frac{f(x) - f(a)}{x - a} - \frac{f'(a)(x - a)}{x - a} \right)$$
$$= \lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x - a)}{x - a}.$$

If we try to take $\nabla f(\mathbf{a})$ out of the limit to get an expression like that for $f'(a)$, we immediately run into trouble, since we end up with the term $\nabla f(\mathbf{a}) \cdot \left(\dfrac{\mathbf{x} - \mathbf{a}}{\|\mathbf{x} - \mathbf{a}\|} \right)$: in the numerator we have the dot product with the vector $\mathbf{x} - \mathbf{a}$, while in the denominator we have the length of that vector. The only time that these cancel is when $n = 1$. (You might be tempted to write the dot product as $\|\nabla f(\mathbf{a})\| \|\mathbf{x} - \mathbf{a}\| \cos\theta$, so that the $\|\mathbf{x} - \mathbf{a}\|$ terms cancel, but we would still have the problem that $\cos\theta$ depends on the way in which $\mathbf{x}$ approaches $\mathbf{a}$.)

Recall that in one variable, the derivative is often written instead in terms of $h = x - a$, so that

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

In more than one variable, we can define $h_i = x_i - a_i$, for $i = 1, \ldots, n$, or the corresponding vector $\mathbf{h} = \mathbf{x} - \mathbf{a}$. The definition of differentiability then can be written as

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{h}}{\|\mathbf{h}\|} = 0.$$

Note that we want the difference between $f(\mathbf{a} + \mathbf{h})$ and $L_{\mathbf{a}}(\mathbf{h})$ to go to zero faster than $\|\mathbf{h}\|$ goes to zero, and that it only makes sense to divide by the length of $\mathbf{h}$, since division by a vector (or the corresponding point) is not defined.

Finally, since the algebra didn't fit on the blackboard all that well, here is a proof that every differentiable function is continuous:

Theorem 6. *If $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a} \in D$, then f is continuous at $\mathbf{a}$.*

Proof. Suppose that f is differentiable at $\mathbf{a}$. Then we know that

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - L_{\mathbf{a}}(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} = \lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \langle \mathbf{x} - \mathbf{a} \rangle}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

By the definition of continuity, we need to show that $\lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$. We have that

$$\begin{aligned}
f(\mathbf{x}) &= f(\mathbf{a}) + (f(\mathbf{x}) - f(\mathbf{a})) \\
&= f(\mathbf{a}) + (f(\mathbf{x}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) \\
&= f(\mathbf{a}) + \left(\frac{f(\mathbf{x}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})}{\|\mathbf{x} - \mathbf{a}\|} \right) (\|\mathbf{x} - \mathbf{a}\|) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}).
\end{aligned}$$

Thus, taking limits of the above as $\mathbf{x} \to \mathbf{a}$, we find $\lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$, since the first term is a constant $(f(\mathbf{a}))$, the second is the product of two terms that both go to zero (the first term is zero by the definition of differentiability, and clearly $\lim_{\mathbf{x} \to \mathbf{a}} \|\mathbf{x} - \mathbf{a}\| = 0$), and the last term vanishes since it's linear (and thus continuous) in $\mathbf{x}$, and so, by direct substitution, $\lim_{\mathbf{x} \to \mathbf{a}} \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) = \nabla f(\mathbf{a}) \cdot (\mathbf{a} - \mathbf{a}) = 0$. $\square$

2 Vector-valued functions

It's possible to extend Definition 3 to functions $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$. If we apply the convention of representing points in terms of their position vectors to the codomain as well as the domain, we can express such a function as $f = \langle f_1, \ldots, f_n \rangle$, where each function f_i is a real-valued function of n variables.

Remark 7. In this section we need to be a bit more careful about working with vectors. Let us agree that the notation $\langle a_1, a_2, \ldots, a_n \rangle$ is equivalent notation for the **column vector** $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$. The former is preferable within a paragraph since it fits on a single line. The column vector notation is necessary in the context of matrix multiplication: we want an $m \times n$ matrix A to define a function $f : \mathbb{R}^n \to \mathbb{R}^m$ according to

$$f(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

Using column vectors, the most general linear[1] function $L : \mathbb{R}^n \to \mathbb{R}^m$ can be written as

$$L(\mathbf{x}) = A\mathbf{x} + \mathbf{b},$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, and A is an $m \times n$ matrix, so that $A\mathbf{x}$ is an $m \times 1$ matrix; that is, a column vector in $\mathbb{R}^m$.

We again define differentiability to mean the existence of a "good" linear approximation. Since $f(\mathbf{x})$ and $L_{\mathbf{a}}(\mathbf{x})$ are now vectors, saying that $L_{\mathbf{a}}(\mathbf{x})$ is a good approximation to $f(\mathbf{x})$ is the same as saying that the magnitude $\| f(\mathbf{x}) - L_{\mathbf{a}}(\mathbf{x}) \|$ is small relative to $\| \mathbf{x} - \mathbf{a} \|$; that is,

$$lim_{\mathbf{x} \to \mathbf{a}} \frac{\| f(\mathbf{x}) - L_{\mathbf{a}}(\mathbf{x}) \|}{\| \mathbf{x} - \mathbf{a} \|} = 0.$$

With a bit of work, we can see what the function $L_{\mathbf{a}}$ has to be. First of all, we need to have $L\mathbf{a}(\mathbf{a}) = f(\mathbf{a})$, or else the numerator won't go to zero as $\mathbf{x} \to \mathbf{a}$. This forces us to take (exercise) $L_{\mathbf{a}}(\mathbf{x}) = A(\mathbf{x} - \mathbf{a}) + f(\mathbf{a})$. Thus, we have

$$f(\mathbf{x}) - L_{\mathbf{a}}(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{a}) - A(\mathbf{x} - \mathbf{a}).$$

This should be reminding you a bit of the numerator in the one variable case: $f(x) - f(a) - f'(a)(x - a)$. With that in mind, this matrix A, whatever it is, should be thought of as *the* derivative in the general case of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. To see what it should be, we try

---

[1] Here, we're using the calculus meaning of linear, generalizing from the linear function $f(x) = ax + b$ in one variable. In linear algebra, linear functions are required to preserve the zero vector (origin) so $\mathbf{b} = \mathbf{0}$ is required in that context.

letting $\mathbf{x} \to \mathbf{a}$ along different paths. If we consider the path $x_1 = a_1 + t, x_2 = a_2, \ldots, x_n = a_n$ (that is, varying x_1 while holding the other variables constant) then

$$\mathbf{x} - \mathbf{a} = \begin{bmatrix} a_1 + t \\ a_2 \\ \vdots \\ a_n \end{bmatrix} - \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = t \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

so $A(\mathbf{x} - \mathbf{a})$ gives us t times the first column of A (check this). Rearranging the limit gives us

$$\lim_{t \to 0} \left[\frac{f(a_1 + t, a_2, \ldots, a_n) - f(a_1, a_2, \ldots, a_n)}{t} - \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} \right] = 0.$$

If we write $f(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$, then we have

$$\lim_{t \to 0} \frac{f(a_1 + t, a_2, \ldots, a_n) - f(a_1, a_2, \ldots, a_n)}{t} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) \\ \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) \end{bmatrix},$$

and this should equal to the first column of A! Repeating this for each variable, we see that the matrix A is exactly the Jacobian matrix of partial derivatives. That is,

$$A = D_{\mathbf{a}} f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{x_n} \end{bmatrix}.$$

In particular, note that for a function $f : \mathbb{R}^n \to \mathbb{R}$, we recover the gradient vector. (Technically, the derivative in this case is a **row** vector, not a column vector. The gradient is the corresponding column vector, so it's not the "true" derivative, but dual to it. Note that multiplying a row vector by a column vector is the same as taking the dot product of two column vectors.

This definition also accounts for parametric curves, viewed as vector-valued functions of one variable. If $\mathbf{r} : \mathbb{R} \to \mathbb{R}^n$ defines a parametric curve, then the derivative $\mathbf{r}'(t) = \begin{bmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}$ as introduced in Calculus III is the same as the one obtained using this definition.