Proof that Continuously differentiable implies differentiable

Sean Fitzpatrick

June 13, 2018

As mentioned in Assignment #2, every continuously differentiable function is differentiable. A proof of this is available in the appendix of Stewart's textbook; however, Stewart only deals with the case of a real-valued function of two variables, so just for fun, I'll do the most general case. It's a good example of how our definition of differentiability and the method of $\epsilon - \delta$ proofs can be put to work.

Note: this is an update of an old handout, and I didn't have time to update the notation. On this handout the notation $\mathbf{D}f(\mathbf{x})$ corresponds to the derivative matrix, which I wrote as $D_{\mathbf{x}}f$ on the assignment.

Theorem 0.1. *Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a given function defined on an open subset U of $\mathbb{R}^n$. Suppose that all partial derivatives of f exist and are continuous on an open neighbourhood of a point $\mathbf{x}_0 \in U$. Then f is differentiable at $\mathbf{x}_0$.*

Proof. Write $f = \langle f_1, \ldots, f_m \rangle$ in terms of its components with respect to the standard unit basis vectors in $\mathbb{R}^m$. We need to show that if each partial derivative $\dfrac{\partial f_i}{\partial x_j}$ is continuous in a neighbourhood of $\mathbf{x}_0$, then

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0) - \mathbf{D}f(\mathbf{x}_0)\cdot\mathbf{h}\|}{\|\mathbf{h}\|} = 0.$$

Let $\epsilon > 0$ be given. Let $\mathbf{X}(\mathbf{h}) \in \mathbb{R}^m$ be the vector-valued function of $\mathbf{h}$ in the numerator of the above limit, defined by

$$\mathbf{X}(\mathbf{h}) = f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0) - \mathbf{D}f(\mathbf{x}_0)\cdot\mathbf{h}.$$

Since $\|\mathbf{X}(\mathbf{h})\| \leq |X_1(\mathbf{h})| + \cdots + |X_m(\mathbf{h})|$, where the X_i denote the components of $\mathbf{X}$, it is enough to show that for each $i = 1, \ldots m$, we can make

$$\frac{|X_i(\mathbf{h})|}{\|\mathbf{h}\|} < \epsilon/m,$$

by choosing $\mathbf{h}$ such that $\|\mathbf{h}\|$ is sufficiently small. Note that

$$X_i(\mathbf{h}) = f_i(\mathbf{x}_0+\mathbf{h}) - f_i(\mathbf{x}_0) - \nabla f_i(\mathbf{x}_0)\cdot\mathbf{h}$$

for each $i = 1, \dots m$. Let us denote the components of $\mathbf{x}_0$ and $\mathbf{h}$ by $\mathbf{x}_0 = \langle x_{0,1}, x_{0,2}, \dots, x_{0,n}\rangle$ and $\mathbf{h} = \langle h_1, \dots, h_n\rangle$, respectively. (There's no real way to do this without the notation getting messy.) We can then re-write the difference $f_i(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)$ as follows:

$$\begin{aligned} f_i(x_{0,1} + h_1, \dots x_{0,n} + h_n) - f_i(x_{0,1}, \dots, x_{0,n}) = &\, f_i(x_{0,1} + h_1, \dots x_{0,n} + h_n) \\ &- f_i(x_{0,1}, x_{0,2} + h_2, \dots, x_{0,n} + h_n) \\ &+ f_i(x_{0,1}, x_{0,2} + h_2, \dots, x_{0,n} + h_n) \\ &- f_i(x_{0,1}, x_{0,2}, x_{0,3} + h_3, \dots x_{0,n} + h_n) \\ &\vdots \\ &+ f_i(x_{0,1}, \dots, x_{0,n-1}, x_{0,n} + h_n) \\ &- f_i(x_{0,1}, \dots, x_{0,n}). \end{aligned}$$

(I warned you it would get messy!) For each $j = 1, \dots n$, we set

$$g_{ij}(y) = f_i(x_{0,1}, \dots, x_{0,j-i}, y, x_{0,j+1} + h_{j+1}, \dots, x_{0,n} + h_n).$$

Notice that the above expansion of $f_i(x_{0,1} + h_1, \dots x_{0,n} + h_n) - f_i(x_{0,1}, \dots, x_{0,n})$ can then be written in the form

$$\begin{aligned} f_i(x_{0,1} + h_1, \dots x_{0,n} + h_n) - f_i(x_{0,1}, \dots, x_{0,n}) = &\, (g_{i1}(x_{0,1} + h_1) - g_{i1}(x_{0,1})) + \\ &\cdots + (g_{in}(x_{0,n} + h_n) - g_{in}(x_{0,n}))\,. \end{aligned}$$

Now, since the partial derivatives of each of the f_i are defined in an open neighbourhood N of $\mathbf{x}_0$, there is a $\delta_1 > 0$ such that $\|\mathbf{h}\| < \delta$ implies that $\mathbf{x}_0 + \mathbf{h} \in N$, and therefore each function $g_{ij}(y)$ is differentiable for any y between $x_{0,j}$ and $x_{0,j} + h_j$. Applying the mean value theorem, we can find some $\tilde{x}_j$ between $x_{0,j}$ and $x_{0,j} + h_j$ such that

$$g_{ij}(x_{0,j} + h_j) - g_{ij}(x_{0,j}) = {g_{ij}}'(\tilde{x}_j)h_j = \frac{\partial f_i}{\partial x_j}(\tilde{\mathbf{x}}_j),$$

where $\tilde{\mathbf{x}}_j = \langle x_{0,1}, \dots, x_{0,j-1}, \tilde{x}_j, x_{0,j+1} + h_j, \dots, x_{0,n} + h_n\rangle$. Thus we can write

$$f_i(x_{0,1} + h_1, \dots x_{0,n} + h_n) - f_i(x_{0,1}, \dots, x_{0,n}) = \frac{\partial f_i}{\partial x_1}(\tilde{\mathbf{x}}_1)h_1 + \cdots \frac{\partial f_i}{\partial x_n}(\tilde{\mathbf{x}}_n)h_n.$$

Now, the other part of the component $X_i(\mathbf{h})$ defined above is given by

$$\nabla f_i(\mathbf{x}_0) \cdot \mathbf{h} = \frac{\partial f_i}{\partial x_1}(\mathbf{x}_0)h_1 + \cdots + \frac{\partial f_i}{\partial x_n}(\mathbf{x}_0)h_n.$$

Combining these, we find that

$$\frac{|X_i(\mathbf{h})|}{\|\mathbf{h}\|} = \left|\left(\frac{\partial f_i}{\partial x_1}(\tilde{\mathbf{x}}_1) - \frac{\partial f_i}{\partial x_1}(\mathbf{x}_0)\right)\frac{h_1}{\|\mathbf{h}\|} + \cdots + \left(\frac{\partial f_i}{\partial x_n}(\tilde{\mathbf{x}}_n) - \frac{\partial f_i}{\partial x_n}(\mathbf{x}_0)\right)\frac{h_n}{\|\mathbf{h}\|}\right|. \tag{1}$$

Now, since each partial derivative $\dfrac{\partial f_i}{\partial x_j}$ is continuous in a neighbourhood of $\mathbf{x}_0$, there exists a $\delta_2 > 0$ such that $\|\tilde{\mathbf{x}}_j - \mathbf{x}_0\| \leq \|\mathbf{h}\| < \delta_2$ implies that

$$\left|\frac{\partial f_i}{\partial x_j}(\tilde{\mathbf{x}}_j) - \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)\right| < \frac{\epsilon}{mn},$$

for each $j = 1, \ldots, n$. (Notice that by the way the numbers $\tilde{x}_j$ were defined via the Mean Value Theorem, we must have $\|\tilde{\mathbf{x}}_j - \mathbf{x}_0\| \leq \|\mathbf{h}\| < \delta_2$ for each $j = 1, \ldots n$. For each of the partial derivatives we might have a different δ that yields the above inequality; we take δ_2 to be the smallest of these.)

Finally, we use the triangle inequality on (1), together with the fact that $\dfrac{|h_j|}{\|\mathbf{h}\|} \leq 1$, to get

$$\begin{aligned}\frac{|X_i(\mathbf{h})|}{\|\mathbf{h}\|} &\leq \left|\frac{\partial f_i}{\partial x_1}(\tilde{\mathbf{x}}_1) - \frac{\partial f_i}{\partial x_1}(\mathbf{x}_0)\right| \frac{|h_1|}{\|\mathbf{h}\|} + \cdots \left|\frac{\partial f_i}{\partial x_n}(\tilde{\mathbf{x}}_n) - \frac{\partial f_i}{\partial x_n}(\mathbf{x}_0)\right| \frac{|h_n|}{\|\mathbf{h}\|} \\ &< \frac{\epsilon}{nm} + \cdots + \frac{\epsilon}{nm} = \frac{\epsilon}{m}.\end{aligned}$$

Thus, if we take $\|\mathbf{h}\| < \delta$, where $\delta = \min\{\delta_1, \delta_2\}$, we get that $\dfrac{X_i(\mathbf{h})}{\|\mathbf{h}\|} < \epsilon/m$ for each $i = 1, \ldots, m$, and thus $\dfrac{\|\mathbf{X}(\mathbf{h})\|}{\|\mathbf{h}\|} < \epsilon$.

Since $\epsilon > 0$ was arbitrary, it must be the case that $\dfrac{\|\mathbf{X}(\mathbf{h})\|}{\|\mathbf{h}\|} \to 0$ as $\mathbf{h} \to \mathbf{0}$. $\square$