Differentiation and Linear Approximation

Let $U \subset \mathbb{R}^n$ and let $f : U \to \mathbb{R}^m$ be a given function. Write $f = (f_1, f_2, \ldots, f_m)$ where each f_i is a real-valued function with domain U. We defined f to be differentiable at $\mathbf{x}_0 \in \mathbb{R}^n$ provided the partial derivatives

$$\frac{\partial f_i}{\partial x_i}(\mathbf{x}_0),\ 1 \leq i \leq m,\ 1 \leq j \leq n$$

all exist and

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

where

$$\mathbf{D}f(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

all partial derivatives being evaluated at $\mathbf{x}_0$. The point of this note is to give some justification for this definition using linear approximations.

For a function of a single variable $f : (a, b) \to \mathbb{R}$ that is differentiable at a point $x_0 \in (a, b)$ we have the linear approximation

$$L(x) = f(x_0) + f'(x_0)(x - x_0).$$

This is exactly the same as the tangent line. Heuristically, for x near x_0, $L(x)$ should approximate the value of $f(x)$. This is because $f'(x_0)$ represents the rate of change of f at x_0, so if we move away from x_0 by a small amount h then f should change by about $f'(x_0)h$. In symbols

$$f(x_0 + h) \approx f(x_0) + f'(x_0)h.$$

If we write $x = x_0 + h$ then this becomes

$$f(x) \approx L(x).$$

But this argument is not precise. Let's make it a little more rigorous. We want to consider the difference $f(x) - L(x)$. We have

$$\begin{aligned} f(x) - L(x) &= f(x) - f(x_0) - f'(x_0)(x - x_0) \\ &= (x - x_0)\left(\frac{f(x) - f(x_0)}{x - x_0} - f'(x_0)\right). \end{aligned}$$

Dividing both sides by $(x - x_0)$ we see that

$$\lim_{x \to x_0} \frac{f(x) - L(x)}{x - x_0} = \lim_{x \to x_0} \left(\frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right) = 0$$

by the definition of the derivative. The point of this? The difference $f(x) - L(x)$ approaches 0 as x approaches x_0. In fact, it goes to 0 so fast that we still get 0 in the limit when we divide by $x - x_0$. So $L(x)$ is indeed a pretty good approximation to $f(x)$, at least for x near x_0.

So let's go back to our general function $f : U \to \mathbb{R}^n$ and the general definition of differentiable and try to interpret it as saying that a certain linear approximation to f is "good". First of all, how should we approximate f? Here's a heuristically reasonable way. Let's assume first that f is real-valued. Suppose we change $\mathbf{x}_0 \in \mathbb{R}^n$ by $\mathbf{h} \in \mathbb{R}^n$. That is, we move from $\mathbf{x}_0$ to $\mathbf{x}_0 + \mathbf{h}$. How can we gauge how much f has changed? Well, if we write

$$\mathbf{h} = (h_1, h_2, \ldots, h_n)$$

then in moving from $\mathbf{x}_0$ to $\mathbf{x}_0 + \mathbf{h}$ we change the ith coordinate of x_0 by h_i. Since $\partial f / \partial x_i$ gives the rate of change of f in the x_i-direction, increasing the ith coordinate by h_i should change f by $(\partial f / \partial x_i) h_i$. If we add the contribution of moving in each coordinate direction, we find that we would expect to have

$$f(\mathbf{x}_0 + \mathbf{h}) \approx f(\mathbf{x}_0) + \frac{\partial f}{\partial x_1} h_1 + \frac{\partial f}{\partial x_2} h_2 + \cdots + \frac{\partial f}{\partial x_n} h_n.$$

Note that we have used the same reasoning as in the single variable case, but we used it in each coordinate direction.

Now for the general situation. As in the introduction, let $f = (f_1, f_2, \cdots, f_m)$. Applying the heuristic reasoning of the previous paragraph to each f_i we expect to have (thinking of our vectors as columns)

$$\begin{aligned} f(\mathbf{x}_0 + \mathbf{h}) &\approx \begin{pmatrix} f_1(\mathbf{x}_0) \\ f_2(\mathbf{x}_0) \\ \vdots \\ f_m(\mathbf{x}_0) \end{pmatrix} + \begin{pmatrix} \frac{\partial f_1}{\partial x_1} h_1 + \frac{\partial f_1}{\partial x_2} h_2 + \cdots + \frac{\partial f_1}{\partial x_n} h_n \\ \frac{\partial f_2}{\partial x_1} h_1 + \frac{\partial f_2}{\partial x_2} h_2 + \cdots + \frac{\partial f_2}{\partial x_n} h_n \\ \vdots \\ \frac{\partial f_m}{\partial x_1} h_1 + \frac{\partial f_m}{\partial x_2} h_2 + \cdots + \frac{\partial f_m}{\partial x_n} h_n \end{pmatrix} \\ &= f(\mathbf{x}_0) + \mathbf{D} f(\mathbf{x}_0) \mathbf{h}. \end{aligned}$$

Thus, writing $\mathbf{x} = \mathbf{x}_0 + \mathbf{h}$, it seems reasonable to take as our linear approximation

$$L(\mathbf{x}) = f(\mathbf{x}_0) + \mathbf{D} f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$$

(compare this to the linear approximation of a real-valued function of a single variable).

Let's look at the definition of differentiable again. With the definition of $L(\mathbf{x})$ we have just made, being differentiable is the same thing as

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{||f(\mathbf{x}) - L(\mathbf{x})||}{||\mathbf{x} - \mathbf{x}_0||} = 0.$$

As in the single variable case, we can interpret this as saying that if f is differentiable at $\mathbf{x}_0$ then the "reasonable" approximation $L(\mathbf{x})$ to f is "good" near $\mathbf{x}_0$.