Notes for Section 2.3

Let us learn about the derivatives and the differentiability of multivariable functions. For starters, let us look at functions $f : S \to T$ where $S \subseteq \mathbb{R}^n$ and $T \subseteq \mathbb{R}$. You can think of such functions as having an input variable $\vec{x}$ that is a vector in $\mathbb{R}^n$ or as having n real number inputs $(x_1, \ldots, x_n)$, while the output is just a real number.

The section starts off with partial derivatives, which is a pretty easy concept to digest. They tell us about the rate of change of f with respect to changing one of its input variables while keeping the other $n-1$ constant. In other words, we are treating f as a function of a single variable while the other $n-1$ variables become constant parameters. For this reason, the definition of the partial derivative $\frac{\partial f}{\partial x_i}$ on p. 106 is very similar to the definition of the derivative of a single variable function. Calculating partial derivatives is also a straightforward task because it can be done the same way as finding the derivative of a single variable function. Unfortunately, partial derivatives are only part of the story about differentiation in the multivariable setting. This is because the partial derivatives encode only limited information about how the output values of f respond to changes in the input. The input $\vec{x}$ to f can change by having several of its coordinates change at the same time.

In single variable calculus, you defined a function f to be differentiable at a point x_0 if the derivative $f'(x_0)$ exists. The corresponding notion in multivariable calculus would be the existence of the partial derivatives. But as I explained above, that is not a strong enough condition to say that the behavior of f near the point $\vec{x}_0$ is described well by the partial derivatives. To find the right way to extend differentiability to multivariable functions, we need to look at it in a different way. In single variable calculus, if f is differentiable at x_0, then $f'(x_0)$ exists, and so does the tangent line

$$y = f'(x_0)(x - x_0) + f(x_0)$$

to f at x_0. In fact, the basic idea of differential calculus is that the behavior of the tangent line near x_0 is a good approximation to how f behaves near x_0. To make this more precise, let $l(x) = f'(x_0)(x - x_0) + f(x_0)$ be the linear function whose graph is the tangent line and consider the difference $E(x) = f(x) - l(x)$. You can think about it as the error in using $l(x)$ to approximate the value of $f(x)$, hence the letter E. It is quite clear that

$$\lim_{x \to x_0} E(x) = \lim_{x \to x_0} [f(x) - l(x)] = 0$$

since f and l are both continuous functions and their values both approach the same number $f(x_0) = l(x_0)$ as $x \to x_0$. But any line that passes through the point $(x_0, f(x_0))$ would do that. What makes the tangent line special is that

$$\begin{aligned}
\lim_{x \to x_0} \frac{E(x)}{x - x_0} &= \lim_{x \to x_0} \frac{f(x) - l(x)}{x - x_0} \\
&= \lim_{x \to x_0} \frac{f(x) - [f'(x_0)(x - x_0) + f(x_0)]}{x - x_0} \\
&= \lim_{x \to x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} \\
&= \lim_{x \to x_0} \left[\frac{f(x) - f(x_0)}{x - x_0} - \frac{f'(x_0)(x - x_0)}{x - x_0} \right] \\
&= \lim_{x \to x_0} \left[\frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right]
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} - \lim_{x \to x_0} f'(x_0) \\
&= f'(x_0) - f'(x_0) \\
&= 0.
\end{aligned}
$$

Think about what this says: as x gets close to x_0, the error in using l to approximate f decreases faster than the difference between x and x_0. So when x is close to x_0, and so $x - x_0$ is a small number (close to 0), $E(x)$ must be a really tiny number. This is what we mean when we say $l(x)$ is a linear function that is a good approximation to $f(x)$ near x_0.

So far, what we have shown is that a function of real numbers f that has a derivative at x_0 also has a good linear approximation near x_0. We can turn this around and define that f is differentiable at x_0 if there a linear function $l(x) = mx + b$ such that l is a good approximation to f near x_0 in the sense that the error $E(x) = f(x) - l(x)$ satisfies

$$
0 = \lim_{x \to x_0} \frac{E(x)}{x - x_0} = \lim_{x \to x_0} \frac{f(x) - l(x)}{x - x_0}.
$$

We have already shown that if $f'(x_0)$ exists, then such a linear function exists. It is easy enough to show that the converse is also true: if such a linear function exists, then $f'(x_0)$ also exists. In fact, it will not come as a surprise that $f'(x_0) = m$, the slope of l. So in single variable calculus, the existence of a good linear approximation at a point x_0 is equivalent to the existence of the derivative at x_0. Therefore either can be used as the definition of differentiability. The existence of the derivative is less abstract and therefore it is what students in single variable calculus classes are typically taught.

It is the idea of a good linear approximation that is the right way to extend differentiability to multivariable functions. What we expect from differential calculus is to allow us to approximate the behavior of a well-behaved function by a linear function. A multivariable function is well-behaved at a point $\vec{x}_0$ if its graph looks like the graph of a linear function when we look at it close enough. Roughly speaking, it should not have holes, breaks, kinks, infinite limits because linear functions do not do any of these things. First, what is a linear function in the multivariable setting? In the current context of functions from $\mathbb{R}^n$ to $\mathbb{R}$, it is a function of the form

$$
l(x_1, x_2, \ldots, x_n) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n + b
$$

for some coefficients $a_1, \ldots, a_n, b \in \mathbb{R}$. Let me digress a bit to note that you could also express this using vectors as

$$
l(\vec{x}) = \vec{a} \cdot \vec{x} + b.
$$

Technically, a better name for such a function is affine, not linear, because linear maps have a definition in linear algebra which conflicts with this definition. But in calculus, linear is the typical term used, and we won't get hung up on the subtleties of terminology. So l is a good approximation to f at $\vec{x}_0$ if the error $E(\vec{x}) = f(\vec{x}) - l(\vec{x})$ gets tiny even relative to the difference between $\vec{x}$ and $\vec{x}_0$ as $\vec{x} \to \vec{x}_0$. That is f is *differentiable* at $\vec{x}_0$ if there is a linear function l such that the error $E(\vec{x}) = f(\vec{x}) - l(\vec{x})$ satisfies

$$
0 = \lim_{\vec{x} \to \vec{x}_0} \frac{E(\vec{x})}{||\vec{x} - \vec{x}_0||} = \lim_{\vec{x} \to \vec{x}_0} \frac{f(\vec{x}) - l(\vec{x})}{||\vec{x} - \vec{x}_0||}.
$$

Notice that the denominator is not $\vec{x} - \vec{x}_0$ but $||\vec{x} - \vec{x}_0||$. This is because $\vec{x} - \vec{x}_0$ is a vector in $\mathbb{R}^n$ and it does not make sense to divide by a vector, and in any case, $||\vec{x} - \vec{x}_0||$ is an appropriate measure of the size of the difference between $\vec{x}$ and $\vec{x}_0$, or the distance between $\vec{x}$ and $\vec{x}_0$.

Let us look at an example to make these abstract ideas easier to digest. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function $f(x, y) = x^2 + y^2$. We will show that there is a good linear approximation l to f near the point $\vec{x}_0 = (1, 2)$. We will let $\vec{x} = (x, y)$, so

$$l(\vec{x}) = l(x, y) = ax + by + c$$

for some $a, b, c \in \mathbb{R}$. To find good candidates for a, b, and c, we will take a hint from single variable calculus and set a and b equal to the partial derivatives at $(1, 2)$ and calculate c so that $f(1, 2) = l(1, 2)$:

$$\begin{aligned} \frac{\partial f}{\partial x} = 2x &\implies a = \frac{\partial f}{\partial x}(1, 2) = 2 \\ \frac{\partial f}{\partial y} = 2y &\implies a = \frac{\partial f}{\partial y}(1, 2) = 4 \end{aligned}$$

and for c

$$\begin{aligned} f(1, 2) &= l(1, 2) \\ 5 &= 2(1) + 4(2) + c \\ c &= -5. \end{aligned}$$

So $l(x, y) = 2x + 4y - 5$. Graph f and l in your favorite 3-D graphing app to see what these surfaces look like. The graph of l is a plane that is tangent to the graph of f at $(1, 2)$. Let us show that l is a good approximation to f near $(1, 2)$.

$$\begin{aligned} \lim_{\vec{x} \to \vec{x}_0} \frac{E(\vec{x})}{||\vec{x} - \vec{x}_0||} &= \lim_{(x,y) \to (1,2)} \frac{f(x, y) - l(x, y)}{\sqrt{(x-1)^2 + (y-2)^2}} \\ &= \lim_{(x,y) \to (1,2)} \frac{x^2 + y^2 - (2x + 4y - 5)}{\sqrt{(x-1)^2 + (y-2)^2}} \\ &= \lim_{(x,y) \to (1,2)} \frac{x^2 - 2x + y^2 - 4y + 5}{\sqrt{(x-1)^2 + (y-2)^2}} \\ &= \lim_{(x,y) \to (1,2)} \frac{x^2 - 2x + 1 + y^2 - 4y + 4}{\sqrt{(x-1)^2 + (y-2)^2}} \\ &= \lim_{(x,y) \to (1,2)} \frac{(x-1)^2 + (y-2)^2}{\sqrt{(x-1)^2 + (y-2)^2}} \\ &= \lim_{(x,y) \to (1,2)} \sqrt{(x-1)^2 + (y-2)^2} \end{aligned}$$

Now you can either convince yourself that

$$\lim_{(x,y) \to (1,2)} [(x-1)^2 + (y-2)^2] = 0,$$

and hence

$$\lim_{(x,y) \to (1,2)} \sqrt{(x-1)^2 + (y-2)^2} = 0,$$

which I will leave as an exercise for you, or you could note that

$$\lim_{(x,y) \to (1,2)} \sqrt{(x-1)^2 + (y-2)^2} = \lim_{\vec{x} \to \vec{x}_0} ||\vec{x} - \vec{x}_0|| = 0$$

as $\vec{x} \to \vec{x}_0$ means exactly that the distance between $\vec{x}$ and $\vec{x}_0$ is approaching 0.

Let me digress a bit and mention that another convenient notation for the partial derivatives in this context, especially when you want to talk about their values at a specific point, is f_x and f_y. So for example, $\frac{\partial f}{\partial x}(1, 2) = f_x(1, 2)$. If you have n input variables, you get $f_{x_1}, \ldots, f_{x_n}$. I wanted to mention this just in case you look in another textbook or on the internet. So the tangent line of single variable calculus becomes a tangent plane for functions from $\mathbb{R}^2$ to $\mathbb{R}$ and in higher dimensions, it becomes what is called a hyperplane. In general, it can be shown that if a function

$f : S \to T$ where $S \subseteq \mathbb{R}^n$ and $T \subseteq \mathbb{R}$ has a good linear approximation $l : \mathbb{R}^n \to \mathbb{R}$ near $\vec{x}_0$ then its partial derivatives must all exist at $\vec{x}_0$ and l must be

$$l(x_1, \ldots, x_n) = \frac{\partial f}{\partial x_1}(\vec{x}_0)x_1 + \cdots + \frac{\partial f}{\partial x_n}(\vec{x}_0)x_n + b$$

where b is such that $f(\vec{x}_0) = l(\vec{x}_0)$. Or more explicitely,

$$l(x_1, \ldots, x_n) = \frac{\partial f}{\partial x_1}(\vec{x}_0)(x_1 - x_1') + \cdots + \frac{\partial f}{\partial x_n}(\vec{x}_0)(x_n - x_n') + f(\vec{x}_0),$$

where $\vec{x}_0 = (x_1', \ldots, x_n')$. We can write this in vector notation as

$$l(\vec{x}) = \big(\frac{\partial f}{\partial x_1}(\vec{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\vec{x}_0)\big) \cdot (\vec{x} - \vec{x}_0) + f(\vec{x}_0),$$

which resembles the equation of the tangent line from single variable calculus. Note that according to what we said, the existence of the partial derivatives of f at $\vec{x}_0$ is a necessary condition of f to be differentiable at $\vec{x}_0$. It turns out it is not a sufficient condition and there are functions whose partial derivatives all exist at a point yet the function fails to differentiable there. I will show you such an example later.

The *gradient* of f is a convenient way to gather all of the partial derivatives of f into a vector:

$$\mathbf{D}\, f(\vec{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right).$$

Note that this is a vector of functions, and in fact, you can think about $\mathbf{D}\, f(\vec{x})$ itself as a multivariable function from $\mathbb{R}^n$ to $\mathbb{R}^n$. So you can evalute $\mathbf{D}\, f(\vec{x})$ at a given point, say at $\vec{x}_0$ and the result is then a vector in $\mathbb{R}^n$. In terms of the gradient, the linear approximation to f is

$$l(\vec{x}) = \mathbf{D}\, f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) + f(\vec{x}_0).$$

Other common notations for the gradient are ∇f and $\operatorname{grad} f$. The symbol ∇ is pronounced nabla.

It is not difficult to extend these ideas to a function $f : S \to T$ where $S \subseteq \mathbb{R}^n$ and $T \subseteq \mathbb{R}^m$. A linear (really affine) function $l : \mathbb{R}^n \to \mathbb{R}^m$ would be one of the form

$$l(\vec{x}) = A\vec{x} + \vec{b}$$

where A is an m by n matrix and $\vec{b} \in \mathbb{R}^m$. So f has a good linear approximation if the error function $E(\vec{x}) = f(\vec{x}) - l(\vec{x})$ satisfies

$$0 = \lim_{\vec{x} \to \vec{x}_0} \frac{||E(\vec{x})||}{||\vec{x} - \vec{x}_0||} = \lim_{\vec{x} \to \vec{x}_0} \frac{||f(\vec{x}) - l(\vec{x})||}{||\vec{x} - \vec{x}_0||}.$$

Notice that we now have the magnitude of $E(\vec{x})$ in the numerator. Since E is now a vector-valued function, it makes sense to measure the size of the error by taking the magnitude of the output. It this case, the right choice of the matrix A turns out to be

$$A = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{x}_0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\vec{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{x}_0) \end{pmatrix}$$

of partial derivatives of f at $\vec{x}_0$, where the $f_1, \ldots, f_m$ are the component functions of f. In general, the matrix

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

contains the partial derivatives of f and you can think about it as playing the same role in multivariable calculus as the derivative function f' of f in single variable calculus. The book calls this matrix of functions $\mathbf{D}\, f(\vec{x})$ the *derivative* of f or the differential of f. It also goes by the names Jacobian or Jacobian matrix of f and total derivative of f. Study Example 6 to see what such a matrix looks like. Using this notation, the linear approximation of f at $\vec{x}_0$ is still

$$l(\vec{x}) = \mathbf{D}\, f(\vec{x}_0)(\vec{x} - \vec{x}_0) + f(\vec{x}_0),$$

only now you need to interpret the vectors $\vec{x}$ and $\vec{x}_0$ as column vectors to be able to make sense of the matrix multiplication.

This section presents two theorems, without proofs. Theorem 8 is the multivariable generalization of a theorem you are familiar with from single variable calculus: a function that is differentiable at a point is also continuous there. In fact, the same proof works in multivariable calculus, other than having to adjust to the more elaborate notation and definition of differentiability. Theorem 9 deserves your attention. I pointed out that having partial derivatives at a point is not enough to guarantee that a function is differentiable there. But having partial derivatives that are continuous near the point is enough to guarantee differentiability. In fact, it is more than enough: having continuous partial derivatives is a stronger condition on a function than differentiability. It is possible to construct a function that is differentiable at a point but its partial derivatives at that point are not continuous in any neighborhood of that point. Theorem 9 can make your life easier when you need to prove that a multivariable function is differentiable: it is often easier to show that the partial derivatives are continuous than to mess with limits to prove that the linear approximation is a good one. See Example 10 for how this works.

I promised you a function whose partial derivatives exist at a point but the function is not differentiable there. The book also presents such an example in Example 9. Note that the function in that example has partial derivatives at $(0, 0)$ but is not continuous there, so it cannot be differentiable at $(0, 0)$ either. But it is a rather contrived example. Here is one that may seem quite tame at first. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = \sqrt{|xy|}$. I will show that f has partial derivatives at $(0, 0)$ but is not differentiable there. First, the partial derivatives:

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{h \to 0} \frac{f(h, 0) - f(0, 0)}{h} = \lim_{h \to 0} \frac{0 - 0}{h} = 0$$

Similarly,

$$\frac{\partial f}{\partial y}(0, 0) = \lim_{h \to 0} \frac{f(0, h) - f(0, 0)}{h} = \lim_{h \to 0} \frac{0 - 0}{h} = 0$$

So the linear approximation to f is

$$l(x, y) = 0(x - 0) + 0(y - 0) + f(0, 0) = 0.$$

But is it a good one? We have $E(x, y) = f(x, y) - l(x, y) = \sqrt{|xy|}$ and

$$\lim_{(x,y)\to(0,0)} \frac{E(x, y)}{\sqrt{(x - 0)^2 + (y - 0)^2}} = \lim_{(x,y)\to(0,0)} \frac{\sqrt{|xy|}}{\sqrt{x^2 + y^2}}.$$

I claim this limit is not 0. Actually, it does not even exist. I will prove to you that it is not 0. If it were 0, then for every neighborhood N of 0 there would have to be a corresponding neighborhood U of $(0, 0)$ such that $f(x, y) \in N$ whenever $(x, y) \in U$. We will show thatfor $N = (-1/2, 1/2)$ there is no such neighborhood U by showing that any neighborhood U of $(0, 0)$ contains some point (x, y) such that $f(x, y) \notin N$. So let U be any neighborhood of $(0, 0)$. Then there must be some $\delta > 0$ such that the open disk $D_\delta(0, 0) \subseteq U$. Choose a number $0 < x < \frac{\delta}{\sqrt{2}}$ and let $y = x$. Then the point (x, y) is in $D_\delta(0, 0) \subseteq U$ since $\sqrt{x^2 + y^2} < \delta$. But

$$\frac{\sqrt{|xy|}}{\sqrt{x^2 + y^2}} = \frac{\sqrt{|x^2|}}{\sqrt{x^2 + x^2}} = \frac{x}{\sqrt{2}x} = \frac{1}{\sqrt{2}} \notin N.$$