February 11, 2019

LECTURE 5: THE RULES OF DIFFERENTIATION.

110.211 HONORS MULTIVARIABLE CALCULUS
PROFESSOR RICHARD BROWN

Synopsis. Here, we bring back the rules for differentiation (used to derive new functions constructed using various combinations of other functions) from Calculus I and use them in our new context. The basic frame for this discussion is, "the rules are the same, but only precisely when they actually make sense." What this means is the focus of this class. Also, we will look at higher derivatives and the notion of a function being differentiable more than once. This involves defining the kth partial of a real-valued (and vector-valued) function and what it means for mixed partials to be equal. The differentiable class of a function is discussed, along with just what kind of object the kth derivative of a real-valued function on n variables is and how it encompasses its n^k partials.

The Rules of Differentiation. The nice thing about calculating derivatives in multivariable calculus is that, in many ways, they follow the same rules as they did in single variable calculus, suitably generalized, of course.

The Constant Multiple Rule. Multiplying a function $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ by a real constant $c \in \mathbb{R}$ affects only the functions values, as in Calculus I functions. The new function $(c\mathbf{f}) : X \subset \mathbb{R}^n \to \mathbb{R}^m$ has a vector output, and a constant times a vector means simply multiplying each component by that constant. Indeed,

$$(c\mathbf{f})(\mathbf{x}) = c \cdot \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} (cf_1)(\mathbf{x}) \\ (cf_2)(\mathbf{x}) \\ \vdots \\ (cf_m)(\mathbf{x}) \end{bmatrix}.$$

The partial derivatives of each f_i are single variable derivatives, where the Constant Multiple Rule held, so

$$\begin{aligned} \frac{\partial (cf_i)}{\partial x_j}(\mathbf{x}) &= \lim_{h \to 0} \frac{(cf_i)(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - (cf)(\mathbf{x})}{h} \\ &= \lim_{h \to 0} \frac{c\left(f_i(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - f(\mathbf{x})\right)}{h} \\ &= c \cdot \left(\lim_{h \to 0} \frac{f_i(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - f(\mathbf{x})}{h} \right) = c \cdot \left(\frac{\partial f_i}{\partial x_j}(\mathbf{x}) \right). \end{aligned}$$

The effect is that the entire derivative matrix is multiplied by c, just as in matrix multiplication by a scalar, so that

$$D(c\mathbf{f})(\mathbf{x}) = c \cdot D\mathbf{f}(\mathbf{x}).$$

The Sum/Difference Rule. The Sum Rule (and hence the Difference Rule when one of the functions is multiplied by -1), is also precisely the same as that for single variable calculus, except that the two functions in the sum must have the same domain and codomain. Only then will the two derivative matrices have the same dimensions, allowing us to actually sum the derivative matrices. So consider $\mathbf{f}, \mathbf{g} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ to m-vector-valued functions on $X \subset \mathbb{R}^n$, and let $\mathbf{h} = \mathbf{f} + \mathbf{g}$. Then

$$D\mathbf{h}(\mathbf{x}) = D\left(\mathbf{f} + \mathbf{g}\right)(\mathbf{x}) = D\mathbf{f}(\mathbf{x}) + D\mathbf{g}(\mathbf{x}).$$

The Product Rule. The Product Rule, in multivariable calculus, can be a bit trickier, given that the product of two vectors may or may not be a vector of the same size: It is for the cross product in $\mathbb{R}^3$). But the dot product of two vectors in $\mathbb{R}^n$ is a scalar. And sometimes, the output of a product of vectors may not even be a vector of any size (look up *outer product*, for example). The tricky part is to ensure that, if two vector-valued functions are differentiable, then so should be the product of those two functions, however, that is defined. And further, to find a rule to write the derivative of the product using the derivatives of the factors. The nice part of all this is that the Product Rule will always hold, as long as the parts and the product make sense.

Indeed, let $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g} : X \subset \mathbb{R}^n \to \mathbb{R}^p$ be two vector-valued functions, possibly of different sizes, but definitely defined on the same domain (why is this necessary?) The define $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})$. However that product is defined, it does need to make sense. But when it does, it means that the $m \times 1$-matrix of outputs $\mathbf{f}$ and the $p \times 1$-matrix of outputs of $\mathbf{g}$ are multiplied together in a well-defined way. But then the Product Rule is

$$D\mathbf{h}(\mathbf{x}) = D\left(\mathbf{f} \cdot \mathbf{g}\right)(\mathbf{x}) = D\mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot D\mathbf{g}(\mathbf{x}).$$

Following the dimensions, at least, we would get that, if one could multiply an m-vector and a p-vector, then one can also multiple a $m \times n$-matrix to a p-vector, and add to that the product of an m-vector with a $p \times n$-matrix. Write this out to verify, but the genral idea is that an $m \times n$ matrix is just a collection of n m-vectors. Here are some examples:

Example 5.1. Let $p = 1$. Then $g(x)$ is a scalar function, and $Dg(\mathbf{x})$ is a $1 \times n$-matrix. Then $\mathbf{h} = \mathbf{f} \cdot g$ makes sense, as the output is the product of an m-vector $\mathbf{f}(\mathbf{x})$ with a scalar $g(\mathbf{x})$ at every input. We get $\mathbf{h} : X \subset \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x})g(\mathbf{x})$, and

$$\underbrace{D\mathbf{h}(\mathbf{x})}_{m\times n} = \underbrace{D\mathbf{f}(\mathbf{x})}_{m\times n}\underbrace{g(\mathbf{x})}_{\text{scalar}} + \underbrace{\mathbf{f}(\mathbf{x})}_{m\times 1}\underbrace{Dg(\mathbf{x})}_{1\times n}.$$

Here is a concrete example: Suppose that we wanted to create a function that was the product of $\mathbf{f}(x, y, z) = \begin{bmatrix} xy + y^2 z \\ x^4 z \end{bmatrix}$, and $g(x, y, z) = \ln(yz)$. Then we can certainly define

$$\mathbf{h}(x, y, z) = \mathbf{f}(x, y, z)g(x, y, z) = \begin{bmatrix} (xy + y^2 z)\ln(yz) \\ (x^4 z)\ln(yz) \end{bmatrix},$$

but we have to carefully choose our domain so that $\mathbf{h}$ makes sense. $\mathbf{f}$ is defined on all of $\mathbb{R}^3$, but the largest domain of g is the set

$$X = \left\{(x, y, z) \in \mathbb{R}^3 \;\middle|\; yz > 0\right\}.$$

Then $\mathbf{h} : X \subset \mathbb{R}^3 \to \mathbb{R}^2$ is defined as above. And on this open domain X, $\mathbf{h}$ will be differentiable (in fact, it is the product of differentiable functions). So we calculate the derivative in two ways: Directly, and via the Product Rule. Directly,

$$\begin{aligned} D\mathbf{h}(\mathbf{x}) &= \begin{bmatrix} \frac{\partial \mathbf{h_1}}{\partial x}(\mathbf{x}) & \frac{\partial \mathbf{h_1}}{\partial y}(\mathbf{x}) & \frac{\partial \mathbf{h_1}}{\partial z}(\mathbf{x}) \\ \frac{\partial \mathbf{h_2}}{\partial x}(\mathbf{x}) & \frac{\partial \mathbf{h_2}}{\partial y}(\mathbf{x}) & \frac{\partial \mathbf{h_2}}{\partial z}(\mathbf{x}) \end{bmatrix} \\ &= \begin{bmatrix} y\ln(yz) & (x+2yz)\ln(yz)+x+yz & y^2\ln(yz)+\frac{xy}{z}+y^2 \\ 4x^3z & \frac{x^4z}{y} & x^4\ln(yz)+x^4 \end{bmatrix}. \end{aligned}$$

Via the Product Rule, we have

$$\begin{aligned} D\mathbf{h}(\mathbf{x}) &= D\mathbf{f}(\mathbf{x}) \cdot g(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \cdot Dg(\mathbf{x}) \\ &= \begin{bmatrix} \frac{\partial \mathbf{f_1}}{\partial x}(\mathbf{x}) & \frac{\partial \mathbf{f_1}}{\partial y}(\mathbf{x}) & \frac{\partial \mathbf{f_1}}{\partial z}(\mathbf{x}) \\ \frac{\partial \mathbf{f_2}}{\partial x}(\mathbf{x}) & \frac{\partial \mathbf{f_2}}{\partial y}(\mathbf{x}) & \frac{\partial \mathbf{f_2}}{\partial z}(\mathbf{x}) \end{bmatrix} \cdot g(\mathbf{x}) + \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial g}{\partial x}(\mathbf{x}) \\ \frac{\partial g}{\partial y}(\mathbf{x}) \\ \frac{\partial g}{\partial z}(\mathbf{x}) \end{bmatrix} \\ &= \begin{bmatrix} y & x+2yz & y^2 \\ 4x^3z & 0 & x^4 \end{bmatrix} \ln(yz) + \begin{bmatrix} xy+y^2z \\ x^4z \end{bmatrix} \cdot \begin{bmatrix} 0 & \frac{1}{y} & \frac{1}{z} \end{bmatrix}. \end{aligned}$$

I will leave it to the reader to see that these two matrices of functions are the same.

Example 5.2. Now let $p = m > 1$, with the Dot Product on vectors. Note that, for ease of calculation here, we let $n = 1$. Then, for $\mathbf{f}, \mathbf{g} : X \subset \mathbb{R} \to \mathbb{R}^m$, the dot-product function is $h : X \subset \mathbb{R} \to \mathbb{R}$, where $h(x) = \mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})$. Note that the product function h here is scalar-valued, but still has n inputs. Then

$$Dh(\mathbf{x}) = \begin{bmatrix} D_{x_1}h(\mathbf{x}) & \cdots & D_{x_n}h(\mathbf{x}) \end{bmatrix},$$

with

$$\begin{aligned} D_{x_i}h(\mathbf{x}) &= \frac{\partial}{\partial x_i}h(\mathbf{x}) = \frac{\partial}{\partial x_i}\left[\sum_{j=1}^m f_j(\mathbf{x}) \cdot g_j(\mathbf{x})\right] \\ &= \sum_{j=1}^n \frac{\partial}{\partial x_i}\left[f_j(\mathbf{x})g_j(\mathbf{x})\right] \qquad \text{(Sum Rule)} \\ &= \sum_{j=1}^m \frac{\partial f_j}{\partial x_i}(\mathbf{x})g_j(\mathbf{x}) + f_j(\mathbf{x})\frac{\partial g_j}{\partial x_i}(\mathbf{x}) \qquad \text{(Calc I Product Rule)} \\ &= \sum_{j=1}^m \frac{\partial f_j}{\partial x_i}(\mathbf{x})g_j(\mathbf{x}) + \sum_{j=1}^m f_j(\mathbf{x})\frac{\partial g_j}{\partial x_i}(\mathbf{x}) \qquad \text{(Sum Rule)} \\ &= Df(\mathbf{x}) \cdot g(\mathbf{x}) + f(\mathbf{x}) \cdot Dg(\mathbf{x}), \end{aligned}$$

where each of the four pieces in this last sum of products is an m-vector. Notice that in the middle of this last calculation, we were simply multiplying together scalar-valued functions, so there was no $\cdot$ present.

Now, as a special note of caution: Be careful with vector products. The two examples above are symmetric products, named because

$$f(x) \cdot g(x) = g(x) \cdot f(x)$$

If the product is not symmetric, then the order of the factors matters. But the Product Rule will still work correctly. One jsut needs to pay attention to the order of the factors inside the product rule. For example, for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^4$, the cross product is called *antisymmetric*, since

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}.$$

Perhaps you already know this via a detailed calculation. However, we will see why this is true geometrically in a while. Hence for $\mathbf{f}, \mathbf{g} : X \subset \mathbb{R} \to \mathbb{R}^3$, and $\mathbf{h}(x) = \mathbf{f}(x) \times \mathbf{g}(x)$, we have

$$D\mathbf{h}(x) = D\left(\mathbf{f} \times \mathbf{g}\right)(x) = D\mathbf{f}(x) \times \mathbf{g}(x) + \mathbf{f}(x) \times D\mathbf{g}(x).$$

And lastly, on this note, the Quotient Rule also hold, but only where it makes sense. One thing to keep in mind for the Quotient Rule is that the denonimator function myust be scalar-valued for even a quotient of functions to make sense. (Why?) At that point, the square of the denominator function in the Quotient Rule will also make sense.

A Note on Partial Derivatives. Given a differentiable real-valued function $f : X \subset \mathbb{R}^3 \to \mathbb{R}$, say, we know that all of

$$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} : X \to \mathbb{R}$$

are all continuous in a neighborhood of $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$. In this case, where all partials of a function are continuous on a domain, we say that the function is a C^1-*function*, or write $f \in C^1$.

Definition 5.3. A *second partial derivative* of f with respect to a variable (x, say) is any one of

$$\begin{aligned} &\tfrac{\partial}{\partial x}\left(\tfrac{\partial f}{\partial x}\right) : X \to \mathbb{R}, \\ &\tfrac{\partial}{\partial x}\left(\tfrac{\partial f}{\partial y}\right) : X \to \mathbb{R}, \\ &\tfrac{\partial}{\partial x}\left(\tfrac{\partial f}{\partial z}\right) : X \to \mathbb{R}. \end{aligned}$$

We also write

$$\begin{aligned} \tfrac{\partial}{\partial x}\left(\tfrac{\partial f}{\partial x}\right) &= \tfrac{\partial^2 f}{\partial x^2} = f_{xx} \\ \tfrac{\partial}{\partial x}\left(\tfrac{\partial f}{\partial y}\right) &= \tfrac{\partial^2 f}{\partial x \partial y} = f_{yx}, \\ \tfrac{\partial}{\partial x}\left(\tfrac{\partial f}{\partial z}\right) &= \tfrac{\partial^2 f}{\partial x \partial z} = f_{zx}. \end{aligned}$$

Pay attention to the order of the variables, and hence the derivatives here. Indeed, the order of differentiation is written differently in the two notations, fractional and subscript-wise. Be careful here. Now If all 9 of these second partial derivative of f exist and are continuous on the domain X, then we say that $f \in C^2$.

Generalizing, let $f : X \subset \mathbb{R}^n \to \mathbb{R}$. For $i_1, i_2, \ldots, i_k \in \{1, 2, \ldots, n\}$, the kth partial derivative of f with respect to $x_{i_1}, \ldots, x_{i_k}$, is

$$\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}(\mathbf{x}) = \frac{\partial}{\partial x_{i_k}}\left(\cdots\left(\frac{\partial f}{\partial x_{i_1}}\right)\cdots\right) = f_{x_{i_1} \cdots x_{i_k}}.$$

Example 5.4. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be defined by $f(x, y, z) = z\cos(2xy)$. It should be readily apparent, and can be rigorously shown, that polynomials in many variables are continuous everywhere, and differentiable everywhere. The same it true for the cosine function. And since f is a product of a polynomial and a composition of the cosine function and a polynomial, $f \in C^1$, and we can calculate

$$\begin{aligned} f_x(x, y, z) &= -z\sin(2xy)2y = -2yz\sin(2xy), \\ f_y(x, y, z) &= -z\sin(2xy)2x = -2xz\sin(2xy), \quad \text{and} \\ f_z(x, y, z) &= \cos(2xy). \end{aligned}$$

But all three of these partial derivatives of f are also sums, products and compositions of differentiable functions, so that $f \in C^2$ also. Thus, we find

$$\begin{aligned} f_{xy}(x, y, z) &= \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = -2z\sin(2xy) - 4xyz\cos(2xy), \quad \text{and} \\ f_{yx}(x, y, z) &= \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = -2z\sin(2xy) - 4xyz\cos(2xy). \end{aligned}$$

Notice immediately that these two functions are the same. It turns out that this is true always when $f \in C^2$:

Theorem 5.5. *Suppose $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is C^2 on an open X. Then, for any choice of $i_1, i_2 \in \{1, 2, \ldots, n\}$,*

$$\frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2}} = \frac{\partial^2 f}{\partial x_{i_2} \partial x_{i_1}}.$$

The proof is constructive and in the book. We will not do it in class.

Definition 5.6. A function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is of class C^k, $k \in \mathbb{N}$ if it has continuous partial derivative up to and including order k. A function $\mathbf{g} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ is of class C^k if each component function $g_i : X \to \mathbb{R}$ is of class C^k.

And finally, a function like the above is of class C^∞ if it is *smooth*. This means that it has continuous partial derivatives of all orders. Some notes:

- This should be an obvious fact, but worth stating explicitly: If $f \in C^k$, then $f \in C^\ell$ for all $\ell < k$.
- A continuous function is said to be of class C^0.

So here is a thought experiment: Suppose $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is of class C^∞, so it is a smooth function. Then we know the following:

(1) f has n first partial derivatives, and
(2) f has n^2 second partials, since each of the n first partials has n second partials.
(3) Thus f has n^k kth partials, for each $k \in \mathbb{N}$.

Now $Df(\mathbf{x})$ is a row matrix with n entries, each entry a real-valued function on n variables. Each of these entries is also differentiable. Plug in a value to evaluate the derivative of f at a point, and one gets a matrix of numbers. But without evaluation, $Df(\mathbf{x})$ is a (row)-matrix of functions. Just for a moment, view this row matrix as a column matrix. Then, in a way, $Df : X \subset \mathbb{R}^n \to \mathbb{R}^n$. And then $D\left(Df\right) = D^2 f(\mathbf{x})$ will be an $n \times n$ matrix of functions, with each entry, $\frac{\partial^2 f}{\partial x_i \partial x_j}$ a real-valued function on n variables. If we, for the moment think of $D^2 f$ as a function on X, then what is its codomain?

And, since f is smooth, the object $D\left(D^2 f\right) = D^3 f$ exists! What kind of object is this?

And in general, what kind of object is $D^k f(\mathbf{x})$, for $k \in \mathbb{N}$?

These objects will play a role in the multivariable Taylor expansion of a function like f, since Taylor series' of functions exist in multivariable calculus and will (must) account for all of the derivatives of a function. Think about this....