February 4, 2019

LECTURE 4: THE DERIVATIVE.

110.211 HONORS MULTIVARIABLE CALCULUS
PROFESSOR RICHARD BROWN

Synopsis. Today, we finish our discussion on limits and pass through the concept of continuity. Really, there is little to add to the mix since the only new idea is that the limit of a function not only exists but equals the function value at a point of continuity. But there are a few rules and extensions that we talk about here. Then on to differentiability, where things start to diverge from single variable calculus. Here we define what differentiability is for a vector-valued function on more than one variable, both from an analytical as well as geometric perspective, and start the discussion on its properties. The accompanying Mathematica notebook gives some geometric meaning to the derivative of a real-valued function on two variables and how the tangent plane to its graph in three space is defined and constructed.

Helpful Documents. Mathematica: `PartialDerivatives`.

The Derivative. A *partial derivative* of a real-valued function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ taken at a point is really a single variable calculus concept, where one studies how a function is changing in a particular direction:

Definition 4.1. Let $\mathbf{a} \in X$ be an interior point, and $f : X \subset \mathbb{R}^n \to \mathbb{R}$ a real-valued function on X. Then the *partial derivative* of f, with respect to the coordinate x_i at the point $\mathbf{x} = \mathbf{a}$ is the real number

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \to 0} \frac{f(a_1, \dots, a_{i-1}, a_i + h, a_{i+1}, \dots, a_n) - f(\mathbf{a})}{h}.$$

The *partial derivative* of f with respect to x_i is the real-valued function

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n) - f(\mathbf{x})}{h}.$$

.

Notes:

- It is simply the ordinary (read: Calculus I) derivative of f with respect to x_i, found by fixing all coordinates x_j, for $j \neq i$, and varying only x_i.
- Alternate notation: $D_{x_i} f(\mathbf{x})$, or $f_{x_i}(\mathbf{x})$.
- Geometrically, given $f(x, y)$ and $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, the y-slice through $\mathbf{graph}(f)$ at $y = b$ is a 1-dimensional curve inside the xz-plane at $y = b$. Then $\frac{\partial f}{\partial x}(a, b)$ is the slope of this curve inside the slice, evaluated at (a, b):

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

  In this case, a varies, but b is held constant. In this case,, we say that the partial derivative of f with respect to x, evaluated at $(x, y) = (a, b)$ is the slope of the line tangent to that portion of the **graph**(f) that intersects the xz-plane at $y = b$. item In turn,

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

is the slope of the line tangent to that portion of the **graph**(f) that intersects the xy-plane at $x = a$. These partial derivatives, as regular single variable calculus derivatives in a single direction, satisfy all of the rules that one developed in Calculus I.

Now, for a point (a, b) in the domain where these two quantities exist, the two tangent lines sitting in $\mathbb{R}^3$, cross at the point $(a, b, f(a, b)) \in \mathbb{R}^3$ and are perpendicular (form a right angle). They will determine a plane in $\mathbb{R}^3$: choose a non-zero vector inside each line, based at the corssing point. The plane determined by these two crossing lines is then the set of all linear combinations (in $\mathbb{R}^3$) of these two vectors.

Example 4.2. In higher, dimensions, this setup generalizes well: For $f : X \subset \mathbb{R}^n \to \mathbb{R}$, with

$$\mathbf{graph}(f) = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ z \end{bmatrix} \in \mathbb{R}^{n+1} \;\middle|\; z = f(x_1, \ldots, x_n) \right\},$$

fix a point $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in X$. Now allow the ith coordinte to vary. Then the slice formed by fixing

$$x_1 = a_1, \ldots, x_{i-1} = a_{i-1}, x_{i+1} = a_{i+1}, \ldots, x_n = a_n$$

(note that this set of equations comprise $n-1$ equations in $\mathbb{R}^{n+1}$, where the graph of f lives), forms a two-dimensional space in $\mathbb{R}^{n+1}$ parameterized by the variable x_1 and z, which we will call the x_iz-plane at $\mathbf{a}$. Here, the intersection

$$\mathbf{graph}(f) \cap \{x_iz\text{-plane at } \mathbf{x} = \mathbf{a}\}$$

is a 1-dimensional curve. If $\frac{\partial f}{\partial x_i}(\mathbf{a})$ exists, then its value represents the slope of the line tangent to this curve in the x_iz-plane at $\mathbf{a}$. As a line in $\mathbb{R}^{n+1}$, it passes through $(\mathbf{a}, f(\mathbf{a}) \in \mathbb{R}^{n+1}$. Now if tangent lines exist for each of the variables x_i, for $i = 1, \ldots, n$, they will form an n-dimensional space inside $\mathbb{R}^{n+1}$ passing through the point $(\mathbf{a}, f(\mathbf{a}) \in \mathbb{R}^{n+1}$. This space will be of vital importance to us.

Back to our two dimensional case, the plane formed by the two tangent lines to the slices of $f(x, y)$ at the point (a, b) is called the *tangent plane* to the graph of f at (a, b). So what is the equation defining this 2-dimensional plane in $\mathbb{R}^3$, what is its equation?

- It will consist of one linear equation in the three variables x, y, and z.
- One choice of vector in the tangent line to the curve in the xz-plane corresponding to $y = b$ will be based at (a, b) and have components $\begin{bmatrix} 1 \\ f_x(a,b) \end{bmatrix}$. (Why is this?) So,

as a vector in $\mathbb{R}^3$, this vector will have components $\begin{bmatrix} 1 \\ 0 \\ f_x(a,b) \end{bmatrix}$, and be based at the point $(x,y,z) = (a,b,f(a,b))$.

- The other vector in the tangent line to the intersection of the yz-plane at $x = a$ with the graph of f, will have components $\begin{bmatrix} 0 \\ 1 \\ f_y(a,b) \end{bmatrix}$, again based at $(a,b,f(a,b))$.
- The tangent plane is then the set of all linear combinations of vectors, based at $(a,b,f(a,b0)$ that have components

$$c_1 \begin{bmatrix} 1 \\ 0 \\ f_x(a,b) \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ f_y(a,b) \end{bmatrix}.$$

A little cumbersome, but well-defined.

There is a better way to describe the tangent plane: The vector

$$\mathbf{n} = \begin{bmatrix} 1 \\ 0 \\ f_x(a,b) \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ f_y(a,b) \end{bmatrix} = \begin{bmatrix} -f_x(a,b) \\ -f_y(a,b) \\ 1 \end{bmatrix}$$

is normal to both of the tangent vectors. Thus it is also normal to every linear combination of these tangent vectors. In fact, then, one can *define* the tangent space defined by these two tangent vectors as the space of vectors normal to $\mathbf{n}$, so with the dot product, we have

$$\underbrace{\begin{bmatrix} x-a \\ y-b \\ z-f(a,b) \end{bmatrix}}_{\text{generic vector at } (a,b,f(a,b))} \bullet \underbrace{\begin{bmatrix} -f_x(a,b) \\ -f_y(a,b) \\ 1 \end{bmatrix}}_{\text{normal vector to all}} = 0.$$

This works out to

$$-f_x(a,b)(x-a) - f_y(a,b)(y-b) + z - f(a,b) = 0,$$

or, with a bit of rearranging of terms

$$z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b).$$

Now, call the right hand side of this last equation $h(x,y)$, so that

$$z = h(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$$

is a linear function $h : \mathbb{R}^2 \to \mathbb{R}$. It's graph $z = h(x,y)$ is then a linear equation representing a plane in $\mathbb{R}^3$ that passes through $(a,b,f(a,b))$, and has precisely the partials $h_x(a,b) = f_x(a,b)$ and $h_y(a,b) = f_y(a,b)$, when it is defined, that is. When it is defined, the graph of this function becomes the *best* linear approximation to $\mathbf{graph}(f)$ at the point $(a,b) \in X$. So what does "best" actually mean in this context? It means:

(1) $z = h(x,y)$ is a linear function in the variables x, y, and z, and
(2) at (a,b), all of the following are true:
   - The functions are equal, so $h(a,b) = f(a,b)$;

- the derivatives are equal, so $\frac{\partial h}{\partial x}(a,b) = h_x(a,b) = f_x(a,b) = \frac{\partial f}{\partial x}(a,b)$, and $\frac{\partial h}{\partial y}(a,b) = h_y(a,b) = f_y(a,b) = \frac{\partial f}{\partial y}(a,b)$.

Example 4.3. Let $f(x,y) = x^2 + y^2$, and choose $(a,b) = (1,2)$. We can go directly to Definition 4.1 here and compute

$$\begin{aligned}\frac{\partial f}{\partial x}(1,2) &= \lim_{h\to 0} \frac{f(1+h,2) - f(1,2)}{h} \\ &= \lim_{h\to 0} \frac{((1+h)^2 + 2^2) - (1^2 + 2^2)}{h} \\ &= \lim_{h\to 0} \frac{(1 + 2h + h^2 + 4 - (1+4))}{h} = \lim_{h\to 0} \frac{2h + h^2}{h} = \lim_{h\to 0} 2 + h = 2.\end{aligned}$$

Similarly, $\frac{\partial f}{\partial y}(1,2) = 4$. Then

$$\begin{aligned} z = h(x,y) &= f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \\ &= 5 + 2(x-1) + 4(y-2) \end{aligned}$$

is the equation in x, y, and z, whose solutions comprise the tangent plane to the graph of $f(x,y)$ in $\mathbb{R}^3$. Of course, as mentioned in the first note after Definition 4.1, one can think directly that partial derivative are really single variable derivatives, as far as calculation goes. What this means is that we can sidestep the definition, and simply write

$$\frac{\partial f}{\partial x}(1,2) = \frac{\partial f}{\partial x}(x,y)\bigg|_{(x,y)=(1,2)} = \frac{\partial}{\partial x}\left[x^2 + y^2\right]\bigg|_{(x,y)=(1,2)} = (2x+0)\bigg|_{(x,y)=(1,2)} = 2.$$

There is a major caveat that we need to mention here: Just because the individual limits $\frac{\partial f}{\partial x}(a,b)$ and $\frac{\partial f}{\partial y}(a,b)$ may both exist, it does not automatically mean that f is differentiable at (a,b)! Example 4, on page 121 of the text is a great example of why the existence of these limits is not enough. I call this the rooftop function:

Example 4.4. Let $g(x,y) = \Big||x| - |y|\Big| - |x| - |y|$. Here,

$$\frac{\partial g}{\partial x}(0,0) = \lim_{h\to 0} \frac{g(0+h,0) - g(0,0)}{h} = \lim_{h\to 0} \frac{\Big||h| - |0|\Big| - |h| - |0| - 0}{h} = 0.$$

Similarly, $\frac{\partial g}{\partial y}(0,0) = 0$. However, step off of the axes, and one can see the sharp edges of the graph. In fact, if one sliced the graph of g along the $x = y$ line (diagonally, with respect to the two axes), then the limits would not exist! Indeed, slice **graph**(g) along the line $y = x$. Call the plane forming this slice

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

Then the piece of the graph of g inside P_x can be written as

$$z = g(x,x) = g(x) = \Big||x| - |x|\Big| - |x| - |x| = -2|x|.$$

But, as already known from Calculus I, This function has no derivative at $x = 0$, since

$$g'(0) = \lim_{x\to 0} \frac{g(x) - g(0)}{x - 0} = \frac{-2|x|}{x}.$$

This limit does not exist, and one can see the corner at the origin of the graph of $g(x)$. Note: This idea of slicing a graph of a function along a line that is different from an axis in the domain will be an important tool in studying the properties of functions of more than one variable. This is the idea of a *directional derivative*, which we will explore soon.

The existence of a proper tangent space to the graph of a function relies on its ability to well-approximate the function from ALL directions. The best way to construct this is, again, to use the limit!

Let $f : \mathbb{R} \to \mathbb{R}$, and notice how we can rewrite

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}, \quad \text{as} \quad \lim_{x \to a} \frac{f(x) - (f(a) + f'(a)(x - a))}{x - a} = 0$$

when (and only when) the limit actually exists.

Exercise 1. Show that this is true.

But this means that, for $h(x) = f(a) + f'(a)(x - a)$, we can say that f is differentiable at $x = a$ precisely when the tangent line $y = h(x)$ to $y = f(x)$ at $x = a$ exists, so precisely when

$$\lim_{x \to a} \frac{f(x) - h(x)}{x - a}.$$

This is important, and establishes an alternate way to define differentiability for a function; A function $f(x)$ is differentiable

Example 4.5. In 2-dimensions, this setup generalizes well: For $X \subset \mathbb{R}^2$ open, with $f : X \to \mathbb{R}$, f is *differentiable* at $(a, b) \in X$ if

- Both $\frac{\partial f}{\partial x}(a, b)$ and $\frac{\partial f}{\partial y}(a, b)$ exist, and
- if $h(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$ satisfies

$$\lim_{(x,y) \to (a,b)} \frac{f(x, y) - h(x, y)}{||(x, y) - (a, b)||} = 0.$$

Note that, again, when $h(x, y)$ exists and satisfies the limit above, then $z = h(x, y)$ is the tangent plane to $\mathbf{graph}(f)$ at $(a, b, f(a, b)) \in \mathbf{graph}(f) \subset \mathbb{R}^3$.

More notes:

- An alternate, but equivalent, notion of differentiability: For $X \subset \mathbb{R}^2$ open, and $f : X \to \mathbb{R}$, f is differentiable at (a, b) if $f_x(x, y)$ and $f_y(x, y)$ are continuous in a neighborhood of (a, b) in X.
- Like in Calculus I, differentiability always implies continuity.
- Also true in n-dimensions: Given $X \subset \mathbb{R}^n$ open, and $f : X \to \mathbb{R}$, f is *differentiable* at $\mathbf{a} \in X$ if
  - Each of $\frac{\partial f}{\partial x_i}(\mathbf{a})$ exist for $i = 1, \ldots, n$, and
  - if

$$h(\mathbf{x}) = f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})(x_i - a_i) \quad \text{satisfies} \quad \lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - h(\mathbf{x})}{||\mathbf{x} - \mathbf{a}||} = 0.$$

There is an easier way to write this:

Definition 4.6. For $f : X \subset \mathbb{R}^n \to \mathbb{R}$ differentiable at $\mathbf{a} \in X$, the *derivative* of f at $\mathbf{a}$ is the $1 \times n$ matrix

$$Df(\mathbf{a}) = \left[\begin{array}{cccc} f_{x_1}(\mathbf{a}) & f_{x_2}(\mathbf{a}) & \dots & f_{x_n}(\mathbf{a}) \end{array} \right].$$

And the *derivative function* is the $1 \times n$ matrix of functions

$$Df(\mathbf{x}) = \left[\begin{array}{cccc} f_{x_1}(\mathbf{x}) & f_{x_2}(\mathbf{x}) & \dots & f_{x_n}(\mathbf{x}) \end{array} \right].$$

Knowing this, the tangent linear function, using the above notation and definition, can be written

$$\begin{aligned} h(\mathbf{x}) &= f(\mathbf{a}) + \sum_{i=1}^{n} f_{x_i}(\mathbf{a})(x_i - a_i) \\ &= f(\mathbf{a}) + \left[\begin{array}{ccc} f_{x_1}(\mathbf{a}) & \dots & f_{x_n}(\mathbf{a}) \end{array} \right] \left[\begin{array}{c} x_1 - a_1 \\ x_2 - a_2 \\ \vdots \\ x_n - a_n \end{array} \right] \\ &= f(\mathbf{a}) + Df(\mathbf{a})\,(\mathbf{x} - \mathbf{a}) \end{aligned}$$

where $Df(\mathbf{a})$ is a $n \times 1$ (row) matrix, and $(\mathbf{x} - \mathbf{a})$ is an $1 \times n$ matrix (an n-vector). The result is a number, as it should. Hence the limit, in the definition becomes

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

Now what about $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$? Here, for $\mathbf{x} \in X$, we have $\mathbf{f}(\mathbf{x}) - \left[\begin{array}{c} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{array} \right]$. with n input variables and m output variables. If the derivative is to exist, then each component real-valued function $f_i : X \to \mathbb{R}$ must have a derivative (including all of the partials). We have

$$D\mathbf{f}(\mathbf{a}) = \left[\begin{array}{c} Df_1(\mathbf{a}) \\ \vdots \\ Df_n(\mathbf{a}) \end{array} \right],$$

where each element in this matrix is, itself, a $1 \times n$ matrix. Hence

$$D\mathbf{f}(\mathbf{a}) = \left[\begin{array}{cccc} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{a}) & \frac{\partial f_n}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{a}) \end{array} \right],$$

an $m \times n$ matrix with $\frac{\partial f_i}{\partial x_j}(\mathbf{a})$ as the ijth entry.

So our most general definition is:

Definition 4.7. Let $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ be a vector-valued function on an open X, and let $\mathbf{a} \in X$. $\mathbf{f}$ is differentiable at $\mathbf{x} = \mathbf{a}$ if

(1) $\frac{\partial f_i}{\partial x_j}(\mathbf{a})$ all exist, for $i = 1, \dots, n$ and $j = 1, \dots, n$, and

(2) the linear map $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})$ satisfies

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

Some final notes:

- In Definition 4.7, the term $||f(\mathbf{x}) - h(\mathbf{x})||$ measures the distance between $\mathbf{f}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ near $\mathbf{a}$, as vector-valued functions.
- $D\mathbf{f}(\mathbf{a})$, as a matrix of numbers, represents a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. It's entries vary as $\mathbf{a}$ varies, but it represents the *best* linear map approximating $\mathbf{f}(\mathbf{x})$ near $\mathbf{x} = \mathbf{a}$.
- $D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}) \in \mathbb{R}^m$ is an m-vector for each value of $\mathbf{x}$ and represents a catalog of ways that moving around near $\mathbf{a}$ affects functions values in the codomain.
- $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})$ defines an affine map $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ (a linear map with a translation). Recall that for $m = 1$, $z = h(\mathbf{x})$ has a graph in $\mathbb{R}^{n+1}$ which is tangent to $\mathbf{graph}(f)$ at the point $(\mathbf{a}, f(\mathbf{a}))$. It is the same for $m > 1$, once one understands the nature of a graph with more than one output, but geometrically, it is far less easy to "see".