February 13, 2019

LECTURE 6: THE CHAIN RULE.

110.211 HONORS MULTIVARIABLE CALCULUS
PROFESSOR RICHARD BROWN

Synopsis. Here, we define and discuss the Chain Rule in the differential calculus of vector-valued functions of more than one independent variable. One can use the Calculus I version to define the multivariable calculus version, which works in the same fashion. However, care must be taken for two reasons: (1) the derivatives of functions here are not the same kinds of functions as the original functions, and (2) composition is tricky when the domains and codomains can be of different sizes. We discuss this at length here.

The Chain Rule.

6.0.1. *The Chain Rule in single variable calculus.* Recall from Calculus I: For $f, g : \mathbb{R} \to \mathbb{R}$, where $f, g \in C^1$,

$$\frac{d}{dx}\left(f \circ g\right)(x) = f'\left(g(x)\right) \cdot g'(x).$$

In essence, the derivative of a composition of functions is the product of the derivatives..., (but with a definite *twist*! - The derivative of the "outside" function is evaluated at the image of x under the "inside" function. This leads to the immediate question of just how the domain of a composition depends on the domains of the constituent pieces in the composition. To see this, let $f : J \subset \mathbb{R} \to \mathbb{R}$ and $g : I \subset \mathbb{R} \to \mathbb{R}$ be defined, but only on the subsets of the real line. Then

$$\textbf{domain}\left(f \circ g\right) = \left\{x \in I \,\middle|\, g(x) \in J\right\} = g^{-1}(J) \subset I.$$

Be careful here, though, as $g^{-1}(J)$ is the set inverse of g, which makes sense even if g does not have an inverse as a function.

Example 6.1. Let $f(x) = \sqrt{x}$ and $g(x) = 2 - x^2$. Of course, without specifying a domain, the domain of each of these is automatically the largest set on which the function makes sense. In these cases, adn using the notation of the above discussion, $f : J \to \mathbb{R}$, with $J = [0, \infty)$, and $g : I \to \mathbb{R}$, where $I = \mathbb{R}$. So what is the domain of $(f \circ g)$? One way to see this is to simply construct the function:

$$\left(f \circ g\right)(x) = f(g(x)) = f(2 - x^2) = \sqrt{2 - x^2}.$$

With this, the domain can only include points that satisfy $2 - x^2 \geq 0$, so $x \in \left[-\sqrt{2}, \sqrt{2}\right]$. And thinking of this in terms of sets alone, we can calculate

$$\begin{aligned}\textbf{domain}\left(f \circ g\right)(x) = g^{-1}(J) = g^{-1}\left([0, \infty)\right) &= \left\{x \in I \,\middle|\, g(x) \in J\right\} \\ &= \left\{x \in \mathbb{R} \,\middle|\, 2 - x^2 \in [0, \infty)\right\} = \left[-\sqrt{2}, \sqrt{2}\right].\end{aligned}$$

But here is an issue: What is the domain of $(g \circ f)$? Calculating the function, we get

$$\left(g \circ f\right)(x) = g(f(x)) = g(\sqrt{x}) = 2 - \left(\sqrt{x}\right)^2 = 2 - x.$$

Without paying attention, one may wrongly assume that the domain is all of $\mathbb{R}$, since $(g \circ f)(x) = 2 - x$ is a degree-1 polynomial. However, the "inside" function has as its domain only the non-negative reals $[0, \infty)$. Hence so does $(g \circ f)$!

To use the set notation, note that J and I are switched here, and

$$\begin{aligned}\textbf{domain}\,(g \circ f)(x) = f^{-1}(I) = f^{-1}(\mathbb{R}) &= \left\{x \in J \;\middle|\; f(x) \in I\right\} \\ &= \left\{x \in [0, \infty) \;\middle|\; \sqrt{x} \in \mathbb{R}\right\} = [0, \infty).\end{aligned}$$

Note: In Leibniz notation, let $z = g(y)$, and $y = f(x)$, so that $z = (g \circ f)(x) = g(f(x))$. Then z is considered a function of x, and the Chain Rule looks like

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}.$$

One can directly and easily again see this notion that the derivative of a product of functions is, in fact, the product of the derivatives. However, when evaluated the derivative of a composition at a point, the "twist" in the product again becomes clear, and

$$\left.\frac{dz}{dx}\right|_{x=a} = \left.\frac{dz}{dy}\right|_{y=g(a)} \cdot \left.\frac{dy}{dx}\right|_{x=a}.$$

Example 6.2. Back to the previous example and translating into Leibniz notation, we have $y = f(x) = \sqrt{x}$, and $z = g(y) = 2 - y^2$. Then

$$z = (g \circ f)(x) = g(f(x)) = g(\sqrt{x}) = 2 - \left(\sqrt{x}\right)^2 = 2 - x, \quad \text{on} \quad [0, \infty).$$

Its derivative, defined on $(0, \infty)$, should be $\frac{dz}{dx} = -1$ everywhere. Here

$$\frac{dz}{dx} = \left.\frac{dz}{dy}\right|_{y=f(x)} \cdot \frac{dy}{dx} = \left.-2y\right|_{y=\sqrt{x}} \cdot \left(\frac{1}{2\sqrt{x}}\right) = \left(-2\sqrt{x}\right)\left(\frac{1}{2\sqrt{x}}\right) = -1.$$

6.1. **The Chain Rule in multivariable calculus.** In vector calculus, the Chain Rule still holds:

Theorem (Theorem 2.5.3 in text)**.** *Suppose $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are open, and $\mathbf{f} : Y \to \mathbb{R}^p$ and $\mathbf{g} : X \to \mathbb{R}^m$ are defined so that $\mathbf{g}(X) \subset Y$. Then, if $\mathbf{g}$ is differentiable at $\mathbf{x}_0 \in X$, and $\mathbf{f}$ is differentiable at $\mathbf{y}_0 = \mathbf{g}(\mathbf{x}_0) \in Y$, then $(\mathbf{f} \circ \mathbf{g})$ is differentiable at $\mathbf{x}_0$, with*

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{x}_0) = D\mathbf{f}(\mathbf{g}(\mathbf{x}_0))\, D\mathbf{g}(\mathbf{x}_0).$$

Example 6.3. Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$, $\mathbf{f}(x, y) = (x^2 y, 1, e^{xy})$, and $g : \mathbb{R}^3 \to \mathbb{R}$, with $g(x, y, z) = xyz$. We calculate $D(g \circ \mathbf{f})(x, y)$ in two ways:

(1) **Composition before derivative.** Here, $(g \circ \mathbf{f}) : \mathbb{R}^2 \to \mathbb{R}$, and

$$(g \circ \mathbf{f})(x, y) = g(\mathbf{f}(x, y)) = g(x^2, 1, e^{xy}) = x^2 y e^{xy}.$$

Then

$$\begin{aligned}D(g \circ \mathbf{f})(x, y) &= \begin{bmatrix} \frac{\partial (g \circ \mathbf{f})}{\partial x}(x, y) & \frac{\partial (g \circ \mathbf{f})}{\partial y}(x, y) \end{bmatrix} \\ &= \begin{bmatrix} 2xye^{xy} + x^2 y^2 e^{xy} & x^2 e^{xy} + x^3 y e^{xy} \end{bmatrix}.\end{aligned}$$

(2) **Via the Chain Rule.** The derivatives of the constituent functions are

$$D\mathbf{f}(x,y) = \begin{bmatrix} 2xy & x^2 \\ 0 & 0 \\ ye^{xy} & xe^{xy} \end{bmatrix} \quad \text{and} \quad Dg(x,y,z) = \begin{bmatrix} yz & xz & xy \end{bmatrix}.$$

So $Dg\left(\mathbf{f}(x,y)\right) = Dg(x^2y, 1, e^{xy}) = \begin{bmatrix} e^{xy} & x^2ye^{xy} & x^2y \end{bmatrix}$. With the Chain Rule, we get

$$\begin{aligned} D\left(g \circ \mathbf{f}\right)(x,y) &= Dg\left(\mathbf{f}(x,y)\right) \cdot D\mathbf{f}(x,y) \\ &= \begin{bmatrix} e^{xy} & x^2ye^{xy} & x^2y \end{bmatrix} \begin{bmatrix} 2xy & x^2 \\ 0 & 0 \\ ye^{xy} & xe^{xy} \end{bmatrix} \\ &= \begin{bmatrix} 2xye^{xy} + x^2y^2e^{xy} & x^2e^{xy} + x^3ye^{xy} \end{bmatrix} \end{aligned}$$

as before.

Now you may be thinking that the variables can be confusing here, with x and y included in the two domains, $\mathbb{R}^2$ for $\mathbf{f}$, and $\mathbb{R}^3$ for g. In a very important way, they are not the same, and should not be considered so! One way to correct this error of notation, and also to make things much more clear, is to switch the names of the variables, using different variables for each domain. Indeed, Let us denote the function $\mathbf{f}$ as before, but noticing explicitly that it has three component functions

$$\mathbf{f}(x,y) = (f_1(x,y), f_2(x,y), f_3(x,y)) = (x^2y, 1, e^{xy}),$$

and now define $g(u,v,w) = uvw$, the same function as before, but with new variable names. Then the two derivatives are, as before, but look like

$$Dg(u,v,w) = \begin{bmatrix} \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} & \frac{\partial g}{\partial w} \end{bmatrix}, \quad \text{and} \quad D\mathbf{f}(x,y) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} \end{bmatrix}.$$

Now, within the composition, we know that

$$\begin{aligned} u &= f_1(x,y) = x^2y \\ v &= f_2(x,y) = 1 \\ w &= f_3(x,y) = e^{xy}. \end{aligned}$$

Hence the derivative of the composition, which is

$$Dg\left(\mathbf{f}(x,y)\right) \cdot D\mathbf{f}(x,y) = \begin{bmatrix} \frac{\partial (g\circ\mathbf{f})}{\partial x}(x,y) & \frac{\partial (g\circ\mathbf{f})}{\partial y}(x,y) \end{bmatrix},$$

where by direct matrix multiplication

$$\begin{aligned}
\frac{\partial\left(g\circ\mathbf{f}\right)}{\partial x}(x,y) &= \frac{\partial g}{\partial u}\cdot\frac{\partial f_1}{\partial x}+\frac{\partial g}{\partial v}\cdot\frac{\partial f_2}{\partial x}+\frac{\partial g}{\partial w}\cdot\frac{\partial f_3}{\partial x} \\
&= \frac{\partial g}{\partial u}\cdot\frac{\partial u}{\partial x}+\frac{\partial g}{\partial v}\cdot\frac{\partial v}{\partial x}+\frac{\partial g}{\partial w}\cdot\frac{\partial w}{\partial x} \\
&= vw\Big|_{\substack{v=1\\ w=e^{xy}}}\cdot(2xy)+uw\Big|_{\substack{u=x^2y\\ w=e^{xy}}}\cdot(0)+uv\Big|_{\substack{u=x^2y\\ v=1}}\cdot\left(ye^{xy}\right) \\
&= 2xye^{xy}+x^2y^2e^{xy}.
\end{aligned}$$

Here, the products of the partials in these derivative of compositions are always understood to have the "twist", as mentioned earlier, so that

$$\frac{\partial g}{\partial u}\cdot\frac{\partial u}{\partial x}=\frac{\partial g}{\partial u}\Big|_{\mathbf{u}=\mathbf{f}(\mathbf{x})}\cdot\frac{\partial u}{\partial x}\Big|_{\mathbf{x}},\quad \text{where}\quad \mathbf{u}=\begin{bmatrix} u\\ v\\ w\end{bmatrix},\quad\text{and}\quad \mathbf{x}=\begin{bmatrix} x\\ y\end{bmatrix}.$$

Here is one more example:

Example 6.4. Let $\mathbf{c}:\mathbb{R}\to\mathbb{R}^3$ be a C^1-curve in three-space, and $f:\mathbb{R}^3\to\mathbb{R}$ be a C^1-scalar-valued function on R^3. Then the composition $g=f\circ\mathbf{c}:\mathbb{R}\to\mathbb{R}$ is just f evaluated along the curve, and looks like a function from $\mathbb{R}$ to $\mathbb{R}$. One often writes

$$g=f\big|_{\mathbf{c}}.$$

In this sense, $g'(t)=\frac{df}{dt}(t)$ along $\mathbf{c}$. We calculate this quantity via the Chain Rule:

Here $\mathbf{c}(t)=\begin{bmatrix} x(t)\\ y(t)\\ z(t)\end{bmatrix}$ is C^1, and

$$\frac{d\mathbf{c}}{dt}(t)=\mathbf{c}'(t)=\begin{bmatrix} x'(t)\\ y'(t)\\ z'(t)\end{bmatrix}=\begin{bmatrix} \frac{dx}{dt}\\ \frac{dy}{dt}\\ \frac{dz}{dt}\end{bmatrix},$$

while $Df(x,y,z)=\begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z}\end{bmatrix}$.

Hence

$$\begin{aligned}
g'(t)=D\left(f\circ\mathbf{c}\right)(t)=\frac{df\big|_{\mathbf{c}}}{dt} &= Df\left(\mathbf{c}(t)\right)\cdot D\mathbf{c}(t) \\
&= \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z}\end{bmatrix}\begin{bmatrix} \frac{dx}{dt}\\ \frac{dy}{dt}\\ \frac{dz}{dt}\end{bmatrix} \\
&= \frac{\partial f}{\partial x}\cdot\frac{dx}{dt}+\frac{\partial f}{\partial y}\cdot\frac{dy}{dt}+\frac{\partial f}{\partial z}\cdot\frac{dz}{dt} \\
&= \frac{\partial f}{\partial x}\Big|_{\mathbf{c}(t)}\cdot x'(t)+\frac{\partial f}{\partial y}\Big|_{\mathbf{c}(t)}\cdot y'(t)+\frac{\partial f}{\partial z}\Big|_{\mathbf{c}(t)}\cdot z'(t).
\end{aligned}$$