Notes for Section 2.5

This section is about properties of derivatives of multivariable functions, which you can use to differentiate more complex multivariable functions: the Constant Multiple Rule, the Sum Rule, the Product Rule, the Quotient Rule, and the Chain Rule. All of these take forms very similar to the corresponding rules in single variable calculus, only they need to be interpreted in terms of addition and multiplication by a scalar of vectors and matrices, and multiplication of matrices since derivatives of multivariable functions are typically vectors or matrices. Theorem 10 lists all except the Chain Rule. Pay attention to details, such as f and g need to share the same domain and codomain to be added. Think about why these restrictions are required. Also, spend some time thinking about what $\mathbf{D}\,f$ and $\mathbf{D}\,g$ are in the various rules. Are they vectors? In what vector space? Are they matrices? What size matrices? Think about each multiplication. Is it a product of numbers, or multiplication of a vector by a scalar or of a matrix by a scalar, or a product of two matrices? If you get confused,ttrying the rule on a specific example may help to clear things up.

Notice that the formulas in Theorem 10 follow directly from single variable calculus. This is because the derivative of a multivariable function $f : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix whose entries are the partial derivatives of the component functions $\frac{\partial f_i}{\partial x_j}$. The partial derivative $\frac{\partial f_i}{\partial x_j}$ is the single variable derivative of f_i if you treat f_i as a function of only x_j and the other input variables as constant parameters. So it makes sense that partial derivatives behave just like derivatives in single variable calculus and so they obey the Constant Multiple Rule, the Sum Rule, the Product Rule, and the Quotient Rule from single variable calculus. For example, the Sum Rule

$$\mathbf{D}(f+g)(x_0) = \mathbf{D}\,f(x_0) + \mathbf{D}\,g(x_0)$$

just says that the ij-entry of the matrix $\mathbf{D}(f+g)(x_0)$ is the sum of the ij-entries of the matrices $\mathbf{D}\,f(x_0)$ and $\mathbf{D}\,g(x_0)$. Those entries are the partial derivatives of the i-th component functions with respect to x_j, so all we are saying is that

$$\mathbf{D}(f+g)(x_0) = \frac{\partial}{\partial x_j}(f_i + g_i) = \frac{\partial f_i}{\partial x_j} + \frac{\partial g_i}{\partial x_j} = \mathbf{D}\,f(x_0) + \mathbf{D}\,g(x_0)$$

for each i and j. If this sounds like a lot of abstract nonsense, try this on the functions $f, g : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$f(x,y,z) = (x^2 + 3yz, xy + yz), \qquad g(x,y,z) = (\sin(x + 2y + 3z), \cos(xyz)).$$

Write down the derivative matrix $\mathbf{D}(f+g)$, and notice as you are doing it that you are using the Sum Rule from single variable calculus each time you calculate one of the six entries.

The interesting part of Theorem 10 is not the formulas, but the assertions of differentiability. For example, we are not just saying that

$$\mathbf{D}(f+g)(x_0) = \mathbf{D}\,f(x_0) + \mathbf{D}\,g(x_0).$$

We are also saying that if f and g are differentiable functions at x_0, that is

$$l_1(x) = \mathbf{D}\,f(x_0)(x - x_0) + f(x_0)$$

and

$$l_2(x) = \mathbf{D}\,g(x_0)(x - x_0) + g(x_0)$$

are good linear approximations for f and g near x_0 in the sense that

$$\lim_{x \to x_0} \frac{||f(x) - l_1(x)||}{||x - x_0||} = 0$$

$$\lim_{x \to x_0} \frac{||g(x) - l_2(x)||}{||x - x_0||} = 0$$

then

$$l(x) = \mathbf{D}(f+g)(x_0)(x-x_0) + (f+g)(x_0) = \big(\mathbf{D}\, f(x_0) + \mathbf{D}\, g(x_0)\big)(x-x_0) + f(x_0) + g(x_0)$$

is also a good linear approximation to $f+g$ near x_0. That is

$$\lim_{x \to x_0} \frac{||(f+g)(x) - l(x)||}{||x - x_0||} = 0$$

as well. This may seem very abstract, but is in fact easy enough to prove. Let us give it a try. First, note that

$$\begin{aligned} l(x) &= \big(\mathbf{D}\, f(x_0) + \mathbf{D}\, g(x_0)\big)(x-x_0) + f(x_0) + g(x_0) \\ &= \mathbf{D}\, f(x_0)(x-x_0) + f(x_0) + \mathbf{D}\, g(x_0)(x-x_0) + g(x_0) \\ &= l_1(x) + l_2(x). \end{aligned}$$

Now

$$\begin{aligned} ||(f+g)(x) - l(x)|| &= ||f(x) + g(x) - (l_1(x) + l_2(x))|| \\ &= ||f(x) - l_1(x) + g(x) - l_2(x)|| \\ &\le ||f(x) - l_1(x)|| + ||g(x) - l_2(x)|| \end{aligned}$$

by the Triangle Inequality. Since $||x - x_0|| > 0$, it is also true that

$$\frac{||(f+g)(x) - l(x)||}{||x-x_0||} \le \frac{||f(x) - l_1(x)|| + ||g(x) - l_2(x)||}{||x-x_0||} = \frac{||f(x) - l_1(x)||}{||x-x_0||} + \frac{||g(x) - l_2(x)||}{||x-x_0||}.$$

Obviously,

$$0 \le \frac{||(f+g)(x) - l(x)||}{||x-x_0||}$$

since the numerator is nonnegative and the denominator is positive. By the Squeeze Theorem (multivariable version of course, but completely analogous to the single variable version):

$$\lim_{x\to x_0} 0 \le \lim_{x\to x_0} \frac{||(f+g)(x) - l(x)||}{||x-x_0||} \le \lim_{x\to x_0} \left[\frac{||f(x) - l_1(x)||}{||x-x_0||} + \frac{||g(x) - l_2(x)||}{||x-x_0||}\right].$$

It is clear that

$$\lim_{x\to x_0} 0 = 0,$$

and we also know

$$\lim_{x\to x_0} \left[\frac{||f(x) - l_1(x)||}{||x-x_0||} + \frac{||g(x) - l_2(x)||}{||x-x_0||}\right] = \lim_{x\to x_0} \frac{||f(x) - l_1(x)||}{||x-x_0||} + \lim_{x\to x_0} \frac{||g(x) - l_2(x)||}{||x-x_0||} = 0 + 0 = 0.$$

We can now conclude that

$$\lim_{x\to x_0} \frac{||(f+g)(x) - l(x)||}{||x-x_0||} = 0,$$

that is l is a good linear approximation to $f+g$, and hence $f+g$ is differentiable at x_0.

This last proof is the same as that in the textbook on the top of p. 126, with a few more details spelled out. But I think the textbook does a rather poor job of explaining to you what it is that they are trying to prove. It is not the formula

$$\mathbf{D}(f+g)(x_0) = \mathbf{D}\, f(x_0) + \mathbf{D}\, g(x_0).$$

That follows from the properties of partial derivatives. It is the fact that the differentiability of f and g at x_0 guarantees the differentiability of $f+g$ at x_0.

The other main result in this section is the multivariable version of the Chain Rule. Let me state the Chain Rule in a way that it will really look analogous to the Chain Rule from single variable calculus. I do not want to be distracted by domain compatibility issues, so I will state this for functions whose domain is the entire vector space, but you should understand that the result is valid as long as the outputs of g are interior points of the domain of f.

Theorem. *The Chain Rule: Let $f : \mathbb{R}^m \to \mathbb{R}^p$ and $g : \mathbb{R}^n \to \mathbb{R}^m$. Suppose g is differentiable at $x \in \mathbb{R}^n$ and f is differentiable at $g(a)$. Then $f \circ g$ is differentiable at a and*

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

I think the textbook pulls a fast one on us here. It claims to prove the Chain Rule, at least in the somewhat special case when all of the partial derivatives of f are continuous at $g(x_0)$. I think what it does prove is that the partial derivatives of $f \circ g$ all exist and the formula for $\mathbf{D}(f \circ g)$ given in the theorem is the right formula for the derivative of $f \circ g$. But remember that the existence of the derivative matrix does not guarantee that $f \circ g$ is actually differentiable at x_0 and I do not see where that is proven. I think that to show $f \circ g$ is differentiable at x_0 if f and g satisfy the conditions in the theorem, we would have to get our hands dirty with some limits and error functions. I am happy to leave that to a more advanced analysis course, if you choose to take one.

As far as showing that the formula of the Chain Rule is correct, I would say the textbook does a pretty good job at reasoning it out. I will let you read it, but of course, I am always happy to talk about it in class if you want me to clarify something. The examples that follow show you how to use the Chain Rule to differentiuate concrete composite functions. It takes a bit of practice to learn the pattern, but it is not particularly tricky.

One last note, on the Product Rule and the Quotient Rule. The way these are stated in Theorem 10, applies only to real-valued functions. But both can be generalized somewhat. In the Product Rule, one of the two functions, either one, could be vector-valued, by interpreting the multiplication as multiplication of a vector by a scalar. This works because you can just use the Product Rule as stated in Theorem 10 for each component function of the vector-valued function among f and g. But you need to be careful about one thing if you want to do this: you need to write the value of the vector-valued function as a column vector in $\mathbb{R}^m$ and you need to expand partial derivatives along rows. So for example, if $f : \mathbb{R}^3 \to \mathbb{R}$ and $g : \mathbb{R}^3 \to \mathbb{R}^4$, then $h = fg : \mathbb{R}^3 \to \mathbb{R}^4$, so $\mathbf{D}\, h(x_0)$ is a 4 by 3 matrix, $g(x_0)$ is a column vector in $\mathbb{R}^4$, $\mathbf{D}\, f(x_0) = \nabla f(x_0)$ is a row vector in $\mathbb{R}^3$, $f(x_0)$ is a real number, and $\mathbf{D}\, g(x_0)$ is a 4 by 3 matrix. So both $g(x_0)\, \mathbf{D}\, f(x_0)$ and $f(x_0)\, \mathbf{D}\, g(x_0)$ are 4 by 3 matrices. If you are not careful about this, you will find that you are supposed to add apples (or scalars) to oranges (or matrices).