MA 105 D3 Lecture 19

Ravi Raghunathan

Department of Mathematics

September 5, 2017

Problems involving the gradient

Recap - the total derivative

Problems involving the gradient, continued

Exercise 3: Find $\nabla_u F(2,2,1)$ where $\nabla_u F$ denotes the directional derivative of the function $F(x,y,z) = 3x - 5y + 2z$ and u is the unit vector in the outward normal to the sphere $x^2 + y^2 + z^2 = 9$ at the point $(2,2,1)$.

Solution: The unit outward normal to the sphere $g(x,y,z) = 9$ at $(2,2,1)$ is given by

$$\frac{\nabla g(2,2,1)}{\|\nabla g(2,2,1)\|}.$$

We see that $\nabla g(2,1,1) = (4,4,2)$ so the corresponding unit vector is $(2,2,1)/3$.

To get the directional derivative we simply take the dot product of ∇F with u:

$$(3,-5,2) \cdot (2,2,1)/3 = -2/3$$

Comments: Also, there is no need to compute the gradient to find the normal vector to the sphere - it is obviously the radial vector at the point $(2,2,1)$!.

Problems involving the gradient, continued

Exercise 4: Find the equations of the tangent plane and the normal line to the surface

$$F(x, y, z) := x^2 + 2xy - y^2 + z^2 = 7$$

at $(1, -1, 3)$.

Solution: We first compute the gradient of F to get $\nabla F(x, y, z) = (2x + 2y, 2x - 2y, 2z)$. At $(1, -1, 3)$, this yields the vector $\lambda(0, 4, 6)$ which is normal to the given surface at $(1, -1, 3)$. The point $(1, 3, 9)$ also lies on the normal line so its equation are

$$x = 1, \frac{y + 1}{4} = \frac{z - 3}{6}.$$

The equation of the tangent plane is given by

$$4(y + 1) + 6(z - 3) = 0,$$

since it consists of all lines orthogonal to the normal and passing through the point $(1, -1, 3)$.

The derivative for $f : U \to \mathbb{R}^n$

We now define the derivative for a function $f : U \to \mathbb{R}^n$, where U is a subset of $\mathbb{R}^m$.

The function f is said to be differentiable at a point x if there exists an $n \times m$ matrix $Df(x)$ such that

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - Df(x) \cdot h\|}{\|h\|} = 0.$$

Here $x = (x_1, x_2, \ldots, x_m)$ and $h = (h_1, h_2, \ldots h_m)$ are vectors in $\mathbb{R}^m$.

The matrix $Df(x)$ is usually called the total derivative of f. It is also referred to as the Jacobian matrix. What are its entries?

From our experience in the 2×1 case we might guess (correctly!) that the entries will be the partial derivatives.

Here is the total derivative or the derivative matrix written out fully.

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_m}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \cdots & \frac{\partial f_2}{\partial x_m}(x) \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \frac{\partial f_n}{\partial x_1}(x) & \frac{\partial f_n}{\partial x_2}(x) & \cdots & \frac{\partial f_n}{\partial x_m}(x) \end{pmatrix}$$

In the 2×2 case we get

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) \end{pmatrix}.$$

As before, the derivative may be viewed as a linear map, this time from $\mathbb{R}^m$ to $\mathbb{R}^n$ (or, in the case just above, from $\mathbb{R}^2$ to $\mathbb{R}^2$).

Rules for the total derivative

Just like in the one variable case, it is easy to prove that

$$D(f+g)(x) = Df(x) + Dg(x).$$

Somewhat harder, but only because the notation gets more cumbersome, is the Chain rule:

$$D(f \circ g)(x) = Df(g(x)) \circ Dg(x),$$

where $\circ$ on the right hand side denotes matrix multiplication. Theorem 26 holds in this greater generality - a function from $\mathbb{R}^m$ to $\mathbb{R}^n$ is differentiable at a point x_0 if all the partial derivatives $\frac{\partial f_i}{\partial x_j}$ $1 \leq i \leq n$, $1 \leq j \leq m$, are continuous in a neighborhood of x_0 (define a neighborhood of x_0 in $\mathbb{R}^m$!).

A diversion: How to calculate powers of e in your head?

From Richard Feynman's "Surely you're joking Mr. Feynman!" (pages 173-174):

One day at Princeton I was sitting in the lounge and overheard some mathematicians talking about the series for e to the x power which is $1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$. Each term you get by multiplying the preceding term by x and dividing by the next number. For example, to get the next term after $x^3/3!$ you multiply that term by x and divide by 4. It's very simple.

When I was a kid I was excited by series, and had played with this thing. I had computed e using that series, and had seen how quickly the new terms became very small.

I mumbled something about how it was easy to calculate e to any power using that series (you just substitute the power for x).
"Oh yeah?" they said. "Well, the what's e to the 3.3?" said some joker - I think it was Tukey.
I say, "That's easy. It's 27.11."

Feynman's anecdote continued

Tukey knows it isn't so easy to compute all that in your head.
"Hey! How'd you do that?"
Another guy says, "You know Feynman, he's just faking it. It's not really right."
They go to get a table, and while they're doing that, I put on a few more figures: "27.1126," I say.
They find it in the table. "It's right! But how'd you do it!"
"I just summed the series."
"Nobody can sum the series that fast. You must just happen to know that one. How about e^3?"
"Look," I say. "It's hard work! Only one a day!"
"Hah! It's a fake!" they say, happily.
"All right," I say, "It's 20.085."
They look in the book as I put a few more figures on. They're all excited now, because I got another one right.

Here are these great mathematicians of the day, puzzled at how I can compute e to any power! One of them says, "He just can't be substituting and summing - it's too hard. There's some trick. You couldn't do just any old number like e to the 1.4."
I say, "It's hard work, but for you, OK. It's 4.05."
As they're looking it up, I put on a few more digits and say, "And that's the last one for the day!" and walk out.

What happened was this: I happened to know three numbers - the logarithm of 10 to the base *e* (needed to convert numbers from base 10 to base *e*), which is 2.3026 (so I knew that *e* to the 2.3 is very close to 10), and because of radioactivity (mean-life and half-life), I knew the log of 2 to the base *e*, which is .69315 (so I also knew that *e* to the .7 is nearly equal to 2). I also knew *e* (to the 1), which is 2.71828.

The first number they gave me was *e* to the 3.3, which is *e* to the 2.3 (10) times *e*, or 27.18. While they were sweating about how I was doing it, I was correcting for the extra .0026 - 2.3026 is a little high.

I knew I couldn't do another one; that was sheer luck. But then the guy said *e* to the 3: that's *e* to the 2.3 times *e* to the .7, or ten times two. So I knew it was 20.something, and while they were worrying how I did it, I adjusted for the .693.

Now I was sure I couldn't do another one, because the last one was again by sheer luck. But the guy said *e* to the 1.4 which is *e* to the .7 times itself. So all I had to do is fix up 4 a little bit!

They never did figure out how I did it.

Higher derivatives

Just as we repeatedly differentiated a function of one variable to get higher derivatives, we can also look at higher partial derivatives.

However, we now have more freedom. If we have a function $f(x_1, x_2)$ of two variables, we could first take the partial derivative with respect to x_1, then with respect to x_2, then again with respect to x_2, and so on. Does the order in which we differentiate matter?

Theorem 28: Let $f : U \to R$ be a function such that the partial derivatives $\frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j}(f)\right)$ exist and are continuous for every $1 \leq i, j \leq m$. Then,

$$\frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j}(f)\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial}{\partial x_i}(f)\right).$$

Functions $f : U \to \mathbb{R}$ for which the mixed partial derivatives of order 2 (that is, the $\frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j}(f)\right)$) are all continuous are called $\mathcal{C}^2$ functions. Theorem 28 says that for $\mathcal{C}^2$ functions, the order in which one takes partial derivatives does not matter.

From now on we will use use the following notation. By

$$\frac{\partial^n f}{\partial x_1^{n_1} \partial x_2^{n_2} \dots \partial x_k^{n_k}},$$

we mean: first take the partial derivative of f n_1 times with respect to x_1, then n_2 times with respect to x_2, and so on. The number n is nothing but $n_1 + n_2 + \dots + n_k$. It is called the order of the mixed partial derivative.

Finally, we say that a function is $\mathcal{C}^k$ if all mixed partial derivatives of order k exist and are continuous. A function $f : U \to \mathbb{R}^n$ will be said to be $\mathcal{C}^k$ if each of the functions $f_1, f_2, \dots, f_n$ are $\mathcal{C}^k$ functions.

From the preceding slide we see the we can talk about $\mathcal{C}^k$ functions for any function from (a subset of) $\mathbb{R}^m$ to $\mathbb{R}^n$. As in the one variable case we can also talk of smooth functions - these are functions for which all partial derivatives of all orders exist. In particular, the notion of a smooth vector field makes perfect sense. There are many interesting facts about smooth vector fields. I will mention just one:
You cannot comb a porcupine.

Or, in more mathematical terms, every smooth vector field on the sphere will vanish at at least one point.

Note that we require at each point on the sphere, the vector we assign must lie in the plane tangent to the sphere at that point.