MA 105 D1 Lecture 12

Ravi Raghunathan

Department of Mathematics

August 31, 2017

Recap

The Chain Rule

The Chain Rule and gradients

Differentiability for functions of two variables

We let $(x, y) = (x_0, y_0) + (h, k) = (x_0 + h, y_0 + k)$. The function $f(x, y)$ is said to be differentiable at (x_0, y_0) if both partial derivatives exist at that point and if

$$\left| f((x_0, y_0) + (h, k)) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k \right| = p(h, k)\|(h, k)\|$$

where $p(h, k)$ is a function that goes to 0 as $\|(h, k)\| \to 0$.This form of differentiability now looks exactly like the one variable version case. In matrix notation we have:

Definition: The function $f(x, y)$ is said be differentiable at a point (x_0, y_0) if there exists a matrix denoted $Df((x_0, y_0))$ with the property that

$$f((x_0, y_0) + (h, k)) - f(x_0, y_0) - Df(x_0, y_0)\begin{pmatrix} h \\ k \end{pmatrix} = p(h, k)\|(h, k)\|,$$

for some function $p(h, k)$ which goes to zero as (h, k) goes to zero.

A criterion for differentiability

Before we state the criterion, we note that with our definition of differentiability, every differentiable function is continuous.

Theorem 26: Let $f : U \to \mathbb{R}$. If the partial derivatives $\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y)$ exist and are continuous in a neighbourhood of a point (x_0, y_0) (that is in a region of the plane of the form $\{(x, y) \mid \|(x, y) - (x_0, y_0)\| < r\}$ for some $r > 0$). Then f is differentiable at (x_0, y_0).

We omit the proof of this theorem. However, we note that a function whose partial derivatives exist and are continuous is said to be continuously differentiable or of class $\mathcal{C}^1$. The theorem says that every $\mathcal{C}^1$ function is differentiable.

The Chain Rule

We now study the situation where we have composition of functions. We assume that $x, y : I \to \mathbb{R}$ are differentiable functions from some interval (open or closed) to $\mathbb{R}$. Thus the pair $(x(t), y(t))$ defines a function from I to $\mathbb{R}^2$. Suppose we have a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is differentiable. We would like to study the derivative of the composite function $z(t) = f(x(t), y(t))$ from I to $\mathbb{R}$.

Theorem 27: With notation as above

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

For a function $w = f(x, y, z)$ in three variables the chain rule takes the form

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt}$$

Clarifications on the notation

This slide was not part of my presentation in class. However, I did mention what I have put on this slide orally during the lecture.

The form in which I have written the chain rule is the standard one used in many books (both in engineering and mathematics). However, it is not very good notation. For instance, in the formula

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

the letter z is being used for two different functions: both as a function $z(t)$ from $\mathbb{R}$ to $\mathbb{R}$ on the left hand side, and as a function $z(x, y)$ from $\mathbb{R}^2$ to $\mathbb{R}$. If one wants to be precise one should write the chain rule as

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Similarly, for the function w we should write

$$\frac{dw}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}$$

Verifying the chain rule in a simple case

Example: Let us verify this rule in a simple case. Let $z = xy$, $x = t^3$ and $y = t^2$.

Then $z = t^5$ so $z'(t) = 5t^4$. On the other hand, using the chain rule we get

$$z'(t) = y \cdot 3t^2 + x \cdot 2t = 3t^4 + 2t^4 = 5t^4.$$

I have modified my slides from class slightly for better clarity. The changes have been added in green.

Example: A continuous mapping $c : I \to \mathbb{R}^n$ of an interval I to $\mathbb{R}$ is called a path or curve in $\mathbb{R}^n$, $(n = 2, 3)$.
In what follows, we will assume that all the curves we have are actually differentiable, not just continuous. We will say what this means below.

An application to tangents of curves

Let us consider a curve $c(t)$ in $\mathbb{R}^3$. Each point on the curve will be given by a triple of coordinates which will depend on t. That is, the curve can be described by a triple of functions $(g(t), h(t), k(t))$. Saying that $c(t)$ is a differentiable function of t, means that each of $g(t), h(t), k(t)$ are differentiable functions from $\mathbb{R} \to \mathbb{R}$. If we write

$$c(t) = g(t)\mathbf{i} + h(t)\mathbf{j} + k(t)\mathbf{k}, \quad \text{then} \quad c'(t_0) = g'(t_0)\mathbf{i} + h'(t_0)\mathbf{j} + k'(t_0)\mathbf{k},$$

represents its tangent or velocity vector at the point $c(t_0)$.

Tangents to curves on surfaces

So far our example has nothing to do with the chain rule. Suppose $z = f(x, y)$ is a surface, and $c(t) = (g(t), h(t), f(g(t), h(t))$ lies on the $z = f(x, y)$. (Here we are assuming that $f : \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function!) Let us compute the tangent vector to the curve at $c(t_0)$. It is given by

$$c'(t_0) = g'(t_0)\mathbf{i} + h'(t_0)\mathbf{j} + k'(t_0)\mathbf{k},$$

where $k(t) = (f(g(t), h(t))$. Using the chain rule we see that

$$k'(t_0) = \frac{\partial f}{\partial x} g'(t_0) + \frac{\partial f}{\partial y} h'(t_0).$$

Someone in class suggested that one should write $\frac{\partial f}{\partial g}$ rather than $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial h}$ rather than $\frac{\partial f}{\partial y}$ in the formula above. This would certainly be acceptable. However $\frac{\partial f}{\partial x}$ is standard notation for the partial derivative w.r.t the first variable and $\frac{\partial f}{\partial y}$ is notation for the partial derivative w.r.t. the second variable, so what I have written above is also standard.

We can further show that this tangent vector lies on the tangent plane to the surface $z = f(x, y)$. Indeed we have already seen that the tangent plane has the equation

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0).$$

A normal vector to this plane is given by

$$\left(-\frac{\partial f}{\partial x}(x_0, y_0), -\frac{\partial f}{\partial y}(x_0, y_0), 1 \right).$$

Thus, to verify that the tangent vector lies on the plane, we need only check that its dot product with normal vector is 0. But this is now clear.

Just to give a concrete example of what we are talking about, take a curve $(g(t), h(t))$ in the unit disc $x^2 + y^2 \leq 1$ in the xy plane. Then $\left(g(t), h(t), \sqrt{1 - g(t)^2 - h(t)^2}\right)$ lies on the upper hemisphere $z = \sqrt{1 - x^2 - y^2}$. For concreteness, we can take $I = \left[0, \frac{1}{\sqrt{2}}\right]$, $g(t) = t$ and $h(t) = t^2$.

Another application: Directional derivatives

Let $U \subset \mathbb{R}^3$ and let $f : U \to \mathbb{R}$ be differentiable. We want to relate the directional derivative to the gradient,

We consider the (differentiable) curve $c(t) = (x_0, y_0, z_0) + tv$, where $v = (v_1, v_2, v_3)$ is a unit vector. We can rewrite $c(t)$ as $c(t) = (x_0 + tv_1, y_0 + tv_2, z_0 + tv_3)$. We apply the chain rule to compute the derivative of the function $f(c(t))$:

$$\frac{df}{dt} = \frac{\partial f}{\partial x} v_1 + \frac{\partial f}{\partial y} v_2 + \frac{\partial f}{\partial z} v_3.$$

But the left hand side is nothing but the directional derivative in the direction v. Hence,

$$\nabla_v f = \frac{df}{dt} = \nabla f \cdot v.$$

Of course, the same argument works when $U \subset \mathbb{R}^2$ and f is a function of two variables. (Again, we have abused notation here. We should really write $\frac{d(f \circ c)}{dt}$ on the left hand side of the first equation instead of $\frac{df}{dt}$.)

The Chain Rule and Gradients

The preceding argument is a special case of a more general fact. Let $c(t)$ be any curve in $\mathbb{R}^3$. Then, clearly by the chain rule we have

$$\frac{df}{dt} = \nabla f(c(t)) \cdot c'(t).$$

I leave this to you as a simple exercise.

Going back to the directional derivative, we can ask ourselves the following question. In what direction is f changing fastest at a given point (x_0, y_0, z_0)? In other words, in which direction does the directional derivative attain its largest value?

Using what we have just learnt, we are looking for a unit vector $v = (v_1, v_2, v_3)$ such that

$$\nabla f(x_0, y_0, z_0) \cdot v$$

is as large as possible

We rewrite the preceding dot product as

$$\nabla f(x_0, y_0, z_0) \cdot v = \|\nabla f(x_0, y_0, z_0)\| \|v\| \cos\theta.$$

where θ is the angle between v and $\nabla f(x_0, y_0, z_0)$.
Since v is a unit vector this gives

$$\nabla f(x_0, y_0, z_0) \cdot v = \|\nabla f(x_0, y_0, z_0)\| \cos\theta.$$

The maximum value on the right hand side is obviously attained when $\theta = 0$, that is, when v points in the direction of ∇f. In other words the function is increasing fastest in the direction v given by ∇f. Thus the unit vector that we seek is

$$v = \frac{\nabla f(x_0, y_0, z_0)}{\|\nabla f(x_0, y_0, z_0)\|}.$$

Surfaces defined implicitly

So far we have only been considering surfaces of the form $z = f(x, y)$, where f was a function on a subset of $\mathbb{R}^2$. We now consider a more general type of surface S defined implicitly:

$$S = \{(x, y, z) \mid f(x, y, z) = b\},$$

where b is a constant. Most surfaces we have come across are usually described in this form, for instance, the sphere which is given by $x^2 + y^2 + z^2 = r^2$ or the right circular cone $x^2 + y^2 - z^2 = 0$. Let us try to understand what a tangent plane is more precisely.

If S is a surface, a tangent plane to S at a point $s \in S$ (if it exists) is a plane that contains the tangent lines at s to all curves passing through s and lying on S. In class, many of you (Poorvi, for one:-$y^3 = x^2$) gave examples of surfaces that had tangent planes under this definition (but were clearly unsatisfactory). If we assume that function f and the curves c are all differentiable, these examples don't work anymore.)

For instance, with the definition above, it is clear that a tangent plane to the right circular cone does not exist at the origin, since such a plane would have to contain the lines $x = 0, y = z$, $x = 0, y = -z$ and $y = 0, x = z$. Clearly no such plane exists.

If $c(t)$ is an curve on the surface S given by $f(x, y, z) = b$, we see that

$$\frac{d}{dt} f(c(t)) = 0.$$

On the other hand, by the chain rule,

$$0 = \frac{d}{dt} f(c(t)) = \nabla f(c(t)) \cdot c'(t).$$

Thus, if $s = c(t_0)$ is a point on the surface, we see that

$$\nabla f(c(t_0)) \cdot c'(t_0) = 0,$$

for every curve $c(t)$ on the surface S passing through t_0. Hence, if $\nabla f(c(t_0)) \neq 0$, then $\nabla f(c(t_0))$ is perpendicular to the tangent plane of S at s_0.

Let $\mathbf{r}$ denote the position vector

$$x\mathbf{i} + y\mathbf{j} + z\mathbf{k},$$

of a point $P = (x, y, z)$ in $\mathbb{R}^3$. Instead of writing $\|\mathbf{r}\|$, it is customary to write r. This notation is very useful. For instance, Newton's Law of Gravitation can be expressed as

$$\mathbf{F} = -\frac{GMm}{r^3} \cdot \mathbf{r},$$

where the mass M is assumed to be at the origin, $\mathbf{r}$ denotes the position vector of the mass m, G is a constant and $\mathbf{F}$ denotes the gravitational force between the two (point) masses.

A simple computation shows that

$$\nabla \left(\frac{1}{r} \right) = -\frac{\mathbf{r}}{r^3}.$$

Thus the gravitational force at any point can be expressed as the gradient of a function. Moreover, it is clear that

$$\left\| \nabla \left(\frac{1}{r} \right) \right\| = \left\| -\frac{\mathbf{r}}{r^3} \right\| = \frac{1}{r^2}.$$

Keeping our previous discussion in mind, we know that if $V = GMm/r$, $\mathbf{F} = \nabla V$.

What are the level surfaces of V? Clearly, r must be a constant on these level sets, so the level surfaces are spheres. Since $\mathbf{F}$ is a multiple of $-\mathbf{r}$, we see that F points towards the origin and is thus orthogonal to the sphere.

In order to make our notation less cumbersome, we introduce the notation f_x for the partial derivative $\frac{\partial f}{\partial x}$. The notations f_y and f_z will have the obvious meanings.

Since we know that the gradient of f is normal to the level surface S given by $f(x, y, z) = c$ (provided the gradient is non zero), it allows us to write down the equation of the tangent plane of S at the point $s = (x_0, y_0, z_0)$. The equation of this plane is

$$f_x(x_0, y_0, z_0)(x-x_0)+f_y(x_0, y_0, z_0)(y-y_0)+f_z(x_0, y_0, z_0)(z-z_0) = 0.$$

For the curve $f(x, y) = c$ we can similarly write down the equation of the tangent passing through (x_0, y_0):

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) = 0.$$

Note that the fact that the gradient of f is normal to the level surface $f(x, y, z) = c$ is true only for implicitly defined surfaces. If the surface is given as $z = f(x, y)$, then we cannot simply take the gradient of f and make the same statement. We must first convert our explicit surface to the implicit surface S given by $g(x, y, z) = z - f(x, y) = 0$. Then ∇g will be normal to S.

The proof of the chain rule

How does one actually prove the chain rule for a function $f(x, y)$ of two variables? We can write

$$f(x(t+h), y(t+h)) = f(x(t)+h[x'(t)+p_1(h)], y(t)+h[y'(t)+p_2(h)])$$

for functions p_1 and p_2 that go to zero as h goes to zero. Here we are simply using the differentiability of x and y as functions of t. Now we can write the right hand side as

$$f(x(t), y(t)) + Df(h[x'(t) + p_1(h)], h[y'(t) + p_2(h)]) + p_3(h)h$$

by using the differentiability of f, for some other function $p_3(h)$ which goes to zero as h goes to zero (you may need to think about this step a little). This gives

$$f(x(t+h), y(t+h)) - f(x(t), y(t)) - f_x x'(t)h - f_y y'(t)h = p(h)h.$$