MA 105 D1 Lecture 11

Ravi Raghunathan

Department of Mathematics

Autumn 2014, IIT Bombay, Mumbai

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

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

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 $f(c(t))$:

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

and this can be rewritten

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

Of course, the same argument works when $U \subset \mathbb{R}^2$ and f is a function of two variables.

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 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.

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, if $r \neq 0$. Moreover, it is clear that

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

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.

The equation of the tangent plane

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):

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

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

$$f(x(t+h), y(t+h)) = f(x(t)+h[x'(t)+o_1(h)], y(t)+h[y'(t)+o_2(h)])$$

for functions o_1 and o_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

$$f(x(t), y(t)) + Df(h[x'(t) + o_1(h)], h[y'(t) + o_2(h)]) + o_3(h)h$$

by using the differentiability of f, for some other function $o_3(h)$ which goes to zero as h goes to zero (you may need to think about this step a little). In turn, this can be rewritten

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

We can now divide by h to get the desired result.

Functions from $\mathbb{R}^m \to \mathbb{R}^n$

So far we have only studied functions whose range was a subset of $\mathbb{R}$. Let us now allow the range to be $\mathbb{R}^n$, $n = 1, 2, 3, \ldots$. Can we understand what continuity, differentiability etc. mean?

Let U be a subset of $\mathbb{R}^m$ $(m = 1, 2, 3, \ldots)$ and let $f : U \to \mathbb{R}^n$ be a function. If $x = (x_1, x_2, \ldots, x_m) \in U$, $f(x)$ will be an n-tuple where each coordinate is a function of x. Thus, we can write $f(x) = (f_1(x), f_2(x), \ldots, f_n(x))$, where each $f_i(x)$ is a function from U to $\mathbb{R}$.

Functions which take values in $\mathbb{R}$ are called scalar valued functions, which functions which take values in $\mathbb{R}^n$, $n > 1$ are usually called vector valued functions.

Vector fields

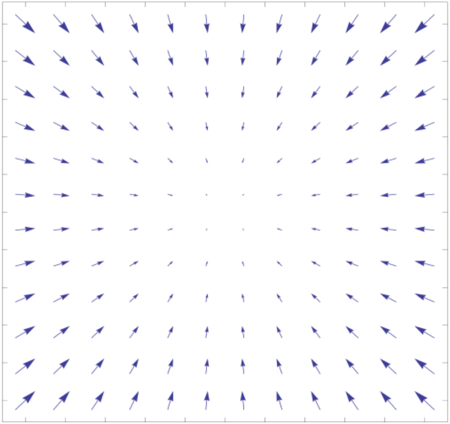
When $m = n$, vector valued functions are often called vector fields. We will study vector fields in slightly greater detail when $m = n = 2$ and $m = n = 3$.

We have already seen one example of a vector field - the gravitational force field $-\frac{GMm}{r^3} \cdot \mathbf{r}$ felt by a mass m whose position vector with respect to a mass M at the origin is $\mathbf{r}$. In this particular case we showed the the force field arose as the gradient of a scalar valued function (the potential $V = GMm/r$).

One of the most important questions in calculus is the following: Given a vector field, when does it arise as the gradient of a scalar function?. In physics, vector force fields that arise from a scalar potential function are called Conservative.

Some pictures of vector fields

We can actually visualize two dimensional vector fields as follows. At each point in $\mathbb{R}^2$ we can draw an arrow starting at that point pointing in the direction of the image vector and with size proportional to the magnitude of the image vector.



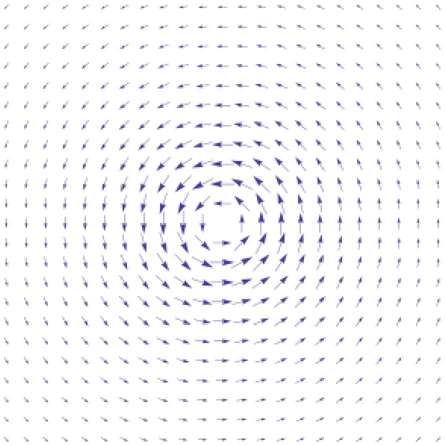
What function from $\mathbb{R}^2$ to $\mathbb{R}^2$ does this picture represent?

$$f(x, y) = (-x, -y)$$

the the radial vector field.

http://en.wikipedia.org/wiki/File:Radial_vector_field_dense.svg

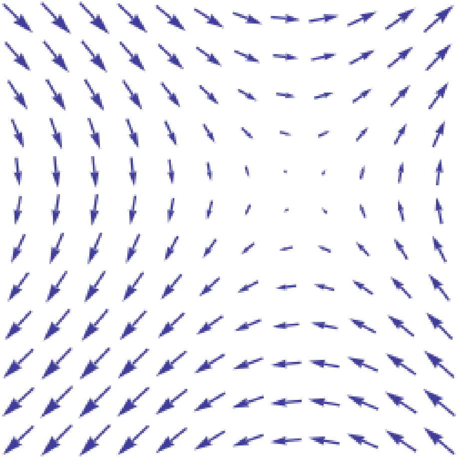
How about this one?



$$f(x,y) = \left(-\frac{y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$$

This is an example of an irrotational vector field.
It cannot be written as the gradient of a potential function.

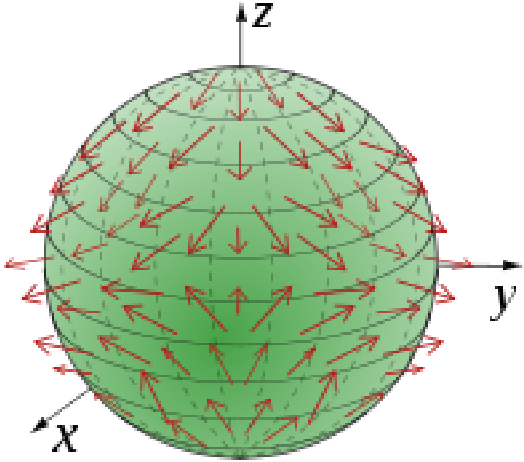
Here is another (more complicated one)



$$f(x, y) = (\sin y, \sin x)$$

http://en.wikipedia.org/wiki/File:VectorField.svg

One can also talk about two dimensional vector fields on any two dimensional surface. Here is a picture of a vector field on a sphere.



http://en.wikipedia.org/wiki/File:Vector_sphere.svg

Vector fields in the real world

Many real world phenomena can be understood using the language of vector fields. In physics, apart from gravitation, electromagnetic forces can also be represented by vector fields. That is, to each point in space we attach the vector representing the force at that point. Such fields are called force fields.

Fluids flowing are also often modeled using vector fields, with each point being mapped to the vector representing the velocity of the fluid flow. For instance, the velocity of winds in the atmosphere can be represented as a vector field. Such fields are called velocity fields.

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 a 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 $\mathbb{R}^m$!).