

9.4.11 The implicit function theorem

Back in §4.3 (lecture 4) there were remarks that there was a theorem to guarantee that implicit differentiation is justifiable as long as it give a sensible answer (sensible meaning no division by zero). We can shed a little more light on that theorem now, and perhaps explain a little better what is missing in the ‘proof’ above that the gradient vector is perpendicular to the level curve.

In implicit differentiation we assume that there was a function $y = y(x)$ that had a derivative and satisfied an equation in x and y . We could write that equation in x and y in terms of a function of two variables. Say the equation is

$$g(x, y) = 0$$

(when we put everything on the left of the equation and use the equation to define $g(x, y)$). Now we see that this relates to a level curve of $g(x, y)$, the level curve $g(x, y) = c$ where $c = 0$.

What we assumed then was that there was a differentiable function $y = y(x)$ so that

$$g(x, y(x)) = 0. \tag{1}$$

If we use the chain rule for partial derivatives we find that

$$\begin{aligned} \frac{d}{dx}g(x, y(x)) &= \frac{\partial g}{\partial x} \frac{dx}{dx} + \frac{\partial g}{\partial y} \frac{dy}{dx} \\ &= \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} \frac{dy}{dx} \end{aligned}$$

But, because of (1), we must get

$$\frac{d}{dx}g(x, y(x)) = 0$$

and so we have

$$\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} \frac{dy}{dx} = 0.$$

Solving that gives

$$\frac{dy}{dx} = -\frac{\partial g / \partial x}{\partial g / \partial y},$$

which gives dy/dx in terms of x and y .

In implicit differentiation we have to assume we know y as well as x . That is we need a $y_0 = y(x_0)$ or one point (x_0, y_0) that satisfies $g(x, y) = 0$.

There is a theorem called the *implicit function theorem* that says:

If $g(x, y)$ is such that the partial derivatives $g_x(x, y)$ and $g_y(x, y)$ are continuous functions, and if we know both

$$g(x_0, y_0) = 0 \text{ and } g_y(x_0, y_0) \neq 0,$$

then there is a differentiable function $y = y(x)$ defined for x close to x_0 so that

$$g(x, y(x)) = 0 \quad (\text{for all } x \text{ in the domain of } y(x)).$$

Moreover the graph $y = y(x)$ fills the level curve $g(x, y) = 0$ near (x_0, y_0) and

$$\frac{dy}{dx} = y'(x) = -\frac{g_x(x, y(x))}{g_y(x, y(x))}.$$

This is quite tricky to prove, but it is the meaning of the statement made previously in §4.3.

We could make a similar statement where the roles of x and y are reversed. So if $g(x_0, y_0) = 0$ and $g_x(x_0, y_0) \neq 0$, then there is a differentiable function $x = x(y)$ defined for y near y_0 that solves

$$g(x(y), y) = 0.$$

For that function we will get

$$\frac{dx}{dy} = -\frac{g_y(x(y), y)}{g_x(x(y), y)}.$$

Putting these two versions together, we see that if $\nabla g|_{(x_0, y_0)} \neq 0$ then we can solve the equation

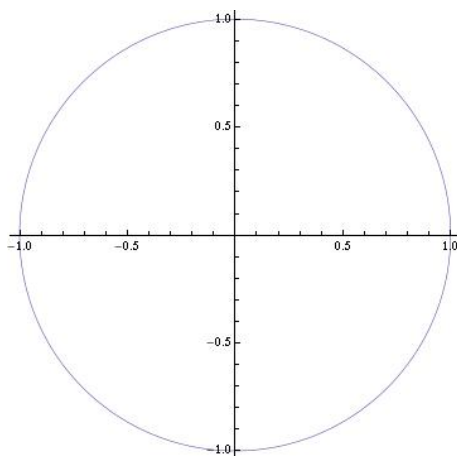
$$g(x, y) = 0$$

either for $y = y(x)$ as a differentiable function of x *or* for $x = x(y)$ as a differentiable function of y . We are not saying we can solve it everywhere, but just nearby the point we start with.

A very simple example comes from $g(x, y) = x^2 + y^2 - 1$. The level curve $g(x, y) = 0$ is the unit circle in this case and

$$\nabla g = (2x, 2y)$$

is never zero on the circle.



If we take $(x_0, y_0) = (0, -1)$ we have $g_y(x_0, y_0) = -2 \neq 0$. So according to the implicit function theorem we must be able to solve $g(x, y) = 0$ for y in terms of x near $x = 0$ so as to have $y(0) = -1$. Indeed, this is quite easy here, as we can just write down a solution

$$y = y(x) = -\sqrt{1 - x^2}$$

and it will be differentiable for $-1 < x < 1$ (which certainly includes $x = 0$).

However, if we took $(x_0, y_0) = (1, 0)$ we cannot solve for y as a function of x near $x = 1$ (both sides of 1). We can solve for x in terms of y though. We can again write the solution down as

$$x = x(y) = \sqrt{1 - y^2}$$

(which is differentiable for $-1 < y < 1$).

The importance of the implicit function theorem is not for simple examples like this (where we can actually solve the equation), but rather that for complicated equations it guarantees that a bit (at least a small bit) of the level curve $g(x, y) = 0$ will be the graph of a differentiable function.

To justify the remarks in the previous subsection §12.4.10 about parametrising the level curve, we would also need the implicit function theorem or something closely related to that theorem.

9.4.12 Gradient is direction of steepest ascent

The word ‘gradient vector’ is justified by the fact that it relates to the direction of steepest ascent.

In a way we know this already because we know from §9.4.7 that the directional derivative $D_{\mathbf{u}}f(x_0, y_0)$ is largest (for a fixed point (x_0, y_0)) when \mathbf{u} is the unit vector in the direction of the gradient $\nabla f|_{(x_0, y_0)}$. We also know that the largest possible value of the directional derivative is the magnitude of the gradient vector.

We can relate this to the graph $z = f(x, y)$. We can think of the graph as a landscape. Suppose we are starting from the point on the graph where $(x, y) = (x_0, y_0)$. The directional derivative $D_{\mathbf{u}}f(x_0, y_0)$ gives us the slope we would encounter if we travelled from position (x_0, y_0) in the direction \mathbf{u} (or think of compass heading \mathbf{u}). We get the largest slope in the direction of the gradient, the smallest slope in the opposite direction.

That means that the gradient vector gives the most uphill direction.

The level curve direction is the direction perpendicular to the gradient. That is what we just figured out above when we showed that the gradient vector is perpendicular to (the tangent line) to the level curve. It also makes sense that if you travel along the level curve you will encounter zero slope.

9.5 Functions of 3 variables

We look now at functions $f(x, y, z)$ or $f(x_1, x_2, x_3)$ of 3 variables.

A lot of what we have to say will be similar to what we did before for functions of two variables, but there will also be some differences.

The first thing that is the same (or hardly any different) is that the function is a well-defined rule associated a single value $f(x, y, z)$ to each point (x, y, z) in its domain. Usually, but not necessarily, the functions we will consider will be specified by giving a formula to find the values. Examples could be

$$\begin{aligned} f(x, y, z) &= x^2 + y^2 + z^2 \\ f(x, y, z) &= x \cos(y + z) + \frac{1}{x} \\ f(x_1, x_2, x_3) &= \frac{x_1 x_2 x_3}{\sqrt{x_1^2 + x_2^2 + x_3^2 + 1}} \end{aligned}$$

As with the earlier case of functions of two variables, a function $f(x, y, z)$ will be defined for points (x, y, z) in some domain, some subset of \mathbb{R}^3 or for all points of \mathbb{R}^3 if the domain is the whole of \mathbb{R}^3 . As with functions of two variables we will sometimes assume that the domain is as big as it can be for the formula for $f(x, y, z)$ to make sense. In the second of the three examples above, we should exclude points with $x = 0$ from the domain as we don't want to be dividing by zero. For the other two examples above we don't need any restrictions.

However, when we start to think of graphs, we find a problem. For a graph of a function of one variable (a graph $y = f(x)$) we need to draw it in the plane (so we need two dimensions, one for the domain and one more for the values). Similarly when we looked at functions of two variables we needed 3 dimensions to fit in the graph $z = f(x, y)$.

A graph of a function $f(x, y, z)$ would need 4 dimensions. It already takes 3 dimensional space to keep track of the domain points (x, y, z) and we would need to find a fourth dimension to fit in the values. So we can't look at it in reality. We can do it theoretically, but then we can only work by analogy with what we are familiar with.

9.5.1 Level surfaces

The level surface approach $f(x, y, z) = c$ is available. It is harder perhaps to think about than the level curves we had in the plane (for functions of two variables), but the level surfaces are in space where we can see (or visualise) them.

For example in the case $f(x, y, z) = x^2 + y^2 + z^2$, we can describe the level surfaces quite easily. For $c > 0$, the equation $f(x, y, z) = c$ means the distance squared from (x, y, z) to the origin is c . So the points (x, y, z) satisfying $f(x, y, z) = c$ are exactly the points on the sphere of radius \sqrt{c} around the origin. [By the sphere we mean the surface. We generally use the word ball for the inside of the sphere, or the surface together with the inside.] When $c = 0$ there is only the one point $(x, y, z) = (0, 0, 0)$ (the origin) on the 'level surface' $f(x, y, z) = c$. We see then that the level set is not always a 'surface' but it typically is a surface if there are any points at all on it. In this example, when $c < 0$ there are no points satisfying $f(x, y, z) = c$.

9.5.2 Derivatives of functions of 3 variables

Quite a number of the things we said about derivatives of functions of two variables will also work (more or less in the same way) for functions of 3 variables. There are more variables and so more things to take care about, but most of the ideas are quite similar.

9.5.3 Partial derivatives

The partial derivative of $f(x, y, z)$ with respect to x is denoted $\frac{\partial f}{\partial x}$ — it is computed by differentiating $f(x, y, z)$ with respect to x while treating y and z (the other variables) as though they were constant. We could define it more formally as

$$\frac{\partial f}{\partial x} \big|_{(x_0, y_0, z_0)} = \frac{d}{dx} \big|_{x=x_0} f(x, y_0, z_0) = \frac{d}{dt} \big|_{t=0} f(x_0 + t, y_0, z_0).$$

The other partial derivatives $\frac{\partial f}{\partial y}$ and $\frac{\partial f}{\partial z}$ work in a similar way.

There is also the notation $f_x(x_0, y_0, z_0)$ for the value of $\partial f / \partial x$ at the point (x_0, y_0, z_0) .

9.5.4 Directional derivatives

These are again similar to the two variable case.

For a unit vector $\mathbf{u} \in \mathbb{R}^3$ the directional derivative of $f(x, y, z)$ at (x_0, y_0, z_0) in the direction \mathbf{u} is

$$D_{\mathbf{u}}f(x_0, y_0, z_0) = \frac{d}{dt} \big|_{t=0} f((x_0, y_0, z_0) + t\mathbf{u})$$

and we compute it using partial derivatives

$$D_{\mathbf{u}}f(x_0, y_0, z_0) = \left(\frac{\partial f}{\partial x} \big|_{(x_0, y_0, z_0)} \right) u_1 + \left(\frac{\partial f}{\partial y} \big|_{(x_0, y_0, z_0)} \right) u_2 + \left(\frac{\partial f}{\partial z} \big|_{(x_0, y_0, z_0)} \right) u_3$$

(where u_1 , u_2 and u_3 are the components of the unit vector $\mathbf{u} = (u_1, u_2, u_3) = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$).

The justification of this formula for the directional derivative depends on the validity of the *linear approximation formula*

$$\begin{aligned} f(x, y, z) &\cong f(x_0, y_0, z_0) + \\ &\quad \left(\frac{\partial f}{\partial x} \big|_{(x_0, y_0, z_0)} \right) (x - x_0) + \left(\frac{\partial f}{\partial y} \big|_{(x_0, y_0, z_0)} \right) (y - y_0) + \left(\frac{\partial f}{\partial z} \big|_{(x_0, y_0, z_0)} \right) (z - z_0) \end{aligned}$$

To be more accurate, what we need is that $f(x, y, z)$ should have a *derivative* at the point (x_0, y_0, z_0) , which means that there should be constants a_1 , a_2 and a_3 so that

$$\lim_{(x, y, z) \rightarrow (x_0, y_0, z_0)} \frac{f(x, y, z) - (f(x_0, y_0, z_0) + a_1(x - x_0) + a_2(y - y_0) + a_3(z - z_0))}{\text{dist}((x, y, z), (x_0, y_0, z_0))} = 0$$

As in the two variable case we will not go into the definition of a limit in space, but roughly the limit above being 0 means that if (x, y, z) is close enough to (x_0, y_0, z_0) but $(x, y, z) \neq (x_0, y_0, z_0)$, then the fraction above is small.

If the function has a derivative, then the numbers a_1 , a_2 and a_3 have to be equal to the partial derivatives $\partial f/\partial x$, $\partial f/\partial y$ and $\partial f/\partial z$ evaluated at (x_0, y_0, z_0) . Moreover the formula in the box above for directional derivatives will be valid.

How do we know a function has a derivative? Well, as in the 2 variable case there is a theorem that guarantees that most of the functions we can easily write down have derivatives (at least at most points). The theorem says that if the 3 partial derivatives exist and are *continuous functions* at (x_0, y_0, z_0) , then the function has a derivative. We won't go into continuous functions much, but a function $g(x, y, z)$ is called continuous at a point (x_0, y_0, z_0) of its domain if

$$\lim_{(x,y,z) \rightarrow (x_0,y_0,z_0)} g(x, y, z) = g(x_0, y_0, z_0).$$

The function is called continuous (with no mention of a specific point) if it is continuous at each point of its domain. There are theorems about limits of products, quotients and such combinations of functions that allow one to be sure that most formulae define continuous functions (apart perhaps from problem points).

9.5.5 The gradient

In 3 variables the gradient of $f(x, y, z)$ is the vector

$$\nabla f|_{(x_0,y_0,z_0)} = \left(\frac{\partial f}{\partial x} |_{(x_0,y_0,z_0)}, \frac{\partial f}{\partial y} |_{(x_0,y_0,z_0)}, \frac{\partial f}{\partial z} |_{(x_0,y_0,z_0)} \right)$$

or

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)$$

The formula in the box above for the directional derivative can be restated with the gradient vector

$$D_{\mathbf{u}}f(x_0, y_0, z_0) = (\nabla f|_{(x_0,y_0,z_0)}) \cdot \mathbf{u}$$

So, as before in 2 variables, we can say that the directional derivative is

$$D_{\mathbf{u}}f(x_0, y_0, z_0) = \|\nabla f|_{(x_0,y_0,z_0)}\| \cos \theta$$

where θ is the angle between the gradient vector and the direction (unit vector) \mathbf{u} . From this we can say that the directional derivative is largest when \mathbf{u} is the unit vector in the direction of the gradient, that is when

$$\mathbf{u} = \frac{1}{\|\nabla f|_{(x_0,y_0,z_0)}\|} \nabla f|_{(x_0,y_0,z_0)}.$$

And the largest possible value of the directional derivative is the magnitude of the gradient vector. The smallest possible directional derivative occurs when \mathbf{u} is in the exact opposite direction to the gradient vector ($\theta = \pi$ and $\cos \theta = -1$) and this smallest directional derivative is minus the magnitude of the gradient vector.

Another way to put that is that the gradient vector is a vector in the direction of steepest increase of the function and the length of the gradient vector is the rate of increase in that direction.

An application of this that arises fairly frequently is when the function represents potential energy. If a charged particle is subject to an electric field it would have a potential energy $V(x, y, z)$ that depends on its position in the field. If left unchecked by any other influence, the particle would tend to go in the direction of steepest decrease of potential energy. That means that the force acting on the particle would be $\mathbf{F} = -\nabla V$ (and would depend on position (x, y, z)). The same idea would apply if the particle has some mass and is in a gravitational field (caused by the earth, the moon or some other body). The gravitational force acting on it would be minus the gradient (thus in the direction of free fall). If a gas like the atmosphere had a variable pressure then the gas will tend to fill in the low pressure areas. So it will be influenced to go towards the low pressure areas in the direction of fastest decrease of pressure.

9.5.6 Gradient perpendicular to the level surface

An important thing to mention is the property that the gradient is perpendicular to the level surface. This fits with the fact that the directional derivative is zero in the directions perpendicular to the level surface.

We will go on to give some examples of this, but we won't really explain it further. A reasonably good explanation could be given by considering curves $\mathbf{p}(t) = (x(t), y(t), z(t))$ that lie entirely in a level surface $f(x, y, z) = c$ and satisfy $p(0) = (x_0, y_0, z_0)$. Then differentiate the equation

$$f(x(t), y(t), z(t)) = c$$

with respect to t using the chain rule. This is similar to an explanation we outlined in §12.4.10 for level curves of functions of two variables. A difference is that when we are dealing with a level surface there are many possible curves on the surface that pass through the point (x_0, y_0, z_0) and we should look at all of them to see all possible tangent directions. (For each direction in the tangent plane to the surface you can expect to find a curve in the surface that is tangent to that direction as it passes (x_0, y_0, z_0) .)

We did not say yet what the chain rule is in this case, but it is like the two variable version with one extra term:

$$\frac{d}{dt}f(x(t), y(t), z(t)) = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial z} \frac{dz}{dt}$$

Examples 9.5.6.1. (i) Find the equation of the tangent plane to the surface

$$x^4 + y^2 + z^6 = 26$$

at the point $(2, 3, 1)$.

We can see that we are dealing with a level surface to the function $f(x, y, z) = x^4 + y^2 + z^6$, the level surface $f(x, y, z) = c$ where $c = 26$. According to what we said above the normal vector to the level surface is $\nabla f|_{(2,3,1)}$ and so we can set about computing that gradient vector.

$$\begin{aligned}\frac{\partial f}{\partial x} &= 4x^3 \\ \frac{\partial f}{\partial y} &= 2y \\ \frac{\partial f}{\partial z} &= 6z^5\end{aligned}$$

Evaluating these at the point $(2, 3, 1)$ we get our normal vector

$$\nabla f|_{(2,3,1)} = (32, 6, 6)$$

So the equation of the plane through $(2, 3, 1)$ perpendicular to that is

$$32(x - 2) + 6(y - 3) + 6(z - 1) = 0$$

or $32x + 6y + 6z = 88$.

- (ii) We can treat a graph $z = f(x, y)$ of a function $f(x, y)$ of 2 variables as a level surface $g(x, y, z) = 0$ of the function $g(x, y, z) = z - f(x, y)$ of 3 variables.

This allows us to calculate the tangent plane to the graph in a different way to the way we did previously. Say we want the tangent plane at the point in the graph where $(x, y) = (x_0, y_0)$. We will need to notice before we finish that the z coordinate at that point on the graph is $z_0 = f(x_0, y_0)$. We have

$$\nabla g = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z} \right) = \left(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1 \right)$$

and we should evaluate this at (x_0, y_0, z_0) to get the normal vector to the tangent plane. This gives the equation to the tangent plane to the graph is

$$\left(-\frac{\partial f}{\partial x} \Big|_{(x_0, y_0)} \right) (x - x_0) + \left(-\frac{\partial f}{\partial y} \Big|_{(x_0, y_0)} \right) (y - y_0) + (z - z_0) = 0.$$

We can rearrange that as

$$z = z_0 + \left(\frac{\partial f}{\partial x} \Big|_{(x_0, y_0)} \right) (x - x_0) + \left(\frac{\partial f}{\partial y} \Big|_{(x_0, y_0)} \right) (y - y_0)$$

and this should look familiar.