If you try more examples though, you will begin to see that we seem to always have this

$$D_{\mathbf{u}}f(x_0, y_0) = \left(\frac{\partial f}{\partial x} \Big|_{(x_0,y_0)} \right) u_1 + \left(\frac{\partial f}{\partial y} \Big|_{(x_0,y_0)} \right) u_2 \qquad (1)$$

Or we might be better to use the other notation for partials

$$D_{\mathbf{u}}f(x_0, y_0) = f_x(x_0, y_0)u_1 + f_y(x_0, y_0)u_2$$

or to just write the formula at an arbitrary (x, y) (without trying to write down that both sides have to be evaluated at the same point):

$$D_{\mathbf{u}}f = \frac{\partial f}{\partial x}u_1 + \frac{\partial f}{\partial y}u_2.$$

Anyhow, it turns out that this formula is usually true. Not quite always, but usually in examples we can easily compute with. We will explain a little more carefully later when it is a valid formula.

The upshot of all this is that the values of the partial derivatives contain the information needed to find all the directional derivatives.

2.4.4 Tangent plane to a graph

In order to explain what it is that makes the formula for directional derivatives given above true, and the circumstances where we can be sure it is true, we explain about the tangent plane to a graph $z = f(x, y)$ at a point (x_0, y_0, z_0) on the graph.

Just as in one variable there are some graphs with no tangent line, or functions that are not differentiable at certain points, there are cases of graphs $z = f(x, y)$ that do not have a tangent plane. Before we can make that more clear, we will need to say what we mean by a tangent plane.

What we will do first is work out what the equation for the tangent plane should be, assuming there is a tangent plane. Then we will be able to say what the definition of a tangent plane is.

So we look for the tangent plane to a graph $z = f(x, y)$ at a point (x_0, y_0, z_0) on the graph. One simple thing that we need in practice is that if the point is on the graph then $z_0 = f(x_0, y_0)$. (We use this to work out the z-coordinate if given the x- and y-coordinates.)

Then we consider the sections of the graph in the x- and y-directions, the same sections we looked at when we were describing partial derivatives. Rather than considering the graphs $z = f(x, y_0)$ (where we fix $y = y_0$) and $z = f(x_0, y)$ (where we fix $x = x_0$) as graphs in a vertical plane we look at them as curves in space. We can parametrise them as

$$\begin{aligned} x &= t \\ y &= y_0 \\ z &= f(t, y_0) \end{aligned}$$

and

$$
\begin{aligned}
x &= x_0 \\
y &= t \\
z &= f(x_0, t_0).
\end{aligned}
$$

Look at the first of these

$$(x, y, z) = (t, y_0, f(t, y_0))$$

which passes by our point (x_0, y_0, z_0) at $t = x_0$. It will have a tangent vector at that point

$$\frac{d}{dt} \mid_{t=x_0} (t, y_0, f(t, y_0)) = \left(1, 0, \frac{d}{dt} f(t, y_0)\right) \mid_{t=x_0} = \left(1, 0, \frac{\partial f}{\partial x} \mid_{(x_0, y_0)}\right)$$

Whatever the tangent plane to the graph $z = f(x, y)$ is, it seems reasonable to believe that this tangent vector should be in the plane (or parallel to it).

Similarly working with the curve we get along the y-direction, we have

$$(x, y, z) = (x_0, t, f(x_0, t))$$

and the tangent vector to that (at $t = y_0$) will come out as

$$\frac{d}{dt} \mid_{t=y_0} (x_0, t, f(x_0, t)) = \left(0, 1, \frac{d}{dt} f(x_0, t)\right) \mid_{t=y_0} = \left(0, 1, \frac{\partial f}{\partial y} \mid_{(x_0, y_0)}\right).$$

This vector should be in the plane also, for the same reason.

So our tangent plane should be parallel to both the vectors

$$\mathbf{A} = (1, 0, f_x(x_0, y_0)) \text{ and } \mathbf{B} = (0, 1, f_y(x_0, y_0)).$$

From properties of the cross product, we know then that $\mathbf{A} \times \mathbf{B}$ should be normal to our tangent plane. We can work that out:

$$
\begin{aligned}
\mathbf{A} \times \mathbf{B} &= \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & f_x(x_0, y_0) \\ 0 & 1 & f_y(x_0, y_0) \end{pmatrix} \\
&= \mathbf{i} \det \begin{pmatrix} 0 & f_x(x_0, y_0) \\ 1 & f_y(x_0, y_0) \end{pmatrix} - \mathbf{j} \det \begin{pmatrix} 1 & f_x(x_0, y_0) \\ 0 & f_y(x_0, y_0) \end{pmatrix} + \mathbf{k} \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= -f_x(x_0, y_0)\mathbf{i} - f_y(x_0, y_0)\mathbf{j} + \mathbf{k} \\
&= (-f_x(x_0, y_0), -f_y(x_0, y_0), 1)
\end{aligned}
$$

We can then write down that the equation of the plane through (x_0, y_0, z_0) perpendicular to that vector should be

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

Rearranging it to have fewer minus signs, we get

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

Recalling that $z_0 = f(x_0, y_0)$, we find that the tangent plane should have this equation

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

(if there is a tangent plane).

2.4.5   Linear approximation and the derivative

We turn now to the definition of what it means for $z = f(x, y)$ to have a tangent plane at (x_0, y_0, z_0). The idea is fairly simple — the graph $z = f(x, y)$ should be close to the graph of the plane near the point (x_0, y_0). Or in slightly more precise terms, the two variable analogue of the linear approximation formula should hold.

We should be able to say that there is a plane

$$z = z_0 + a_1(x - x_0) + a_2(y - y_0)$$

(or graph of a function with only constant and linear terms) through (x_0, y_0, z_0) so that it is true that

$$f(x, y) \cong z_0 + a_1(x - x_0) + a_2(y - y_0)$$

is a good approximation when we keep (x, y) not too far from (x_0, y_0). Unless we are a bit more precise about what good means, this will still be ambiguous. We want to insist on something that can't be true of two different planes and is still often achievable.

What we do is insist that the error in the above approximation should be small compared to the distance from (x, y) to (x_0, y_0) (when that distance is itself small).

Definition 2.4.5.1. A function $z = f(x, y)$ is *differentiable* at (x_0, y_0) if it is possible to find constants a_1 and a_2 so that

$$\lim_{(x,y)\to(x_0,y_0)} \frac{f(x, y) - (f(x_0, y_0) + a_1(x - x_0) + a_2(y - y_0))}{\text{dist}((x, y), (x_0, y_0))} = 0.$$

We could rewrite that

$$\lim_{(x,y)\to(x_0,y_0)} \frac{f(x, y) - (f(x_0, y_0) + a_1(x - x_0) + a_2(y - y_0))}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} = 0. \tag{2}$$

Sometimes the term *total derivative* is used for this rather than derivative.

We are being a bit cavalier here in bringing in limits of functions of two variables with no explanation. The book by Thomas has a lot more than we are going to do about this. One thing to point out is that when we let (x, y) approach (x_0, y_0), we never actually take (x, y) to be equal to (x_0, y_0). That means we don't end up having to worry about interpreting $0/0$ in the above limit. Another important thing is that (x, y) does not just approach (x_0, y_0) from one direction, but from all directions at once. The interpretation of the limit being 0 is that the thing we are taking the limit of will be close to 0 as long as we take (x, y) within some small distance of (x_0, y_0). But we must not have to make the small distance be 0, and we should never consider $(x, y) = (x_0, y_0)$.

Proposition 2.4.5.2. *If (2) holds, then*

$$D_{\mathbf{u}}f(x_0, y_0) = a_1 u_1 + a_2 u_2$$

must be true (for all unit vectors $\mathbf{u} = (u_1, u_2)$*).*

The number a_1 *in the definition (2) must always be the partial derivative with respect to* x *at* (x_0, y_0) *(and the number* a_2 *must be the partial with respect to* y*).*

Proof. For this we go back to first principles, and write the directional derivative as a limit

$$D_{\mathbf{u}}f(x_0, y_0) = \frac{d}{dt}\,|_{t=0}\, f(x_0 + tu_1, y_0 + tu_2) = \lim_{t \to 0} \frac{f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0)}{t}$$

What we are going to prove is that, if we know (2), then $D_{\mathbf{u}}f(x_0, y_0)$ must exist and must be the number $a_1 u_1 + a_2 u_2$. We do this by looking at the difference

$$\frac{f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0)}{t} - (a_1 u_1 + a_2 u_2)$$

$$= \frac{f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0) - (a_1 u_1 + a_2 u_2)t}{t}$$

$$= \frac{f(x_0 + tu_1, y_0 + tu_2) - (f(x_0, y_0) + a_1 tu_1 + a_2 tu_2)}{t}$$

and its absolute value

$$\frac{|f(x_0 + tu_1, y_0 + tu_2) - (f(x_0, y_0) + a_1 tu_1 + a_2 tu_2)|}{|t|}.$$

Notice that

$$|t| = \text{dist}((x_0 + tu_1, y_0 + tu_2), (x_0, y_0))$$

since $\mathbf{u}$ is a unit vector. Therefore, from (2) we can say that

$$\lim_{t \to 0} \frac{f(x_0 + tu_1, y_0 + tu_2) - (f(x_0, y_0) + a_1 tu_1 + a_2 tu_2)}{t} = 0.$$

Hence $D_{\mathbf{u}}f(x_0, y_0) = a_1 u_1 + a_2 u_2$.

Taking $\mathbf{u} = (1, 0)$ gives $f_x(x_0, y_0) = a_1$ and taking $\mathbf{u} = (0, 1)$ gives $f_y(x_0, y_0) = a_2$. □

The following definition is **not** really adequate in all situations.

Definition 2.4.5.3. We say that a function $f(x, y)$ is *continuous at a point* (x_0, y_0) if

$$\lim_{(x,y) \to (x_0, y_0)} f(x, y) = f(x_0, y_0).$$

We say that $f(x, y)$ is *continuous* if it is continuous at each point of its domain.

Theorem 2.4.5.4. *If a function* $z = f(x, y)$ *has partial derivatives* $\partial f / \partial x$ *and* $\partial f / \partial y$ *that are continuous functions at* (x_0, y_0)*, then* f *has a derivative at* (x_0, y_0)*.*

We will not attempt to prove this, but this is the meaning of the earlier claim that (1) is usually true.

Most functions that you can easily write down (via a formula) are continuous except perhaps at some bad points. Thus, we can use (1) routinely.