The tangent plane and directional derivative

Eccentric Ed and Nutty Ned are hikers. They love hiking up and down hills—but only along paths that lie in vertical planes: Ed insists on walking exclusively in the direction of due East (though he doesn't care which point (a, y, z) he starts from in the plane $x = a$ perpendicular to his direction of motion), and Ned insists on walking exclusively in the direction of due North (though he doesn't care which point (x, b, z) he starts from in the plane $y = b$ perpendicular to his direction of motion)—well, that's what's eccentric or nutty about them. For the sake of argument, let's suppose that Ed and Ned are hiking on the surface $z = f(x, y)$ when, by an amazing coincidence, they arrive simultaneously at the point with coordinates (a, b, c) and hence position vector

$$\boldsymbol{\rho}_0 \;=\; a\,\mathbf{i} + b\,\mathbf{j} + c\,\mathbf{k}. \tag{1}$$

Note that, of necessity,

$$c \;=\; f(a, b), \tag{2}$$

because $z = f(x, y)$ at every point on the surface.

In which directions are Ed and Ned heading—three-dimensionally, that is? (We already know that, two-dimensionally—in the sense of parallel to the sea-level plane—Ed is moving due east, and Ned is moving due north.) At the instant in question, Ed is walking on a path whose slope is $f_x(a, b)$. What does that mean? It means that, *if* Ed kept going in the same fixed direction (which in reality, of course, he can't, because he's following the surface), then for every unit he went along he'd have to go $f_x(a, b)$ units up (= down if $f_x(a, b)$ is negative). Another way of saying precisely the same thing is that Ed is moving parallel to

$$\mathbf{i} + f_x(a, b)\mathbf{k}. \tag{3}$$

So the tangent line to Ed's trajectory at the instant in question is the line through (a, b, c) that is parallel to $\mathbf{i} + f_x(a, b)\mathbf{k}$. This line has the property that it meets the surface in a single point in the vicinity of (a, b, c).[1] An analogous argument for Ned reveals that he is moving parallel to

$$\mathbf{j} + f_y(a, b)\mathbf{k}. \tag{4}$$

So the tangent line to Ned's trajectory at the instant in question is the line through (a, b, c) that is parallel to $\mathbf{j} + f_y(a, b)\mathbf{k}$. This line also has the property that it meets the surface in a single point in the vicinity of (a, b, c).

But these two lines determine a plane, which must therefore also meet the surface in a single point in the vicinity of (a, b, c). We call this plane the tangent plane. What is its equation? We need to find a vector that's normal to the plane, say $\mathbf{n}$. But if $\mathbf{n}$ is normal to the tangent plane, then it must be normal to both Ed's tangent line and Ned's tangent line

[1] It may eventually meet the surface again in some distant point, because the concavity of the surface in Ed's direction has changed sign—but distant, by definition, means outside the vicinity of.

(because both lie in the plane). So we can take

$$
\begin{aligned}
\mathbf{n} &= \{\mathbf{i} + f_x(a,b)\mathbf{k}\} \times \{\mathbf{j} + f_y(a,b)\mathbf{k}\} \\
&= \mathbf{i} \times \mathbf{j} + \mathbf{i} \times f_y(a,b)\mathbf{k} + f_x(a,b)\mathbf{k} \times \mathbf{j} + f_x(a,b)\mathbf{k} \times f_y(a,b)\mathbf{k} \\
&= \mathbf{i} \times \mathbf{j} + f_y(a,b)\mathbf{i} \times \mathbf{k} + f_x(a,b)\mathbf{k} \times \mathbf{j} + f_x(a,b)\, f_y(a,b)\, \mathbf{k} \times \mathbf{k} \qquad (5) \\
&= \mathbf{k} + f_y(a,b)\, \{-\mathbf{j}\} + f_x(a,b)\{-\mathbf{i}\} + f_x(a,b)\, f_y(a,b)\, \mathbf{0} \\
&= -f_x(a,b)\mathbf{i} - f_y(a,b)\mathbf{j} + \mathbf{k}.
\end{aligned}
$$

But if $\boldsymbol{\rho} = x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$, then the plane through $\boldsymbol{\rho}_0$ with normal $\mathbf{n}$ has equation $\mathbf{n}\cdot(\boldsymbol{\rho}-\boldsymbol{\rho}_0) = 0$; and (1) implies that $\boldsymbol{\rho} - \boldsymbol{\rho}_0 = (x-a)\mathbf{i} + (y-b)\mathbf{j} + (z-c)\mathbf{k}$. So, putting everything together, the equation of the tangent plane is

$$-f_x(a,b)\{x-a\} - f_y(a,b)\{y-b\} + 1.\{z-c\} \;=\; 0 \qquad (6)$$

or (because $c = f(a,b)$)

$$z \;=\; f(a,b) + f_x(a,b)\{x-a\} + f_y(a,b)\{y-b\}. \qquad (7)$$

Now, if you move from the point with coordinates (a,b,c) on the surface $z = f(x,y)$ to a neighboring point on the surface with coordinates $(x,y,z) = (a+dx, b+dy, c+dz)$, then the change in your elevation (whether positive or negative) is given precisely by $dz = z-c = z - f(a,b)$ where z means *height of surface*. But if the neighboring point is very close to (a,b,c)—i.e., if dx, dy and hence dz are all small—then there won't be much difference between the height of the surface and the height of its tangent plane at the original point. So the change in elevation (whether positive or negative) is approximately given by $z - f(a,b)$ where z means *height of tangent plane*. In other words, on using (7), for the change in elevation we have

$$dz \approx f_x(a,b)dx + f_y(a,b)dy \qquad (8)$$

The right-hand side of (8), i.e., the quantity

$$f_x(a,b)\, dx + f_y(a,b)\, dy \qquad (9)$$

is known as the differential of f at (a,b) in the domain of f. The related quantity $df = f_x\, dx + f_y\, dy$ is the differential of f at (x,y) in the domain of f, i.e., at a variable point, as opposed to at a fixed one. General differentials are best regarded, at least for now, as simply a compact way of incorporating both of a function's partial derivatives into a single equation. For example, for $f(x,y) = \sin(x^2y)$, we can regard the statement

$$df \;=\; 2xy\, \cos(x^2y)\, dx + x^2 \cos(x^2y)\, dy \;=\; x\, \cos(x^2y)\{2ydx + xdy\} \qquad (10)$$

as nothing more than a compact way of stating that

$$f_x \;=\; 2xy\, \cos(x^2y), \qquad f_y \;=\; x^2 \cos(x^2y). \qquad (11)$$

Now, we have already established that the surface graph $z = f(x,y)$ of the function f at the point with coordinates $(a,b,c) = (a,b,f(a,b))$ has slope $f_x(a,b)$ when moving across the domain at $0°$ to due east, or in the direction of $\mathbf{i}$; and the surface graph has slope f_y when moving across the domain at $90°$ to due east (= north), or in the direction of $\mathbf{j}$. In both cases, the slope of the graph in the given direction is the slope of the tangent plane in the given direction. But what is the slope of the graph when moving across the domain at an angle θ to due east, or in the direction of

$$\hat{\mathbf{s}} \;=\; \cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j}? \tag{12}$$

The answer is still the same: the slope of the graph in the direction of $\hat{\mathbf{s}}$ is the slope of the tangent plane in the given direction. Therefore, to find the rate of change of the function f in the direction of $\hat{\mathbf{s}}$, we must find the slope of the tangent plane in that direction. How do we do that? Same as always: go one unit along in the domain, and see how far that takes you up (or down, of course, which is simply up with a negative sign). But going one unit in the direction of $\hat{\mathbf{s}}$ is the same as going $\cos(\theta)$ units in the direction of $\mathbf{i}$ followed by $\sin(\theta)$ units in the direction of $\mathbf{j}$: it takes you from (a, b) to (x, y), where $x = a + \cos(\theta)$ and $y = b + \sin(\theta)$). But how far does that take you up? All we have to do is substitute for x and y in (7): it takes you up $z - c = z - f(a, b) =$

$$f_x(a,b)\{x-a\} + f_y(a,b)\{y-b\} \;=\; f_x(a,b)\,\cos(\theta) + f_y(a,b)\,\sin(\theta). \tag{13}$$

What have we just discovered? That the rate of change of the function f in the direction of $\hat{\mathbf{s}}$ is $\cos(\theta)f_x(a, b) + \sin(\theta)f_y(a, b)$. What would be a good notation for this quantity? Well, the rate of change of f in the direction of $\mathbf{i}$ is called f_x or $\frac{\partial f}{\partial x}$ because x measures distance in that direction. Similarly, the rate of change of f in the direction of $\mathbf{j}$ is called f_y or $\frac{\partial f}{\partial y}$ because y measures distance in that direction. What measures distance in the direction of $\hat{\mathbf{s}}$? Whatever we choose to call it. But because the direction of $\hat{\mathbf{s}}$ is precisely the same as that of $\mathbf{s}$ and $\mathbf{s}$ has magnitude s, it is surely most natural to use s. Then, from (13), the rate of change of $z = f(x, y)$ in the direction of $\hat{\mathbf{s}}$ is

$$\frac{\partial z}{\partial s} \;=\; f_s \;=\; \cos(\theta)\,f_x(a,b) + \sin(\theta)\,f_y(a,b) \tag{14}$$

at (a, b) in the domain of f. We call this quantity a directional derivative. The related quantity

$$f_s \;=\; \cos(\theta)\,f_x + \sin(\theta)\,f_y \tag{15}$$

is the corresponding directional derivative of f at (x, y) in the domain of f, i.e., at a variable point as opposed to at a fixed one.

We now observe from (12) that (15) may be rewritten as

$$f_s \;=\; (\cos(\theta)\,\mathbf{i} + \sin(\theta)\,\mathbf{j})\cdot(f_x\,\mathbf{i} + f_y\,\mathbf{j}) \;=\; \hat{\mathbf{s}}\cdot(f_x\,\mathbf{i} + f_y\,\mathbf{j}). \tag{16}$$

The vector $f_x\,\mathbf{i} + f_y\,\mathbf{j}$ is called the gradient of f.[2] It is convenient to define a vector operator—called the gradient operator, or grad or ∇ for short—by

$$\nabla \;=\; \mathbf{i}\,\frac{\partial}{\partial x} + \mathbf{j}\,\frac{\partial}{\partial y} \tag{17}$$

[2] At (x, y) on its domain. It is assumed, of course that f is differentiable.

and interpret $\nabla f = \left(\mathbf{i}\,\frac{\partial}{\partial x} + \mathbf{j}\,\frac{\partial}{\partial y}\right) f = \mathbf{i}\,\frac{\partial f}{\partial x} + \mathbf{j}\,\frac{\partial f}{\partial y}$ to mean $f_x\,\mathbf{i} + f_y\,\mathbf{j}$. Then, from (16), the (directional) derivative of f in the direction of $\hat{\mathbf{s}}$ is

$$f_s \;=\; \hat{\mathbf{s}} \cdot \nabla f \tag{18}$$

All of these ideas generalize in the obvious way to functions of three (or more) variables. That is, if $\phi = \phi(x, y, z)$ is a sufficiently smooth function of three variables, then its gradient is defined by

$$\nabla\phi \;=\; \left(\mathbf{i}\,\frac{\partial}{\partial x} + \mathbf{j}\,\frac{\partial}{\partial y} + \mathbf{k}\,\frac{\partial}{\partial z}\right)\phi \equiv \frac{\partial\phi}{\partial x}\,\mathbf{i} + \frac{\partial\phi}{\partial y}\,\mathbf{j} + \frac{\partial\phi}{\partial z}\,\mathbf{k} \tag{19}$$

and the (directional) derivative of ϕ in the direction of $\hat{\mathbf{s}}$ is

$$\phi_s \;=\; \hat{\mathbf{s}} \cdot \nabla\phi. \tag{20}$$

Because the quantity $\phi_s = \hat{\mathbf{s}} \cdot \nabla\phi$ is greatest when $\hat{\mathbf{s}}$ is parallel to $\nabla\phi$ (and so the angle between vectors is zero and has cosine 1), we see that $\nabla\phi$ represents the maximum rate of change of ϕ in both magnitude and direction.