Directional Derivatives

Suppose you want to find the rate of change of a function f. This rate of change should depend on *where you are* and *in what direction* you're moving.

Think of standing on the side of a hill. The rate of change certainly depends on where on the hill you're standing. But even if you know where you're standing, the rate of change — the slope of the hill — depends on what direction you move in. If you move straight uphill, the rate of change is large and positive; if you move straight downhill, the rate of change is large and negative. If you move along the hill "sideways", the rate of change is 0, because your altitude doesn't change.

You can say "where you are" by giving a *point*; you can say "what direction you're moving in" by giving a *vector*.

Suppose, then, that you have a function f with multiple inputs but only one output. How rapidly is f changing at a point p in the direction of the vector v?

You can use the same procedure that you use to define the ordinary derivative: Move a little bit, measure the average change, then take the limit as the amount you move goes to 0. Here, then, is the definition of the **directional derivative** of f at p in the direction of v:

$$Df_{\vec{v}}(p) = \lim_{h \to 0} \frac{f(p + h\vec{v}) - f(p)}{h\|\vec{v}\|}.$$

Notice that $h|\vec{v}|$ is just the amount you moved to get from p to $p + h\vec{v}$.

I'll clean this up, and also obtain a formula which is better for computation. Let $s = h\|\vec{v}\|$, so $h = \dfrac{s}{\|\vec{v}\|}$. As $h \to 0$, I have $s \to 0$, so

$$Df_{\vec{v}}(p) = \lim_{s \to 0} \frac{f\left(p + s\dfrac{\vec{v}}{\|\vec{v}\|}\right) - f(p)}{s}.$$

This formula is a little better. Notice that $\dfrac{\vec{v}}{\|\vec{v}\|}$ is a *unit* vector. This makes sense, because the only function of $\vec{v}$ is to "point out" a direction. You wouldn't want a bigger $\vec{v}$ to give a bigger rate of change.

However, you can use the Chain Rule to do better. Let

$$\sigma(s) = p + s\frac{\vec{v}}{\|\vec{v}\|}.$$

Notice that $\sigma(0) = p$. Then

$$Df_{\vec{v}}(p) = \lim_{s \to 0} \frac{f(\sigma(s)) - f(\sigma(0))}{s} = (f \circ \sigma)'(0) = D(f \circ \sigma)(0) = Df(\sigma(0)) \circ \sigma'(0) = \nabla f(p) \cdot \frac{\vec{v}}{\|\vec{v}\|}.$$

The formula is

$$Df_{\vec{v}}(p) = \nabla f(p) \cdot \frac{\vec{v}}{\|\vec{v}\|}.$$

You could also write the formula above as

$$Df_{\vec{v}}(p) = \operatorname{comp}_{\vec{v}} \nabla f(p).$$

Here is an important interpretation.

Proposition. (a) ∇f points in the direction of most rapid increase of f, and $-\nabla f$ points in the direction of most rapid decrease of f.

(b) $\|\nabla f\|$ is the rate of most rapid increase of f, and $-\|\nabla f\|$ is the rate of most rapid decrease of f.

(c) The gradient vector at a point is perpendicular to the level curve (or level surface, or in general, the **level set** $f = c$) of the function.

Proof. If θ is the angle between ∇f and $\vec{v}$, I have

$$Df_{\vec{v}}(p) = \nabla f(p) \cdot \frac{\vec{v}}{\|\vec{v}\|} = \|\nabla f\| \left\| \frac{\vec{v}}{\|\vec{v}\|} \right\| \cos\theta = \|\nabla f\| \cos\theta.$$

$\left(\left\| \dfrac{\vec{v}}{\|\vec{v}\|} \right\| = 1$ because $\dfrac{\vec{v}}{\|\vec{v}\|}$ is a unit vector.$)$

The last expression is largest when $\cos\theta = 1$ — that is when $\theta = 0$. So the directional derivative is largest when $\vec{v}$ points in the same direction as ∇f. And in that direction (since $\cos\theta = 1$), the value of the directional derivative is $Df_{\vec{v}}(p) = \|\nabla f\|$.

This establishes (a) and (b) for "most rapid increase", and similar reasoning gives the statements for "most rapid decrease".

For simplicity, I'll consider (c) in the case of a level curve. Suppose $z = f(x, y)$ is a function of 2 variables. A level curve is a curve $f(x, y) = c$. Suppose I parametrize this curve by

$$x = g(t), \quad y = h(t).$$

Thus, $f(g(t), h(t)) = c$.

Then differentiating this equation using the Chain Rule, I have

$$\nabla f(g(t), h(t)) \cdot (g'(t), h'(t)) = 0.$$

Since $(g'(t), h'(t))$ is the tangent vector to the curve, the last equation says that the gradient and the tangent vector are perpendicular. This is what it means for the gradient vector to be perpendicular to the curve. □

In other words, the gradient gives the "biggest rate of change". If you want the rate of change in another direction, you scale the gradient down by taking the scalar component in the desired direction.

First of all, the formula above gives us a convenient way of computing the directional derivative.

Example. The picture below shows that graph of $f(x, y) = x^3 - 3x^2y$. This surface is called the **monkey saddle**. (Do you see why?)
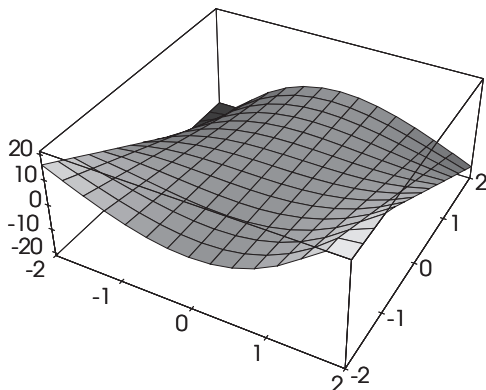
Find the directional derivative of f at $(2, 2)$ in the direction of $\vec{v} = (3, -4)$.

The gradient of f is

$$\nabla f = (3x^2 - 6xy, -3x^2), \quad \text{so} \quad \nabla f(2, 2) = (-12, -12).$$

Note that $\|(3, -4)\| = \sqrt{3^2 + (-4)^2} = 5$. So the directional derivative at $(2, 2)$ in the direction of $\vec{v} = (3, -4)$ is

$$Df_{\vec{v}}(2, 2) = \nabla f(2, 2) \cdot \frac{\vec{v}}{\|\vec{v}\|} = (-12, -12) \cdot \frac{(3, -4)}{5} = \frac{12}{5}.$$

This is the rate of change at f at the point $(2, 2)$ in the direction of $\vec{v}$.

Tou can think of the directional derivative as being obtained from the gradient by "scaling down" the gradient in the direction of $\vec{v}$. Here "scaling down" means that you take the scalar component of $\nabla f(2, 2)$ on $\vec{v}$.

The picture below illustrates this idea. Each segment in the picture below has been scaled so its length is the magnitude of the directional derivative in the direction in which the segment points.
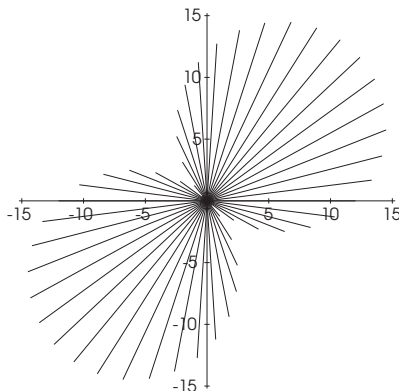
Notice that the largest segments are in the directions $(-1, -1)$ and $(1, 1)$ — the directions of the gradient and its negative, respectively. The segment shrink in the perpendicular directions, because the rates of change perpendicular to the gradient — i.e. along level curves — are 0. □

Example. Find the rate of change of $f(x, y, z) = xy - yz + xz$ at the point $(1, -2, -2)$ in the direction toward the origin. Is f increasing or decreasing in this direction?

First, compute the gradient at the point:

$$\nabla f = (y + z, x - z, -y + x), \quad \text{so} \quad \nabla f(1, -2, -2) = (-4, 3, 3).$$

Next, determine the direction vector. I'm at $P(1, -2, -2)$ and I'm looking toward the origin $Q(0, 0, 0)$. Therefore,

$$\vec{PQ} = (-1, 2, 2).$$

Make this into a unit vector by dividing by its length:

$$\frac{\vec{PQ}}{\|\vec{PQ}\|} = \frac{1}{3}(-1, 2, 2).$$

Finally, take the dot product of the unit vector with the gradient:

$$Df_{\vec{v}}(1, -2, -2) = \nabla f(1, -2, -2) \cdot \frac{\vec{PQ}}{\|\vec{PQ}\|} = (-4, 3, 3) \cdot \frac{1}{3}(-1, 2, 2) = \frac{16}{3}.$$

f is increasing in this direction, since the directional derivative is positive. □

Example. For a differentiable function $z = f(x, y)$, it is known that the directional derivative at $(2, 3)$ in the direction of $(12, 5)$ is $\dfrac{62}{13}$, and the directional derivative at $(2, 3)$ in the direction of the point $Q(4, 4)$ is $\dfrac{10}{\sqrt{5}}$.

What is $\nabla f(2, 3)$?

Write

$$\nabla f(1, 2) = (a, b).$$

Then

$$Df_{\vec{v}}(2, 3) = \nabla f(2, 3) \cdot \frac{\vec{v}}{\|\vec{v}\|} = (a, b) \cdot \frac{\vec{v}}{\|\vec{v}\|}.$$

First, the directional derivative in the direction of $(12, 5)$ is $\dfrac{62}{13}$, so

$$(a, b) \cdot \frac{(12, 5)}{\|(12, 5)\|} = \frac{62}{13}$$

$$(a, b) \cdot \frac{(12, 5)}{13} = \frac{62}{13}$$

$$12a + 5b = 62$$

Likewise, the directional derivative in the direction of the point $Q(4, 4)$ is $\dfrac{10}{\sqrt{5}}$. The vector from $P(2, 3)$ to $Q(4, 4)$ is $\overrightarrow{PQ} = (2, 1)$, so

$$(a, b) \cdot \frac{(2, 1)}{\|(2, 1)\|} = \frac{10}{\sqrt{5}}$$

$$(a, b) \cdot \frac{(2, 1)}{\sqrt{5}} = \frac{10}{\sqrt{5}}$$

$$2a + b = 10$$

From $2a + b = 10$ I get $b = -2a + 10$. Plug this into $12a + 5b = 62$ and solve for a:

$$12a + 5(-2a + 10) = 62$$
$$2a + 50 = 62$$
$$2a = 12$$
$$a = 6$$

Then $b = -2 \cdot 6 + 10 = -2$. Thus, $\nabla(2, 3) = (6, -2)$. □

∇f is actually not a *vector*, but a *function* which produces a vector when you plug in a point. For example, if $f(x, y) = x^2 + y^3$, then

$$\nabla f(x, y) = (2x, 3y).$$

So $\nabla f(2, 1) = (4, 3)$, but $\nabla f(-5, 2) = (-10, 6)$, and so on. You get a different vector at each point.

A function which takes a point as input and produces a vector as output is called a **vector field**. In this case, you can think of ∇f as "attaching" a vector $\nabla f(x, y)$ to each point (x, y) in the plane.

Example. Sketch the graph of $z = x^2 + y^2$ together with its gradient field.

The gradient is $\nabla f = (2x, 2y)$. Here are the graph of $z = x^2 + y^2$ with the gradient field "underneath":
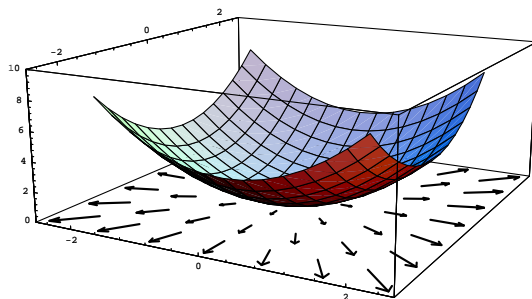
Notice that the arrows in the field point in the steepest uphill direction for the surface above: The gradient points in the direction of most rapid increase.
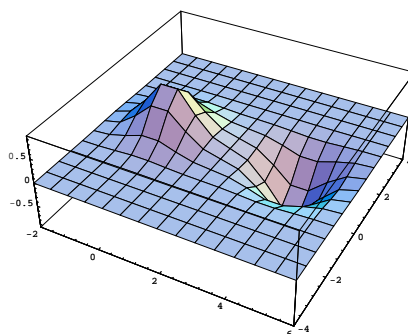
Notice also that the *size* of an arrow in the picture varies from point to point. The size of an arrow is the rate of most rapid increase. ☐

Example. Sketch the gradient field, the graph, and the level curves for $f(x, y) = e^{-x^2 - y^2} - e^{-(x-4)^2 - y^2}$.
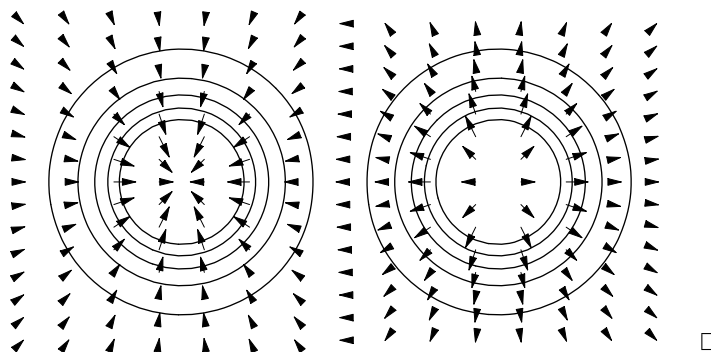
This is the gradient field for $f(x, y) = e^{-x^2 - y^2} - e^{-(x-4)^2 - y^2}$:
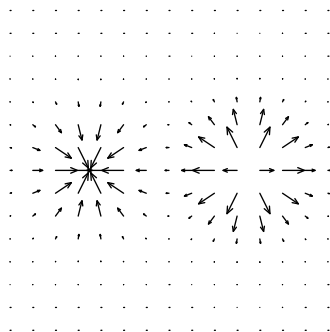
This is the graph of $f(x, y)$:

5

Here are the level curves for $f(x, y)$ together with the gradient field:

At points (x, y) "on the hill", the gradient points *toward* the top of the hill. At points (x, y) "in the valley", the gradient points *away from* the bottom of the valley. In general, the gradient points in the *direction most rapid increase*; that is, in the *steepest uphill direction* at a point.

The reason for the quotes in the last paragraph — and a common source of confusion — is that the graph (the surface) is a 3-dimensional object, while the gradient is a 2-dimensional object. If the graph is like the surface of the earth, then the gradient field is like a *map* of the surface. □

Here is an application of the gradient to constructing tangent planes. It's useful when you can't easily express a surface either parametrically or as a graph of a function.

Example. Find the equation of the tangent plane to

$$x^5 + y^5 + z^5 = 2x + y + z - 1 \quad \text{at} \quad (1, 1, -1).$$

If you set $x = 1$, $y = 1$, $z = 1$, the equation is satisfied. Hence, the point *is* on the surface.

Since it's hard to solve the equation for z, you can't use the normal vector formula

$$\vec{N} = \left(-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1 \right).$$

But you can get a normal using the gradient and a little trick.

Define

$$w = x^5 + y^5 + z^5 - 2x - y - z + 1.$$

w is a function of 3 variables. If I set $w = 0$, I get the original equation. Thus, the original surface is the **level surface** $w = 0$ for the new function w. Therefore, ∇w will be perpendicular to the level surface at a point.

The gradient is

$$\nabla w = (5x^4 - 2, 5y^4 - 1, 5z^4 - 1), \quad \text{so} \quad \nabla w(1, 1, -1) = (3, 4, 4).$$

Thus, $\nabla w(1, 1, -1) = (3, 4, 4)$ is perpendicular to the level surface $w = 0$ — that is, to the original surface — at $(1, 1, -1)$.

Hence, the tangent plane is

$$3(x - 1) + 4(y - 1) + 4(z + 1) = 0 \quad \text{or} \quad 3x + 4y + 4z = 3. \quad □$$