Directional Derivatives and the Gradient.

What to know:

1. Be able to compute directional derivatives
2. Know the definition of the gradient, its two most important properties and be able to use them (not their proofs).

In Calculus I, we saw that a derivative can be used to understand how a function changes near a given point. In Calculus 3, we talked about partial derivatives of a function $f(x, y)$, which capture a similar kind of information: how a function changes along two very specific directions, namely $x = const.$ and $y = const.$ (if you want to think about it more geometrically, the partial derivatives capture the rate of change of the curves we find once we slice the graph of f with planes parallel to the axes).

The question is, what if we'd like to see what happens in another direction? Is knowing just the partial derivatives at a point enough to determine what happens in other directions?

To answer this question, we introduce the directional derivative.

Definition 1. *If $\vec{u} = \langle a, b\rangle$ is a* ***unit vector*** *and $f(x, y)$ is differentiable, then the directional derivative in the direction of $\vec{u}$ is defined as*

$$D_{\vec{u}}f(x_0, y_0) = \lim_{t\to 0} \frac{f(x_0 + at, y_0 + bt) - f(x_0, y_0)}{t}.$$

As usual, this definition is not very useful for practical purposes. We can do a little computation and use the chain rule, to find that

$$D_{\vec{u}}f(x_0, y_0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b = \langle f_x(x_0, y_0), f_y(x_0, y_0)\rangle \cdot \langle a, b\rangle. \tag{1}$$

This is the formula we'll use in practice!

Remark: It is very important in the definition that the vector $\vec{u}$ has unit length. If we are given a vector that is not of unit length and we are asked to find the directional derivative in the direction it determines, we have to turn it into a unit vector first by dividing it by its magnitude before we use it in the above formula.

Example 1. *Let $\theta \in [0, 2\pi)$ be fixed. Then the unit vector that forms angle θ with the positive x axis is $\vec{u} = \langle\cos(\theta), \sin(\theta)\rangle$. So if, for example $f(x, y) = 3x^2 + 2yx + y^4$ and $\theta = \frac{\pi}{3}$, we have*

$$D_{\vec{u}}f(x, y) = (6x + 2y)\frac{1}{2} + (2x + 4y^3)\frac{\sqrt{3}}{2}.$$

Evaluating at (1,2), we'd find

$$D_{\vec{u}}(1, 2) = 5 + 17\sqrt{2}.$$

The Gradient

From the formula in (1), we see that the expression $\langle f_x(x_0, y_0), f_y(x_0, y_0)\rangle$ includes all the information we need to determine the directional derivative in any direction at $x_0, y_0)$, so it's important enough to have a name. We define:

Definition 2. *If $f(x, y)$ is a differentiable function, its gradient at (x_0, y_0) is defined to be the vector*

$$\nabla f(x_0, y_0) = \langle f_x(x_0, y_0), f_y(x_0, y_0)\rangle.$$

An analogous definition holds in any dimension, for example:

Definition 3. *If $f(x, y, z)$ is a differentiable function, its gradient at (x_0, y_0, z_0) is defined to be the vector*

$$\nabla f(x_0, y_0, z_0) = \langle f_x(x_0, y_0, z_0), f_y(x_0, y_0, z_0), f_z(x_0, y_0, z_0)\rangle.$$

Remark: Other notations for the gradient include grad f and df.

We will now see two properties of the gradient that make it a really important object. They are stated in three dimensions, but they hold unchanged in any dimension.

The gradient gives the direction of maximum net rate of change

The first reason why the gradient is important is the following theorem, which says that the ∇f points to the direction where f increases the fastest, whereas $-\nabla f$ points towards the direction in which f decreases the fastest :

Theorem 1. *Let $f(x, y, z)$ be a differentiable function, and suppose that $\nabla f(x_0, y_0, z_0) \neq 0$. Then for any unit vector $\vec{u}$ we have*

$$D_{-\frac{\nabla f(x_0,y_0,z_0)}{|\nabla f(x_0,y_0,z_0)|}} f(x_0, y_0, z_0) \leq D_{\vec{u}} f(x_0, y_0, z_0) \leq D_{\frac{\nabla f(x_0,y_0,z_0)}{|\nabla f(x_0,y_0,z_0)|}} f(x_0, y_0, z_0).$$

Moreover, the maximum net rate of change at (x_0, y_0, z_0) is $|\nabla f(x_0, y_0, z_0)|$ and the minimum net rate of change is $-|\nabla f(x_0, y_0, z_0)|$.

Proof. Take any unit vector $\vec{u}$. Then

$$D_{\vec{u}} f = \nabla f \cdot \vec{u} = |\nabla f||\vec{u}| \cos(\theta) = |\nabla f| \cos(\theta),$$

where θ is the angle between $\nabla f(x_0, y_0, z_0)$ and $\vec{u}$, and we also used that $\vec{u}$ is a unit vector. Since $-1 \leq \cos(\theta) \leq 1$, we find

$$-|\nabla f| \leq |\nabla f| \cos(\theta) \leq |\nabla f|, \tag{2}$$

or

$$-|\nabla f| \leq D_{\vec{u}} f \leq |\nabla f|. \tag{3}$$

However,

$$D_{\frac{\nabla f}{|\nabla f|}} f = \frac{\nabla f}{|\nabla f|} \cdot \nabla f = \frac{|\nabla f|^2}{|\nabla f|} = |\nabla f|,$$

and

$$D_{-\frac{\nabla f}{|\nabla f|}} f = -\frac{\nabla f}{|\nabla f|} \cdot \nabla f = -\frac{|\nabla f|^2}{|\nabla f|} = -|\nabla f|,$$

Together with (3), we find

$$D_{-\frac{\nabla f}{|\nabla f|}} f \leq D_{\vec{u}} f \leq D_{\frac{\nabla f}{|\nabla f|}} f$$

and this shows the theorem. □

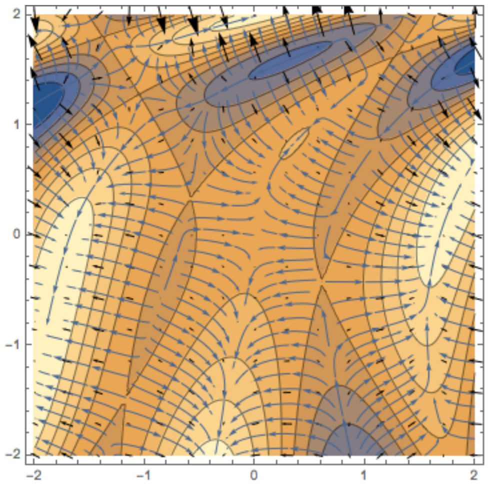

Figure 1: The level curves of a function, plotted together with its gradient at each point. Warm colors correspond to large values, cold colors correspond to small values.

Example 2. *Find the maximum rate of change of* $f(x,y) = x^2y + yx$ *at* $(1,2)$ *and the direction in which it happens.*

Solution. We find the gradient:

$$\nabla f(1,2) = \langle 2xy + y, x^2 + x\rangle_{|(1,2)} = \langle 6, 2\rangle.$$

This tells us that the maximum rate of change is

$$D_{\vec{u}} f = 2\sqrt{10}.$$

in the direction determined by $\langle 6, 2\rangle$. □

Example 3. *The Gradient Descent Algorithm: a widely used iterative algorithm to find local minima of a function* f *of many variables. It starts with some initial* $\mathbf{x}^{(0)}$ *and performs the iteration*

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - \alpha \nabla f(\mathbf{x}^{(n)}),$$

where α *is a fixed parameter (in some versions of the algorithm you can allow it to change at each step). Can you explain why such an algorithm would lead us towards a local minimum?*

The gradient is orthogonal to the level curves/surfaces.

The second reason why the gradient is a very important object is the following theorem:

Theorem 2. *If* $f(x,y,z)$ *is a differentiable function then the gradient* $\nabla f(x_0,y_0,z_0)$ *is orthogonal to the tangent plane to the level surface of* f *at* (x_0,y_0,z_0).

Proof. We will show that the gradient is perpendicular to the initial velocity of every curve $\gamma(t) = (x(t), y(t), z(t))$ in the level surface through (x_0,y_0,z_0) that starts at (x_0,y_0,z_0). Suppose that $f(x_0,y_0,z_0) = c$, so that the level set through (x_0,y_0,z_0) is

$$S = \{(x,y,z) : f(x,y,z) = c\}.$$

Then $f(\gamma(t)) = c$ for all t because we assumed that it is a curve inside the level set S, and $\gamma(0) = (x_0, y_0, z_0)$. Therefore,

$$\frac{d}{dt}(f(\gamma(t))_{|t=0} = 0 \implies (\frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial t})_{|t=0} = 0$$
$$\implies \nabla f(\gamma(0)) \cdot \gamma'(0) = 0.$$

This is saying that $\nabla f(x_0, y_0, z_0)$ is perpendicular to $\gamma'(0)$, and γ was arbitrary. This shows that $\nabla f(x_0, y_0, z_0)$ is orthogonal to the level set S at (x_0, y_0, z_0). □

Tangent planes

We can use this property of the gradient to compute tangent planes of level sets easily. Recall from Math 126 that given a point $p = (x_0, y_0, z_0)$ and a vector $v = (a, b, c)$, the plane through p with normal vector v was given by

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

So, the tangent plane of a level set

$$S = \{(x, y, z) : F(x, y, z) = c\}$$

at (x_0, y_0, z_0) is given by

$$\partial_x F(x_0, y_0, z_0)(x - x_0) + \partial_y F(x_0, y_0, z_0)(y - y_0) + \partial_z F(x_0, y_0, z_0)(z - z_0) = 0$$

or

$$\nabla F(x_0, y_0, z_0) \cdot \langle x - x_0, y - y_0, z - z_0 \rangle = 0. \tag{4}$$

Exercise 1. *Use* (4) *to find the level set of the function* $F(x, y, z) = y^2z^3 + 3zy + 2xz + 2$ *at* $(-6, 2, 1)$. *Compare your answer to the last example of the previous section.*