

DERIVATIVES ALONG VECTORS AND DIRECTIONAL DERIVATIVES

MATH 225

DERIVATIVES ALONG VECTORS

Suppose that f is a function of two variables, that is, $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, or, if we are thinking without coordinates, $f: E^2 \rightarrow \mathbb{R}$. The function f could be the distance to some point or curve, the altitude function for some landscape, or temperature (assumed to be static, *i.e.*, not changing with time). Let $P \in E^2$, and assume some disk centered at p is contained in the domain of f , that is, assume that P is an interior point of the domain of f . This allows us to move in any direction from P , at least a little, and stay in the domain of f . We want to ask how fast $f(X)$ changes as X moves away from P , and to express it as some kind of derivative.

The answer clearly depends on which direction you go. If f is not constant near P , then $f(X)$ increases in some directions, and decreases in others. If we move along a level curve of f , then $f(X)$ doesn't change at all (that's what a level curve means—a curve on which the function is constant).

The answer also depends on how fast you go. Suppose that f measures temperature, and that some particle is moving along a path through P . The particle experiences a change of temperature. This happens not because the temperature is a function of time, but rather because of the particle's motion. Now suppose that a second particle moves along the same path in the same direction, but faster. It will experience a faster change of temperature. It should seem reasonable (I hope!) that if the second particle is moving twice as fast as the first, then it will feel the temperature change twice as fast as the first.

The notion of a *derivative along a vector* makes this idea precise. Suppose we are moving along a line through P with constant velocity \mathbf{v} , that is, $X = P + t\mathbf{v}$. Then $f(X) = f(P + t\mathbf{v})$. This is a Math 111 function of t (the input and output are both real numbers), and so it makes sense to take its derivative.

Definition. The *derivative of f at P along the vector \mathbf{v}* is the number

$$(1) \quad D_{\mathbf{v}}f(P) = \left. \frac{d}{dt} \right|_0 f(P + t\mathbf{v}) = \lim_{t \rightarrow 0} \frac{f(P + t\mathbf{v}) - f(P)}{t},$$

provided this limit exists.

Example 1. Let $f(x, y) = x^2 + 2y^2$, and let's compute the derivative of f at the point $P = (2, -1)$ along the vector $\mathbf{v} = \mathbf{i} + 3\mathbf{j}$. To compute $D_{\mathbf{v}}f(P)$ we can use either part of the definition. Simply plug in and simplify until you get to a point where you can take the derivative or evaluate the limit:

$$\begin{aligned} D_{\mathbf{v}}f(P) &= \left. \frac{d}{dt} \right|_0 f(P + t\mathbf{v}) = \left. \frac{d}{dt} \right|_0 f((2, -1) + t(\mathbf{i} + 3\mathbf{j})) = \left. \frac{d}{dt} \right|_0 f(2 + t, -1 + 3t) \\ &= \left. \frac{d}{dt} \right|_0 ((2 + t)^2 + 2(-1 + 3t)^2) = 2(2) + 4(-1)(3) = -8. \end{aligned}$$

(Note that you can streamline the computation a bit by taking the derivative *and* plugging in $t = 0$ at the same time.) Here it is again using the limit version of the definition:

$$\begin{aligned} D_{\mathbf{v}}f(P) &= \lim_{t \rightarrow 0} \frac{f(P + t\mathbf{v}) - f(P)}{t} = \lim_{t \rightarrow 0} \frac{f(2 + t, -1 + 3t) - f(2, -1)}{t} \\ &= \lim_{t \rightarrow 0} \frac{(19t^2 - 8t + 6) - 6}{t} = \lim_{t \rightarrow 0} (19t - 8) = -8. \end{aligned}$$

This second method of evaluating $D_{\mathbf{v}}f(P)$, with the limit definition, is generally *not* the one to use unless the first one doesn't apply for some reason.

Note that the expression $f(P + t\mathbf{v})$ restricts f to the line $P + t\mathbf{v}$, that is, the inputs to the function are only along that line. Thus the values of f that the derivative “sees” are only those on the line. We are effectively using P and \mathbf{v} to make the line into a number line in which P is 0 on the line and \mathbf{v} determines the positive direction and “unit length.” The values of f restricted to this line determine the Math 111 function $h(t) = f(P + t\mathbf{v})$, and this is the function we are really differentiating, that is, $D_{\mathbf{v}}f(P) = h'(0)$. The reason for evaluating this derivative at $t = 0$ is that when you are moving along the line $P + t\mathbf{v}$, you're at P when $t = 0$. If you evaluated the derivative at some other t you would get a derivative at some point on the line other than P .

It's important to note that partial derivatives are special cases of derivatives along vectors. If $f: E^2 \rightarrow \mathbb{R}$, and (x, y) are coordinates on E^2 , then $\frac{\partial f}{\partial x}(P) = D_{\mathbf{i}}f(P)$ and $\frac{\partial f}{\partial y}(P) = D_{\mathbf{j}}f(P)$. This works even if the coordinate system isn't Euclidean (in which case \mathbf{i} and \mathbf{j} aren't necessarily orthonormal).

The definition of $D_{\mathbf{v}}f(P)$ is the best way to think about derivatives along vectors, but it's clearly a cumbersome way to compute them. On the other hand, partial derivatives are easy to compute. The following shows that the partial derivatives can be used to compute general derivatives along vectors.

Example 1, continued. We have $\frac{\partial f}{\partial x}(P) = 4$ and $\frac{\partial f}{\partial y}(P) = -4$, as you can check. Let $\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j}$ be an arbitrary vector. We have

$$\begin{aligned} D_{\mathbf{v}}f(P) &= \left. \frac{d}{dt} \right|_0 f(P + t\mathbf{v}) = \left. \frac{d}{dt} \right|_0 f((2, -1) + t(v_1\mathbf{i} + v_2\mathbf{j})) = \left. \frac{d}{dt} \right|_0 f(2 + v_1t, -1 + v_2t) \\ &= \left. \frac{d}{dt} \right|_0 ((2 + v_1t)^2 + 2(-1 + v_2t)^2) = 4v_1 - 4v_2. \end{aligned}$$

What you should observe is that $D_{\mathbf{v}}f(P)$ is a linear combination of $\frac{\partial f}{\partial x}(P)$ and $\frac{\partial f}{\partial y}(P)$ with the same coefficients as \mathbf{v} . In other words,

$$(2) \quad D_{\mathbf{v}}f(P) = v_1 \frac{\partial f}{\partial x}(P) + v_2 \frac{\partial f}{\partial y}(P).$$

This shows, at least for this example, that if you know the partial derivatives, you can use them to compute the derivative along any vector. You probably also realize that this expression is the dot product of $\nabla f(P)$ and \mathbf{v} . This turns out to be the case more generally, which is a consequence of the chain rule, and is discussed below.

It's worth comparing this notion of derivative with the Math 111 version. As noted earlier, asking how $f(X)$ changes when X moves away from P depends on both direction and speed. Different directions and speeds are not generally considered when computing the derivative of a Math 111 function $g: \mathbb{R} \rightarrow \mathbb{R}$. In $y = g(x)$ we always think of x as increasing, that is, moving from left to right. If we think about speed at all, we think of moving at one unit of distance per unit of time. The derivative of g at x_0 is

$$g'(x_0) = \left. \frac{d}{dx} \right|_{x_0} g(x) = \lim_{h \rightarrow 0} \frac{g(x_0 + h) - g(x_0)}{h},$$

provided the limit exists. We can make this look like (1) by writing it as

$$g'(x_0) = \left. \frac{d}{dt} \right|_0 g(x_0 + t) = \lim_{t \rightarrow 0} \frac{g(x_0 + t) - g(x_0)}{t}.$$

You should convince yourself that these really are equal. We can go twice as fast through x_0 if we like by considering $g(x_0 + 2t)$. We get

$$\left. \frac{d}{dt} \right|_0 g(x_0 + 2t) = 2g'(x_0 + 2t)|_{t=0} = 2g'(x_0),$$

in which the “extra” factor of 2 appears because of the chain rule. What happens if we go backwards, that is, how is $\left. \frac{d}{dt} \right|_0 g(x_0 - t)$ related to $g'(x_0)$? The chain rule implies that $\left. \frac{d}{dt} \right|_0 g(x_0 - t) = -g'(x_0)$.

In much the same way, you can show that $D_{2\mathbf{v}}f(P) = 2D_{\mathbf{v}}f(P)$ and $D_{-\mathbf{v}}f(P) = -D_{\mathbf{v}}f(P)$. The first of these says that if two particles move along a line through P in the same direction and one travels twice as fast as the other, then the faster one senses the values of f changing twice as fast. The second says that if two particles are traveling in opposite directions along the line through P at the same speed, the rates at which they sense the values of f changing are negatives of each other. This makes sense—if the function increases in one direction, it decreases in the opposite direction at the same (absolute) rate. More generally, if α is any scalar, then $D_{\alpha\mathbf{v}}f(p) = \alpha D_{\mathbf{v}}f(p)$.

You should try computing $D_{2\mathbf{v}}f(P)$ and $D_{-\mathbf{v}}f(P)$ directly for f , P , and \mathbf{v} as in Example 1 to see that you get -16 and 8 respectively.

DIRECTIONAL DERIVATIVES

Sometimes when asking how $f(X)$ changes when X moves away from P , we want an answer that depends only on direction. In this case we would ask for the rate at which $f(X)$ changes *per unit distance moved* in a particular direction. If \mathbf{u} is a unit vector pointing in that direction, then this rate is given by $D_{\mathbf{u}}f(P)$ and this is called a *directional derivative*. More precisely, we have the following definition.

Definition. The *directional derivative* of f at P in the direction \mathbf{v} is $D_{\mathbf{u}}f(P)$, where \mathbf{u} is the unit vector pointing in the direction of \mathbf{v} , provided this derivative exists.

This definition makes sense only if $\mathbf{v} \neq \mathbf{0}$, in which case $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|$.

Warnings.

- (1) Our text deals only with directional derivatives. The reason for considering the more general derivatives along arbitrary vectors is that $D_{\mathbf{v}}f(P)$ thought of as a function of \mathbf{v} turns out to have nice some properties that aren't clear if \mathbf{v} is restricted to be a unit vector.
- (2) Some texts call $D_{\mathbf{v}}f(P)$ a directional derivative even if \mathbf{v} isn't a unit vector. In this case a “directional” derivative depends on more than just direction! This terminology has always seemed a bit odd to me, but it's in common use, and so you have to be aware of it.
- (3) The terminology “derivative along a vector” is not standard.

DERIVATIVES ALONG VECTORS FOR MAPPINGS

Suppose $F: E^2 \rightarrow E^3$ is a differentiable mapping. Typically, the image of F is some surface in E^3 . Two examples we have seen are the maps that have a sphere and a torus

as images,

$$S(\theta, \varphi) = \mathbf{v}(\theta, \varphi) \quad \text{and} \quad T(\theta, \varphi) = 3\mathbf{u}(\theta) + \mathbf{v}(\theta, \varphi),$$

where

$$\mathbf{u}(\theta) = (\cos \theta)\mathbf{i} + (\sin \theta)\mathbf{j} \quad \text{and} \quad \mathbf{v}(\theta, \varphi) = (\cos \varphi)\mathbf{u}(\theta) + (\sin \varphi)\mathbf{k}.$$

If P is a point in E^2 and \mathbf{v} is a vector in V^2 (the vector space of displacement vectors on E^2), we can define the derivative of F at P along \mathbf{v} in the same way as we did for $f: E^n \rightarrow \mathbb{R}$, namely,

$$D_{\mathbf{v}}F(P) = \left. \frac{d}{dt} \right|_0 F(P + t\mathbf{v}).$$

There is an important difference, however. Note that $\Gamma(t) = F(P + t\mathbf{v})$ is a curve in E^3 that passes through $F(P)$ at $t = 0$; in fact, it's a curve in the image of F . Then $D_{\mathbf{v}}F(P) = \Gamma'(0)$ is the velocity vector of the curve as it passes through P . This vector is tangent to the image of F at P , and it is an element of V^3 , the displacement vectors on E^3 . As before, if r and s are coordinates in E^2 , the domain of F , the partial derivatives $\frac{\partial F}{\partial r}$ and $\frac{\partial F}{\partial s}$ are special cases of this, namely $\frac{\partial F}{\partial r}(P) = D_{\mathbf{i}}F(P)$ and $\frac{\partial F}{\partial s}(P) = D_{\mathbf{j}}F(P)$, where \mathbf{i} and \mathbf{j} are the unit vectors in the r and s directions relative to the (r, s) coordinate system (we are assuming this is a Euclidean coordinate system). Note, however, that the partial derivatives don't have to be computed using $\left. \frac{d}{dt} \right|_0 F(P + t\mathbf{v})$. They are computed more easily the "usual" way, namely, thinking of one variable as a constant and differentiating with respect to the other variable. If \mathbf{v} is a unit vector, we can call $D_{\mathbf{v}}F(P)$ a directional derivative, just as we did for functions $f: E^n \rightarrow \mathbb{R}$.

Example 2. Suppose $F: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ is defined by $F(r, s) = (r^2 - s^3, r/s, rs)$, $\mathbf{v} = 2\mathbf{i} - 3\mathbf{j}$, and $P = (-1, 1)$. Then

$$\begin{aligned} \frac{\partial F}{\partial r}(P) &= 2r\mathbf{i} + \frac{1}{s}\mathbf{j} + s\mathbf{k} \Big|_P = -2\mathbf{i} + \mathbf{j} + \mathbf{k}, \\ \frac{\partial F}{\partial s}(P) &= -3s^2\mathbf{i} - \frac{r}{s^2}\mathbf{j} + r\mathbf{k} \Big|_P = -3\mathbf{i} + \mathbf{j} - \mathbf{k}, \end{aligned}$$

and

$$\begin{aligned} D_{\mathbf{v}}F(P) &= \left. \frac{d}{dt} \right|_0 F(P + t\mathbf{v}) = \left. \frac{d}{dt} \right|_0 F(-1 + 2t, 1 - 3t) \\ &= \left. \frac{d}{dt} \right|_0 \left((-1 + 2t)^2 - (1 - 3t)^3, \frac{-1 + 2t}{1 - 3t}, (-1 + 2t)(1 - 3t) \right) \\ &= (2(-1)2 - 3(1)(-3))\mathbf{i} + \frac{2(1) - (-1)(-3)}{1^2}\mathbf{j} + (2(1) + (-1)(-3))\mathbf{k} \\ &= 5\mathbf{i} - \mathbf{j} + 5\mathbf{k}. \end{aligned}$$

You might notice that \mathbf{i} and \mathbf{j} are playing two roles here. On one hand, they are the two standard basis vectors in \mathbb{R}^2 , the domain of F . On the other hand, they are two of the three standard basis vectors in \mathbb{R}^3 , the range of F . Technically we should use different notation for these different uses, but as long as you are aware of this, it's usually not a problem.

Given $F: E^2 \rightarrow E^3$ and $P \in E^2$, we can define a map $L: V^2 \rightarrow V^3$ that represents all of the derivatives along vectors at P , namely,

$$L(\mathbf{v}) = D_{\mathbf{v}}F(P) = \left. \frac{d}{dt} \right|_0 F(P + t\mathbf{v}).$$

A good interpretation of the map L is that it takes velocity vectors of parameterized lines in E^2 passing through P to velocity vectors of parameterized curves in the image of F passing through $F(P)$: if a moving particle in E^2 passes through P with velocity \mathbf{v} , the corresponding moving particle in the image of F passes through $F(P)$ with velocity $L(\mathbf{v})$. Since the velocity of a moving particle as it passes through P is completely determined by its motion in an arbitrarily small region around P , you can think of L as representing what F does on an infinitesimal region centered at P . Think about this!

Example 2, continued. Let F and P be as above, but let $\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j}$ be arbitrary. Then

$$\begin{aligned} L(\mathbf{v}) = D_{\mathbf{v}}F(P) &= \left. \frac{d}{dt} \right|_0 F(P + t\mathbf{v}) = \left. \frac{d}{dt} \right|_0 F(-1 + v_1t, 1 + v_2t) = \dots \\ &= (-2v_1 - 3v_2)\mathbf{i} + (v_1 + v_2)\mathbf{j} + (v_1 - v_2)\mathbf{k}. \end{aligned}$$

(I'll let you fill in the missing steps.) If you rearrange this, you get

$$L(\mathbf{v}) = D_{\mathbf{v}}F(P) = v_1(-2\mathbf{i} + \mathbf{j} + \mathbf{k}) + v_2(-3\mathbf{i} + \mathbf{j} - \mathbf{k}) = v_1 \frac{\partial F}{\partial r}(P) + v_2 \frac{\partial F}{\partial s}(P).$$

Note that this is the same as the formula in (2), but with a different interpretation (one is a scalar and the other is a vector). Once again, if you know the partial derivatives, you can use them to compute the derivative along an arbitrary vector: $D_{\mathbf{v}}F(P)$ is a linear combination of $\frac{\partial F}{\partial r}(P)$ and $\frac{\partial F}{\partial s}(P)$ with the same coefficients as \mathbf{v} . You may have also observed that L is linear. This is no accident, as we will see in the next section.

Two final comments about derivatives of mappings along vectors:

- To really understand this, you need to look at pictures. Some examples are worked out for the torus mapping in the *Mathematica* notebook Sphere&Torus.nb in the Math 225 folder. They are in the sections “A Curve on the Torus and its Tangent Vector at $t = 0$,” “Several Curves and Tangent Vectors on the Torus,” and “A Plane Tangent to the Torus.”
- Everything in this section is equally valid for mappings from E^k to E^ℓ .

THE CHAIN RULE AND LINEAR ALGEBRA TO THE RESCUE!

Suppose $f: E^n \rightarrow \mathbb{R}$ is differentiable at $P \in E^n$. If you look at $D_{\mathbf{v}}f(P) = \left. \frac{d}{dt} \right|_0 f(P+t\mathbf{v})$ and think that you should be able to use the chain rule, you're right! It's the derivative of f composed with the parameterized line $\gamma(t) = P + t\mathbf{v}$:

$$D_{\mathbf{v}}f(P) = (f \circ \gamma)'(0).$$

The chain rule says that this should be the product of the derivative of f at $P = \gamma(0)$, which is represented by $\nabla f(P)$, and the derivative of γ at 0, which is $\gamma'(0) = \mathbf{v}$. The appropriate product is dot product, and so we have

$$D_{\mathbf{v}}f(P) = \nabla f(P) \cdot \mathbf{v}.$$

This formula simplifies the computation in Example 1.

Example 1 redone. The general gradient of f is $\nabla f = 2x\mathbf{i} + 4y\mathbf{j}$. At $P = (2, -1)$ we have $\nabla f(P) = 4\mathbf{i} - 4\mathbf{j}$. Thus,

$$D_{\mathbf{v}}f(P) = \nabla f(P) \cdot \mathbf{v} = (4\mathbf{i} - 4\mathbf{j}) \cdot (\mathbf{i} + 3\mathbf{j}) = -8.$$

How easy!

Now suppose $F: E^2 \rightarrow E^3$ is differentiable at $P \in E^2$. Again letting $\gamma(t) = P + t\mathbf{v}$, we have

$$D_{\mathbf{v}}F(P) = (F \circ \gamma)'(0).$$

This time, the derivative of F at $P = \gamma(0)$ is represented by the linear map $DF(P)$ from the definition of differentiability: $DF(P)$ is the linear part of the best affine approximation of F at P ,

$$F(P + \mathbf{h}) \approx F(P) + DF(P)(\mathbf{h}).$$

The chain rule then tells us that $(F \circ \gamma)'(0) = DF(\gamma(0))(\gamma'(0)) = DF(P)(\mathbf{v})$. Thus, we have the appealing, but rather abstract, formula $D_{\mathbf{v}}F(P) = DF(P)(\mathbf{v})$. What you want to realize is that the right side of this is a linear map evaluated at a vector, which can be represented by matrix multiplication. When thinking of it in this way, I like to write the formula as

$$D_{\mathbf{v}}F(P) = (DF_P)\mathbf{v}.$$

Before your eyes glaze over (maybe it's already too late!), you should realize that the two sides of this equation have different definitions, but the chain rule tells us that they compute the same value. The formula expresses the conceptual directional derivative on the left side as the concrete product of the matrix DF_P and the vector \mathbf{v} , suitably interpreted. Redoing Example 2 should help, and it will illustrate how much easier the computation is with matrices.

Example 2 redone. Since $F: \mathbb{R}^2 \rightarrow \mathbb{R}^3$, the linear map $DF(P)$ is represented by a 3×2 matrix:

$$DF_P = \begin{pmatrix} 2r & -3s^2 \\ 1/s & -r/s^2 \\ s & r \end{pmatrix}_P = \begin{pmatrix} -2 & -3 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Notice that the partial derivatives $\frac{\partial F}{\partial r}$ and $\frac{\partial F}{\partial s}$ are the *columns* of the matrix. We then get

$$D_{\mathbf{v}}F(P) = (DF_P)\mathbf{v} = \begin{pmatrix} -2 & -3 \\ 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ -1 \\ 5 \end{pmatrix} = 5\mathbf{i} - \mathbf{j} + 5\mathbf{k}.$$

ROBERT L. FOOTE, FALL 1998
UPDATED, FALL 2006