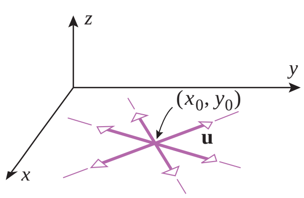
13.6 DIRECTIONAL DERIVATIVES AND GRADIENTS

The partial derivatives $f_x(x, y)$ and $f_y(x, y)$ represent the rates of change of $f(x, y)$ in directions parallel to the x- and y-axes. In this section we will investigate rates of change of $f(x, y)$ in other directions.

DIRECTIONAL DERIVATIVES

In this section we extend the concept of a *partial* derivative to the more general notion of a *directional* derivative. We have seen that the partial derivatives of a function give the instantaneous rates of change of that function in directions parallel to the coordinate axes. Directional derivatives allow us to compute the rates of change of a function with respect to distance in *any* direction.

Suppose that we wish to compute the instantaneous rate of change of a function $f(x, y)$ with respect to distance from a point (x_0, y_0) in some direction. Since there are infinitely many different directions from (x_0, y_0) in which we could move, we need a convenient method for describing a specific direction starting at (x_0, y_0). One way to do this is to use a unit vector



▲ Figure 13.6.1

$$\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j}$$

that has its initial point at (x_0, y_0) and points in the desired direction (Figure 13.6.1). This vector determines a line l in the xy-plane that can be expressed parametrically as

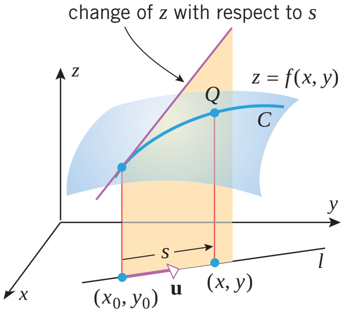
$$x = x_0 + su_1, \quad y = y_0 + su_2 \tag{1}$$

Since u is a unit vector, s is the arc length parameter that has its reference point at (x_0, y_0) and has positive values in the direction of $\mathbf{u}$. For $s = 0$, the point (x, y) is at the reference point (x_0, y_0), and as s increases, the point (x, y) moves along l in the direction of $\mathbf{u}$. On the line l the variable $z = f(x_0 + su_1, y_0 + su_2)$ is a function of the parameter s. The value of the derivative dz/ds at $s = 0$ then gives an instantaneous rate of change of $f(x, y)$ with respect to distance from (x_0, y_0) in the direction of $\mathbf{u}$.

13.6.1 DEFINITION If $f(x, y)$ is a function of x and y, and if $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j}$ is a unit vector, then the ***directional derivative of f in the direction of*** $\mathbf{u}$ at (x_0, y_0) is denoted by $D_{\mathbf{u}}f(x_0, y_0)$ and is defined by

$$D_{\mathbf{u}}f(x_0, y_0) = \frac{d}{ds}\left[f(x_0 + su_1, y_0 + su_2)\right]_{s=0} \tag{2}$$

provided this derivative exists.

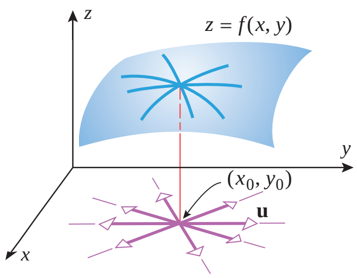


▲ Figure 13.6.2

Geometrically, $D_{\mathbf{u}}f(x_0, y_0)$ can be interpreted as the ***slope of the surface z = f(x, y) in the direction of*** $\mathbf{u}$ at the point $(x_0, y_0, f(x_0, y_0))$ (Figure 13.6.2). Usually the value of $D_{\mathbf{u}}f(x_0, y_0)$ will depend on both the point (x_0, y_0) and the direction $\mathbf{u}$. Thus, at a fixed point the slope of the surface may vary with the direction (Figure 13.6.3). Analytically, the directional derivative represents the ***instantaneous rate of change of f(x, y) with respect to distance in the direction of*** $\mathbf{u}$ at the point (x_0, y_0).

▶ **Example 1** Let $f(x, y) = xy$. Find and interpret $D_{\mathbf{u}}f(1, 2)$ for the unit vector

$$\mathbf{u} = \frac{\sqrt{3}}{2}\mathbf{i} + \frac{1}{2}\mathbf{j}$$



The slope of the surface varies with the direction of **u**.

▲ Figure 13.6.3

Solution. It follows from Equation (2) that

$$D_{\mathbf{u}} f(1,2) = \frac{d}{ds}\left[f\left(1+\frac{\sqrt{3}s}{2}, 2+\frac{s}{2}\right)\right]_{s=0}$$

Since

$$f\left(1+\frac{\sqrt{3}s}{2}, 2+\frac{s}{2}\right) = \left(1+\frac{\sqrt{3}s}{2}\right)\left(2+\frac{s}{2}\right) = \frac{\sqrt{3}}{4}s^2 + \left(\frac{1}{2}+\sqrt{3}\right)s + 2$$

we have

$$\begin{aligned} D_{\mathbf{u}} f(1,2) &= \frac{d}{ds}\left[\frac{\sqrt{3}}{4}s^2 + \left(\frac{1}{2}+\sqrt{3}\right)s + 2\right]_{s=0} \\ &= \left[\frac{\sqrt{3}}{2}s + \frac{1}{2} + \sqrt{3}\right]_{s=0} = \frac{1}{2}+\sqrt{3} \end{aligned}$$

Since $\frac{1}{2}+\sqrt{3} \approx 2.23$, we conclude that if we move a small distance from the point $(1, 2)$ in the direction of **u**, the function $f(x, y) = xy$ will increase by about 2.23 times the distance moved. ◀

The definition of a directional derivative for a function $f(x, y, z)$ of three variables is similar to Definition 13.6.1.

13.6.2 DEFINITION If $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$ is a unit vector, and if $f(x, y, z)$ is a function of x, y, and z, then the ***directional derivative of f in the direction of* u** at (x_0, y_0, z_0) is denoted by $D_{\mathbf{u}} f(x_0, y_0, z_0)$ and is defined by

$$D_{\mathbf{u}} f(x_0, y_0, z_0) = \frac{d}{ds}\left[f(x_0+su_1, y_0+su_2, z_0+su_3)\right]_{s=0} \tag{3}$$

provided this derivative exists.

Although Equation (3) does not have a convenient geometric interpretation, we can still interpret directional derivatives for functions of three variables in terms of instantaneous rates of change in a specified direction.

For a function that is differentiable at a point, directional derivatives exist in every direction from the point and can be computed directly in terms of the first-order partial derivatives of the function.

13.6.3 THEOREM

(*a*) *If $f(x, y)$ is differentiable at (x_0, y_0), and if $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j}$ is a unit vector, then the directional derivative $D_{\mathbf{u}} f(x_0, y_0)$ exists and is given by*

$$D_{\mathbf{u}} f(x_0, y_0) = f_x(x_0, y_0)u_1 + f_y(x_0, y_0)u_2 \tag{4}$$

(*b*) *If $f(x, y, z)$ is differentiable at (x_0, y_0, z_0), and if $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$ is a unit vector, then the directional derivative $D_{\mathbf{u}} f(x_0, y_0, z_0)$ exists and is given by*

$$D_{\mathbf{u}} f(x_0, y_0, z_0) = f_x(x_0, y_0, z_0)u_1 + f_y(x_0, y_0, z_0)u_2 + f_z(x_0, y_0, z_0)u_3 \tag{5}$$

PROOF We will give the proof of part (*a*); the proof of part (*b*) is similar and will be omitted. The function $z = f(x_0 + su_1, y_0 + su_2)$ is the composition of the function $z = f(x, y)$ with the functions

$$x = x(s) = x_0 + su_1 \quad \text{and} \quad y = y(s) = y_0 + su_2$$

As such, the chain rule in Formula (5) of Section 13.5 immediately gives

$$\begin{aligned} D_{\mathbf{u}} f(x_0, y_0) &= \frac{d}{ds}\left[f(x_0 + su_1, y_0 + su_2)\right]_{s=0} \\ &= \left.\frac{dz}{ds}\right|_{s=0} = f_x(x_0, y_0)u_1 + f_y(x_0, y_0)u_2 \quad \blacksquare \end{aligned}$$

We can use Theorem 13.6.3 to confirm the result of Example 1. For $f(x, y) = xy$ we have $f_x(1, 2) = 2$ and $f_y(1, 2) = 1$ (verify). With

$$\mathbf{u} = \frac{\sqrt{3}}{2}\mathbf{i} + \frac{1}{2}\mathbf{j}$$

Equation (4) becomes

$$D_{\mathbf{u}} f(1, 2) = 2\left(\frac{\sqrt{3}}{2}\right) + \frac{1}{2} = \sqrt{3} + \frac{1}{2}$$

which agrees with our solution in Example 1.

Recall from Formula (13) of Section 11.2 that a unit vector $\mathbf{u}$ in the xy-plane can be expressed as

$$\mathbf{u} = \cos\phi\,\mathbf{i} + \sin\phi\,\mathbf{j} \tag{6}$$

where ϕ is the angle from the positive x-axis to $\mathbf{u}$. Thus, Formula (4) can also be expressed as

$$D_{\mathbf{u}} f(x_0, y_0) = f_x(x_0, y_0)\cos\phi + f_y(x_0, y_0)\sin\phi \tag{7}$$

▶ **Example 2** Find the directional derivative of $f(x, y) = e^{xy}$ at $(-2, 0)$ in the direction of the unit vector that makes an angle of $\pi/3$ with the positive x-axis.

Solution. The partial derivatives of f are

$$\begin{aligned} &f_x(x, y) = ye^{xy}, \quad &&f_y(x, y) = xe^{xy} \\ &f_x(-2, 0) = 0, \quad &&f_y(-2, 0) = -2 \end{aligned}$$

The unit vector $\mathbf{u}$ that makes an angle of $\pi/3$ with the positive x-axis is

$$\mathbf{u} = \cos(\pi/3)\mathbf{i} + \sin(\pi/3)\mathbf{j} = \frac{1}{2}\mathbf{i} + \frac{\sqrt{3}}{2}\mathbf{j}$$

Thus, from (7)

$$\begin{aligned} D_{\mathbf{u}} f(-2, 0) &= f_x(-2, 0)\cos(\pi/3) + f_y(-2, 0)\sin(\pi/3) \\ &= 0(1/2) + (-2)(\sqrt{3}/2) = -\sqrt{3} \quad ◀ \end{aligned}$$

Note that in Example 3 we used a ***unit vector*** to specify the direction of the directional derivative. This is required in order to apply either Formula (4) or Formula (5).

▶ **Example 3** Find the directional derivative of $f(x, y, z) = x^2y - yz^3 + z$ at the point $(1, -2, 0)$ in the direction of the vector $\mathbf{a} = 2\mathbf{i} + \mathbf{j} - 2\mathbf{k}$.

Solution. The partial derivatives of f are

$$f_x(x, y, z) = 2xy, \quad f_y(x, y, z) = x^2 - z^3, \quad f_z(x, y, z) = -3yz^2 + 1$$
$$f_x(1, -2, 0) = -4, \quad f_y(1, -2, 0) = 1, \quad f_z(1, -2, 0) = 1$$

Since $\mathbf{a}$ is not a unit vector, we normalize it, getting

$$\mathbf{u} = \frac{\mathbf{a}}{\|\mathbf{a}\|} = \frac{1}{\sqrt{9}}(2\mathbf{i} + \mathbf{j} - 2\mathbf{k}) = \frac{2}{3}\mathbf{i} + \frac{1}{3}\mathbf{j} - \frac{2}{3}\mathbf{k}$$

Formula (5) then yields

$$D_{\mathbf{u}} f(1, -2, 0) = (-4)\left(\frac{2}{3}\right) + \frac{1}{3} - \frac{2}{3} = -3 \quad ◀$$

THE GRADIENT

Formula (4) can be expressed in the form of a dot product as

$$\begin{aligned} D_{\mathbf{u}} f(x_0, y_0) &= (f_x(x_0, y_0)\mathbf{i} + f_y(x_0, y_0)\mathbf{j}) \cdot (u_1\mathbf{i} + u_2\mathbf{j}) \\ &= (f_x(x_0, y_0)\mathbf{i} + f_y(x_0, y_0)\mathbf{j}) \cdot \mathbf{u} \end{aligned}$$

Similarly, Formula (5) can be expressed as

$$D_{\mathbf{u}} f(x_0, y_0, z_0) = (f_x(x_0, y_0, z_0)\mathbf{i} + f_y(x_0, y_0, z_0)\mathbf{j} + f_z(x_0, y_0, z_0)\mathbf{k}) \cdot \mathbf{u}$$

In both cases the directional derivative is obtained by dotting the direction vector $\mathbf{u}$ with a new vector constructed from the first-order partial derivatives of f.

Remember that ∇f is not a product of ∇ and f. Think of ∇ as an "operator" that acts on a function f to produce the gradient ∇f.

13.6.4 DEFINITION

(a) If f is a function of x and y, then the ***gradient of f*** is defined by

$$\nabla f(x, y) = f_x(x, y)\mathbf{i} + f_y(x, y)\mathbf{j} \tag{8}$$

(b) If f is a function of x, y, and z, then the ***gradient of f*** is defined by

$$\nabla f(x, y, z) = f_x(x, y, z)\mathbf{i} + f_y(x, y, z)\mathbf{j} + f_z(x, y, z)\mathbf{k} \tag{9}$$

The symbol ∇ (read "del") is an inverted delta. (It is sometimes called a "nabla" because of its similarity in form to an ancient Hebrew ten-stringed harp of that name.)

Formulas (4) and (5) can now be written as

$$D_{\mathbf{u}} f(x_0, y_0) = \nabla f(x_0, y_0) \cdot \mathbf{u} \tag{10}$$

and

$$D_{\mathbf{u}} f(x_0, y_0, z_0) = \nabla f(x_0, y_0, z_0) \cdot \mathbf{u} \tag{11}$$

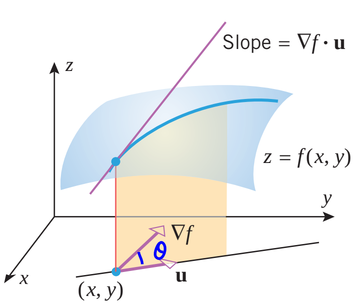
respectively. For example, using Formula (11) our solution to Example 3 would take the form

$$D_{\mathbf{u}} f(1, -2, 0) = \nabla f(1, -2, 0) \cdot \mathbf{u} = (-4\mathbf{i} + \mathbf{j} + \mathbf{k}) \cdot \left(\tfrac{2}{3}\mathbf{i} + \tfrac{1}{3}\mathbf{j} - \tfrac{2}{3}\mathbf{k}\right)$$

sqrt((2/3)^2+(1/3)^2+(2/3)^2)
= 1 because u is unit vector.

$$= (-4)\left(\tfrac{2}{3}\right) + \tfrac{1}{3} - \tfrac{2}{3} = -3$$

Formula (10) can be interpreted to mean that the slope of the surface $z = f(x, y)$ at the point (x_0, y_0) in the direction of $\mathbf{u}$ is the dot product of the gradient with $\mathbf{u}$ (Figure 13.6.4).



▲ **Figure 13.6.4**

PROPERTIES OF THE GRADIENT

The gradient is not merely a notational device to simplify the formula for the directional derivative; we will see that the length and direction of the gradient ∇f provide important information about the function f and the surface $z = f(x, y)$. For example, suppose that $\nabla f(x, y) \neq \mathbf{0}$, and let us use Formula (4) of Section 11.3 to rewrite (10) as

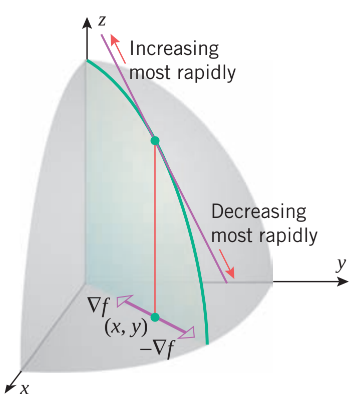
u is a unit vector whose magnitude is 1 no matter what u_1, u_2,..u_n combinations are. See examples above as in pg. 963

$$D_{\mathbf{u}} f(x, y) = \nabla f(x, y) \cdot \mathbf{u} = \|\nabla f(x, y)\| \|\mathbf{u}\| \cos\theta = \|\nabla f(x, y)\| \cos\theta \qquad (12)$$

where θ is the angle between $\nabla f(x, y)$ and $\mathbf{u}$. Equation (12) tells us that the maximum value of $D_{\mathbf{u}} f$ at the point (x, y) is $\|\nabla f(x, y)\|$, and this maximum occurs when $\theta = 0$, that is, when $\mathbf{u}$ is in the direction of $\nabla f(x, y)$. Geometrically, this means:

direction of gradient = normal to level curve.

At (x, y), the surface $z = f(x, y)$ has its maximum slope in the direction of the gradient, and the maximum slope is $\|\nabla f(x, y)\|$.



▲ Figure 13.6.5

That is, the function $f(x, y)$ increases most rapidly in the direction of its gradient (Figure 13.6.5).

Similarly, (12) tells us that the minimum value of $D_{\mathbf{u}} f$ at the point (x, y) is $-\|\nabla f(x, y)\|$, and this minimum occurs when $\theta = \pi$, that is, when $\mathbf{u}$ is oppositely directed to $\nabla f(x, y)$. Geometrically, this means:

At (x, y), the surface $z = f(x, y)$ has its minimum slope in the direction that is opposite to the gradient, and the minimum slope is $-\|\nabla f(x, y)\|$.

That is, the function $f(x, y)$ decreases most rapidly in the direction opposite to its gradient (Figure 13.6.5).

Finally, in the case where $\nabla f(x, y) = \mathbf{0}$, it follows from (12) that $D_{\mathbf{u}} f(x, y) = 0$ in all directions at the point (x, y). This typically occurs where the surface $z = f(x, y)$ has a "relative maximum," a "relative minimum," or a saddle point.

A similar analysis applies to functions of three variables. As a consequence, we have the following result.

13.6.5 THEOREM *Let f be a function of either two variables or three variables, and let P denote the point $P(x_0, y_0)$ or $P(x_0, y_0, z_0)$, respectively. Assume that f is differentiable at P.*

(a) *If $\nabla f = \mathbf{0}$ at P, then all directional derivatives of f at P are zero.*

(b) *If $\nabla f \neq \mathbf{0}$ at P, then among all possible directional derivatives of f at P, the derivative in the direction of ∇f at P has the largest value. The value of this largest directional derivative is $\|\nabla f\|$ at P.*

(c) *If $\nabla f \neq \mathbf{0}$ at P, then among all possible directional derivatives of f at P, the derivative in the direction opposite to that of ∇f at P has the smallest value. The value of this smallest directional derivative is $-\|\nabla f\|$ at P.*

▶ **Example 4** Let $f(x, y) = x^2 e^y$. Find the maximum value of a directional derivative at $(-2, 0)$, and find the unit vector in the direction in which the maximum value occurs.

Solution. Since

$$\nabla f(x, y) = f_x(x, y)\mathbf{i} + f_y(x, y)\mathbf{j} = 2xe^y\mathbf{i} + x^2e^y\mathbf{j}$$

the gradient of f at $(-2, 0)$ is

$$\nabla f(-2, 0) = -4\mathbf{i} + 4\mathbf{j}$$

By Theorem 13.6.5, the maximum value of the directional derivative is

$$\|\nabla f(-2, 0)\| = \sqrt{(-4)^2 + 4^2} = \sqrt{32} = 4\sqrt{2}$$

This maximum occurs in the direction of $\nabla f(-2, 0)$. The unit vector in this direction is

$$\mathbf{u} = \frac{\nabla f(-2, 0)}{\|\nabla f(-2, 0)\|} = \frac{1}{4\sqrt{2}}(-4\mathbf{i} + 4\mathbf{j}) = -\frac{1}{\sqrt{2}}\mathbf{i} + \frac{1}{\sqrt{2}}\mathbf{j} \blacktriangleleft$$

What would be the minimum value of a directional derivative of

$$f(x, y) = x^2e^y$$

at $(-2, 0)$?

GRADIENTS ARE NORMAL TO LEVEL CURVES

We have seen that the gradient points in the direction in which a function increases most rapidly. For a function $f(x, y)$ of two variables, we will now consider how this direction of maximum rate of increase can be determined from a contour map of the function. Suppose that (x_0, y_0) is a point on a level curve $f(x, y) = c$ of f, and assume that this curve can be smoothly parametrized as

$$x = x(s), \quad y = y(s) \tag{13}$$

where s is an arc length parameter. Recall from Formula (6) of Section 12.4 that the unit tangent vector to (13) is

$$\mathbf{T} = \mathbf{T}(s) = \left(\frac{dx}{ds}\right)\mathbf{i} + \left(\frac{dy}{ds}\right)\mathbf{j}$$

Since $\mathbf{T}$ gives a direction along which f is nearly constant, we would expect the instantaneous rate of change of f with respect to distance in the direction of $\mathbf{T}$ to be 0. That is, we would expect that

$$D_{\mathbf{T}}f(x, y) = \nabla f(x, y) \cdot \mathbf{T}(s) = 0$$

To show this to be the case, we differentiate both sides of the equation $f(x, y) = c$ with respect to s. Assuming that f is differentiable at (x, y), we can use the chain rule to obtain

$$\frac{\partial f}{\partial x}\frac{dx}{ds} + \frac{\partial f}{\partial y}\frac{dy}{ds} = 0$$

which we can rewrite as

$$\left(\frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}\right) \cdot \left(\frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j}\right) = 0$$

or, alternatively, as

$$\nabla f(x, y) \cdot \mathbf{T} = 0$$

Therefore, if $\nabla f(x, y) \neq \mathbf{0}$, then $\nabla f(x, y)$ should be normal to the level curve $f(x, y) = c$ at any point (x, y) on the curve.

It is proved in advanced courses that if $f(x, y)$ has continuous first-order partial derivatives, and if $\nabla f(x_0, y_0) \neq \mathbf{0}$, then near (x_0, y_0) the graph of $f(x, y) = c$ is indeed a smooth curve through (x_0, y_0). Furthermore, we also know from Theorem 13.4.4 that f will be differentiable at (x_0, y_0). We therefore have the following result.

Show that the level curves for

$$f(x, y) = x^2 + y^2$$

are circles and verify Theorem 13.6.6 at $(x_0, y_0) = (3, 4)$.

13.6.6 THEOREM *Assume that $f(x, y)$ has continuous first-order partial derivatives in an open disk centered at (x_0, y_0) and that $\nabla f(x_0, y_0) \neq \mathbf{0}$. Then $\nabla f(x_0, y_0)$ is normal to the level curve of f through (x_0, y_0).*

When we examine a contour map, we instinctively regard the distance between adjacent contours to be measured in a normal direction. If the contours correspond to equally spaced