52. In this problem, you will establish rigorously that

$$\lim_{(x,y)\to(0,0)} \frac{x^3+y^3}{x^2+y^2} = 0.$$

(a) Show that $|x| \le \|(x, y)\|$ and $|y| \le \|(x, y)\|$.

(b) Show that $|x^3 + y^3| \le 2(x^2 + y^2)^{3/2}$. (Hint: Begin with the triangle inequality, and then use part (a).)

(c) Show that if $0 < \|(x, y)\| < \delta$, then $|(x^3 + y^3)/(x^2 + y^2)| < 2\delta$.

(d) Now prove that $\lim_{(x,y)\to(0,0)}(x^3 + y^3)/(x^2 + y^2) = 0$.

53. (a) If a and b are any real numbers, show that $2|ab| \le a^2 + b^2$.

(b) Let

$$f(x, y) = xy\left(\frac{x^2-y^2}{x^2+y^2}\right).$$

Use part (a) to show that if $0 < \|(x, y)\| < \delta$, then $|f(x, y)| < \delta^2/2$.

(c) Prove that $\lim_{(x,y)\to(0,0)} f(x, y)$ exists, and find its value.

2.3 The Derivative

Our goal for this section is to define the derivative of a function $\mathbf{f}: X \subseteq \mathbf{R}^n \to \mathbf{R}^m$, where n and m are arbitrary positive integers. Predictably, the derivative of a vector-valued function of several variables is a more complicated object than the derivative of a scalar-valued function of a single variable. In addition, the notion of differentiability is quite subtle in the case of a function of more than one variable.

We first define the basic computational tool of partial derivatives. After doing so, we can begin to understand differentiability via the geometry of tangent planes to surfaces. Finally, we generalize these relatively concrete ideas to higher dimensions.

Partial Derivatives

Recall that if $F: X \subseteq \mathbf{R} \to \mathbf{R}$ is a scalar-valued function of one variable, then the **derivative** of F at a number $a \in X$ is

$$F'(a) = \lim_{h\to 0} \frac{F(a+h) - F(a)}{h}. \tag{1}$$

Moreover, F is said to be **differentiable at** a precisely when the limit in equation (1) exists.

DEFINITION 3.1 Suppose $f: X \subseteq \mathbf{R}^n \to \mathbf{R}$ is a scalar-valued function of n variables. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote a point of $\mathbf{R}^n$. A **partial function** F **with respect to the variable** x_i is a one-variable function obtained from f by holding all variables constant except x_i. That is, we set x_j equal to a constant a_j for $j \ne i$. Then the partial function in x_i is defined by

$$F(x_i) = f(a_1, a_2, \ldots, x_i, \ldots, a_n).$$

EXAMPLE 1 If $f(x, y) = (x^2 - y^2)/(x^2 + y^2)$, then the partial functions with respect to x are given by

$$F(x) = f(x, a_2) = \frac{x^2 - a_2^2}{x^2 + a_2^2},$$

where a_2 may be any constant. If, for example, $a_2 = 0$, then the partial function is

$$F(x) = f(x, 0) = \frac{x^2}{x^2} \equiv 1.$$

Geometrically, this partial function is nothing more than the restriction of f to the horizontal line $y = 0$. Note that since the origin is not in the domain of f, 0 should not be taken to be in the domain of F. (See Figure 2.46.) ◆

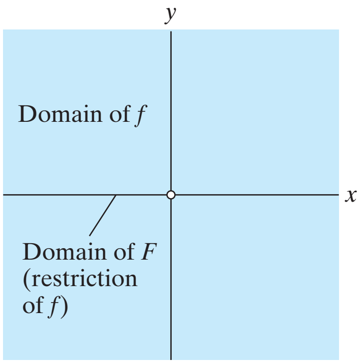

Figure 2.46 The function f of Example 1 is defined on $\mathbf{R}^2 - \{(0,0)\}$, while its partial function F along $y = 0$ is defined on the x-axis minus the origin.

Remark In practice, we usually do not go to the notational trouble of explicitly replacing the x_j's ($j \neq i$) by constants when working with partial functions. Instead, we make a mental note that the partial function is obtained by allowing only one variable to vary, while all the other variables are held fixed.

DEFINITION 3.2 The **partial derivative of f with respect to x_i** is the (ordinary) derivative of the partial function with respect to x_i. That is, the partial derivative with respect to x_i is $F'(x_i)$, in the notation of Definition 3.1. Standard notations for the partial derivative of f with respect to x_i are

$$\frac{\partial f}{\partial x_i}, \quad D_{x_i} f(x_1, \ldots, x_n), \quad \text{and} \quad f_{x_i}(x_1, \ldots, x_n).$$

Symbolically, we have

$$\frac{\partial f}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}. \tag{2}$$

By definition, the partial derivative is the (instantaneous) rate of change of f when all variables, except the specified one, are held fixed. In the case where f is a (scalar-valued) function of two variables, we can understand

$$\frac{\partial f}{\partial x}(a, b)$$

geometrically as the slope at the point $(a, b, f(a, b))$ of the curve obtained by intersecting the surface $z = f(x, y)$ with the plane $y = b$, as shown in Figure 2.47. Similarly,

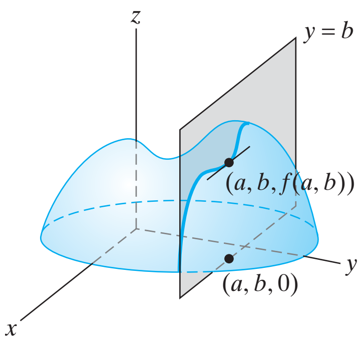

Figure 2.47 Visualizing the partial derivative $\frac{\partial f}{\partial x}(a, b)$.

$$\frac{\partial f}{\partial y}(a, b)$$

is the slope at $(a, b, f(a, b))$ of the curve formed by the intersection of $z = f(x, y)$ and $x = a$, shown in Figure 2.48.

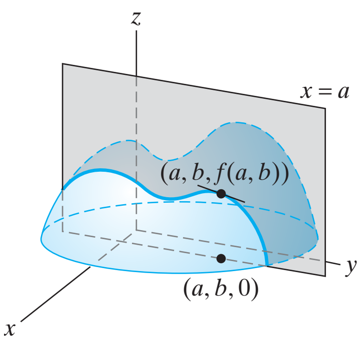

Figure 2.48 Visualizing the partial derivative $\frac{\partial f}{\partial y}(a, b)$.

EXAMPLE 2 For the most part, partial derivatives are quite easy to compute, once you become adept at treating variables like constants. If

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

then we have

$$\frac{\partial f}{\partial x} = 2xy - \sin(x + y).$$

(Imagine y to be a constant throughout the differentiation process.) Also

$$\frac{\partial f}{\partial y} = x^2 - \sin(x + y).$$

(Imagine x to be a constant.) Similarly, if $g(x, y) = xy/(x^2 + y^2)$, then, from the quotient rule of ordinary calculus, we have

$$g_x(x, y) = \frac{(x^2 + y^2)y - xy(2x)}{(x^2 + y^2)^2} = \frac{y(y^2 - x^2)}{(x^2 + y^2)^2},$$

and

$$g_y(x, y) = \frac{(x^2 + y^2)x - xy(2y)}{(x^2 + y^2)^2} = \frac{x(x^2 - y^2)}{(x^2 + y^2)^2}.$$

Note that, of course, neither g nor its partial derivatives are defined at $(0, 0)$. ◆

EXAMPLE 3 Occasionally, it is necessary to appeal explicitly to limits to evaluate partial derivatives. Suppose $f: \mathbf{R}^2 \to \mathbf{R}$ is defined by

$$f(x, y) = \begin{cases} \dfrac{3x^2y - y^3}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}.$$

Then, for $(x, y) \neq (0, 0)$, we have

$$\frac{\partial f}{\partial x} = \frac{8xy^3}{(x^2 + y^2)^2} \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{3x^4 - 6x^2y^2 - y^4}{(x^2 + y^2)^2}.$$

But what should $\dfrac{\partial f}{\partial x}(0, 0)$ and $\dfrac{\partial f}{\partial y}(0, 0)$ be? To find out, we return to Definition 3.2 of the partial derivatives:

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{h \to 0} \frac{f(0 + h, 0) - f(0, 0)}{h} = \lim_{h \to 0} \frac{0 - 0}{h} = 0,$$

and

$$\frac{\partial f}{\partial y}(0, 0) = \lim_{h \to 0} \frac{f(0, 0 + h) - f(0, 0)}{h} = \lim_{h \to 0} \frac{-h - 0}{h} = \lim_{h \to 0} -1 = -1.$$ ◆

Tangency and Differentiability

If $F: X \subseteq \mathbf{R} \to \mathbf{R}$ is a scalar-valued function of one variable, then to have F differentiable at a number $a \in X$ means precisely that the graph of the curve $y = F(x)$ has a tangent line at the point $(a, F(a))$. (See Figure 2.49.) Moreover, this tangent line is given by the equation

$$y = F(a) + F'(a)(x - a). \tag{3}$$

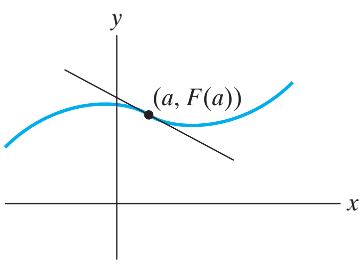

Figure 2.49 The tangent line to $y = F(x)$ at $x = a$ has equation $y = F(a) + F'(a)(x - a)$.

If we define the function $H(x)$ to be $F(a) + F'(a)(x - a)$ (i.e., $H(x)$ is the right side of equation (3) that gives the equation for the tangent line), then H has two properties:

1. $H(a) = F(a)$
2. $H'(a) = F'(a)$.

In other words, the line defined by $y = H(x)$ passes through the point $(a, F(a))$ and has the same slope at $(a, F(a))$ as the curve defined by $y = F(x)$. (Hence, the term "tangent line.")

Now suppose $f: X \subseteq \mathbf{R}^2 \to \mathbf{R}$ is a scalar-valued function of two variables, where X is open in $\mathbf{R}^2$. Then the graph of f is a surface. What should the **tangent plane** to the graph of $z = f(x, y)$ at the point $(a, b, f(a, b))$ be? Geometrically,

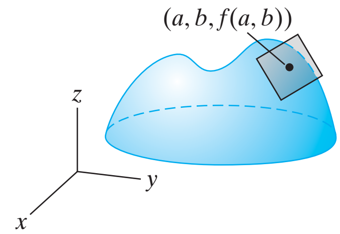

Figure 2.50 The plane tangent to $z = f(x, y)$ at $(a, b, f(a, b))$.

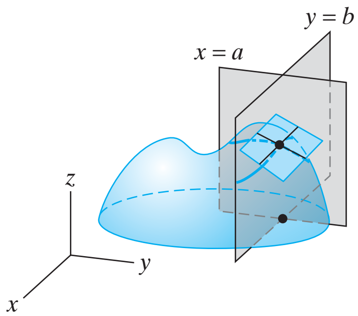

Figure 2.51 The tangent plane at $(a, b, f(a, b))$ contains the lines tangent to the curves formed by intersecting the surface $z = f(x, y)$ by the planes $x = a$ and $y = b$.

the situation is as depicted in Figure 2.50. From our earlier observations, we know that the partial derivative $f_x(a, b)$ is the slope of the line tangent at the point $(a, b, f(a, b))$ to the curve obtained by intersecting the surface $z = f(x, y)$ with the plane $y = b$. (See Figure 2.51.) This means that if we travel along this tangent line, then for every unit change in the positive x-direction, there's a change of $f_x(a, b)$ units in the z-direction. Hence, by using formula (1) of §1.2, the tangent line is given in vector parametric form as

$$\mathbf{l}_1(t) = (a, b, f(a, b)) + t(1, 0, f_x(a, b)).$$

Thus, a vector parallel to this tangent line is

$$\mathbf{u} = \mathbf{i} + f_x(a, b)\,\mathbf{k}.$$

Similarly, the partial derivative $f_y(a, b)$ is the slope of the line tangent at the point $(a, b, f(a, b))$ to the curve obtained by intersecting the surface $z = f(x, y)$ with the plane $x = a$. (Again see Figure 2.51.) Consequently, the tangent line is given by

$$\mathbf{l}_2(t) = (a, b, f(a, b)) + t(0, 1, f_y(a, b)),$$

so a vector parallel to this tangent line is

$$\mathbf{v} = \mathbf{j} + f_y(a, b)\,\mathbf{k}.$$

Both of the aforementioned tangent lines must be contained in the plane tangent to $z = f(x, y)$ at $(a, b, f(a, b))$, if one exists. Hence, a vector $\mathbf{n}$ *normal* to the tangent plane must be perpendicular to both $\mathbf{u}$ and $\mathbf{v}$. Therefore, we may take $\mathbf{n}$ to be

$$\mathbf{n} = \mathbf{u} \times \mathbf{v} = -f_x(a, b)\,\mathbf{i} - f_y(a, b)\,\mathbf{j} + \mathbf{k}.$$

Now, use equation (1) of §1.5 to find that the equation for the tangent plane—that is, the plane through $(a, b, f(a, b))$ with normal $\mathbf{n}$—is

$$(-f_x(a, b), -f_y(a, b), 1) \cdot (x - a, y - b, z - f(a, b)) = 0$$

or, equivalently,

$$-f_x(a, b)(x - a) - f_y(a, b)(y - b) + z - f(a, b) = 0.$$

By rewriting this last equation, we have shown the following result:

THEOREM 3.3 If the graph of $z = f(x, y)$ has a tangent plane at $(a, b, f(a, b))$, then that tangent plane has equation

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b). \tag{4}$$

Note that if we define the function $h(x, y)$ to be equal to $f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$ (i.e., $h(x, y)$ is the right side of equation (4)), then h has the following properties:

1. $h(a, b) = f(a, b)$
2. $\dfrac{\partial h}{\partial x}(a, b) = \dfrac{\partial f}{\partial x}(a, b)$ and $\dfrac{\partial h}{\partial y}(a, b) = \dfrac{\partial f}{\partial y}(a, b)$.

In other words, h and its partial derivatives agree with those of f at (a, b).

It is tempting to think that the surface $z = f(x, y)$ has a tangent plane at $(a, b, f(a, b))$ as long as you can make sense of equation (4), that is, as long as the

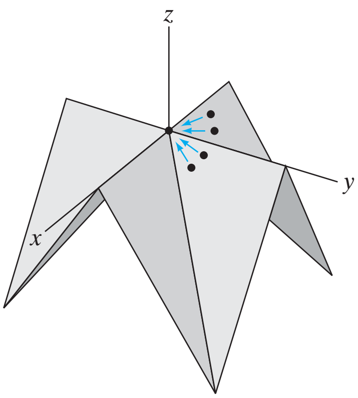

Figure 2.52 If two points approach (0, 0, 0) while remaining on one face of the surface described in Example 4, the limiting plane they and (0, 0, 0) determine is different from the one determined by letting the two points approach (0, 0, 0) while remaining on another face.

partial derivatives $f_x(a, b)$ and $f_y(a, b)$ exist. Indeed, this would be analogous to the one-variable situation where the existence of the derivative and the existence of the tangent line mean exactly the same thing. However, it is possible for a function of two variables to have well-defined partial derivatives (so that equation (4) makes sense) yet *not* have a tangent plane.

EXAMPLE 4 Let $f(x, y) = ||x| - |y|| - |x| - |y|$ and consider the surface defined by the graph of $z = f(x, y)$ shown in Figure 2.52. The partial derivatives of f at the origin may be calculated from Definition 3.2 as

$$f_x(0, 0) = \lim_{h\to 0} \frac{f(0+h, 0) - f(0, 0)}{h} = \lim_{h\to 0} \frac{||h|| - |h|}{h} = \lim_{h\to 0} 0 = 0$$

and

$$f_y(0, 0) = \lim_{h\to 0} \frac{f(0, 0+h) - f(0, 0)}{h} = \lim_{h\to 0} \frac{|-|h|| - |h|}{h} = \lim_{h\to 0} 0 = 0.$$

(Indeed, the partial functions $F(x) = f(x, 0)$ and $G(y) = f(0, y)$ are both identically zero and, thus, have zero derivatives.) Consequently, *if* the surface in question has a tangent plane at the origin, then equation (4) tells us that it has equation $z = 0$. But there is no geometric sense in which the surface $z = f(x, y)$ has a tangent plane at the origin. If we think of a tangent plane as the geometric limit of planes that pass through the point of tangency and two other "moving" points on the surface as those two points approach the point of tangency, then Figure 2.52 shows that there is no uniquely determined limiting plane. ◆

Example 4 shows that the existence of a tangent plane to the graph of $z = f(x, y)$ is a stronger condition than the existence of partial derivatives. It turns out that such a stronger condition is more useful in that theorems from the calculus of functions of a single variable carry over to the context of functions of several variables. What we must do now is find a suitable analytic definition of differentiability that captures this idea. We begin by looking at the definition of the one-variable derivative with fresh eyes.

By replacing the quantity $a + h$ by the variable x, the limit equation in formula (1) may be rewritten as

$$F'(a) = \lim_{x\to a} \frac{F(x) - F(a)}{x - a}.$$

This is equivalent to the equation

$$\lim_{x\to a} \left(\frac{F(x) - F(a)}{x - a} \right) - F'(a) = 0.$$

The quantity $F'(a)$ does not depend on x and therefore may be brought inside the limit. We thus obtain the equation

$$\lim_{x\to a} \left\{ \frac{F(x) - F(a)}{x - a} - F'(a) \right\} = 0.$$

Finally, some easy algebra enables us to conclude that the function F is differentiable at a if there is a number $F'(a)$ such that

$$\lim_{x\to a} \frac{F(x) - [F(a) + F'(a)(x - a)]}{x - a} = 0. \tag{5}$$

What have we learned from writing equation (5)? Note that the expression in brackets in the numerator of the limit expression in equation (5) is the function

$H(x)$ that was used to define the tangent line to $y = F(x)$ at $(a, F(a))$. Thus, we may rewrite equation (5) as

$$\lim_{x \to a} \frac{F(x) - H(x)}{x - a} = 0.$$

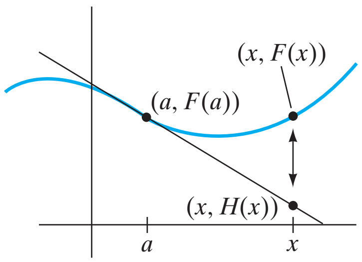

Figure 2.53 If F is differentiable at a, the vertical distance between $F(x)$ and $H(x)$ must approach zero faster than the horizontal distance between x and a does.

For the limit above to be zero, we certainly must have that the limit of the numerator is zero. But since the limit of the denominator is also zero, we can say even more, namely, that the difference between the y-values of the graph of F and of its tangent line must approach zero faster than x approaches a. This is what is meant when we say that "H is a good linear approximation to F near a." (See Figure 2.53.) Geometrically, it means that, near the point of tangency, the graph of $y = F(x)$ is approximately straight like the graph of $y = H(x)$.

If we now pass to the case of a scalar-valued function $f(x, y)$ of two variables, then to say that $z = f(x, y)$ has a tangent plane at $(a, b, f(a, b))$ (i.e., that f is differentiable at (a, b)) should mean that the vertical distance between the graph of f and the "candidate" tangent plane given by

$$z = h(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

must approach zero faster than the point (x, y) approaches (a, b). (See Figure 2.54.) In other words, near the point of tangency, the graph of $z = f(x, y)$ is approximately flat just like the graph of $z = h(x, y)$. We can capture this geometric idea with the following formal definition of differentiability:

DEFINITION 3.4 Let X be open in $\mathbf{R}^2$ and $f: X \subseteq \mathbf{R}^2 \to \mathbf{R}$ be a scalar-valued function of two variables. We say that f is **differentiable at** $(a, b) \in X$ if the partial derivatives $f_x(a, b)$ and $f_y(a, b)$ exist and if the function

$$h(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

is a good linear approximation to f near (a, b)—that is, if

$$\lim_{(x,y)\to(a,b)} \frac{f(x, y) - h(x, y)}{\|(x, y) - (a, b)\|} = 0.$$

Moreover, if f is differentiable at (a, b), then the equation $z = h(x, y)$ defines the **tangent plane** to the graph of f at the point $(a, b, f(a, b))$. If f is differentiable at all points of its domain, then we simply say that f is **differentiable**.

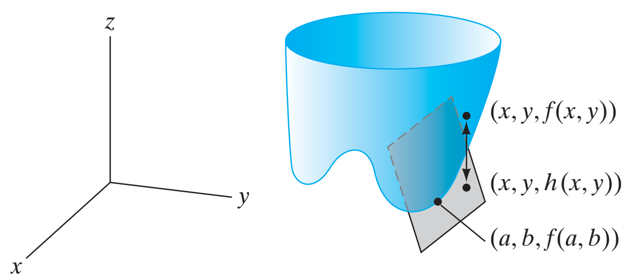

Figure 2.54 If f is differentiable at (a, b), the distance between $f(x, y)$ and $h(x, y)$ must approach zero faster than the distance between (x, y) and (a, b) does.

EXAMPLE 5 Let us return to the function $f(x, y) = ||x| - |y|| - |x| - |y|$ of Example 4. We already know that the partial derivatives $f_x(0, 0)$ and $f_y(0, 0)$ exist and equal zero. Thus, the function h of Definition 3.4 is the zero function. Consequently, f will be differentiable at (0,0) just in case

$$\lim_{(x,y)\to(0,0)} \frac{f(x, y) - h(x, y)}{\|(x, y) - (0, 0)\|} = \lim_{(x,y)\to(0,0)} \frac{f(x, y)}{\|(x, y)\|}$$

$$= \lim_{(x,y)\to(0,0)} \frac{||x| - |y|| - |x| - |y|}{\sqrt{x^2 + y^2}}$$

is zero. However, it is not hard to see that the limit in question fails to exist. Along the line $y = 0$, we have

$$\frac{f(x, y)}{\|(x, y)\|} = \frac{||x| - 0| - |x| - |0|}{\sqrt{x^2}} = \frac{0}{|x|} = 0,$$

but along the line $y = x$, we have

$$\frac{f(x, y)}{\|(x, y)\|} = \frac{||x| - |x|| - |x| - |x|}{\sqrt{x^2 + x^2}} = \frac{-2|x|}{\sqrt{2}|x|} = -\sqrt{2}.$$

Hence, f fails to be differentiable at $(0, 0)$ and has no tangent plane at $(0, 0, 0)$. ◆

The limit condition in Definition 3.4 can be difficult to apply in practice. Fortunately, the following result, which we will not prove, simplifies matters in many instances. Recall from Definition 2.3 that the phrase "a **neighborhood** of a point P in a set X" just means an open set containing P and contained in X.

THEOREM 3.5 Suppose X is open in $\mathbf{R}^2$. If $f: X \to \mathbf{R}$ has continuous partial derivatives in a neighborhood of (a, b) in X, then f is differentiable at (a, b).

A proof of a more general result (Theorem 3.10) is provided in the addendum to this section.

EXAMPLE 6 Let $f(x, y) = x^2 + 2y^2$. Then $\partial f/\partial x = 2x$ and $\partial f/\partial y = 4y$, both of which are continuous functions on all of $\mathbf{R}^2$. Thus, Theorem 3.5 implies that f is differentiable everywhere. The surface $z = x^2 + 2y^2$ must therefore have a tangent plane at every point. At the point $(2, -1)$, for example, this tangent plane is given by the equation

$$z = 6 + 4(x - 2) - 4(y + 1)$$

(or, equivalently, by $4x - 4y - z = 6$). ◆

While we're on the subject of continuity and differentiability, the next result is the multivariable analogue of a familiar theorem about functions of one variable.

THEOREM 3.6 If $f: X \subseteq \mathbf{R}^2 \to \mathbf{R}$ is differentiable at (a, b), then it is continuous at (a, b).

EXAMPLE 7 Let the function $f\colon \mathbf{R}^2 \to \mathbf{R}$ be defined by

$$f(x, y) = \begin{cases} \dfrac{x^2y^2}{x^4 + y^4} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}.$$

The function f is not continuous at the origin, since $\lim_{(x,y)\to(0,0)} f(x, y)$ does not exist. (However, f is continuous everywhere else in $\mathbf{R}^2$.) By Theorem 3.6, f therefore cannot be differentiable at the origin. Nonetheless, the partial derivatives of f do exist at the origin, and we have

$$f(x, 0) = \frac{0}{x^4 + 0} \equiv 0 \quad \Longrightarrow \quad \frac{\partial f}{\partial x}(0, 0) = 0,$$

and

$$f(0, y) = \frac{0}{0 + y^4} \equiv 0 \quad \Longrightarrow \quad \frac{\partial f}{\partial y}(0, 0) = 0,$$

since the partial functions are constant. Thus, we see that if we want something like Theorem 3.6 to be true, the existence of partial derivatives alone is not enough. ◆

Differentiability in General

It is not difficult now to see how to generalize Definition 3.4 to three (or more) variables: For a scalar-valued function of three variables to be differentiable at a point (a, b, c), we must have that (i) the three partial derivatives exist at (a, b, c) and (ii) the function $h\colon \mathbf{R}^3 \to \mathbf{R}$ defined by

$$\begin{aligned} h(x, y, z) = f(a, b, c) &+ f_x(a, b, c)(x - a) \\ &+ f_y(a, b, c)(y - b) + f_z(a, b, c)(z - c) \end{aligned}$$

is a good linear approximation to f near (a, b, c). In other words, (ii) means that

$$\lim_{(x,y,z)\to(a,b,c)} \frac{f(x, y, z) - h(x, y, z)}{\|(x, y, z) - (a, b, c)\|} = 0.$$

The passage from three variables to arbitrarily many is now straightforward.

DEFINITION 3.7 Let X be open in $\mathbf{R}^n$ and $f\colon X \to \mathbf{R}$ be a scalar-valued function; let $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in X$. We say that f is **differentiable at a** if all the partial derivatives $f_{x_i}(\mathbf{a})$, $i = 1, \ldots, n$, exist and if the function $h\colon \mathbf{R}^n \to \mathbf{R}$ defined by

$$\begin{aligned} h(\mathbf{x}) = f(\mathbf{a}) &+ f_{x_1}(\mathbf{a})(x_1 - a_1) + f_{x_2}(\mathbf{a})(x_2 - a_2) \\ &+ \cdots + f_{x_n}(\mathbf{a})(x_n - a_n) \end{aligned} \tag{6}$$

is a good linear approximation to f near $\mathbf{a}$, meaning that

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - h(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

We can use vector and matrix notation to rewrite things a bit. Define the **gradient** of a scalar-valued function $f: X \subseteq \mathbf{R}^n \to \mathbf{R}$ to be the *vector*

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

Consequently,

$$\nabla f(\mathbf{a}) = (f_{x_1}(\mathbf{a}), f_{x_2}(\mathbf{a}), \ldots, f_{x_n}(\mathbf{a})).$$

Alternatively, we can use matrix notation and define the **derivative** of f at $\mathbf{a}$, denoted $Df(\mathbf{a})$, to be the row matrix whose entries are the components of $\nabla f(\mathbf{a})$; that is,

$$Df(\mathbf{a}) = \begin{bmatrix} f_{x_1}(\mathbf{a}) & f_{x_2}(\mathbf{a}) & \cdots & f_{x_n}(\mathbf{a}) \end{bmatrix}.$$

Then, by identifying the vector $\mathbf{x} - \mathbf{a}$ with the $n \times 1$ column matrix whose entries are the components of $\mathbf{x} - \mathbf{a}$, we have

$$\begin{aligned} \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) = Df(\mathbf{a})(\mathbf{x} - \mathbf{a}) &= \begin{bmatrix} f_{x_1}(\mathbf{a}) & f_{x_2}(\mathbf{a}) & \cdots & f_{x_n}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} x_1 - a_1 \\ x_2 - a_2 \\ \vdots \\ x_n - a_n \end{bmatrix} \\ &= f_{x_1}(\mathbf{a})(x_1 - a_1) + f_{x_2}(\mathbf{a})(x_2 - a_2) \\ &\quad + \cdots + f_{x_n}(\mathbf{a})(x_n - a_n). \end{aligned}$$

Hence, vector notation allows us to rewrite equation (6) quite compactly as

$$h(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}).$$

Thus, to say that h is a good linear approximation to f near $\mathbf{a}$ in equation (6) means that

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - [f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})]}{\|\mathbf{x} - \mathbf{a}\|} = 0. \tag{7}$$

Compare equation (7) with equation (5). Differentiability of functions of one and several variables should really look very much the same to you. It is worth noting that the analogues of Theorems 3.5 and 3.6 hold in the case of n variables.

The gradient of a function is an extremely important construction, and we consider it in greater detail in §2.6.

You may be wondering what, if any, geometry is embedded in this general notion of differentiability. Recall that the graph of the function $f: X \subseteq \mathbf{R}^n \to \mathbf{R}$ is the **hypersurface** in $\mathbf{R}^{n+1}$ given by the equation $x_{n+1} = f(x_1, x_2, \ldots, x_n)$. (See equation (2) of §2.1.) If f is differentiable at $\mathbf{a}$, then the hypersurface determined by the graph has a **tangent hyperplane** at $(\mathbf{a}, f(\mathbf{a}))$ given by the equation

$$\begin{aligned} x_{n+1} = h(x_1, x_2, \ldots, x_n) &= f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) \\ &= f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x} - \mathbf{a}). \end{aligned} \tag{8}$$

Compare equation (8) with equation (3) for the tangent line to the curve $y = F(x)$ at $(\mathbf{a}, F(\mathbf{a}))$. Although we cannot visualize the graph of a function of more than two variables, nonetheless, we can use vector notation to lend real meaning to tangency in n dimensions.

EXAMPLE 8 Before we drown in a sea of abstraction and generalization, let's do some concrete computation. An example of an "n-dimensional paraboloid" in

$\mathbf{R}^{n+1}$ is given by the equation

$$x_{n+1} = x_1^2 + x_2^2 + \cdots + x_n^2,$$

that is, by the graph of the function $f(x_1, \ldots, x_n) = x_1^2 + x_2^2 + \cdots + x_n^2$. We have

$$\frac{\partial f}{\partial x_i} = 2x_i, \quad i = 1, 2, \ldots, n,$$

so that

$$\nabla f(x_1, \ldots, x_n) = (2x_1, 2x_2, \ldots, 2x_n).$$

Note that the partial derivatives of f are continuous everywhere. Hence, the n-dimensional version of Theorem 3.5 tells us that f is differentiable everywhere. In particular, f is differentiable at the point $(1, 2, \ldots, n)$,

$$\nabla f(1, 2, \ldots, n) = (2, 4, \ldots, 2n),$$

and

$$Df(1, 2, \ldots, n) = \begin{bmatrix} 2 & 4 & \cdots & 2n \end{bmatrix}.$$

Thus, the paraboloid has a tangent hyperplane at the point

$$(1, 2, \ldots, n, 1^2 + 2^2 + \cdots + n^2)$$

whose equation is given by equation (8):

$$\begin{aligned} x_{n+1} &= (1^2 + 2^2 + \cdots + n^2) + \begin{bmatrix} 2 & 4 & \cdots & 2n \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 2 \\ \vdots \\ x_n - n \end{bmatrix} \\ &= (1^2 + 2^2 + \cdots + n^2) + 2(x_1 - 1) + 4(x_2 - 2) + \cdots + 2n(x_n - n) \\ &= (1^2 + 2^2 + \cdots + n^2) + 2x_1 + 4x_2 + \cdots + 2nx_n \\ &\quad - (2 \cdot 1 + 4 \cdot 2 + \cdots + 2n \cdot n) \\ &= 2x_1 + 4x_2 + \cdots + 2nx_n - (1^2 + 2^2 + \cdots + n^2) \\ &= \sum_{i=1}^{n} 2ix_i - \frac{n(n+1)(2n+1)}{6}. \end{aligned}$$

(The formula $1^2 + 2^2 + \cdots + n^2 = n(n+1)(2n+1)/6$ is a well-known identity, encountered when you first learned about the definite integral. It's straightforward to prove using mathematical induction.) ◆

At last we're ready to take a look at differentiability in the most general setting of all. Let X be open in $\mathbf{R}^n$ and let $\mathbf{f}\colon X \to \mathbf{R}^m$ be a vector-valued function of n variables. We define the **matrix of partial derivatives** of $\mathbf{f}$, denoted $D\mathbf{f}$, to be the $m \times n$ matrix whose ijth entry is $\partial f_i/\partial x_j$, where $f_i\colon X \subseteq \mathbf{R}^n \to \mathbf{R}$ is the ith component function of $\mathbf{f}$. That is,

$$D\mathbf{f}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

The ith row of $D\mathbf{f}$ is nothing more than Df_i—and the entries of Df_i are precisely the components of the gradient vector ∇f_i. (Indeed, in the case where $m = 1$, ∇f and Df mean exactly the same thing.)

EXAMPLE 9 Suppose $\mathbf{f}\colon \mathbf{R}^3 \to \mathbf{R}^2$ is given by $\mathbf{f}(x, y, z) = (x \cos y + z, xy)$. Then we have

$$D\mathbf{f}(x, y, z) = \begin{bmatrix} \cos y & -x \sin y & 1 \\ y & x & 0 \end{bmatrix}.$$

◆

We generalize equation (7) and Definition 3.7 in an obvious way to make the following definition:

DEFINITION 3.8 (GRAND DEFINITION OF DIFFERENTIABILITY) Let $X \subseteq \mathbf{R}^n$ be open, let $\mathbf{f}\colon X \to \mathbf{R}^m$, and let $\mathbf{a} \in X$. We say that $\mathbf{f}$ is **differentiable at a** if $D\mathbf{f}(\mathbf{a})$ exists and if the function $\mathbf{h}\colon \mathbf{R}^n \to \mathbf{R}^m$ defined by

$$\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

is a good linear approximation to $\mathbf{f}$ near $\mathbf{a}$. That is, we must have

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - \mathbf{h}(\mathbf{x})\|}{\|\mathbf{x} - \mathbf{a}\|} = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - [\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})]\|}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

Some remarks are in order. First, the reason for having the vector length appearing in the numerator in the limit equation in Definition 3.8 is so that there is a quotient of real numbers of which we can take a limit. (Definition 3.7 concerns scalar-valued functions only, so there is automatically a quotient of real numbers.) Second, the term $D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})$ in the definition of $\mathbf{h}$ should be interpreted as the product of the $m \times n$ matrix $\mathrm{D}\mathbf{f}(\mathbf{a})$ and the $n \times 1$ column matrix

$$\begin{bmatrix} x_1 - a_1 \\ x_2 - a_2 \\ \vdots \\ x_n - a_n \end{bmatrix}.$$

Because of the consistency of our definitions, the following results should not surprise you:

THEOREM 3.9 If $\mathbf{f}\colon X \subseteq \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at $\mathbf{a}$, then it is continuous at $\mathbf{a}$.

THEOREM 3.10 If $\mathbf{f}\colon X \subseteq \mathbf{R}^n \to \mathbf{R}^m$ is such that, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, all $\partial f_i/\partial x_j$ exist and are continuous in a neighborhood of $\mathbf{a}$ in X, then $\mathbf{f}$ is differentiable at $\mathbf{a}$.

THEOREM 3.11 A function $\mathbf{f}\colon X \subseteq \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at $\mathbf{a} \in X$ (in the sense of Definition 3.8) if and only if each of its component functions $f_i\colon X \subseteq \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$, is differentiable at $\mathbf{a}$ (in the sense of Definition 3.7).

The proofs of Theorems 3.9, 3.10, and 3.11 are provided in the addendum to this section. Note that Theorems 3.10 and 3.11 frequently make it a straightforward matter to check that a function is differentiable: Just look at the partial derivatives of the component functions and verify that they are continuous. Thus, in many—but not all—circumstances, we can avoid working directly with the limit in Definition 3.8.

EXAMPLE 10 The function $\mathbf{g}\colon \mathbf{R}^3 - \{(0,0,0)\} \to \mathbf{R}^3$ given by

$$\mathbf{g}(x, y, z) = \left(\frac{3}{x^2 + y^2 + z^2}, xy, xz\right)$$

has

$$D\mathbf{g}(x, y, z) = \begin{bmatrix} \dfrac{-6x}{(x^2+y^2+z^2)^2} & \dfrac{-6y}{(x^2+y^2+z^2)^2} & \dfrac{-6z}{(x^2+y^2+z^2)^2} \\ y & x & 0 \\ z & 0 & x \end{bmatrix}.$$

Each of the entries of this matrix is continuous over $\mathbf{R}^3 - \{(0,0,0)\}$. Hence, by Theorem 3.10, $\mathbf{g}$ is differentiable over its entire domain. ◆

What Is a Derivative?

Although we have defined quite carefully what it means for a function to be differentiable, the derivative itself has really taken a “backseat” in the preceding discussion. It is time to get some perspective on the concept of the derivative.

In the case of a (differentiable) scalar-valued function of a single variable, $f\colon X \subseteq \mathbf{R} \to \mathbf{R}$, the derivative $f'(a)$ is simply a real number, the slope of the tangent line to the graph of f at the point $(a, f(a))$. From a more sophisticated (and slightly less geometric) point of view, the derivative $f'(a)$ is the number such that the function

$$h(x) = f(a) + f'(a)(x - a)$$

is a good linear approximation to $f(x)$ for x near a. (And, of course, $y = h(x)$ is the equation of the tangent line.)

If a function $f\colon X \subseteq \mathbf{R}^n \to \mathbf{R}$ of n variables is differentiable, there must exist n partial derivatives $\partial f/\partial x_1, \ldots, \partial f/\partial x_n$. These partial derivatives form the components of the gradient vector ∇f (or the entries of the $1 \times n$ matrix Df). It is the gradient that should properly be considered to be the derivative of f, but in the following sense: $\nabla f(\mathbf{a})$ is the vector such that the function $h: \mathbf{R}^n \to \mathbf{R}$ given by

$$h(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

is a good linear approximation to $f(\mathbf{x})$ for $\mathbf{x}$ near $\mathbf{a}$. Finally, the derivative of a differentiable vector-valued function $\mathbf{f}: X \subseteq \mathbf{R}^n \to \mathbf{R}^m$ may be taken to be the matrix $D\mathbf{f}$ of partial derivatives, but in the sense that the function $\mathbf{h}: \mathbf{R}^n \to \mathbf{R}^m$ given by

$$\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

is a good linear approximation to $\mathbf{f}(\mathbf{x})$ near $\mathbf{a}$. You should view the derivative $D\mathbf{f}(\mathbf{a})$ not as a "static" matrix of numbers, but rather as a matrix that defines a *linear mapping* from $\mathbf{R}^n$ to $\mathbf{R}^m$. (See Example 5 of §1.6.) This is embodied in the limit equation of Definition 3.8 and, though a subtle idea, is truly the heart of differential calculus of several variables.

In fact, we could have approached our discussion of differentiability much more abstractly right from the beginning. We could have defined a function $\mathbf{f}: X \subseteq \mathbf{R}^n \to \mathbf{R}^m$ to be differentiable at a point $\mathbf{a} \in X$ to mean that there exists some linear mapping $\mathbf{L}: \mathbf{R}^n \to \mathbf{R}^m$ such that

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - [\mathbf{f}(\mathbf{a}) + \mathbf{L}(\mathbf{x} - \mathbf{a})]\|}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

Recall that any linear mapping $\mathbf{L}: \mathbf{R}^n \to \mathbf{R}^m$ is really nothing more than multiplication by a suitable $m \times n$ matrix A (i.e., that $\mathbf{L}(\mathbf{y}) = A\mathbf{y}$). It is possible to show that if there is a linear mapping that satisfies the aforementioned limit equation, then the matrix A that defines it is both uniquely determined and is precisely the matrix of partial derivatives $D\mathbf{f}(\mathbf{a})$. (See Exercises 60–62 where these facts are proved.) However, to begin with such a definition, though equivalent to Definition 3.8, strikes us as less well motivated than the approach we have taken. Hence, we have presented the notions of differentiability and the derivative from what we hope is a somewhat more concrete and geometric perspective.

Addendum: Proofs of Theorems 3.9, 3.10, and 3.11

Proof of Theorem 3.9 We begin by claiming the following: Let $\mathbf{x} \in \mathbf{R}^n$ and $B = (b_{ij})$ be an $m \times n$ matrix. If $\mathbf{y} = B\mathbf{x}$, (so $\mathbf{y} \in \mathbf{R}^m$), then

$$\|\mathbf{y}\| \leq K\|\mathbf{x}\|, \tag{9}$$

where $K = \left(\sum_{i,j} b_{ij}^2\right)^{1/2}$. We postpone the proof of (9) until we establish the main theorem.

To show that $\mathbf{f}$ is continuous at $\mathbf{a}$, we will show that $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})\| \to 0$ as $\mathbf{x} \to \mathbf{a}$. We do so by using the fact that $\mathbf{f}$ is differentiable at $\mathbf{a}$ (Definition 3.8).

We have

$$\begin{aligned} \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})\| &= \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})\| \\ &\leq \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})\| + \|D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})\|, \qquad (10) \end{aligned}$$

using the triangle inequality. Note that the first term in the right side of inequality (10) is the numerator of the limit expression in Definition 3.8. Thus, since $\mathbf{f}$ is