that satisfies Laplace's equation is called a **harmonic function**.[3]

(a) Is $f(x, y, z) = x^2 + y^2 - 2z^2$ harmonic? What about $f(x, y, z) = x^2 - y^2 + z^2$?

(b) We may generalize Laplace's equation to functions of n variables as

$$\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2} = 0.$$

Give an example of a harmonic function of n variables, and verify that your example is correct.

29. The three-dimensional **heat equation** is the partial differential equation

$$k\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) = \frac{\partial T}{\partial t},$$

where k is a positive constant. It models the temperature $T(x, y, z, t)$ at the point (x, y, z) and time t of a body in space.

(a) We examine a simplified version of the heat equation. Consider a straight wire "coordinatized" by x. Then the temperature $T(x, t)$ at time t and position x along the wire is modeled by the one-dimensional heat equation

$$k\frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t}.$$

Show that the function $T(x, t) = e^{-kt} \cos x$ satisfies this equation. Note that if t is held constant at value t_0, then $T(x, t_0)$ shows how the temperature varies along the wire at time t_0. Graph the curves $z = T(x, t_0)$ for $t_0 = 0, 1, 10$, and use them to understand the graph of the surface $z = T(x, t)$ for $t \geq 0$. Explain what happens to the temperature of the wire after a long period of time.

(b) Show that $T(x, y, t) = e^{-kt}(\cos x + \cos y)$ satisfies the two-dimensional heat equation

$$k\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right) = \frac{\partial T}{\partial t}.$$

Graph the surfaces given by $z = T(x, y, t_0)$, where $t_0 = 0, 1, 10$. If we view the function $T(x, y, t)$ as modeling the temperature at points (x, y) of a flat plate at time t, then describe what happens to the temperature of the plate after a long period of time.

(c) Now show that $T(x, y, z, t) = e^{-kt}(\cos x + \cos y + \cos z)$ satisfies the three-dimensional heat equation.

30. Let

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

(a) Find $f_x(x, y)$ and $f_y(x, y)$ for $(x, y) \neq (0, 0)$. (You will find a computer algebra system helpful.)

(b) Either by hand (using limits) or by means of part (a), find the partial derivatives $f_x(0, y)$ and $f_y(x, 0)$.

(c) Find the values of $f_{xy}(0, 0)$ and $f_{yx}(0, 0)$. Reconcile your answer with Theorem 4.3.

A surface that has the least surface area among all surfaces with a given boundary is called a ***minimal surface.*** *Soap bubbles are naturally occurring examples of minimal surfaces. It is a fact that minimal surfaces having equations of the form $z = f(x, y)$ (where f is of class C^2) satisfy the partial differential equation*

$$\left(1 + z_y^2\right) z_{xx} + \left(1 + z_x^2\right) z_{yy} = 2z_x z_y z_{xy}. \tag{6}$$

Exercises 31–33 concern minimal surfaces and equation (6).

31. Show that a plane is a minimal surface.

32. **Scherk's surface** is given by the equation $e^z \cos y = \cos x$.

(a) Use a computer to graph a portion of this surface.

(b) Verify that Scherk's surface is a minimal surface.

33. One way to describe the surface known as the **helicoid** is by the equation $x = y \tan z$.

(a) Use a computer to graph a portion of this surface.

(b) Verify that the helicoid is a minimal surface.

2.5 The Chain Rule

Among the various properties that the derivative satisfies, one that stands alone in both its usefulness and its subtlety is the derivative's behavior with respect to composition of functions. This behavior is described by a formula known as

[3] Laplace did fundamental and far-reaching work in both mathematical physics and probability theory. Laplace's equation and harmonic functions are part of the field of **potential theory,** a subject that Laplace can be credited as having developed. Potential theory has applications to such areas as gravitation, electricity and magnetism, and fluid mechanics, to name a few.

the **chain rule**. In this section, we review the chain rule of one-variable calculus and see how it generalizes to the cases of scalar- and vector-valued functions of several variables.

The Chain Rule for Functions of One Variable: A Review

We begin with a typical example of the use of the chain rule from single-variable calculus.

EXAMPLE 1 Let $f(x) = \sin x$ and $x(t) = t^3 + t$. We may then construct the composite function $f(x(t)) = \sin(t^3 + t)$. The chain rule tells us how to find the derivative of $f \circ x$ with respect to t:

$$(f \circ x)'(t) = \frac{d}{dt}(\sin(t^3 + t)) = (\cos(t^3 + t))(3t^2 + 1).$$

Since $x = t^3 + t$, we have

$$(f \circ x)'(t) = \frac{d}{dx}(\sin x) \cdot \frac{d}{dt}(t^3 + t) = f'(x) \cdot x'(t). \qquad \blacklozenge$$

In general, suppose X and T are open subsets of $\mathbf{R}$ and $f: X \subseteq \mathbf{R} \to \mathbf{R}$ and $x: T \subseteq \mathbf{R} \to \mathbf{R}$ are functions defined so that the composite function $f \circ x: T \to \mathbf{R}$ makes sense. (See Figure 2.57.) In particular, this means that the range of the function x must be contained in X, the domain of f. The key result is the following:

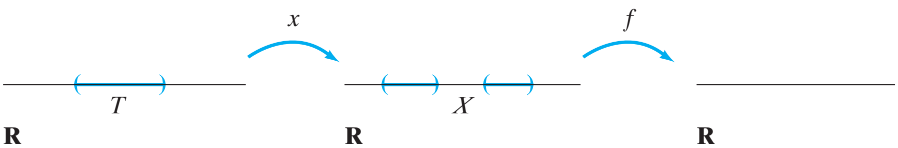

Figure 2.57 The range of the function x must be contained in the domain X of f in order for the composite $f \circ x$ to be defined.

THEOREM 5.1 (The chain rule in one variable) Under the preceding assumptions, if x is differentiable at $t_0 \in T$ and f is differentiable at $x_0 = x(t_0) \in X$, then the composite $f \circ x$ is differentiable at t_0 and, moreover,

$$(f \circ x)'(t_0) = f'(x_0)x'(t_0). \tag{1}$$

A more common way to write the chain rule formula in Theorem 5.1 is

$$\frac{df}{dt}(t_0) = \frac{df}{dx}(x_0)\,\frac{dx}{dt}(t_0). \tag{2}$$

Although equation (2) is most useful in practice, it does represent an unfortunate abuse of notation in that the symbol f is used to denote both a function of x and one of t. It would be more appropriate to define a new function y by $y(t) = (f \circ x)(t)$ so that $dy/dt = (df/dx)(dx/dt)$. But our original abuse of notation is actually a convenient one, since it avoids the awkwardness of having too many variable names appearing in a single discussion. In the name of simplicity, we will therefore continue to commit such abuses and urge you to do likewise.

The formulas in equations (1) and (2) are so simple that little more needs to be said. We elaborate, nonetheless, because this will prove helpful when we generalize to the case of several variables. The chain rule tells us the following: To understand how f depends on t, we must know how f depends on the "intermediate variable" x and how this intermediate variable depends on the "final" independent variable t. The diagram in Figure 2.58 traces the hierarchy of the variable dependences. The "paths" indicate the derivatives involved in the chain rule formula.

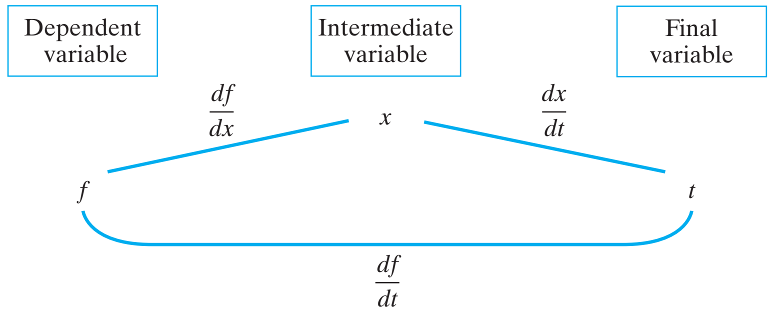

Figure 2.58 The chain rule for functions of a single variable.

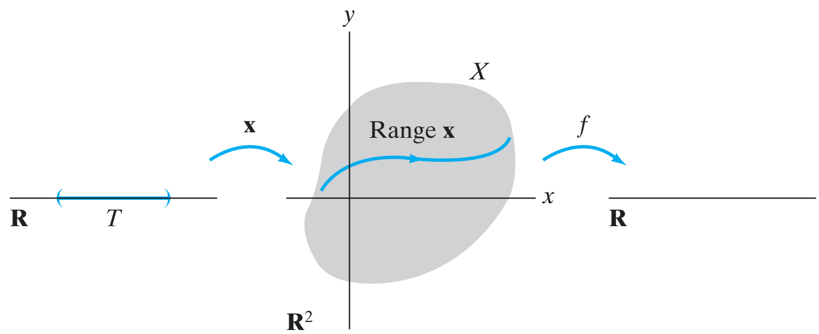

Figure 2.59 The composite function $f \circ \mathbf{x}$.

The Chain Rule in Several Variables

Now let's go a step further and assume $f: X \subseteq \mathbf{R}^2 \to \mathbf{R}$ is a C^1 function of two variables and $\mathbf{x}: T \subseteq \mathbf{R} \to \mathbf{R}^2$ is a differentiable vector-valued function of a single variable. If the range of $\mathbf{x}$ is contained in X, then the composite $f \circ \mathbf{x}: T \subseteq \mathbf{R} \to \mathbf{R}$ is defined. (See Figure 2.59.) It's good to think of $\mathbf{x}$ as describing a parametrized curve in $\mathbf{R}^2$ and f as a sort of "temperature function" on X. The composite $f \circ \mathbf{x}$ is then nothing more than the restriction of f to the curve (i.e., the function that measures the temperature along just the curve). The question is, how does f depend on t? We claim the following:

PROPOSITION 5.2 Suppose $\mathbf{x}: T \subseteq \mathbf{R} \to \mathbf{R}^2$ is differentiable at $t_0 \in T$, and $f: X \subseteq \mathbf{R}^2 \to \mathbf{R}$ is differentiable at $\mathbf{x}_0 = \mathbf{x}(t_0) = (x_0, y_0) \in X$, where T and X are open in $\mathbf{R}$ and $\mathbf{R}^2$, respectively, and range $\mathbf{x}$ is contained in X. If, in addition, f is of class C^1, then $f \circ \mathbf{x}: T \to \mathbf{R}$ is differentiable at t_0 and

$$\frac{df}{dt}(t_0) = \frac{\partial f}{\partial x}(\mathbf{x}_0)\frac{dx}{dt}(t_0) + \frac{\partial f}{\partial y}(\mathbf{x}_0)\frac{dy}{dt}(t_0).$$

Before we prove Proposition 5.2, some remarks are in order. First, notice the mixture of ordinary and partial derivatives appearing in the formula for the

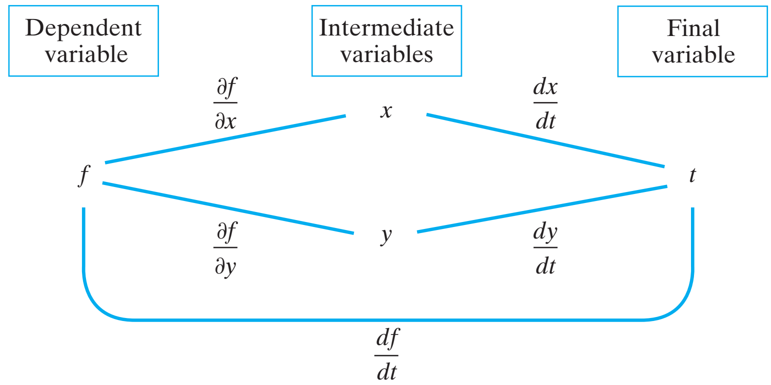

Figure 2.60 The chain rule of Proposition 5.2.

derivative. These terms make sense if we contruct an appropriate "variable hierarchy" diagram, as shown in Figure 2.60. At the intermediate level, f depends on two variables, x and y (or, equivalently, on the vector variable $\mathbf{x} = (x, y)$), so partial derivatives are in order. On the final or composite level, f depends on just a single independent variable t and, hence, the use of the ordinary derivative df/dt is warranted. Second, the formula in Proposition 5.2 is a generalization of equation (2): A product term appears for each of the two intermediate variables.

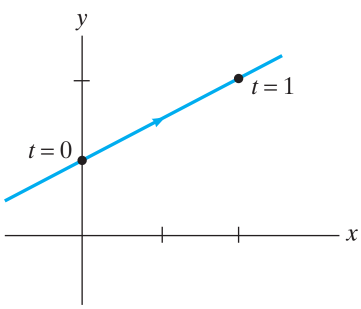

Figure 2.61 The graph of the function $\mathbf{x}$ of Example 2.

EXAMPLE 2 Suppose $f(x, y) = (x + y^2)/(2x^2 + 1)$ is a temperature function on $\mathbf{R}^2$ and $\mathbf{x}(t) = (2t, t + 1)$. The function $\mathbf{x}$ gives parametric equations for a line. (See Figure 2.61.) Then

$$(f \circ \mathbf{x})(t) = f(\mathbf{x}(t)) = \frac{2t + (t+1)^2}{8t^2 + 1} = \frac{t^2 + 4t + 1}{8t^2 + 1}$$

is the temperature function along the line, and we have

$$\frac{df}{dt} = \frac{4 - 14t - 32t^2}{(8t^2 + 1)^2},$$

by the quotient rule. Thus, all the hypotheses of Proposition 5.2 are satisfied and so the derivative formula must hold. Indeed, we have

$$\frac{\partial f}{\partial x} = \frac{1 - 2x^2 - 4xy^2}{(2x^2 + 1)^2},$$

$$\frac{\partial f}{\partial y} = \frac{2y}{2x^2 + 1},$$

and

$$\mathbf{x}'(t) = \left(\frac{dx}{dt}, \frac{dy}{dt}\right) = (2, 1).$$

Therefore,

$$\begin{aligned}\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} &= \frac{1 - 2x^2 - 4xy^2}{(2x^2 + 1)^2} \cdot 2 + \frac{2y}{2x^2 + 1} \cdot 1 \\ &= \frac{2(1 - 8t^2 - 8t(t+1)^2)}{(8t^2 + 1)^2} + \frac{2(t+1)}{8t^2 + 1},\end{aligned}$$

after substitution of $2t$ for x and $t+1$ for y. Hence,

$$\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \frac{2(2-7t-16t^2)}{(8t^2+1)^2},$$

which checks with our previous result for df/dt. ◆

Proof of Proposition 5.2 Denote the composite function $f \circ \mathbf{x}$ by z. We want to establish a formula for dz/dt at t_0. Since z is just a scalar-valued function of one variable, differentiability and the existence of the derivative mean the same thing. Thus, we consider

$$\frac{dz}{dt}(t_0) = \lim_{t\to t_0} \frac{z(t)-z(t_0)}{t-t_0},$$

and see if this limit exists. We have

$$\frac{dz}{dt}(t_0) = \lim_{t\to t_0} \frac{f(x(t), y(t)) - f(x(t_0), y(t_0))}{t-t_0}.$$

The first step is to rewrite the numerator of the limit expression by subtracting and adding $f(x_0, y)$ and to apply a modicum of algebra. Thus,

$$\begin{aligned}\frac{dz}{dt}(t_0) &= \lim_{t\to t_0} \frac{f(x,y) - f(x_0,y) + f(x_0,y) - f(x_0,y_0)}{t-t_0}\\ &= \lim_{t\to t_0} \frac{f(x,y)-f(x_0,y)}{t-t_0} + \lim_{t\to t_0} \frac{f(x_0,y)-f(x_0,y_0)}{t-t_0}.\end{aligned}$$

(Remember that $\mathbf{x}(t_0) = \mathbf{x}_0 = (x_0, y_0)$.) Now, for the main innovation of the proof. We apply the mean value theorem to the partial functions of f. This tells us that there must be a number c between x_0 and x and another number d between y_0 and y such that

$$f(x,y) - f(x_0,y) = f_x(c,y)(x-x_0)$$

and

$$f(x_0,y) - f(x_0,y_0) = f_y(x_0,d)(y-y_0).$$

Thus,

$$\begin{aligned}\frac{dz}{dt}(t_0) &= \lim_{t\to t_0} f_x(c,y)\frac{x-x_0}{t-t_0} + \lim_{t\to t_0} f_y(x_0,d)\frac{y-y_0}{t-t_0}\\ &= \lim_{t\to t_0} f_x(c,y)\frac{x(t)-x(t_0)}{t-t_0} + \lim_{t\to t_0} f_y(x_0,d)\frac{y(t)-y(t_0)}{t-t_0}\\ &= f_x(x_0,y_0)\frac{dx}{dt}(t_0) + f_y(x_0,y_0)\frac{dy}{dt}(t_0),\end{aligned}$$

by the definition of the derivatives

$$\frac{dx}{dt}(t_0) \quad \text{and} \quad \frac{dy}{dt}(t_0)$$

and the fact that $f_x(c,y)$ and $f_y(x_0,d)$ must approach $f_x(x_0,y_0)$ and $f_y(x_0,y_0)$, respectively, as t approaches t_0, by continuity of the partials. (Recall that f was assumed to be of class C^1.) This completes the proof. ■

Proposition 5.2 and its proof are easy to generalize to the case where f is a function of n variables (i.e., $f\colon X \subseteq \mathbf{R}^n \to \mathbf{R}$) and $\mathbf{x}: T \subseteq \mathbf{R} \to \mathbf{R}^n$. The appropriate chain rule formula in this case is

$$\frac{df}{dt}(t_0) = \frac{\partial f}{\partial x_1}(\mathbf{x}_0)\frac{dx_1}{dt}(t_0) + \frac{\partial f}{\partial x_2}(\mathbf{x}_0)\frac{dx_2}{dt}(t_0) + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}_0)\frac{dx_n}{dt}(t_0). \tag{3}$$

Note that the right side of equation (3) can also be written by using matrix notation so that

$$\frac{df}{dt}(t_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \end{bmatrix} \begin{bmatrix} \frac{dx_1}{dt}(t_0) \\ \frac{dx_2}{dt}(t_0) \\ \vdots \\ \frac{dx_n}{dt}(t_0) \end{bmatrix}.$$

Thus, we have shown

$$\boxed{\frac{df}{dt}(t_0) = Df(\mathbf{x}_0)D\mathbf{x}(t_0) = \nabla f(\mathbf{x}_0) \cdot \mathbf{x}'(t_0),} \tag{4}$$

where we use $\mathbf{x}'(t_0)$ as a notational alternative to $D\mathbf{x}(t_0)$. The version of the chain rule given in formula (4) is particularly important and will be used a number of times in our subsequent work.

Let us consider further instances of composition of functions of many variables. For example, suppose X is open in $\mathbf{R}^3$, T is open in $\mathbf{R}^2$, and $f: X \subseteq \mathbf{R}^3 \to \mathbf{R}$ and $\mathbf{x}: T \subseteq \mathbf{R}^2 \to \mathbf{R}^3$ are such that the range of $\mathbf{x}$ is contained in X. Then the composite $f \circ \mathbf{x}: T \subseteq \mathbf{R}^2 \to \mathbf{R}$ can be formed, as shown in Figure 2.62. Note that the range of $\mathbf{x}$, that is, $\mathbf{x}(T)$, is just a surface in $\mathbf{R}^3$, so $f \circ \mathbf{x}$ can be thought of as an appropriate "temperature function" restricted to this surface. If we use $\mathbf{x} = (x, y, z)$ to denote the vector variable in $\mathbf{R}^3$ and $\mathbf{t} = (s, t)$ for the vector variable in $\mathbf{R}^2$, then we can write a plausible chain rule formula from an appropriate variable hierarchy diagram. (See Figure 2.63.) Thus, it is

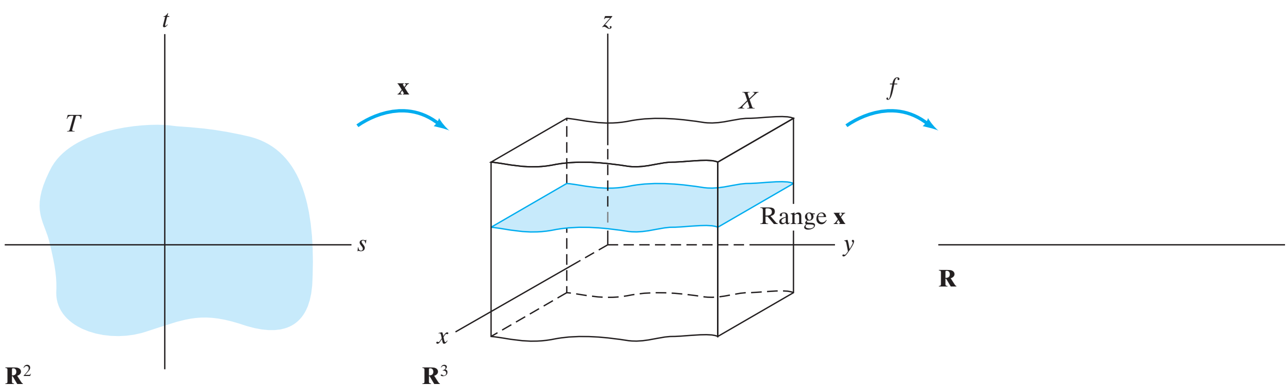

Figure 2.62 The composite $f \circ \mathbf{x}$ where $f: X \subseteq \mathbf{R}^3 \to \mathbf{R}$ and $\mathbf{x}: T \subseteq \mathbf{R}^2 \to \mathbf{R}^3$.

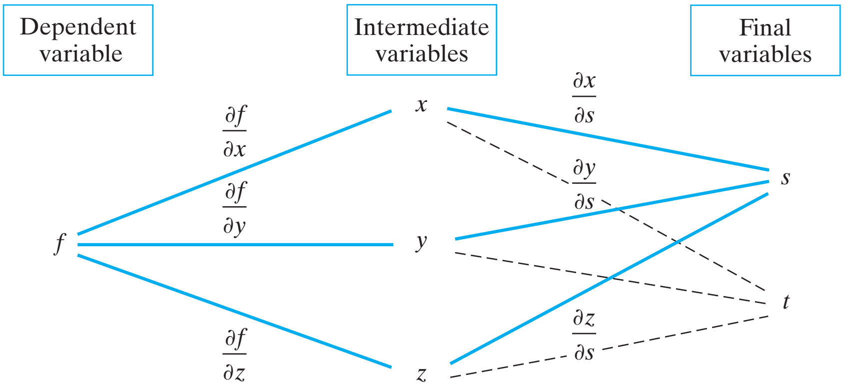

Figure 2.63 The chain rule for $f \circ \mathbf{x}$, where $f: X \subseteq \mathbf{R}^3 \to \mathbf{R}$ and $\mathbf{x}: T \subseteq \mathbf{R}^2 \to \mathbf{R}^3$.

reasonable to expect that the following formulas hold:

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial s}$$

and (5)

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial t}.$$

(Again, we abuse notation by writing both $\partial f/\partial s$, $\partial f/\partial t$ and $\partial f/\partial x$, $\partial f/\partial y$, $\partial f/\partial z$.) Indeed, when f is a function of x, y, and z of class C^1, formula (3) with $n = 3$ applies once we realize that $\partial x/\partial s$, $\partial x/\partial t$, etc., represent ordinary differentiation of the partial functions in s or t.

EXAMPLE 3 Suppose

$$f(x, y, z) = x^2 + y^2 + z^2 \quad \text{and} \quad \mathbf{x}(s, t) = (s\cos t, e^{st}, s^2 - t^2).$$

Then $h(s, t) = f \circ \mathbf{x}(s, t) = s^2\cos^2 t + e^{2st} + (s^2 - t^2)^2$, so that

$$\frac{\partial h}{\partial s} = \frac{\partial (f \circ \mathbf{x})}{\partial s} = 2s\cos^2 t + 2te^{2st} + 4s(s^2 - t^2)$$

$$\frac{\partial h}{\partial t} = \frac{\partial (f \circ \mathbf{x})}{\partial t} = -2s^2\cos t\sin t + 2se^{2st} - 4t(s^2 - t^2).$$

We also have

$$\frac{\partial f}{\partial x} = 2x, \qquad \frac{\partial f}{\partial y} = 2y, \qquad \frac{\partial f}{\partial z} = 2z$$

and

$$\begin{aligned} \frac{\partial x}{\partial s} &= \cos t, & \frac{\partial x}{\partial t} &= -s\sin t, \\ \frac{\partial y}{\partial s} &= te^{st}, & \frac{\partial y}{\partial t} &= se^{st}, \\ \frac{\partial z}{\partial s} &= 2s, & \frac{\partial z}{\partial t} &= -2t. \end{aligned}$$

Hence, we compute

$$\frac{\partial f}{\partial s} = \frac{\partial (f \circ \mathbf{x})}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial s}$$

$$= 2x(\cos t) + 2y(te^{st}) + 2z(2s)$$

$$= 2s\cos t(\cos t) + 2e^{st}(te^{st}) + 2(s^2 - t^2)(2s)$$

$$= 2s\cos^2 t + 2te^{2st} + 4s(s^2 - t^2),$$

just as we saw earlier. We leave it to you to use the chain rule to calculate $\partial f/\partial t$ in a similar manner. ◆

Of course, there is no need for us to stop here. Suppose we have an open set X in $\mathbf{R}^m$, an open set T in $\mathbf{R}^n$, and functions $f: X \to \mathbf{R}$ and $\mathbf{x}: T \to \mathbf{R}^m$ such that $h = f \circ \mathbf{x}: T \to \mathbf{R}$ can be defined. If f is of class C^1 and $\mathbf{x}$ is differentiable, then, from the previous remarks, h must also be differentiable and, moreover,

$$\frac{\partial h}{\partial t_j} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_j} + \cdots + \frac{\partial f}{\partial x_m}\frac{\partial x_m}{\partial t_j}$$

$$= \sum_{k=1}^{m} \frac{\partial f}{\partial x_k}\frac{\partial x_k}{\partial t_j}, \qquad j = 1, 2, \ldots, n.$$

Since the component functions of a vector-valued function are just scalar-valued functions, we can say even more. Suppose $\mathbf{f}: X \subseteq \mathbf{R}^m \to \mathbf{R}^p$ and $\mathbf{x}: T \subseteq \mathbf{R}^n \to \mathbf{R}^m$ are such that $\mathbf{h} = \mathbf{f} \circ \mathbf{x}: T \subseteq \mathbf{R}^n \to \mathbf{R}^p$ can be defined. (As always, we assume that X is open in $\mathbf{R}^m$ and T is open in $\mathbf{R}^n$.) See Figure 2.64 for a representation of the situation. If $\mathbf{f}$ is of class C^1 and $\mathbf{x}$ is differentiable, then the composite $\mathbf{h} = \mathbf{f} \circ \mathbf{x}$ is differentiable and the following general formula holds:

$$\frac{\partial h_i}{\partial t_j} = \sum_{k=1}^{m} \frac{\partial f_i}{\partial x_k}\frac{\partial x_k}{\partial t_j}, \qquad i = 1, 2, \ldots, p; \quad j = 1, 2, \ldots, n. \tag{6}$$

The plausibility of formula (6) is immediate, given the variable hierarchy diagram shown in Figure 2.65.

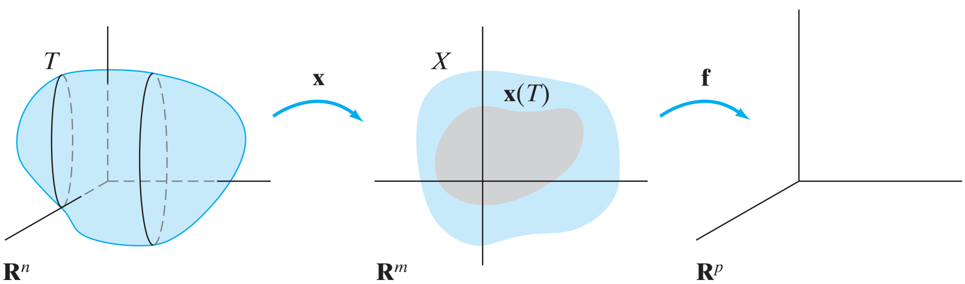

Figure 2.64 The composite $\mathbf{f} \circ \mathbf{x}$ where $\mathbf{f}: X \subseteq \mathbf{R}^m \to \mathbf{R}^p$ and $\mathbf{x}: T \subseteq \mathbf{R}^n \to \mathbf{R}^m$.

Now comes the real "magic." Recall that if A is a $p \times m$ matrix and B is an $m \times n$ matrix, then the product matrix $C = AB$ is defined and is a $p \times n$ matrix. Moreover, the ijth entry of C is given by

$$c_{ij} = \sum_{k=1}^{m} a_{ik}b_{kj}.$$

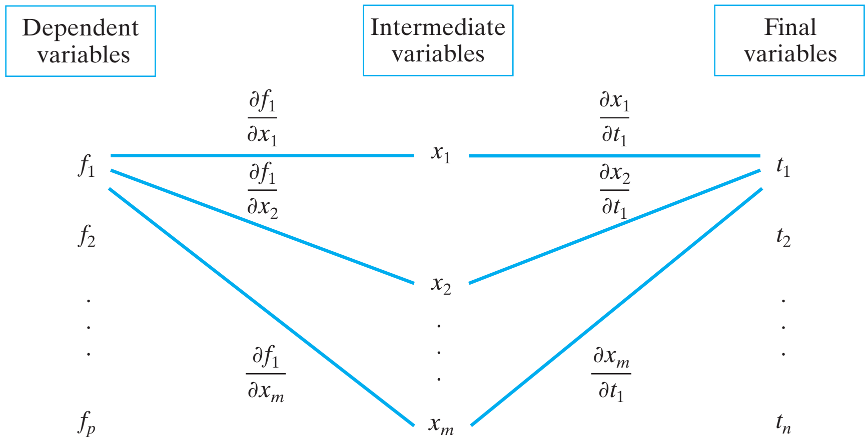

Figure 2.65 The chain rule diagram for $\mathbf{f} \circ \mathbf{x}$, where $\mathbf{f}\colon X \subseteq \mathbf{R}^m \to \mathbf{R}^p$ and $\mathbf{x}\colon T \subseteq \mathbf{R}^n \to \mathbf{R}^m$.

If we recall that the ijth entry of the matrix $D\mathbf{h}(\mathbf{t})$ is $\partial h_i/\partial t_j$, and similarly for $D\mathbf{f}(\mathbf{x})$ and $D\mathbf{x}(\mathbf{t})$, then we see that formula (6) expresses nothing more than the following equation of matrices:

$$D\mathbf{h}(\mathbf{t}) = D(\mathbf{f} \circ \mathbf{x})(\mathbf{t}) = D\mathbf{f}(\mathbf{x})D\mathbf{x}(\mathbf{t}). \tag{7}$$

The similarity between formulas (7) and (1) is striking. One of the reasons (perhaps the principal reason) for defining matrix multiplication as we have is precisely so that the chain rule in several variables can have the elegant appearance that it has in formula (7).

EXAMPLE 4 Suppose $\mathbf{f}\colon \mathbf{R}^3 \to \mathbf{R}^2$ is given by $\mathbf{f}(x_1, x_2, x_3) = (x_1 - x_2, x_1x_2x_3)$ and $\mathbf{x}\colon \mathbf{R}^2 \to \mathbf{R}^3$ is given by $\mathbf{x}(t_1, t_2) = (t_1t_2,\ t_1^2,\ t_2^2)$. Then $\mathbf{f} \circ \mathbf{x}\colon \mathbf{R}^2 \to \mathbf{R}^2$ is given by $(\mathbf{f} \circ \mathbf{x})(t_1, t_2) = (t_1t_2 - t_1^2, t_1^3t_2^3)$, so that

$$D(\mathbf{f} \circ \mathbf{x})(\mathbf{t}) = \begin{bmatrix} t_2 - 2t_1 & t_1 \\ 3t_1^2t_2^3 & 3t_1^3t_2^2 \end{bmatrix}.$$

On the other hand,

$$D\mathbf{f}(\mathbf{x}) = \begin{bmatrix} 1 & -1 & 0 \\ x_2x_3 & x_1x_3 & x_1x_2 \end{bmatrix} \quad \text{and} \quad D\mathbf{x}(t) = \begin{bmatrix} t_2 & t_1 \\ 2t_1 & 0 \\ 0 & 2t_2 \end{bmatrix},$$

so that the product matrix is

$$D\mathbf{f}(\mathbf{x})D\mathbf{x}(\mathbf{t}) = \begin{bmatrix} t_2 - 2t_1 & t_1 \\ x_2x_3t_2 + 2x_1x_3t_1 & x_2x_3t_1 + 2x_1x_2t_2 \end{bmatrix}$$

$$= \begin{bmatrix} t_2 - 2t_1 & t_1 \\ t_1^2t_2^3 + 2t_1^2t_2^3 & t_1^3t_2^2 + 2t_1^3t_2^2 \end{bmatrix},$$

after substituting for x_1, x_2, and x_3. Thus, $D(\mathbf{f} \circ \mathbf{x})(\mathbf{t}) = D\mathbf{f}(\mathbf{x})D\mathbf{x}(\mathbf{t})$, as expected.

Alternatively, we may use the variable hierarchy diagram shown in Figure 2.66 and compute any individual partial derivative we may desire. For example,

$$\frac{\partial f_2}{\partial t_1} = \frac{\partial f_2}{\partial x_1}\frac{\partial x_1}{\partial t_1} + \frac{\partial f_2}{\partial x_2}\frac{\partial x_2}{\partial t_1} + \frac{\partial f_2}{\partial x_3}\frac{\partial x_3}{\partial t_1}$$

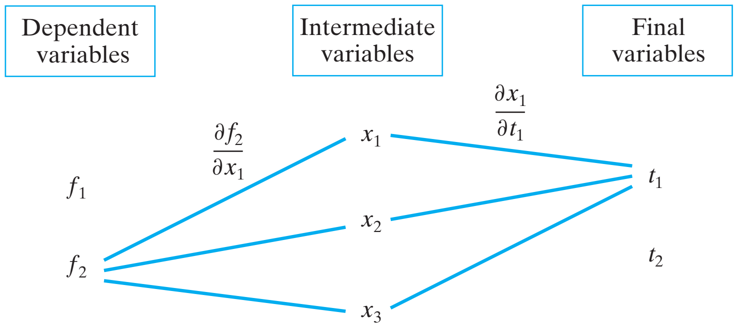

Figure 2.66 The variable hierarchy diagram for Example 4.

by formula (6). Then by abuse of notation,

$$\begin{aligned}\frac{\partial f_2}{\partial t_1} &= (x_2x_3)(t_2) + (x_1x_3)(2t_1) + (x_1x_2)(0)\\ &= (t_1^2t_2^2)(t_2) + (t_1t_2)(t_2^2)(2t_1)\\ &= 3t_1^2t_2^3,\end{aligned}$$

which is indeed the (2, 1) entry of the matrix product. ◆

At last we state the most general version of the chain rule from a technical standpoint; a proof may be found in the addendum to this section.

THEOREM 5.3 (THE CHAIN RULE) Suppose $X \subseteq \mathbf{R}^m$ and $T \subseteq \mathbf{R}^n$ are open and $\mathbf{f}\colon X \to \mathbf{R}^p$ and $\mathbf{x}\colon T \to \mathbf{R}^m$ are defined so that range $\mathbf{x} \subseteq X$. If $\mathbf{x}$ is differentiable at $\mathbf{t}_0 \in T$ and $\mathbf{f}$ is differentiable at $\mathbf{x}_0 = \mathbf{x}(\mathbf{t}_0)$, then the composite $\mathbf{f} \circ \mathbf{x}$ is differentiable at $\mathbf{t}_0$, and we have

$$D(\mathbf{f} \circ \mathbf{x})(\mathbf{t}_0) = D\mathbf{f}(\mathbf{x}_0)D\mathbf{x}(\mathbf{t}_0).$$

The advantage of Theorem 5.3 over the earlier versions of the chain rule we have been discussing is that it requires $\mathbf{f}$ only to be differentiable at the point in question, not to be of class C^1. Note that, of course, Theorem 5.3 includes all the special cases of the chain rule we have previously discussed. In particular, Theorem 5.3 includes the important case of formula (4).

EXAMPLE 5 Let $\mathbf{f}\colon \mathbf{R}^2 \to \mathbf{R}^2$ be defined by $\mathbf{f}(x, y) = (x - 2y + 7, 3xy^2)$. Suppose that $\mathbf{g}\colon \mathbf{R}^3 \to \mathbf{R}^2$ is differentiable at $(0, 0, 0)$ and we know that $\mathbf{g}(0, 0, 0) = (-2, 1)$ and

$$D\mathbf{g}(0, 0, 0) = \begin{bmatrix} 2 & 4 & 5 \\ -1 & 0 & 1 \end{bmatrix}.$$

We use this information to determine $D(\mathbf{f} \circ \mathbf{g})(0, 0, 0)$.

First, note that Theorem 5.3 tells us that $\mathbf{f} \circ \mathbf{g}$ must be differentiable at $(0, 0, 0)$ and, second, that

$$D(\mathbf{f} \circ \mathbf{g})(0, 0, 0) = D\mathbf{f}(\mathbf{g}(0, 0, 0))\, D\mathbf{g}(0, 0, 0) = D\mathbf{f}(-2, 1)D\mathbf{g}(0, 0, 0).$$

Since we know **f** completely, it is easy to compute that

$$D\mathbf{f}(x, y) = \begin{bmatrix} 1 & -2 \\ 3y^2 & 6xy \end{bmatrix} \quad \text{so that} \quad D\mathbf{f}(-2, 1) = \begin{bmatrix} 1 & -2 \\ 3 & -12 \end{bmatrix}.$$

Thus,

$$D(\mathbf{f} \circ \mathbf{g})(0, 0, 0) = \begin{bmatrix} 1 & -2 \\ 3 & -12 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 4 & 3 \\ 18 & 12 & 3 \end{bmatrix}.$$

We remark that we needed the full strength of Theorem 5.3, as we do not know anything about the differentiability of **g** other than at the point $(0, 0, 0)$. ◆

EXAMPLE 6 **(Polar/rectangular conversions)** Recall that in §1.7 we provided the basic equations relating polar and rectangular coordinates:

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}.$$

Now suppose you have an equation defining a quantity w as a function of x and y; that is,

$$w = f(x, y).$$

Then, of course, w may just as well be regarded as a function of r and θ by susbtituting $r\cos\theta$ for x and $r\sin\theta$ for y. That is,

$$w = g(r, \theta) = f(x(r, \theta), y(r, \theta)).$$

Our question is as follows: Assuming all functions involved are differentiable, how are the partial derivatives $\partial w/\partial r$, $\partial w/\partial\theta$ related to $\partial w/\partial x$, $\partial w/\partial y$?

In the situation just described, we have $w = g(r, \theta) = (f \circ \mathbf{x})(r, \theta)$, so that the chain rule implies

$$Dg(r, \theta) = Df(x, y)D\mathbf{x}(r, \theta).$$

Therefore,

$$\begin{aligned} \begin{bmatrix} \dfrac{\partial g}{\partial r} & \dfrac{\partial g}{\partial \theta} \end{bmatrix} &= \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}. \end{aligned}$$

By extracting entries, we see that the various partial derivatives of w are related by the following formulas:

$$\begin{cases} \dfrac{\partial w}{\partial r} = \cos\theta \dfrac{\partial w}{\partial x} + \sin\theta \dfrac{\partial w}{\partial y} \\ \dfrac{\partial w}{\partial \theta} = -r\sin\theta \dfrac{\partial w}{\partial x} + r\cos\theta \dfrac{\partial w}{\partial y} \end{cases}. \tag{8}$$

The significance of (8) is that it provides us with a relation of **differential operators**:

$$\begin{cases} \dfrac{\partial}{\partial r} = \cos\theta\,\dfrac{\partial}{\partial x} + \sin\theta\,\dfrac{\partial}{\partial y} \\[2ex] \dfrac{\partial}{\partial \theta} = -r\sin\theta\,\dfrac{\partial}{\partial x} + r\cos\theta\,\dfrac{\partial}{\partial y} \end{cases}. \tag{9}$$

The appropriate interpretation for (9) is the following: Differentiation with respect to the polar coordinate r is the same as a certain combination of differentiation with respect to both Cartesian coordinates x and y (namely, the combination $\cos\theta\,\partial/\partial x + \sin\theta\,\partial/\partial y$). A similar comment applies to differentiation with respect to the polar coordinate θ. Note that, when $r \neq 0$, we can solve algebraically for $\partial/\partial x$ and $\partial/\partial y$ in (9), obtaining

$$\begin{cases} \dfrac{\partial}{\partial x} = \cos\theta\,\dfrac{\partial}{\partial r} - \dfrac{\sin\theta}{r}\,\dfrac{\partial}{\partial \theta} \\[2ex] \dfrac{\partial}{\partial y} = \sin\theta\,\dfrac{\partial}{\partial r} + \dfrac{\cos\theta}{r}\,\dfrac{\partial}{\partial \theta} \end{cases}. \tag{10}$$

We will have occasion to use the relations in (9) and (10), and the method of their derivation, later in this text. ◆

Addendum: Proof of Theorem 5.3

We begin by noting that the derivative matrices $D\mathbf{f}(\mathbf{x}_0)$ and $D\mathbf{x}(\mathbf{t}_0)$ both exist because $\mathbf{f}$ is assumed to be differentiable at $\mathbf{x}_0$ and $\mathbf{x}$ is assumed to be differentiable at $\mathbf{t}_0$. Thus, the product matrix $D\mathbf{f}(\mathbf{x}_0)D\mathbf{x}(\mathbf{t}_0)$ exists. We need to show that the limit in Definition 3.8 is satisfied by this product matrix, that is, that

$$\lim_{\mathbf{t}\to\mathbf{t}_0} \frac{\|(\mathbf{f}\circ\mathbf{x})(\mathbf{t}) - [(\mathbf{f}\circ\mathbf{x})(\mathbf{t}_0) + D\mathbf{f}(\mathbf{x}_0)D\mathbf{x}(\mathbf{t}_0)(\mathbf{t}-\mathbf{t}_0)]\,\|}{\|\mathbf{t}-\mathbf{t}_0\|} = 0. \tag{11}$$

In view of the uniqueness of the derivative matrix, it then automatically follows that $\mathbf{f}\circ\mathbf{x}$ is differentiable at $\mathbf{t}_0$ and that $D\mathbf{f}(\mathbf{x}_0)D\mathbf{x}(\mathbf{t}_0) = D(\mathbf{f}\circ\mathbf{x})(\mathbf{t}_0)$. Thus, we entirely concern ourselves with establishing the limit (11) above.

Consider the numerator of (11). First, we rewrite

$$\begin{aligned} &(\mathbf{f}\circ\mathbf{x})(\mathbf{t}) - [(\mathbf{f}\circ\mathbf{x})(\mathbf{t}_0) + D\mathbf{f}(\mathbf{x}_0)D\mathbf{x}(\mathbf{t}_0)(\mathbf{t}-\mathbf{t}_0)] \\ &\qquad = (\mathbf{f}\circ\mathbf{x})(\mathbf{t}) - (\mathbf{f}\circ\mathbf{x})(\mathbf{t}_0) - D\mathbf{f}(\mathbf{x}_0)(\mathbf{x}(\mathbf{t}) - \mathbf{x}(t_0)) \\ &\qquad\qquad + D\mathbf{f}(\mathbf{x}_0)(\mathbf{x}(\mathbf{t}) - \mathbf{x}(t_0)) - D\mathbf{f}(\mathbf{x}_0)D\mathbf{x}(\mathbf{t}_0)(\mathbf{t}-\mathbf{t}_0). \end{aligned}$$

Then we use the triangle inequality:

$$\begin{aligned} &\|(\mathbf{f}\circ\mathbf{x})(\mathbf{t}) - [(\mathbf{f}\circ\mathbf{x})(\mathbf{t}_0) + D\mathbf{f}(\mathbf{x}_0)D\mathbf{x}(\mathbf{t}_0)(\mathbf{t}-\mathbf{t}_0)]\,\| \\ &\qquad \le \|(\mathbf{f}\circ\mathbf{x})(\mathbf{t}) - (\mathbf{f}\circ\mathbf{x})(\mathbf{t}_0) - D\mathbf{f}(\mathbf{x}_0)(\mathbf{x}(\mathbf{t}) - \mathbf{x}(t_0))\| \\ &\qquad\qquad + \|D\mathbf{f}(\mathbf{x}_0)(\mathbf{x}(\mathbf{t}) - \mathbf{x}(t_0)) - D\mathbf{f}(\mathbf{x}_0)D\mathbf{x}(\mathbf{t}_0)(\mathbf{t}-\mathbf{t}_0)\| \\ &\qquad = \|(\mathbf{f}\circ\mathbf{x})(\mathbf{t}) - (\mathbf{f}\circ\mathbf{x})(\mathbf{t}_0) - D\mathbf{f}(\mathbf{x}_0)(\mathbf{x}(\mathbf{t}) - \mathbf{x}(t_0))\| \\ &\qquad\qquad + \|D\mathbf{f}(\mathbf{x}_0)\,[(\mathbf{x}(\mathbf{t}) - \mathbf{x}(t_0)) - D\mathbf{x}(\mathbf{t}_0)(\mathbf{t}-\mathbf{t}_0)]\,\|. \end{aligned}$$