12.6 Linear Approximations, Differentiability, and Differentials

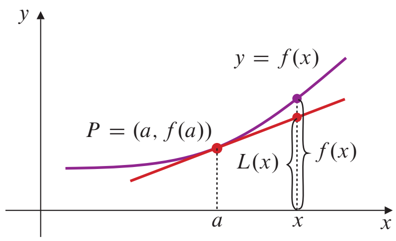

Figure 12.25 The linearization of f at $x = a$

As observed in Section 4.9, the tangent line to the graph $y = f(x)$ at $x = a$ provides a convenient approximation for values of $f(x)$ for x near a (see Figure 12.25):

$$f(x) \approx L(x) = f(a) + f'(a)(x - a).$$

Here, $L(x)$ is the **linearization** of f at a; its graph is the tangent line to $y = f(x)$ there. The mere existence of $f'(a)$ is sufficient to guarantee that the error in the approximation (the vertical distance between the curve and tangent at x) is small compared with the distance $h = x - a$ between a and x, that is,

$$\begin{aligned}\lim_{h\to 0} \frac{f(a+h) - L(a+h)}{h} &= \lim_{h\to 0} \frac{f(a+h) - f(a) - f'(a)h}{h} \\ &= \lim_{h\to 0} \frac{f(a+h) - f(a)}{h} - f'(a) \\ &= f'(a) - f'(a) = 0.\end{aligned}$$

Similarly, the tangent plane to the graph of $z = f(x, y)$ at (a, b) is $z = L(x, y)$, where

$$L(x, y) = f(a, b) + f_1(a, b)(x - a) + f_2(a, b)(y - b)$$

is the **linearization** of f at (a, b). We can use $L(x, y)$ to approximate values of $f(x, y)$ near (a, b):

$$f(x, y) \approx L(x, y) = f(a, b) + f_1(a, b)(x - a) + f_2(a, b)(y - b).$$

EXAMPLE 1 Find an approximate value for $f(x, y) = \sqrt{2x^2 + e^{2y}}$ at $(2.2, -0.2)$.

Solution It is convenient to use the linearization at $(2, 0)$, where the values of f and its partials are easily evaluated:

$$\begin{aligned} & & f(2,0) &= 3, \\ f_1(x,y) &= \frac{2x}{\sqrt{2x^2 + e^{2y}}}, & f_1(2,0) &= \frac{4}{3}, \\ f_2(x,y) &= \frac{e^{2y}}{\sqrt{2x^2 + e^{2y}}}, & f_2(2,0) &= \frac{1}{3}.\end{aligned}$$

Thus, $L(x, y) = 3 + \frac{4}{3}(x - 2) + \frac{1}{3}(y - 0)$, and

$$f(2.2, -0.2) \approx L(2.2, -0.2) = 3 + \frac{4}{3}(2.2 - 2) + \frac{1}{3}(-0.2 - 0) = 3.2\,.$$

(For the sake of comparison, $f(2.2, -0.2) \approx 3.2172$ to 4 decimal places.)

Unlike the single-variable case, the mere existence of the partial derivatives $f_1(a, b)$ and $f_2(a, b)$ does not even imply that f is continuous at (a, b), let alone that the error in the linearization is small compared with the distance $\sqrt{(x - a)^2 + (y - b)^2}$ between (a, b) and (x, y). We adopt this latter condition as our definition of what it means for a function of two variables to be *differentiable* at a point.

DEFINITION 5

We say that the function $f(x, y)$ is **differentiable** at the point (a, b) if

$$\lim_{(h,k)\to(0,0)} \frac{f(a+h, b+k) - f(a,b) - h\, f_1(a,b) - k f_2(a,b)}{\sqrt{h^2+k^2}} = 0.$$

This definition and the following theorems can be generalized to functions of any number of variables in the obvious way. For the sake of simplicity, we state them for the two-variable case only.

The function $f(x, y)$ is differentiable at the point (a, b) if and only if the surface $z = f(x, y)$ has a *nonvertical tangent plane* at (a, b). This implies that $f_1(a, b)$ and $f_2(a, b)$ must exist and that f must be continuous at (a, b). (Recall, however, that the existence of the partial derivatives does *not* even imply that f is continuous, let alone differentiable.) In particular, the function is *continuous* wherever it is differentiable. We will prove a two-variable version of the Mean-Value Theorem and use it to show that functions are differentiable wherever they have *continuous* first partial derivatives.

THEOREM 3

A Mean-Value Theorem

If $f_1(x, y)$ and $f_2(x, y)$ are continuous in a neighbourhood of the point (a, b), and if the absolute values of h and k are sufficiently small, then there exist numbers θ_1 and θ_2, each between 0 and 1, such that

$$f(a+h, b+k) - f(a,b) = hf_1(a+\theta_1 h, b+k) + kf_2(a, b+\theta_2 k).$$

PROOF The proof of this theorem is very similar to that of Theorem 1 in Section 12.4, so we give only a sketch here. The reader can fill in the details. Write

$$f(a+h, b+k) - f(a,b) = \big(f(a+h, b+k) - f(a, b+k)\big) + \big(f(a, b+k) - f(a,b)\big),$$

and then apply the single-variable Mean-Value Theorem separately to $f(x, b+k)$ on the interval between a and $a+h$, and to $f(a, y)$ on the interval between b and $b+k$ to get the desired result.

THEOREM 4

If f_1 and f_2 are continuous in a neighbourhood of the point (a, b), then f is differentiable at (a, b).

PROOF Using Theorem 3 and the facts that

$$\left|\frac{h}{\sqrt{h^2+k^2}}\right| \le 1 \qquad \text{and} \qquad \left|\frac{k}{\sqrt{h^2+k^2}}\right| \le 1,$$

we estimate

$$\begin{aligned}
&\left|\frac{f(a+h, b+k) - f(a,b) - hf_1(a,b) - kf_2(a,b)}{\sqrt{h^2+k^2}}\right| \\
&\quad = \left|\frac{h}{\sqrt{h^2+k^2}}\big(f_1(a+\theta_1 h, b+k) - f_1(a,b)\big) \right.\\
&\qquad\quad \left. + \frac{k}{\sqrt{h^2+k^2}}\big(f_2(a, b+\theta_2 k) - f_2(a,b)\big)\right| \\
&\quad \le \big|f_1(a+\theta_1 h, b+k) - f_1(a,b)\big| + \big|f_2(a, b+\theta_2 k) - f_2(a,b)\big|.
\end{aligned}$$

Since f_1 and f_2 are continuous at (a, b), each of these latter terms approaches 0 as h and k approach 0. This is what we needed to prove.

We illustrate differentiability with an example where we can calculate directly the error in the tangent plane approximation.

EXAMPLE 2 Calculate $f(x+h, y+k) - f(x, y) - f_1(x, y)h - f_2(x, y)k$ if $f(x, y) = x^3 + xy^2$.

Solution Since $f_1(x, y) = 3x^2 + y^2$ and $f_2(x, y) = 2xy$, we have

$$\begin{aligned} &f(x+h, y+k) - f(x, y) - f_1(x, y)h - f_2(x, y)k \\ &\qquad = (x+h)^3 + (x+h)(y+k)^2 - x^3 - xy^2 - (3x^2 + y^2)h - 2xyk \\ &\qquad = 3xh^2 + h^3 + 2yhk + hk^2 + xk^2. \end{aligned}$$

Observe that the result above is a polynomial in h and k with no term of degree less than 2 in these variables. Therefore, this difference approaches zero like the *square* of the distance $\sqrt{h^2 + k^2}$ from (x, y) to $(x+h, y+k)$ as $(h, k) \to (0, 0)$, so the condition for differentiability is certainly satisfied:

$$\lim_{(h,k)\to(0,0)} \frac{3xh^2 + h^3 + 2yhk + hk^2 + xk^2}{\sqrt{h^2 + k^2}} = 0.$$

This quadratic behaviour is the case for any function f with continuous *second* partial derivatives. (See Exercise 23 below.)

Proof of the Chain Rule

We are now able to give a formal statement and proof of a simple but representative case of the Chain Rule for multivariate functions.

THEOREM 5

A Chain Rule

Let $z = f(x, y)$, where $x = u(s, t)$ and $y = v(s, t)$. Suppose that

(i) $u(a, b) = p$ and $v(a, b) = q$,

(ii) the first partial derivatives of u and v exist at the point (a, b), and

(iii) f is differentiable at the point (p, q). Then $z = w(s, t) = f(u(s, t), v(s, t))$ has first partial derivatives with respect to s and t at (a, b), and

$$\begin{aligned} w_1(a, b) &= f_1(p, q)u_1(a, b) + f_2(p, q)v_1(a, b), \\ w_2(a, b) &= f_1(p, q)u_2(a, b) + f_2(p, q)v_2(a, b). \end{aligned}$$

That is,

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \quad \text{and} \quad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

PROOF Define a function E of two variables as follows: $E(0, 0) = 0$, and if $(h, k) \neq (0, 0)$, then

$$E(h, k) = \frac{f(p+h, q+k) - f(p, q) - hf_1(p, q) - kf_2(p, q)}{\sqrt{h^2 + k^2}}.$$

Observe that $E(h,k)$ is continuous at $(0,0)$ because f is differentiable at (p,q). Now,

$$f(p+h,q+k) - f(p,q) = hf_1(p,q) + kf_2(p,q) + \sqrt{h^2+k^2}\,E(h,k).$$

In this formula put $h = u(a+\sigma,b) - u(a,b)$ and $k = v(a+\sigma,b) - v(a,b)$ and divide by σ to obtain

$$\begin{aligned}\frac{w(a+\sigma,b) - w(a,b)}{\sigma} &= \frac{f(u(a+\sigma,b), v(a+\sigma,b)) - f(u(a,b), v(a,b))}{\sigma} \\ &= \frac{f(p+h,q+k) - f(p,q)}{\sigma} \\ &= f_1(p,q)\,\frac{h}{\sigma} + f_2(p,q)\,\frac{k}{\sigma} + \sqrt{\left(\frac{h}{\sigma}\right)^2 + \left(\frac{k}{\sigma}\right)^2}\,E(h,k).\end{aligned}$$

We want to let σ approach 0 in this formula. Note that

$$\lim_{\sigma\to 0}\frac{h}{\sigma} = \lim_{\sigma\to 0}\frac{u(a+\sigma,b) - u(a,b)}{\sigma} = u_1(a,b),$$

and, similarly, $\lim_{\sigma\to 0}(k/\sigma) = v_1(a,b)$. Since $(h,k) \to (0,0)$ if $\sigma \to 0$, we have

$$w_1(a,b) = f_1(p,q)u_1(a,b) + f_2(p,q)v_1(a,b).$$

The proof for w_2 is similar.

Differentials

If the first partial derivatives of a function $z = f(x_1,\ldots,x_n)$ exist at a point, we may construct a **differential** dz or df of the function at that point in a manner similar to that used for functions of one variable:

$$\begin{aligned}dz = df &= \frac{\partial z}{\partial x_1}\,dx_1 + \frac{\partial z}{\partial x_2}\,dx_2 + \cdots + \frac{\partial z}{\partial x_n}\,dx_n \\ &= f_1(x_1,\ldots,x_n)\,dx_1 + \cdots + f_n(x_1,\ldots,x_n)\,dx_n.\end{aligned}$$

Here, the differential dz is considered to be a function of the $2n$ independent variables $x_1, x_2, \ldots, x_n, dx_1, dx_2, \ldots, dx_n$.

For a *differentiable* function f, the differential df is an approximation to the change Δf in value of the function given by

$$\Delta f = f(x_1 + dx_1, \ldots, x_n + dx_n) - f(x_1,\ldots,x_n).$$

The error in this approximation is small compared with the distance between the two points in the domain of f; that is,

$$\frac{\Delta f - df}{\sqrt{(dx_1)^2 + \cdots + (dx_n)^2}} \to 0 \quad \text{if all } dx_i \to 0, \qquad (1 \le i \le n).$$

In this sense, differentials are just another way of looking at linearization.

EXAMPLE 3 Estimate the percentage change in the period $T = 2\pi\sqrt{\dfrac{L}{g}}$ of a simple pendulum if the length, L, of the pendulum increases by 2% and the acceleration of gravity, g, decreases by 0.6%.

Solution We calculate the differential of T:

$$\begin{aligned} dT &= \frac{\partial T}{\partial L}\,dL + \frac{\partial T}{\partial g}\,dg \\ &= \frac{2\pi}{2\sqrt{Lg}}\,dL - \frac{2\pi\sqrt{L}}{2g^{3/2}}\,dg. \end{aligned}$$

We are given that $dL = \frac{2}{100}L$ and $dg = -\frac{6}{1{,}000}g$. Thus,

$$dT = \frac{1}{100}2\pi\sqrt{\frac{L}{g}} - \left(-\frac{6}{1{,}000}\right)\frac{2\pi}{2}\sqrt{\frac{L}{g}} = \frac{13}{1{,}000}T.$$

Therefore, the period T of the pendulum increases by 1.3%.

Functions from n-Space to m-Space

(*This is an optional topic.*) A vector $\mathbf{f} = (f_1, f_2, \ldots, f_m)$ of m functions, each depending on n variables $(x_1, x_2, \ldots, x_n)$, defines a *transformation* (i.e., a function) from $\mathbb{R}^n$ to $\mathbb{R}^m$; specifically, if $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a point in $\mathbb{R}^n$, and

$$\begin{aligned} y_1 &= f_1(x_1, x_2, \ldots, x_n) \\ y_2 &= f_2(x_1, x_2, \ldots, x_n) \\ &\ \vdots \\ y_m &= f_m(x_1, x_2, \ldots, x_n), \end{aligned}$$

then $\mathbf{y} = (y_1, y_2, \ldots, y_m)$ is the point in $\mathbb{R}^m$ that corresponds to $\mathbf{x}$ under the transformation $\mathbf{f}$. We can write these equations more compactly as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}).$$

Information about the rate of change of $\mathbf{y}$ with respect to $\mathbf{x}$ is contained in the various partial derivatives $\partial y_i/\partial x_j$, $(1 \le i \le m,\ 1 \le j \le n)$, and is conveniently organized into an $m \times n$ matrix, $D\mathbf{f}(\mathbf{x})$, called the **Jacobian matrix** of the transformation $\mathbf{f}$:

$$D\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

If the partial derivatives in the Jacobian matrix are continuous, we say that $\mathbf{f}$ is **differentiable** at $\mathbf{x}$. In this case the linear transformation (see Section 10.7) represented by the Jacobian matrix is called **the derivative** of the transformation $\mathbf{f}$.

Remark We can regard the scalar-valued function of two variables, $f(x, y)$ say, as a transformation from $\mathbb{R}^2$ to $\mathbb{R}$. Its derivative is then the linear transformation with matrix

$$Df(x, y) = \big(f_1(x, y),\ f_2(x, y)\big).$$

It is not our purpose to enter into a study of such *vector-valued functions of a vector variable* at this point, but we can observe here that the Jacobian matrix of the composition of two such transformations is the matrix product of their Jacobian matrices.

To see this, let $\mathbf{y} = \mathbf{f}(\mathbf{x})$ be a transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ as described above, and let $\mathbf{z} = \mathbf{g}(\mathbf{y})$ be another such transformation from $\mathbb{R}^m$ to $\mathbb{R}^k$ given by

$$\begin{aligned} z_1 &= g_1(y_1, y_2, \ldots, y_m) \\ z_2 &= g_2(y_1, y_2, \ldots, y_m) \\ &\vdots \\ z_k &= g_k(y_1, y_2, \ldots, y_m), \end{aligned}$$

which has the $k \times m$ Jacobian matrix

$$D\mathbf{g}(\mathbf{y}) = \begin{pmatrix} \dfrac{\partial z_1}{\partial y_1} & \dfrac{\partial z_1}{\partial y_2} & \cdots & \dfrac{\partial z_1}{\partial y_m} \\ \dfrac{\partial z_2}{\partial y_1} & \dfrac{\partial z_2}{\partial y_2} & \cdots & \dfrac{\partial z_2}{\partial y_m} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial z_k}{\partial y_1} & \dfrac{\partial z_k}{\partial y_2} & \cdots & \dfrac{\partial z_k}{\partial y_m} \end{pmatrix}.$$

Then the composition $\mathbf{z} = \mathbf{g} \circ \mathbf{f}(\mathbf{x}) = \mathbf{g}\big(\mathbf{f}(\mathbf{x})\big)$ given by

$$\begin{aligned} z_1 &= g_1\big(f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)\big) \\ z_2 &= g_2\big(f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)\big) \\ &\vdots \\ z_k &= g_k\big(f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)\big) \end{aligned}$$

has, according to the Chain Rule, the $k \times n$ Jacobian matrix

$$\begin{pmatrix} \dfrac{\partial z_1}{\partial x_1} & \dfrac{\partial z_1}{\partial x_2} & \cdots & \dfrac{\partial z_1}{\partial x_n} \\ \dfrac{\partial z_2}{\partial x_1} & \dfrac{\partial z_2}{\partial x_2} & \cdots & \dfrac{\partial z_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial z_k}{\partial x_1} & \dfrac{\partial z_k}{\partial x_2} & \cdots & \dfrac{\partial z_k}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial z_1}{\partial y_1} & \dfrac{\partial z_1}{\partial y_2} & \cdots & \dfrac{\partial z_1}{\partial y_m} \\ \dfrac{\partial z_2}{\partial y_1} & \dfrac{\partial z_2}{\partial y_2} & \cdots & \dfrac{\partial z_2}{\partial y_m} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial z_k}{\partial y_1} & \dfrac{\partial z_k}{\partial y_2} & \cdots & \dfrac{\partial z_k}{\partial y_m} \end{pmatrix} \begin{pmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \cdots & \dfrac{\partial y_1}{\partial x_n} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \cdots & \dfrac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial y_m}{\partial x_1} & \dfrac{\partial y_m}{\partial x_2} & \cdots & \dfrac{\partial y_m}{\partial x_n} \end{pmatrix}$$

This is, in fact, the Chain Rule for compositions of transformations:

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = D\mathbf{g}\big(\mathbf{f}(\mathbf{x})\big)D\mathbf{f}(\mathbf{x}),$$

and exactly mimics the one-variable Chain Rule $D(g \circ f)(x) = Dg\big(f(x)\big)Df(x)$.

The transformation $\mathbf{y} = \mathbf{f}(\mathbf{x})$ also defines a vector $d\mathbf{y}$ of differentials of the variables y_i in terms of the vector $d\mathbf{x}$ of differentials of the variables x_j. Writing $d\mathbf{y}$ and $d\mathbf{x}$ as column vectors we have

$$d\mathbf{y} = \begin{pmatrix} dy_1 \\ dy_2 \\ \vdots \\ dy_m \end{pmatrix} = \begin{pmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \cdots & \dfrac{\partial y_1}{\partial x_n} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \cdots & \dfrac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial y_m}{\partial x_1} & \dfrac{\partial y_m}{\partial x_2} & \cdots & \dfrac{\partial y_m}{\partial x_n} \end{pmatrix} \begin{pmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{pmatrix} = D\mathbf{f}(\mathbf{x})\,d\mathbf{x}.$$

EXAMPLE 4 Find the Jacobian matrix $D\mathbf{f}(1,0)$ for the transformation from $\mathbb{R}^2$ to $\mathbb{R}^3$ given by

$$\mathbf{f}(x,y) = \big(xe^y + \cos(\pi y), x^2, x - e^y\big)$$

and use it to find an approximate value for $\mathbf{f}(1.02, 0.01)$.

Solution $D\mathbf{f}(x,y)$ is the 3×2 matrix whose jth row consists of the partial derivatives of the jth component of $\mathbf{f}$ with respect to x and y. Thus,

$$D\mathbf{f}(1,0) = \begin{pmatrix} e^y & xe^y - \pi\sin(\pi y) \\ 2x & 0 \\ 1 & -e^y \end{pmatrix}\Bigg|_{(1,0)} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 1 & -1 \end{pmatrix}.$$

Since $\mathbf{f}(1,0) = (2,1,0)$ and $d\mathbf{x} = \begin{pmatrix} 0.02 \\ 0.01 \end{pmatrix}$, we have

$$d\mathbf{f} = D\mathbf{f}(1,0)\,d\mathbf{x} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 0.02 \\ 0.01 \end{pmatrix} = \begin{pmatrix} 0.03 \\ 0.04 \\ 0.01 \end{pmatrix}.$$

Therefore, $\mathbf{f}(1.02, 0.01) \approx (2.03, 1.04, 0.01)$.

For transformations between spaces of the same dimension (say from $\mathbb{R}^n$ to $\mathbb{R}^n$), the corresponding Jacobian matrices are square and have determinants. These Jacobian determinants will play an important role in our consideration of implicit functions and inverse functions in Section 12.8 and in changes of variables in multiple integrals in Chapter 14.

Maple's **VectorCalculus** package has a function **Jacobian** that takes two inputs, a list (or vector) of expressions and a list of variables, and produces the Jacobian matrix of the partial derivatives of those expressions with respect to the variables. For example,

```
> with(VectorCalculus):
> Jacobian([x*y*exp(z), (x+2*y)*cos(z)],[x,y,z]);
```

$$\begin{bmatrix} ye^z & xe^z & xye^z \\ \cos(z) & 2\cos(z) & -(x+2y)\sin(z) \end{bmatrix}$$

VectorCalculus has only been included since Maple 8. If you have an earlier release, use `linalg` instead, and the function `jacobian`.

Differentials in Applications

Differentials are sometimes used as an alternative representation for differentiable functions. This is particularly so in the field of thermodynamics. In thermodynamics, physical states of thermodynamic equilibrium are expressed mathematically in terms of the existence of a function,

$$E = E(S, V, N_1, \ldots, N_n),$$

where E is internal energy, S is entropy, V is volume, and the N_i are numbers of atoms or molecules of type i.

These quantities are interpreted physically, but they are just independent variables in a function to which normal mathematical rules apply. Discussion of the physical meaning of a quantity like entropy, for example, is largely beyond the scope of this book. (One might remark that entropy is a logarithmic measure of the number of underlying physical states that appear indistinguishable on human scales, but such a description is completely unnecessary for this discussion.) $E(S, V, N_1, \ldots, N_n)$ is known as a *function of state*. Any explicit equation relating thermodynamic variables is also known as an *equation of state*.

Thermodynamics allows for any number of such variables to define the state. There can be others than those indicated for different physical systems. All such variables are additive in that, for example, the energy of two physical systems together is simply the sum of the energies of each system. The same is true for volume, entropy, and number. These additive variables are called *extensive* variables. In thermodynamics they are referred to as *state variables* or as *state functions*. That is because any one of the other variables can be expressed as a function of E and the remaining variables. For example, $S = S(E, V, N_1, \dots, N_n)$.

Differentials appear in thermodynamics as the normal way to express the existence of a state function. In writing

$$dE = \frac{\partial E}{\partial S}\,dS + \frac{\partial E}{\partial V}\,dV + \frac{\partial E}{\partial N_1}\,dN_1 + \cdots + \frac{\partial E}{\partial N_n}\,dN_n,$$

we are saying that E depends on the variables whose differentials appear on the right side of the equation. In fact, everything is so effectively done with differentials that often no explicit function E is needed or even known.

Historically, the differential was also meant to convey an intuitive sense of change in time, even though mathematically it is simply the differential of a function. In fact, this historical interpretation can be quite confusing, because, paradoxically, the existence of the function of state, and its differential, means the physical system is in *thermodynamic equilibrium*, which can be described as a time-independent condition of a physical system. If it were not in (timeless) thermodynamic equilibrium, there would be no state function and no corresponding differentials. The resolution of the paradox is to stick to the mathematics, remembering that the differential only depicts a change in the values of variables and not any external process.

So, for example, the state equation has nothing to do with whether some process is slow or not. Differentials in this case do not suggest a physical process any more than the differential of any other function does. The differential only expresses the content of the function, so it has nothing to do with the physical processes that cause changes, or with whether any change is carried out slowly (reversible processes) or not.

The partial derivatives that appear in the differential form of the state equation also have explicit physical interpretations: $\frac{\partial E}{\partial S}$ is *temperature* T, $-\frac{\partial E}{\partial V}$ is *pressure* P, and the quantities $\frac{\partial E}{\partial N_i}$ are known as *chemical potentials*, μ_i. These partial derivatives represent slopes on the graph of the function of state, and as such they are not additive. It makes no sense, for example, to add temperatures. Physically, these slopes define a condition rather than an amount. These nonadditive quantities are called *intensive* variables.

With these definitions substituted, the differential form of the equation of state becomes

$$dE = T\,dS - P\,dV + \mu_1\,dN_1 + \cdots + \mu_n\,dN_n,$$

which is known as the *Gibbs equation*. However, despite the special treatment, this expression remains simply the differential of $E(S, V, N_1, \dots, N_n)$. The Gibbs equation is a fundamental starting point in many thermodynamical problems.

Another related, and well-known, equation of differentials is the Gibbs-Duhem equation,

$$0 = S\,dT - V\,dP + N_1\,d\mu_1 + \cdots + N_n\,d\mu_n.$$

This remarkable equation indicates that the intensive variables of thermodynamics are not independent of each other. It holds because the additivity of the extensive variables implies that the function of state, $E = E(S, V, N_1, \dots, N_n)$, is homogeneous of degree 1. (See Exercise 24 at the end of this section.)

Differentials and Legendre Transformations

It is often useful to shift the dependence of a function on one or more of its independent variables to dependence on, instead, the derivatives of the function with respect to these variables. Consider, for example, the function $y = f(x)$, and denote its derivative by p; that is, $p = f'(x)$. If we let $u = px - f(x)$ and calculate the differential of u, treating x and p as independent variables, we obtain

$$du = p\,dx + x\,dp - f'(x)\,dx = p\,dx + x\,dp - p\,dx = x\,dp.$$

Since there is no dx term remaining in this differential, u does not depend explicitly on x, but only on p. Let us therefore define $f^*(p) = u = px - f(x)$. $f^*(p)$ is called the Legendre transformation of $f(x)$ with respect to x, and the two variables x and p are said to be **conjugate** to one another. Observe that

$$f(x) + f^*(p) = px,$$

and the symmetry of this equation indicates that f must also be the Legendre transformation of f^*; $f^{**} = f$. In fact, taking the partial derivatives of the equation with respect to x and p we obtain the symmetric relationships

$$f'(x) = p \quad \text{and} \quad (f^*)'(p) = x$$

from which it is apparent that f' and $(f^*)'$ are inverse functions;

$$f'\big((f^*)'(p)\big) = p, \qquad (f^*)'\big(f'(x)\big) = x.$$

Remark The above definition of f^* clearly shows the symmetry in its relationship with f. An alternative transformation, $-f^*(p)$ (i.e., the function $f(x) - px$) shifts dependence between a variable and the derivative of the function just as effectively, although it does not share this symmetry. In some fields, particularly thermodynamics, this alternative is known as the Legendre transformation instead.

EXAMPLE 5 Calculate the Legendre transformation $f^*(p)$ of the function $f(x) = e^x$.

Solution Here $p = f'(x) = e^x$, so $x = \ln p$. Therefore,

$$f^*(p) = px - f(x) = p \ln p - p.$$

For functions of several variables, Legendre transformations can be taken with respect to one or more of the independent variables. If $u = f(x, y)$, $p = f_1(x, y)$, and $q = f_2(x, y)$, and if $w = px + qy - u$, then

$$dw = p\,dx + x\,dp + q\,dy + y\,dq - f_1(x, y)\,dx - f_2(x, y)\,dy = x\,dp + y\,dq$$

and w does not depend explicitly on x or y, but only on p and q. We can call $w(p, q)$ (or $-w(p, q)$ if we are doing thermodynamics) the Legendre transformation of $f(x, y)$ with respect to x and y, and treat both $\{x, p\}$ and $\{y, q\}$ as conjugate pairs of variables. Observe that

$$\begin{aligned} f_1(x, y) &= p \\ f_2(x, y) &= q \end{aligned} \quad \text{and} \quad \begin{aligned} w_1(p, q) &= x \\ w_2(p, q) &= y. \end{aligned}$$

Returning to thermodynamics, the Gibbs equation tells us that E depends on S, V, and N_i. Since $T = \dfrac{\partial E}{\partial S}$, T and S are conjugate and we can express energy in terms of temperature rather than entropy by using an (alternative) Legendre transformation. Let $F = E - TS$. Then

$$dF = dE - S\,dT - T\,dS = -S\,dT - P\,dV + \mu_1\,dN_1 + \cdots + \mu_n\,dN_n.$$

Thus, $F = F(T, V, N_1, \ldots, N_n)$. F is known as the *Helmholtz free energy*, which is called a *thermodynamic potential*. It can be more practical to use F, which depends explicitly on T, rather than E when an experiment is run at constant temperature.

Legendre transformations can be done in terms of any or all of the conjugate pairs. In the case of the Helmholtz free energy, only the conjugates T and S are used. Other specific Legendre transformations lead to other thermodynamic *potentials*. For example, the *Gibbs free energy*, $G = E - TS + PV$, is widely used in chemistry, where processes normally take place at constant temperature and pressure. (See Exercise 30 below.)

Legendre transformations are very important in other areas of classical and modern physics. Historically, they appear in classical mechanics, where the functional expression of the energy, known as the Hamiltonian, is expressed in terms of Legendre transformations of a function known as the Lagrangian. (See Exercise 32 for a problem developing this relationship.) These notions extend to modern physics, which is often cast in terms of Lagrangians.

EXERCISES 12.6

In Exercises 1–6, use suitable linearizations to find approximate values for the given functions at the points indicated.

1. $f(x, y) = x^2 y^3$ at $(3.1, 0.9)$

2. $f(x, y) = \tan^{-1}\left(\dfrac{y}{x}\right)$ at $(3.01, 2.99)$

3. $f(x, y) = \sin(\pi xy + \ln y)$ at $(0.01, 1.05)$

4. $f(x, y) = \dfrac{24}{x^2 + xy + y^2}$ at $(2.1, 1.8)$

5. $f(x, y, z) = \sqrt{x + 2y + 3z}$ at $(1.9, 1.8, 1.1)$

6. $f(x, y) = x\, e^{y+x^2}$ at $(2.05, -3.92)$

In Exercises 7–10, write the differential of the given function and use it to estimate the value of the function at the given point by starting with a known value at a nearby point.

7. $z = x^2 e^{3y}$, at $x = 3.05, y = -0.02$

8. $g(s, t) = s^2/t$, $g(2.1, 1.9)$

9. $F(x, y, z) = \sqrt{x^2 + y + 2 + z^2}$, $F(0.7, 2.6, 1.7)$

10. $u = x \sin(x + y)$, at $x = \dfrac{\pi}{2} + \dfrac{1}{20}$, $y = \dfrac{\pi}{2} - \dfrac{1}{30}$

11. The edges of a rectangular box are each measured to within an accuracy of 1% of their values. What is the approximate maximum percentage error in

(a) the calculated volume of the box,

(b) the calculated area of one of the faces of the box, and

(c) the calculated length of a diagonal of the box?

12. The radius and height of a right-circular conical tank are measured to be 25 ft and 21 ft, respectively. Each measurement is accurate to within 0.5 in. By about how much can the calculated volume of the tank be in error?

13. By approximately how much can the calculated area of the conical surface of the tank in Exercise 12 be in error?

14. Two sides and the contained angle of a triangular plot of land are measured to be 224 m, 158 m, and 64°, respectively. The length measurements were accurate to within 0.4 m and the angle measurement to within 2°. What is the approximate maximum percentage error if the area of the plot is calculated from these measurements?

15. The angle of elevation of the top of a tower is measured at two points A and B on the ground in the same direction from the base of the tower. The angles are 50° at A and 35° at B, each measured to within 1°. The distance AB is measured to be 100 m with error at most 0.1%. What is the calculated height of the building, and by about how much can it be in error? To which of the three measurements is the calculated height most sensitive?

16. By approximately what percentage will the value of $w = \dfrac{x^2 y^3}{z^4}$ increase or decrease if x increases by 1%, y increases by 2%, and z increases by 3%?

17. Find the Jacobian matrix for the transformation $\mathbf{f}(r, \theta) = (x, y)$, where

$$x = r\cos\theta \quad \text{and} \quad y = r\sin\theta.$$

(Although (r, θ) can be regarded as *polar coordinates* in the xy-plane, they are Cartesian coordinates in their own $r\theta$-plane.)

18. Find the Jacobian matrix for the transformation $\mathbf{f}(R, \phi, \theta) = (x, y, z)$, where

$$x = R\sin\phi\cos\theta, \quad y = R\sin\phi\sin\theta, \quad z = R\cos\phi.$$

Here, (R, ϕ, θ) are *spherical coordinates* in xyz-space, as introduced in Section 10.6.

19. Find the Jacobian matrix $D\mathbf{f}(x, y, z)$ for the transformation of $\mathbb{R}^3$ to $\mathbb{R}^2$ given by

$$\mathbf{f}(x, y, z) = (x^2 + yz, y^2 - x\ln z).$$

Use $D\mathbf{f}(2, 2, 1)$ to help you find an approximate value for $\mathbf{f}(1.98, 2.01, 1.03)$.

20. Find the Jacobian matrix $D\mathbf{g}(1, 3, 3)$ for the transformation of $\mathbb{R}^3$ to $\mathbb{R}^3$ given by

$$\mathbf{g}(r, s, t) = (r^2 s, r^2 t, s^2 - t^2)$$

and use the result to find an approximate value for $\mathbf{g}(0.99, 3.02, 2.97)$.

21. Prove that if $f(x, y)$ is differentiable at (a, b), then $f(x, y)$ is continuous at (a, b).

22. Prove the following version of the Mean-Value Theorem: If $f(x, y)$ has first partial derivatives continuous near every point of the straight line segment joining the points (a, b) and $(a+h, b+k)$, then there exists a number θ satisfying $0 < \theta < 1$ such that

$$\begin{aligned} f(a+h, b+k) = &f(a, b) + hf_1(a+\theta h, b+\theta k) \\ &+ kf_2(a+\theta h, b+\theta k). \end{aligned}$$

(*Hint:* Apply the single-variable Mean-Value Theorem to $g(t) = f(a+th, b+tk)$.) Why could we not have used this result in place of Theorem 3 to prove Theorem 4 and hence the version of the Chain Rule given in this section?

23. Generalize Exercise 22 as follows: show that, if $f(x, y)$ has continuous partial derivatives of second order near the point (a, b), then there exists a number θ satisfying $0 < \theta < 1$ such that, for h and k sufficiently small in absolute value,

$$\begin{aligned} f(a+h, b+k) = &f(a, b) + hf_1(a, b) + kf_2(a, b) \\ &+ h^2 f_{11}(a+\theta h, b+\theta k) \\ &+ 2hkf_{12}(a+\theta h, b+\theta k) \\ &+ k^2 f_{22}(a+\theta h, b+\theta k). \end{aligned}$$

Hence, show that there is a constant K such that for all values of h and k that are sufficiently small in absolute value,

$$\begin{aligned} &\left| f(a+h, b+k) - f(a, b) - hf_1(a, b) - kf_2(a, b) \right| \\ &\qquad \le K(h^2 + k^2). \end{aligned}$$

Thermodynamics and Legendre Transformations

24. Use the Gibbs equation

$$dE = T\,dS - P\,dV + \mu_1\,dN_1 + \cdots + \mu_n\,dN_n$$

and the fact that, being additive in its extensive variables, $E = E(S, V, N_1, \ldots, N_n)$ is necessarily homogeneous of degree 1, to establish the Gibbs-Duhem equation

$$0 = S\,dT - V\,dP + N_1\,d\mu_1 + \cdots + N_n\,d\mu_n.$$

(*Hint:* Use Euler's Theorem, Theorem 2 of Section 12.5.)

25. The equation of state for an ideal gas in the form of $E = E(S, V, N)$, using extensive variables only, is rarely quoted. It is

$$E = \frac{3h^2 N}{4\pi m}\left(\frac{N}{V}\right)^{2/3} e^{\left(\frac{2S}{3Nk} - \frac{5}{3}\right)}.$$

However, it is common to see $PV = NkT$, or $E = \frac{3}{2}NkT$ instead. Here k is the Boltzmann constant, h is Planck's constant, and m is the mass of one atom. Deduce these common forms from the explicit formula for E given as a function of S, V, and N.

26. If $f''(x) > 0$ for all x, show that the Legendre transformation $f^*(p)$ is the maximum value of the function $g(x) = px - f(x)$ considered as a function of x alone with p fixed.

In Exercises 27–29 give an explicit formula for the Legendre transformation $f^*(p)$ of the given function $f(x)$.

27. $f(x) = x^2$

28. $f(x) = x^4$

29. $f(x) = \ln(2 + 3x)$

30. Use differentials to show that the Gibbs free energy, $G = E - TS + PV$, depends on T and P alone when the numbers of molecules of each type are fixed. Determine the partial derivatives of G with respect to the new variables T and P.

31. Entropy can be written as a function, $S = S(E, V, N_1, \cdots, N_n)$. Legendre transformations can be performed on it too, although they are not so well-known. The resulting functions are called *Massieu-Planck functions*. Show that one of these, the Massieu's potential, $\Phi = S - \frac{1}{T}E$, depends on temperature instead of energy.

32. In classical mechanics, the energy of a system is expressed in terms of a function called the *Hamiltonian*. When the energy is independent of time, the Hamiltonian depends only on the positions, q_i, and the momenta, p_i, of the particles in the system, that is, $H = H(q_1, \cdots, q_n, p_1, \cdots, p_n)$. There is also another function, called the *Lagrangian*, that depends on the positions q_i and the velocities $\dot{q}_i$, that is, $L = L(q_1, \cdots, q_n, \dot{q}_1, \cdots, \dot{q}_n)$, such that the Hamiltonian is a Legendre transformation of the Lagrangian with respect to the velocity variables:

$$\begin{aligned} &H(q_1, \cdots, q_n, p_1, \cdots, p_n) \\ &\quad = \sum_i p_i \dot{q}_i - L(q_1, \cdots, q_n, \dot{q}_1, \cdots, \dot{q}_n). \end{aligned}$$

(a) What variables are conjugate in this Legendre transformation? What partial derivatives of L are implicitly determined by it?

(b) In the absence of external forces, the principle of least action requires that $\dfrac{\partial L}{\partial q_i} = \dot{p}_i$. By taking the differential of H and using the result of part (a), show that $\dfrac{\partial H}{\partial q_i} = -\dot{p}_i$ and $\dfrac{\partial H}{\partial p_i} = \dot{q}_i$. These are known as Hamilton's equations.

(c) Use Hamilton's equations to show that the Hamiltonian, $\frac{1}{2}(q^2 + p^2)$, represents a harmonic oscillator because it is equivalent to the differential equation $\ddot{q} + q = 0$.

27. If the equations $x = f(u, v)$, $y = g(u, v)$ can be solved for u and v in terms of x and y, show that

$$\frac{\partial(u, v)}{\partial(x, y)} = 1 \Bigg/ \frac{\partial(x, y)}{\partial(u, v)}.$$

Hint: Use the result of Exercise 26.

28. If $x = f(u, v)$, $y = g(u, v)$, $u = h(r, s)$, and $v = k(r, s)$, then x and y can be expressed as functions of r and s. Verify by direct calculation that

$$\frac{\partial(x, y)}{\partial(r, s)} = \frac{\partial(x, y)}{\partial(u, v)} \frac{\partial(u, v)}{\partial(r, s)}.$$

This is a special case of the Chain Rule for Jacobians.

29. Two functions, $f(x, y)$ and $g(x, y)$, are said to be functionally dependent if one is a function of the other; that is, if there exists a single-variable function $k(t)$ such that $f(x, y) = k\Big(g(x, y)\Big)$ for all x and y. Show that in this case $\partial(f, g)/\partial(x, y)$ vanishes identically. Assume that all necessary derivatives exist.

30. Prove the converse of Exercise 29 as follows: Let $u = f(x, y)$ and $v = g(x, y)$, and suppose that $\partial(u, v)/\partial(x, y) = \partial(f, g)/\partial(x, y)$ is identically zero for all x and y. Show that $(\partial u/\partial x)_v$ is identically zero. Hence u, considered as a function of x and v, is independent of x; that is, $u = k(v)$ for some function k of one variable. Why does this imply that f and g are functionally dependent?

Thermodynamics Problems

31. Use the different versions of the equation of state, presented in this section, to determine explicit functions u and v such that $S = u(E, V, N)$ and $S = v(T, V, N)$.

In Exercises 32–34, verify the given Maxwell relation by using a suitable Legendre transformation (see the Thermodynamics subsection of Section 12.6) to involve the appropriate set of independent variables.

32. $\left(\dfrac{\partial P}{\partial T}\right)_{V,N} = \left(\dfrac{\partial S}{\partial V}\right)_{T,N}$

33. $\left(\dfrac{\partial V}{\partial S}\right)_{P,N} = \left(\dfrac{\partial T}{\partial P}\right)_{S,N}$

34. $\left(\dfrac{\partial S}{\partial P}\right)_{T,N} = -\left(\dfrac{\partial V}{\partial T}\right)_{P,N}$

12.9 Taylor's Formula, Taylor Series, and Approximations

As is the case for functions of one variable, power series representations and their partial sums (Taylor polynomials) can provide an efficient method for determining the behaviour of a smooth function of several variables near a point in its domain. In this section we will look briefly at the extension of Taylor's formula and Taylor series to such functions. We will do this for functions of n variables as it is no more difficult to do this than to treat the special case $n = 2$.

As a starting point, recall Taylor's formula for a function $F(t)$ with continuous derivatives of order up to $m + 1$ on the interval $[0, 1]$. (See Theorem 12 in Section 4.10, and put $f = F$, $a = 0$, $x = h = 1$, and $s = \theta$ in the version of Taylor's formula given there.)

$$F(1) = F(0) + F'(0) + \frac{F''(0)}{2!} + \cdots + \frac{F^{(m)}(0)}{m!} + \frac{F^{(m+1)}(\theta)}{(m+1)!},$$

where θ is some number between 0 and 1. (The last term in the formula is the *Lagrange* form of the remainder.)

To simplify the manipulation of many variables, irrespective of how many there are, it is convenient to introduce the idea of a function of a vector, which is an intuitively straightforward extension from functions of scalars. If $\mathbf{x}$ has components $(x_1, x_2, \ldots, x_n)$, then $f(\mathbf{x})$ just means $f(x_1, x_2, \ldots, x_n)$, a function of n variables.

Now suppose that $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{h} = (h_1, h_2, \ldots, h_n)$ belong to $\mathbb{R}^n$. If f is a function of $\mathbf{x} \in \mathbb{R}^n$ that has continuous partial derivatives of orders up to $m + 1$ in an open set containing the line segment joining $\mathbf{a}$ and $\mathbf{a} + \mathbf{h}$, we can apply the above formula to

$$F(t) = f(\mathbf{a} + t\mathbf{h}), \qquad (0 \le t \le 1).$$

By the Chain Rule we will have

$$\begin{aligned} F'(t) &= h_1 f_{h_1}(\mathbf{a} + t\mathbf{h}) + h_2 f_{h_2}(\mathbf{a} + t\mathbf{h}) + \cdots + h_n f_{h_n}(\mathbf{a} + t\mathbf{h}) \\ &= (\mathbf{h} \bullet \nabla) f(\mathbf{a} + t\mathbf{h}), \end{aligned}$$

where

$$(\mathbf{h} \bullet \nabla) f(\mathbf{a} + t\mathbf{h}) = \Big((h_1 D_1 + h_2 D_2 + \cdots + h_n D_n) f(\mathbf{x})\Big)\Big|_{\mathbf{x}=\mathbf{a}+t\mathbf{h}}$$

and $D_j = \partial/\partial x_j$, $(1 \le p \le n)$. Similarly,

$$\begin{aligned} F''(t) &= h_1 h_1 f_{11}(\mathbf{a} + t\mathbf{h}) + h_1 h_2 f_{12}(\mathbf{a} + t\mathbf{h}) + \cdots + h_n h_n f_{nn}(\mathbf{a} + t\mathbf{h}) \\ &= \big(\mathbf{h} \bullet \nabla\big)^2 f(\mathbf{a} + t\mathbf{h}) \\ &\vdots \\ F^{(j)}(t) &= \big(\mathbf{h} \bullet \nabla\big)^j f(\mathbf{a} + t\mathbf{h}) \end{aligned}$$

Thus, $F(1) = f(\mathbf{a} + \mathbf{h})$, $F(0) = f(\mathbf{a})$, and $F^{(j)}(0) = (\mathbf{h} \bullet \nabla)^j f(\mathbf{a})$. The Taylor formula given above thus says that

$$\begin{aligned} f(\mathbf{a} + \mathbf{h}) &= f(\mathbf{a}) + \mathbf{h} \bullet \nabla f(\mathbf{a}) + \frac{(\mathbf{h} \bullet \nabla)^2 f(\mathbf{a})}{2!} + \cdots + \frac{(\mathbf{h} \bullet \nabla)^m f(\mathbf{a})}{m!} \\ &\quad + \frac{(\mathbf{h} \bullet \nabla)^{m+1} f(\mathbf{a} + \theta\mathbf{h})}{(m+1)!} \\ &= \sum_{j=0}^{m} \frac{(\mathbf{h} \bullet \nabla)^j f(\mathbf{a})}{j!} + \frac{(\mathbf{h} \bullet \nabla)^{m+1} f(\mathbf{a} + \theta\mathbf{h})}{(m+1)!} \\ &= P_m(\mathbf{h}) + R_m(\mathbf{h}, \theta). \end{aligned}$$

This is Taylor's formula for f about $\mathbf{x} = \mathbf{a}$. $P_m(\mathbf{h})$ is a polynomial of degree m in the components of $\mathbf{h}$. $P_m(\mathbf{h})$ is called the mth degree Taylor polynomial of f about $\mathbf{x} = \mathbf{a}$. The term corresponding to j in the summation defining P_m is, if not zero, a polynomial of degree exactly j in the components of $\mathbf{h}$, whose coefficients are jth order partial derivatives of f evaluated at $\mathbf{x} = \mathbf{a}$. The remainder term $R_m(\mathbf{h}, \theta)$ is also a polynomial in the components of $\mathbf{h}$, each of whose terms if not zero has degree exactly $m + 1$, but its coefficients are $(m + 1)$st order partial derivatives of f evaluated at an indeterminate point $\mathbf{a} + \theta\mathbf{h}$ along the line segment between $\mathbf{a}$ and $\mathbf{a} + \mathbf{h}$.

Sometimes it is useful to replace the explicit remainder in Taylor's formula with a Big-O term that is bounded by a multiple of $|\mathbf{h}|^{m+1}$ as $|\mathbf{h}| \to 0$. (See Section 4.10.)

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \mathbf{h} \bullet \nabla f(\mathbf{a}) + \frac{(\mathbf{h} \bullet \nabla)^2 f(\mathbf{a})}{2!} + \cdots + \frac{(\mathbf{h} \bullet \nabla)^m f(\mathbf{a})}{m!} + O(|\mathbf{h}|^{m+1}).$$

If all partial derivatives of f are continuous, and if there exists a positive number r such that whenever $|\mathbf{h}| < r$ we have for all $\theta \in [0, 1]$,

$$\lim_{m\to\infty} R_{m+1}(\mathbf{h}, \theta) = 0,$$

then we can represent $f(\mathbf{a} + \mathbf{h})$ as the sum of the Taylor series

$$f(\mathbf{a} + \mathbf{h}) = \sum_{j=0}^{\infty} \frac{(\mathbf{h} \bullet \nabla)^j f(\mathbf{a})}{j!}.$$

Remark An alternative approach is to develop Taylor's formula with directional derivatives. Following Section 12.7, a function $g(s)$ is introduced, where $s - s_0$ is distance, measured along a line L in direction $\mathbf{u}$, from the point on L corresponding to $s = s_0$. As in Section 4.10, a Taylor formula for $g(s)$ is

$$g(s) = g(s_0) + g'(s_0)(s-s_0) + \frac{1}{2} g''(s_0)(s-s_0)^2 + \cdots + \frac{1}{2} g^{(m)}(s_0)(s-s_0)^2 + O\big(|s-s_0|^{m+1}\big).$$

Since $d/ds = \mathbf{u} \bullet \nabla$ is the directional derivative operation in direction $\mathbf{u}$, the directional derivative extends to all orders in the Taylor expansion in s. We may choose $g(s) = f(\mathbf{a} + (s - s_0)\mathbf{u})$, where $(s - s_0)\mathbf{u} = \mathbf{h}$. It follows that $|\mathbf{h}|^n = |s - s_0|^n$ and

$$g(s) = f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + (\mathbf{h} \bullet \nabla) f(\mathbf{a}) + \frac{(\mathbf{h} \bullet \nabla)^2 f(\mathbf{a})}{2!} \cdots + \frac{(h \bullet \nabla)^m f(\mathbf{a})}{m!} + O(|\mathbf{h}|^{m+1})$$

as above.

We stress that the expression $(hD_1 + kD_2)^j f(a,b)$ means first calculate $(hD_1 + kD_2)^j f(x,y)$ and then evaluate the result at $(x,y) = (a,b)$.

EXAMPLE 1 Let us illustrate the above ideas with a simple special case. If f is a function of two variables, x and y, having continuous partial derivatives of order up to 4 in the disk $(x-a)^2 + (y-b)^2 \le r^2$, then for $\mathbf{h} = (h,k)$ in $\mathbb{R}^2$ satisfying $h^2 + k^2 < r$ we have

$$\begin{aligned}
f(a+h, b+k) &= P_3(h,k) + R_3(h,k,\theta) \\
&= f(a,b) + (hD_1 + kD_2) f(a,b) + \frac{1}{2!}(hD_1 + kD_2)^2 f(a,b) \\
&\quad + \frac{1}{3!}(hD_1 + kD_2)^3 f(a,b) + R_3(h,k,\theta) \\
&= f(a,b) + hf_1(a,b) + kf_2(a,b) \\
&\quad + \frac{1}{2!}\big(h^2 f_{11}(a,b) + 2hk f_{12}(a,b) + k^2 f_{22}(a,b)\big) \\
&\quad + \frac{1}{3!}\big(h^3 f_{111}(a,b) + 3h^2k f_{112}(a,b) + 3hk^2 f_{122}(a,b) + k^3 f_{222}(a,b)\big) \\
&\quad + R_3(h,k,\theta),
\end{aligned}$$

where $R_3(h,k,\theta) = \dfrac{1}{4!}(hD_1 + kD_2)^4 f(a+\theta h, b+\theta k) = O\big((h^2+k^2)^2\big)$. Note that since $0 < \theta < 1$, all the 4th-order partial derivatives of f are bounded on the line segment from (a,b) to $(a+\theta h, b+\theta k)$. This is why the remainder term is $O\big((h^2+k^2)^2\big)$.

As for functions of one variable, the Taylor polynomial of degree m provides the "best" nth-degree polynomial approximation to $f(x,y)$ near (a,b). For $n = 1$ this approximation reduces to the tangent plane approximation

$$f(x,y) \approx f(a,b) + f_1(a,b)(x-a) + f_2(a,b)(y-b).$$

EXAMPLE 2 Find a second-degree polynomial approximation to the function $f(x,y) = \sqrt{x^2+y^3}$ near the point $(1,2)$, and use it to estimate the value of $\sqrt{(1.02)^2 + (1.97)^3}$.

Solution For the second-degree approximation we need the values of the partial derivatives of f up to second order at $(1,2)$. We have

$$\begin{aligned}
f(x,y) &= \sqrt{x^2+y^3} & f(1,2) &= 3 \\
f_1(x,y) &= \frac{x}{\sqrt{x^2+y^3}} & f_1(1,2) &= \frac{1}{3} \\
f_2(x,y) &= \frac{3y^2}{2\sqrt{x^2+y^3}} & f_2(1,2) &= 2 \\
f_{11}(x,y) &= \frac{y^3}{(x^2+y^3)^{3/2}} & f_{11}(1,2) &= \frac{8}{27} \\
f_{12}(x,y) &= \frac{-3xy^2}{2(x^2+y^3)^{3/2}} & f_{12}(1,2) &= -\frac{2}{9} \\
f_{22}(x,y) &= \frac{12x^2y + 3y^4}{4(x^2+y^3)^{3/2}} & f_{22}(1,2) &= \frac{2}{3}.
\end{aligned}$$

Thus,

$$f(1+h, 2+k) \approx 3 + \frac{1}{3}h + 2k + \frac{1}{2!}\Big(\frac{8}{27}h^2 + 2\Big(-\frac{2}{9}\Big)hk + \frac{2}{3}k^2\Big)$$

or, setting $x = 1 + h$ and $y = 2 + k$,

$$f(x, y) = 3 + \frac{1}{3}(x-1) + 2(y-2) + \frac{4}{27}(x-1)^2 - \frac{2}{9}(x-1)(y-2) + \frac{1}{3}(y-2)^2.$$

This is the required second-degree Taylor polynomial for f near $(1, 2)$. Therefore,

$$\begin{aligned}\sqrt{(1.02)^2 + (1.97)^3} &= f(1 + 0.02, 2 - 0.03)\\ &\approx 3 + \frac{1}{3}(0.02) + 2(-0.03) + \frac{4}{27}(0.02)^2\\ &\quad - \frac{2}{9}(0.02)(-0.03) + \frac{1}{3}(-0.03)^2\\ &\approx 2.947\,159\,3\,.\end{aligned}$$

(For comparison purposes the true value is $2.947\,163\,6\ldots$ The approximation is accurate to 6 significant figures.)

As observed for functions of one variable, it is not usually necessary to calculate derivatives in order to determine the coefficients in a Taylor series or Taylor polynomial. It is often much easier to perform algebraic manipulations on known series. For instance, the above example could have been done by writing f in the form

$$\begin{aligned}f(1 + h, 2 + k) &= \sqrt{(1 + h)^2 + (2 + k)^3}\\ &= \sqrt{9 + 2h + h^2 + 12k + 6k^2 + k^3}\\ &= 3\sqrt{1 + \frac{2h + h^2 + 12k + 6k^2 + k^3}{9}}\end{aligned}$$

and then applying the binomial expansion

$$\sqrt{1 + t} = 1 + \frac{1}{2}t + \frac{1}{2!}\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)t^2 + \cdots$$

with $t = \dfrac{2h + h^2 + 12k + 6k^2 + k^3}{9}$ to obtain the terms up to second degree in the variables h and k.

EXAMPLE 3 Find the Taylor polynomial of degree 3 in powers of x and y for the function $f(x, y) = e^{x-2y}$.

Solution The required Taylor polynomial will be the Taylor polynomial of degree 3 for e^t evaluated at $t = x - 2y$:

$$\begin{aligned}P_3(x, y) &= 1 + (x - 2y) + \frac{1}{2!}(x - 2y)^2 + \frac{1}{3!}(x - 2y)^3\\ &= 1 + x - 2y + \frac{1}{2}x^2 - 2xy + 2y^2 + \frac{1}{6}x^3 - x^2y + 2xy^2 - \frac{4}{3}y^3.\end{aligned}$$

Remark Maple can, of course, be used to compute multivariate Taylor polynomials with its function **mtaylor**, which, depending on the Maple version, may have to be read in from the Maple library before it can be used if it is not part of the Maple kernel.

```
> readlib(mtaylor):
```

Arguments fed to `mtaylor` are as follows:

(a) an expression involving the expansion variables

(b) a list whose elements are either variable names or equations of the form `variable=value` giving the coordinates of the point about which the expansion is calculated. (Just naming a variable is equivalent to using the equation `variable=0`.)

(c) (optionally) a positive integer m forcing the order of the computed Taylor polynomial to be less than m. If m is not specified, the value of Maple's global variable "Order" is used. The default value is 6.

A few examples should suffice.

```
> mtaylor(cos(x+y^2),[x,y]);
```

$$1-\frac{1}{2}x^2-y^2x+\frac{1}{24}x^4-\frac{1}{2}y^4+\frac{1}{6}y^2x^3$$

```
> mtaylor(cos(x+y^2),[x=Pi,y],5);
```

$$-1+\frac{1}{2}(x-\pi)^2+y^2(x-\pi)-\frac{1}{24}(x-\pi)^4+\frac{1}{2}y^4$$

```
> mtaylor(g(x,y),[x=a,y=b],3);
```

$$g(a,b)+D_1(g)(a,b)(x-a)+D_2(g)(a,b)(y-b)+\frac{1}{2}D_{1,1}(g)(a,b)(x-a)^2$$
$$+(x-a)D_{1,2}(g)(a,b)(y-b)+\frac{1}{2}D_{2,2}(g)(a,b)(y-b)^2$$

The function `mtaylor` can be a bit quirky. It has a tendency to expand linear terms; for example, in an expansion about $x=1$ and $y=-2$, it may rewrite terms $2+(x-1)+2(y+2)$ in the form $5+x+2y$.

Approximating Implicit Functions

In the previous section we saw how to determine whether an equation in several variables could be solved for one of those variables as a function of the others. Even when such a solution is known to exist, it may not be possible to find an exact formula for it. However, if the equation involves only smooth functions, then the solution will have a Taylor series. We can determine at least the first several coefficients in that series and thus obtain a useful approximation to the solution. The following example shows the technique.

EXAMPLE 4 Show that the equation $\sin(x+y)=xy+2x$ has a solution of the form $y=f(x)$ near $x=0$ satisfying $f(0)=0$, and find the terms up to fourth degree for the Taylor series for $f(x)$ in powers of x.

Solution The given equation can be written in the form $F(x,y)=0$, where

$$F(x,y)=\sin(x+y)-xy-2x.$$

Since $F(0,0)=0$ and $F_2(0,0)=\cos(0)=1\neq 0$, the equation has a solution $y=f(x)$ near $x=0$ satisfying $f(0)=0$ by the Implicit Function Theorem. It is not possible to calculate $f(x)$ exactly, but it will have a Maclaurin series of the form

$$y=f(x)=a_1x+a_2x^2+a_3x^3+a_4x^4+\cdots.$$

(There is no constant term because $f(0)=0$.) We can substitute this series into the given equation and keep track of terms up to degree 4 in order to calculate the coefficients a_1, a_2, a_3, and a_4. For the left side we use the Maclaurin series for sin to obtain

$$\begin{aligned}\sin(x+y) &= \sin\Big((1+a_1)x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots\Big) \\ &= (1+a_1)x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots \\ &\quad - \frac{1}{3!}\Big((1+a_1)x + a_2x^2 + \cdots\Big)^3 + \cdots \\ &= (1+a_1)x + a_2x^2 + \Big(a_3 - \frac{1}{6}(1+a_1)^3\Big)x^3 \\ &\quad + \Big(a_4 - \frac{3}{6}(1+a_1)^2a_2\Big)x^4 + \cdots.\end{aligned}$$

The right side is $xy + 2x = 2x + a_1x^2 + a_2x^3 + a_3x^4 + \cdots$. Equating coefficients of like powers of x, we obtain

$$\begin{aligned} 1 + a_1 &= 2 & a_1 &= 1 \\ a_2 &= a_1 & a_2 &= 1 \\ a_3 - \frac{1}{6}(1+a_1)^3 &= a_2 & a_3 &= \frac{7}{3} \\ a_4 - \frac{1}{2}(1+a_1)^2 a_2 &= a_3 & a_4 &= \frac{13}{3}. \end{aligned}$$

Thus,

$$y = f(x) = x + x^2 + \frac{7}{3}x^3 + \frac{13}{3}x^4 + \cdots.$$

(We could have obtained more terms in the series by keeping track of higher powers of x in the substitution process.)

Remark From the series for $f(x)$ obtained above, we can determine the values of the first four derivatives of f at $x = 0$. Remember that

$$a_k = \frac{f^{(k)}(0)}{k!}.$$

We have, therefore,

$$\begin{aligned} f'(0) &= a_1 = 1 & f''(0) &= 2!a_2 = 2 \\ f'''(0) &= 3!a_3 = 14 & f^{(4)}(0) &= 4!a_4 = 104. \end{aligned}$$

We could have done the example by first calculating these derivatives by implicit differentiation of the given equation and then determining the series coefficients from them. This would have been a much more difficult way to do it. (Try it and see.)

EXERCISES 12.9

In Exercises 1–6, find the Taylor series for the given function about the indicated point.

1. $f(x, y) = \dfrac{1}{2 + xy^2}$, $(0, 0)$

2. $f(x, y) = \ln(1 + x + y + xy)$, $(0, 0)$

3. $f(x, y) = \tan^{-1}(x + xy)$, $(0, -1)$

4. $f(x, y) = x^2 + xy + y^3$, $(1, -1)$

5. $f(x, y) = e^{x^2+y^2}$, $(0, 0)$

6. $f(x, y) = \sin(2x + 3y)$, $(0, 0)$

In Exercises 7–12, find Taylor polynomials of the indicated degree for the given functions near the given point. After calculating them by hand, try to get the same results using Maple's `mtaylor` function.

7. $f(x,y)=\dfrac{1}{2+x-2y}$, degree 3, near $(2,1)$

8. $f(x,y)=\ln(x^2+y^2)$, degree 3, near $(1,0)$

9. $f(x,y)=\displaystyle\int_0^{x+y^2} e^{-t^2}\,dt$, degree 3, near $(0,0)$

10. $f(x,y)=\cos(x+\sin y)$, degree 4, near $(0,0)$

11. $f(x,y)=\dfrac{\sin x}{y}$, degree 2, near $(\frac{\pi}{2},1)$

12. $f(x,y)=\dfrac{1+x}{1+x^2+y^4}$, degree 2, near $(0,0)$

In Exercises 13–14, show that, for x near the indicated point $x=a$, the given equation has a solution of the form $y=f(x)$ taking on the indicated value at that point. Find the first three nonzero terms of the Taylor series for $f(x)$ in powers of $x-a$.

13. $x\sin y=y+\sin x$, near $x=0$, with $f(0)=0$

14. $\sqrt{1+xy}=1+x+\ln(1+y)$, near $x=0$, with $f(0)=0$

15. Show that the equation $x+2y+z+e^{2z}=1$ has a solution of the form $z=f(x,y)$ near $x=0$, $y=0$, where $f(0,0)=0$. Find the Taylor polynomial of degree 2 for $f(x,y)$ in powers of x and y.

16. Use series methods to find the value of the partial derivative $f_{112}(0,0)$ given that $f(x,y)=\arctan(x+y)$.

17. Use series methods to evaluate

$$\frac{\partial^{4n}}{\partial x^{2n}\partial y^{2n}}\frac{1}{1+x^2+y^2}\bigg|_{(0,0)}.$$

CHAPTER REVIEW

Key Ideas

- **What do the following sentences and phrases mean?**
  - ◇ $\mathcal{S}$ is the graph of $f(x,y)$.
  - ◇ $\mathcal{C}$ is a level curve of $f(x,y)$.
  - ◇ $\lim_{(x,y)\to(a,b)} f(x,y)=L$.
  - ◇ $f(x,y)$ is continuous at (a,b).
  - ◇ the partial derivative $(\partial/\partial x)f(x,y)$
  - ◇ the tangent plane to $z=f(x,y)$ at (a,b)
  - ◇ pure second partials ◇ mixed second partials
  - ◇ $f(x,y)$ is a harmonic function.
  - ◇ $L(x,y)$ is the linearization of $f(x,y)$ at (a,b).
  - ◇ the differential of $z=f(x,y)$
  - ◇ $f(x,y)$ is differentiable at (a,b).
  - ◇ the gradient of $f(x,y)$ at (a,b)
  - ◇ the directional derivative of $f(x,y)$ at (a,b) in direction $\mathbf{v}$
  - ◇ the Jacobian determinant $\partial(x,y)/\partial(u,v)$
- **Under what conditions are two mixed partial derivatives equal?**
- **State the Chain Rule for $z=f(x,y)$, where $x=g(u,v)$, and $y=h(u,v)$.**
- **Describe the process of calculating partial derivatives of implicitly defined functions.**
- **What is the Taylor series of $f(x,y)$ about (a,b)?**

Review Exercises

1. Sketch some level curves of the function $x+\dfrac{4y^2}{x}$.

2. Sketch some isotherms (curves of constant temperature) for the temperature function

$$T=\frac{140+30x^2-60x+120y^2}{8+x^2-2x+4y^2}\ ^\circ\text{C}.$$

What is the coolest location?

3. Sketch some level curves of the polynomial function $f(x,y)=x^3-3xy^2$. Why do you think the graph of this function is called a *monkey saddle*?

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

Calculate each of the following partial derivatives or explain why it does not exist: $f_1(0,0)$, $f_2(0,0)$, $f_{21}(0,0)$, $f_{12}(0,0)$.

5. Let $f(x,y)=\dfrac{x^3-y^3}{x^2-y^2}$. Where is $f(x,y)$ continuous? To what additional set of points does $f(x,y)$ have a continuous extension? In particular, can f be extended to be continuous at the origin? Can f be defined at the origin in such a way that its first partial derivatives exist there?

6. The surface $\mathcal{S}$ is the graph of the function $z=f(x,y)$, where $f(x,y)=e^{x^2-2x-4y^2+5}$.

(a) Find an equation of the tangent plane to $\mathcal{S}$ at the point $(1,-1,1)$.

(b) Sketch a representative sample of the level curves of the function $f(x,y)$.

7. Consider the surface $\mathcal{S}$ with equation $x^2+y^2+4z^2=16$.

(a) Find an equation for the tangent plane to $\mathcal{S}$ at the point (a,b,c) on $\mathcal{S}$.

(b) For which points (a,b,c) on $\mathcal{S}$ does the tangent plane to $\mathcal{S}$ at (a,b,c) pass through the point $(0,0,4)$? Describe this set of points geometrically.

(c) For which points (a,b,c) on $\mathcal{S}$ is the tangent plane to $\mathcal{S}$ at (a,b,c) parallel to the plane $x+y+2\sqrt{2}z=97$?

8. Two variable resistors, R_1 and R_2, are connected in parallel so that their combined resistance, R, is given by

$$\frac{1}{R}=\frac{1}{R_1}+\frac{1}{R_2}.$$

If $R_1=100$ ohms $\pm5\%$ and $R_2=25$ ohms $\pm2\%$, by approximately what percentage can the calculated value of their combined resistance $R=20$ ohms be in error?

9. You have measured two sides of a triangular field and the angle between them. The side measurements are 150 m and 200 m, each accurate to within ±1 m. The angle measurement is 30°, accurate to within $\pm2°$. What area do you calculate for the field, and what is your estimate of the maximum percentage error in this area?

10. Suppose that $T(x,y,z)=x^3y+y^3z+z^3x$ gives the temperature at the point (x,y,z) in 3-space.

(a) Calculate the directional derivative of T at $(2, -1, 0)$ in the direction toward the point $(1, 1, 2)$.

(b) A fly is moving through space with constant speed 5. At time $t = 0$ the fly crosses the surface $2x^2 + 3y^2 + z^2 = 11$ at right angles at the point $(2, -1, 0)$, moving in the direction of increasing temperature. Find dT/dt at $t = 0$ as experienced by the fly.

11. Consider the function $f(x, y, z) = x^2y + yz + z^2$.

(a) Find the directional derivative of f at $(1, -1, 1)$ in the direction of the vector $\mathbf{i} + \mathbf{k}$.

(b) An ant is crawling on the plane $x + y + z = 1$ through $(1, -1, 1)$. Suppose it crawls so as to keep f constant. In what direction is it going as it passes through $(1, -1, 1)$?

(c) Another ant crawls on the plane $x + y + z = 1$, moving in the direction of the greatest rate of increase of f. Find its direction as it goes through $(1, -1, 1)$.

12. Let $f(x, y, z) = (x^2 + z^2) \sin \dfrac{\pi xy}{2} + yz^2$. Let P_0 be the point $(1, 1, -1)$.

(a) Find the gradient of f at P_0.

(b) Find the linearization $L(x, y, z)$ of f at P_0.

(c) Find an equation for the tangent plane at P_0 to the level surface of f through P_0.

(d) If a bird flies through P_0 with speed 5, heading directly toward the point $(2, -1, 1)$, what is the rate of change of f as seen by the bird as it passes through P_0?

(e) In what direction from P_0 should the bird fly at speed 5 to experience the greatest rate of increase of f?

13. Verify that for any constant, k, the function $u(x, y) = k\Big(\ln\cos(x/k) - \ln\cos(y/k)\Big)$ satisfies the *minimal surface equation*

$$(1 + u_x^2)u_{yy} - uu_xu_yu_{xy} + (1 + u_y^2)u_{xx} = 0.$$

14. The equations $F(x, y, z) = 0$ and $G(x, y, z) = 0$ can define any two of the variables x, y, and z as functions of the remaining variable. Show that

$$\frac{dx}{dy}\,\frac{dy}{dz}\,\frac{dz}{dx} = 1.$$

15. The equations $\begin{cases} x = u^3 - uv \\ y = 3uv + 2v^2 \end{cases}$ define u and v as functions of x and y near the point P where $(u, v, x, y) = (-1, 2, 1, 2)$.

(a) Find $\dfrac{\partial u}{\partial x}$ and $\dfrac{\partial u}{\partial y}$ at P.

(b) Find the approximate value of u when $x = 1.02$ and $y = 1.97$.

16. The equations $\begin{cases} u = x^2 + y^2 \\ v = x^2 - 2xy^2 \end{cases}$ define x and y implicitly as functions of u and v for values of (x, y) near $(1, 2)$ and values of (u, v) near $(5, -7)$.

(a) Find $\dfrac{\partial x}{\partial u}$ and $\dfrac{\partial y}{\partial u}$ at $(u, v) = (5, -7)$.

(b) If $z = \ln(y^2 - x^2)$, find $\dfrac{\partial z}{\partial u}$ at $(u, v) = (5, -7)$.

Challenging Problems

1. (a) If the graph of a function $f(x, y)$ that is differentiable at (a, b) contains part of a straight line through (a, b), show that the line lies in the tangent plane to $z = f(x, y)$ at (a, b).

(b) If $g(t)$ is a differentiable function of t, describe the surface $z = yg(x/y)$ and show that all its tangent planes pass through the origin.

2. A particle moves in 3-space in such a way that its direction of motion at any point is perpendicular to the level surface of

$$f(x, y, z) = 4 - x^2 - 2y^2 + 3z^2$$

through that point. If the path of the particle passes through the point $(1, 1, 8)$, show that it also passes through $(2, 4, 1)$. Does it pass through $(3, 7, 0)$?

3. **(The Laplace operator in spherical coordinates)** If $u(x, y, z)$ has continuous second partial derivatives and

$$v(R, \phi, \theta) = u(R \sin\phi \cos\theta, R \sin\phi \sin\theta, R \cos\phi),$$

show that

$$\frac{\partial^2 v}{\partial R^2} + \frac{2}{R}\frac{\partial v}{\partial R} + \frac{\cot\phi}{R^2}\frac{\partial v}{\partial \phi} + \frac{1}{R^2}\frac{\partial^2 v}{\partial \phi^2} + \frac{1}{R^2 \sin^2\phi}\frac{\partial^2 v}{\partial \theta^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}.$$

You can do this by hand, but it is a lot easier using computer algebra.

4. **(Spherically expanding waves)** If f is a twice differentiable function of one variable and $R = \sqrt{x^2 + y^2 + z^2}$, show that $u(x, y, z, t) = \dfrac{f(R - ct)}{R}$ satisfies the three-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right).$$

What is the geometric significance of this solution as a function of increasing time t? *Hint:* You may want to use the result of Exercise 3. In this case $v(R, \phi, \theta)$ is independent of ϕ and θ.

In Exercises 25–27, find the terms up to second power in ϵ in the solution y of the given equation.

25. $y + \epsilon \sin \pi y = x$

26. $y^2 + \epsilon e^{-y^2} = 1 + x^2$

27. $2y + \dfrac{\epsilon x}{1 + y^2} = 1$

28. Use perturbation methods to evaluate y with error less than 10^{-8} given that $y + (y^5/100) = 1/2$.

29. Use perturbation methods to find approximate values for x and y from the system $x + 2y + \dfrac{1}{100}e^{-x} = 3$, $x - y + \dfrac{1}{100}e^{-y} = 0$. Calculate all terms up to second order in $\epsilon = 1/100$.

13.7 Newton's Method

A frequently encountered problem in applied mathematics is to determine, to some desired degree of accuracy, a root (i.e., a solution r) of an equation of the form

$$f(r) = 0.$$

Such a root is called a **zero** of the function f. In Section 4.2 we introduced Newton's Method, a simple but powerful method for determining roots of functions that are sufficiently smooth. The method involves *guessing* an approximate value x_0 for a root r of the function f, and then calculating successive approximations $x_1, x_2, \ldots,$ using the formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \qquad n = 0, 1, 2, \cdots.$$

If the initial guess x_0 is not too far from r, and if $|f'(x)|$ is *not too small* and $|f''(x)|$ is *not too large* near r, then the successive approximations $x_1, x_2, \ldots$ will converge very rapidly to r. Recall that each new approximation x_{n+1} is obtained as the x-intercept of the tangent line drawn to the graph of f at the previous approximation, x_n. The tangent line to the graph $y = f(x)$ at $x = x_n$ has equation

$$y - f(x_n) = f'(x_n)(x - x_n).$$

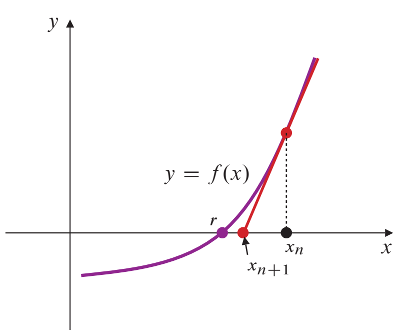

Figure 13.24 x_{n+1} is the x-intercept of the tangent at x_n

(See Figure 13.24.) The x-intercept, x_{n+1}, of this line is determined by setting $y = 0$, $x = x_{n+1}$ in this equation, so is given by the formula in the shaded box above.

Newton's Method can be extended to finding solutions of systems of m equations in m variables. We will show here how to adapt the method to find approximations to a solution (x, y) of the pair of equations

$$\begin{cases} f(x, y) = 0 \\ g(x, y) = 0, \end{cases}$$

starting from an initial guess (x_0, y_0). Under auspicious circumstances, we will observe the same rapid convergence of approximations to the root that typifies the single-variable case.

The idea is as follows. The two surfaces $z = f(x, y)$ and $z = g(x, y)$ intersect in a curve which itself intersects the xy-plane at the point whose coordinates are the desired solution. If (x_0, y_0) is near that point, then the tangent planes to the two surfaces at (x_0, y_0) will intersect in a straight line. This line meets the xy-plane at a point (x_1, y_1) that should be even closer to the solution point than was (x_0, y_0). We can easily determine (x_1, y_1). The tangent planes to $z = f(x, y)$ and $z = g(x, y)$ at (x_0, y_0) have equations

$$z = f(x_0, y_0) + f_1(x_0, y_0)(x - x_0) + f_2(x_0, y_0)(y - y_0),$$
$$z = g(x_0, y_0) + g_1(x_0, y_0)(x - x_0) + g_2(x_0, y_0)(y - y_0).$$

The line of intersection of these two planes meets the xy-plane at the point (x_1, y_1) satisfying

$$f_1(x_0, y_0)(x_1 - x_0) + f_2(x_0, y_0)(y_1 - y_0) + f(x_0, y_0) = 0,$$
$$g_1(x_0, y_0)(x_1 - x_0) + g_2(x_0, y_0)(y_1 - y_0) + g(x_0, y_0) = 0.$$

Solving these two equations for x_1 and y_1, we obtain

$$x_1 = x_0 - \left.\frac{fg_2 - f_2g}{f_1g_2 - f_2g_1}\right|_{(x_0,y_0)} = x_0 - \left.\frac{\begin{vmatrix} f & f_2 \\ g & g_2 \end{vmatrix}}{\begin{vmatrix} f_1 & f_2 \\ g_1 & g_2 \end{vmatrix}}\right|_{(x_0,y_0)},$$

$$y_1 = y_0 - \left.\frac{f_1g - fg_1}{f_1g_2 - f_2g_1}\right|_{(x_0,y_0)} = y_0 - \left.\frac{\begin{vmatrix} f_1 & f \\ g_1 & g \end{vmatrix}}{\begin{vmatrix} f_1 & f_2 \\ g_1 & g_2 \end{vmatrix}}\right|_{(x_0,y_0)}.$$

Observe that the denominator in each of these expressions is the Jacobian determinant $\partial(f, g)/\partial(x, y)\big|_{(x_0,y_0)}$. This is another instance where the Jacobian is the appropriate multivariable analogue of the derivative of a function of one variable.

Continuing in this way, we generate successive approximations (x_n, y_n) according to the formulas

$$x_{n+1} = x_n - \left.\frac{\begin{vmatrix} f & f_2 \\ g & g_2 \end{vmatrix}}{\begin{vmatrix} f_1 & f_2 \\ g_1 & g_2 \end{vmatrix}}\right|_{(x_n,y_n)},$$

$$y_{n+1} = y_n - \left.\frac{\begin{vmatrix} f_1 & f \\ g_1 & g \end{vmatrix}}{\begin{vmatrix} f_1 & f_2 \\ g_1 & g_2 \end{vmatrix}}\right|_{(x_n,y_n)}.$$

We stop when the desired accuracy has been achieved.

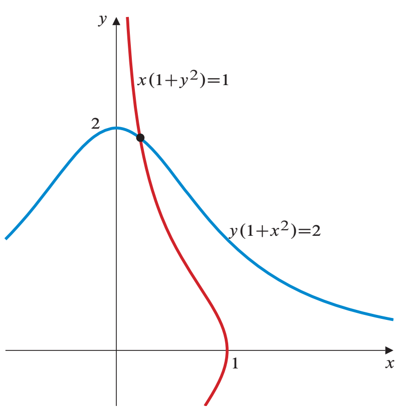

Figure 13.25 The two graphs intersect near (0.2, 1.8)

EXAMPLE 1 Find the root of the system of equations $x(1 + y^2) - 1 = 0$, $y(1 + x^2) - 2 = 0$ with sufficient accuracy to ensure that the left sides of the equations vanish to the sixth decimal place.

Solution A sketch of the graphs of the two equations (see Figure 13.25) in the xy-plane indicates that the system has only one root near the point $(0.2, 1.8)$. Application of Newton's Method requires successive computations of the quantities

$$f(x, y) = x(1 + y^2) - 1, \quad f_1(x, y) = 1 + y^2, \quad f_2(x, y) = 2xy,$$
$$g(x, y) = y(1 + x^2) - 2, \quad g_1(x, y) = 2xy, \quad g_2(x, y) = 1 + x^2.$$

Using a calculator or computer, we can calculate successive values of (x_n, y_n) starting from $x_0 = 0.2$, $y_0 = 1.8$:

Table 1. Root near (0.2, 1.8)

n	x_n	y_n	$f(x_n, y_n)$	$g(x_n, y_n)$
0	0.200 000	1.800 000	−0.152 000	−0.128 000
1	0.216 941	1.911 349	0.009 481	0.001 303
2	0.214 827	1.911 779	−0.000 003	0.000 008
3	0.214 829	1.911 769	0.000 000	0.000 000

The values in Table 1 were calculated sequentially in a spreadsheet by the method suggested below. They were rounded for inclusion in the table, but the unrounded values were used in subsequent calculations. If you actually use the (rounded) values of x_n and y_n given in the table to calculate $f(x_n, y_n)$ and $g(x_n, y_n)$, your results may vary slightly.

The desired approximations to the root are the x_n and y_n values in the last line of the above table. Note the rapidity of convergence. However, many function evaluations are needed for each iteration of the method. For large systems, Newton's Method is computationally too inefficient to be practical. Other methods requiring more iterations but many fewer calculations per iteration are used in practice.

Implementing Newton's Method Using a Spreadsheet

A computer spreadsheet is an ideal environment in which to calculate Newton's Method approximations. For a pair of equations in two unknowns such as the system in Example 1, you can proceed as follows:

(i) In the first nine cells of the first row (A1–I1) put the labels `n`, `x`, `y`, `f`, `g`, `f1`, `f2`, `g1`, and `g2`.

(ii) In cells A2–A9 put the numbers 0, 1, 2, . . . , 7.

(iii) In cells B2 and C2 put the starting values x_0 and y_0.

(iv) In cells D2–I2 put formulas for calculating $f(x, y)$, $g(x, y)$, . . . , $g_2(x, y)$ in terms of values of x and y assumed to be in B2 and C2.

(v) In cells B3 and C3 store the Newton's Method formulas for calculating x_1 and y_1 in terms of the values x_0 and y_0, using values calculated in the second row. For instance, cell B3 should contain the formula

```
+B2-(D2*I2-G2*E2)/(F2*I2-G2*H2).
```

(vi) Replicate the formulas in cells D2–I2 to cells D3–I3.

(vii) Replicate the formulas in cells B3–I3 to the cells B4–I9.

You can now inspect the successive approximations x_n and y_n in columns B and C. To use different starting values, just replace the numbers in cells B2 and C2. To solve a different system of (two) equations, replace the contents of cells D2–I2. You may wish to save this spreadsheet for reuse with the exercises below or other systems you may want to solve later.

Remark While a detailed analysis of the convergence of Newton's Method approximations is beyond the scope of this book, a few observations can be made. At each step in the approximation process we must divide by J, the Jacobian determinant of f and g with respect to x and y evaluated at the most recently obtained approximation. Assuming that the functions and partial derivatives involved in the formulas are continuous, the larger the value of J at the actual solution, the more likely are the approximations to converge to the solution, and to do so rapidly. If J vanishes (or is very small) at the solution, the successive approximations may not converge, even if the initial guess is quite close to the solution. Even if the first partials of f and g are large at the solution, their Jacobian may be small if their gradients are nearly parallel there. Thus, we cannot expect convergence to be rapid when the curves $f(x, y) = 0$ and $g(x, y) = 0$ intersect at a very small angle.

Newton's Method can be applied to systems of m equations in m variables; the formulas are the obvious generalizations of those for two functions given above.

EXERCISES 13.7

Find the solutions of the systems in Exercises 1–6, so that the left-hand sides of the equations vanish up to 6 decimal places. These can be done with the aid of a scientific calculator, but that approach will be very time consuming. It is much easier to program the Newton's Method formulas on a computer to generate the required approximations. In each case try to determine reasonable *initial guesses* by sketching graphs of the equations.

1. $y - e^x = 0, \quad x - \sin y = 0$

2. $x^2 + y^2 - 1 = 0, \quad y - e^x = 0$ (two solutions)

3. $x^4 + y^2 - 16 = 0, \quad xy - 1 = 0$ (four solutions)

4. $x^2 - xy + 2y^2 = 10, \quad x^3y^2 = 2$ (four solutions)

5. $y - \sin x = 0, \quad x^2 + (y+1)^2 - 2 = 0$ (two solutions)

6. $\sin x + \sin y - 1 = 0, \quad y^2 - x^3 = 0$ (two solutions)

7. Write formulas for obtaining successive Newton's Method

$$f(x, y, z) = 0, \quad g(x, y, z) = 0, \quad h(x, y, z) = 0,$$

starting from an initial guess (x_0, y_0, z_0).

8. Use the formulas from Exercise 7 to find the first octant intersection point of the surfaces $y^2 + z^2 = 3$, $x^2 + z^2 = 2$, and $x^2 - z = 0$.

9. The equations $y - x^2 = 0$ and $y - x^3 = 0$ evidently have the solutions $x = y = 0$ and $x = y = 1$. Try to obtain these solutions using the two-variable form of Newton's Method with starting values
(a) $x_0 = y_0 = 0.1$, and (b) $x_0 = y_0 = 0.9$.
How many iterations are required to obtain 6-decimal-place accuracy for the appropriate solution in each case?
How do you account for the difference in the behaviour of Newton's Method for these equations near $(0, 0)$ and $(1, 1)$?

13.8 Calculations with Maple

The calculations involved in solving systems of equations involving several variables can be very lengthy, even if the number of variables is small. In particular, locating critical points of a function of n variables involves solving a system of n (usually nonlinear) equations in n unknowns. In such situations the effective use of a computer algebra system like Maple can be very helpful. In this optional (and brief) section we present examples of how to use Maple's "fsolve" routine to solve systems of nonlinear equations and to find and classify critical points and thereby solve extreme-value problems.

Solving Systems of Equations

Maple has a procedure called **fsolve** built into its kernel (no package needs to be loaded to access it) that attempts to find floating-point real solutions to systems n equations in n variables. (For a single polynomial equation in one variable it will try to find all the real roots, but it may miss some.) For our purposes, an equation consists of either a single expression f in the variables (in which case the equation is taken to be $f = 0$) or else two expressions joined by an equal sign, as in $f = g$. The procedure takes two or three arguments. The first is a set of n equations, enclosed in braces and separated by commas. The second argument is a set (also enclosed in braces) listing the n variables for which the equations are to be solved. (The number of variables in the equations must equal the number of equations.) The elements of the second set may consist of equations of the form "variable = initial guess," where the initial guess is a number we have reason to believe is *close* to the actual solution. It may not always be possible to make a good initial guess at the values of the variables, so, if we like, we can include a third argument specifing intervals of values of the variables in which to search for a solution. For example, to find a solution to the system $x^2 + y^3 = 3$, $x\sin(y) - y\cos(x) = 0$ near $(1, 2)$, we could try

```
> Digits := 6:
> fsolve({x^2+y^2=3, x*sin(y)-y*cos(x)}, {x=1, y=2});
```

$$\{x = 0.909510, \ y = 1.47404\}$$