3

Higher-Order Derivatives: Maxima and Minima

Leonhard Euler (by Emanuel Handman) (1707–1783).

All that is superfluous displeases God and Nature.
All that displeases God and Nature is evil.

Dante Alighieri, circa 1300

... namely, because the shape of the whole universe is most perfect, and, in fact, designed by the wisest creator, nothing in all of the world will occur in which no maximum or minimum rule is somehow shining forth.

Leonhard Euler

In one-variable calculus, to test a function $f(x)$ for a local maximum or minimum, one often uses the second derivative. We look for critical points x_0, that is, points x_0 for which $f'(x_0) = 0$, and at each such point we check the sign of the second derivative $f''(x_0)$. If $f''(x_0) < 0$, $f(x_0)$ is a local maximum of f; if $f''(x_0) > 0$, $f(x_0)$ is a local minimum of f; if $f''(x_0) = 0$, the test fails.

This chapter extends these methods to real-valued functions of several variables. We begin in Section 3.1 with a discussion of iterated and higher-order partial derivatives, and in Section 3.2 we discuss the multivariable form of Taylor's theorem; this is then used in Section 3.3 to derive tests for maxima, minima, and saddle points. As with functions of one variable, such methods help one to visualize the shape of a graph.

In Section 3.4, we study the problem of maximizing a real-valued function subject to supplementary conditions, also referred to as constraints. For example, we might wish to maximize $f(x, y, z)$ among those (x, y, z) constrained to lie on the unit sphere, $x^2 + y^2 + z^2 = 1$. Section 3.5 discusses a technical theorem (the implicit function theorem) useful for studying constraints. It will also be useful later in our study of surfaces.

3.1 Iterated Partial Derivatives

The preceding chapter developed considerable information concerning the derivative of a map and investigated the geometry associated with the derivative of real-valued functions by making use of the gradient. In this section, we proceed to study higher-order derivatives, with the goal of proving the equality of the "mixed second partial derivatives" of a function. We begin by defining the necessary terms.

Let $f: \mathbb{R}^3 \to \mathbb{R}$ be of class C^1. Recall that this means that $\partial f/\partial x$, $\partial f/\partial y$, and $\partial f/\partial z$ exist and are continuous. If these derivatives, in turn, have continuous partial derivatives, we say that f is of ***class*** C^2, or is ***twice continuously differentiable***. Likewise, if we say f is of class C^3, we mean f has continuous iterated partial derivatives of third order, and so on. Here are a few examples of how second-order derivatives are written:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \qquad \frac{\partial^2 f}{\partial x\, \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right), \qquad \frac{\partial^2 f}{\partial z\, \partial y} = \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial y}\right), \qquad \text{etc.}$$

The process can, of course, be repeated for third-order derivatives, and so on. If f is a function of only x and y and $\partial f/\partial x$, $\partial f/\partial y$ are continuously differentiable, then by taking second partial derivatives we get the four functions

$$\frac{\partial^2 f}{\partial x^2}, \qquad \frac{\partial^2 f}{\partial y^2}, \qquad \frac{\partial^2 f}{\partial x\, \partial y}, \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y\, \partial x}.$$

All of these are called ***iterated partial derivatives***, while $\partial^2 f/\partial x\, \partial y$ and $\partial^2 f/\partial y\, \partial x$ are called ***mixed partial derivatives***.

EXAMPLE 1 Find all second partial derivatives of $f(x, y) = xy + (x + 2y)^2$.

SOLUTION The first partials are

$$\frac{\partial f}{\partial x} = y + 2(x + 2y), \qquad \frac{\partial f}{\partial y} = x + 4(x + 2y).$$

Now differentiate each of these expressions with respect to x and y:

$$\frac{\partial^2 f}{\partial x^2} = 2, \qquad \frac{\partial^2 f}{\partial y^2} = 8$$

$$\frac{\partial^2 f}{\partial x\, \partial y} = 5, \qquad \frac{\partial^2 f}{\partial y\, \partial x} = 5. \quad \blacktriangle$$

EXAMPLE 2 Find all second partial derivatives of $f(x, y) = \sin x \sin^2 y$.

SOLUTION We proceed just as in Example 1:

$$\frac{\partial f}{\partial x} = \cos x \sin^2 y, \qquad \frac{\partial f}{\partial y} = 2 \sin x \sin y \cos y = \sin x \sin 2y;$$

$$\frac{\partial^2 f}{\partial x^2} = -\sin x \sin^2 y, \qquad \frac{\partial^2 f}{\partial y^2} = 2 \sin x \cos 2y;$$

$$\frac{\partial^2 f}{\partial x\, \partial y} = \cos x \sin 2y, \qquad \frac{\partial^2 f}{\partial y\, \partial x} = 2 \cos x \sin y \cos y = \cos x \sin 2y. \quad \blacktriangle$$

EXAMPLE 3 Let $f(x, y, z) = e^{xy} + z \cos x$. Then

$$\frac{\partial f}{\partial x} = ye^{xy} - z \sin x, \qquad \frac{\partial f}{\partial y} = xe^{xy}, \qquad \frac{\partial f}{\partial z} = \cos x,$$

$$\frac{\partial^2 f}{\partial z\, \partial x} = -\sin x, \qquad \frac{\partial^2 f}{\partial x\, \partial z} = -\sin x, \qquad \text{etc.} \quad \blacktriangle$$

The Mixed Partials are Equal

In all these examples note that the pairs of mixed partial derivatives, such as $\partial^2 f/\partial x\, \partial y$ and $\partial^2 f/\partial y\, \partial x$, or $\partial^2 f/\partial z\, \partial x$ and $\partial^2 f/\partial x\, \partial z$, are equal. It is a basic and perhaps surprising fact that *this is always the case for C^2 functions*. We shall prove this in the next theorem for functions $f(x, y)$ of two variables, but the proof can be readily extended to functions of n variables.

THEOREM 1: Equality of Mixed Partials If $f(x, y)$ is of class C^2 (is twice continuously differentiable), then the mixed partial derivatives are equal; that is,

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

PROOF Consider the following expression (see Figure 3.1.1):

$$S(\Delta x, \Delta y) = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0 + \Delta x, y_0) - f(x_0, y_0 + \Delta y) + f(x_0, y_0).$$

Holding y_0 and Δy fixed, define

$$g(x) = f(x, y_0 + \Delta y) - f(x, y_0),$$

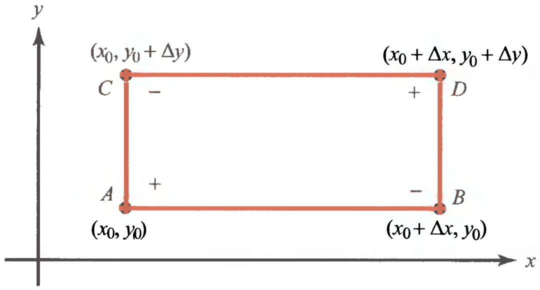

Figure 3.1.1 The algebra behind the equality of mixed partials: writing the difference of differences in two ways.

so that $S(\Delta x, \Delta y) = g(x_0 + \Delta x) - g(x_0)$, *which expresses S as a difference of differences*. By the mean-value theorem for functions of one variable, $g(x_0 + \Delta x) - g(x_0)$ equals $g'(\bar{x})\Delta x$ for some $\bar{x}$ between x_0 and $x_0 + \Delta x$. Hence,

$$S(\Delta x, \Delta y) = \left[\frac{\partial f}{\partial x}(\bar{x}, y_0 + \Delta y) - \frac{\partial f}{\partial x}(\bar{x}, y_0)\right]\Delta x.$$

Applying the mean-value theorem again, there is a $\bar{y}$ between y_0 and $y_0 + \Delta y$ such that

$$S(\Delta x, \Delta y) = \frac{\partial^2 f}{\partial y\, \partial x}(\bar{x}, \bar{y})\, \Delta x\, \Delta y.$$

Because $\partial^2 f/\partial y\, \partial x$ is continuous, it follows that

$$\frac{\partial^2 f}{\partial y\, \partial x}(x_0, y_0) = \lim_{(\Delta x, \Delta y)\to(0,0)} \frac{1}{\Delta x\, \Delta y}[S(\Delta x, \Delta y)].$$

Noting that S is symmetric in Δx and Δy, one shows in a similar way that $\partial^2 f/\partial x\, \partial y$ is given by the *same limit formula*, which proves the result. ■

Historical Note

The equality of mixed partial derivatives is one of the most important results of multivariable calculus. It will reappear on several occasions later in the book, when we study vector identities.

In the next historical note, we will discuss the role of partial derivatives in the formulation of many of the basic equations governing physical phenomena. One of the giants in this era was Leonhard Euler (1707–1783), who developed the equations of fluid mechanics that bear his name—the Euler equations. It was in connection with the needs of this development that he discovered, around 1734, the equality of mixed partial derivatives. Euler was about 27 years old at the time.

In Exercise 11 we ask the reader to deduce from Theorem 1 that for a C^3 function of x, y, and z,

$$\frac{\partial^3 f}{\partial x\,\partial y\,\partial z} = \frac{\partial^3 f}{\partial z\,\partial y\,\partial x} = \frac{\partial^3 f}{\partial y\,\partial z\,\partial x}, \qquad \text{etc.}$$

In other words, we can compute iterated partial derivatives in any order we please.

EXAMPLE 4 Verify the equality of the mixed second partial derivatives for the function

$$f(x, y) = xe^y + yx^2.$$

SOLUTION Here

$$\frac{\partial f}{\partial x} = e^y + 2xy, \qquad \frac{\partial f}{\partial y} = xe^y + x^2,$$

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

and so we have

$$\frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial^2 f}{\partial x\,\partial y}. \quad \blacktriangle$$

Sometimes the notation f_x, f_y, f_z is used for the partial derivatives: $f_x = \partial f/\partial x$, and so on. With this notation, we write $f_{xy} = (f_x)_y$, and so equality of the mixed partials is denoted by $f_{xy} = f_{yx}$. Notice that $f_{xy} = \partial^2 f/\partial y\,\partial x$, so the order of x and y is reversed in the two notations; fortunately, the equality of mixed partials makes this potential ambiguity irrelevant. The following example illustrates this subscript notation.

EXAMPLE 5 Let

$$z = f(x, y) = e^x \sin xy$$

and write $x = g(s, t)$, $y = h(s, t)$ for certain functions g and h. Let

$$k(s, t) = f(g(s, t), h(s, t)).$$

Calculate k_{st}.

SOLUTION By the chain rule,

$$k_s = f_x g_s + f_y h_s = (e^x \sin xy + ye^x \cos xy)g_s + (xe^x \cos xy)h_s.$$

Differentiating in t using the product rules gives

$$k_{st} = (f_x)_t g_s + f_x (g_s)_t + (f_y)_t h_s + f_y (h_s)_t.$$

Applying the chain rule again to $(f_x)_t$ and $(f_y)_t$ gives

$$(f_x)_t = f_{xx} g_t + f_{xy} h_t \qquad \text{and} \qquad (f_y)_t = f_{yx} g_t + f_{yy} h_t,$$

and so k_{st} becomes

$$\begin{aligned} k_{st} &= (f_{xx} g_t + f_{xy} h_t) g_s + f_x g_{st} + (f_{yx} g_t + f_{yy} h_t) h_s + f_y h_{st} \\ &= f_{xx} g_t g_s + f_{xy}(h_t g_s + h_s g_t) + f_{yy} h_t h_s + f_x g_{st} + f_y h_{st}. \end{aligned}$$

Notice that this last formula is symmetric in (s, t), verifying the equality $k_{st} = k_{ts}$. Computing f_{xx}, f_{xy}, and f_{yy}, we get

$$\begin{aligned} k_{st} = {} & (e^x \sin xy + 2ye^x \cos xy - y^2 e^x \sin xy) g_t g_s \\ & + (xe^x \cos xy + e^x \cos xy - xye^x \sin xy)(h_t g_s + h_s g_t) \\ & - (x^2 e^x \sin xy) h_t h_s + (e^x \sin xy + ye^x \cos xy) g_{st} + (xe^x \cos xy) h_{st}, \end{aligned}$$

in which it is understood that $x = g(s, t)$ and $y = h(s, t)$. ▲

Historical Note

Some Partial Differential Equations

> Philosophy [nature] is written in that great book which ever is before our eyes—I mean the universe—but we cannot understand it if we do not first learn the language and grasp the symbols in which it is written. The book is written in mathematical language, and the symbols are triangles, circles and other geometrical figures, without whose help it is impossible to comprehend a single word of it; without which one wanders in vain through a dark labyrinth.
>
> GALILEO

This quotation illustrates the Greek belief, again popular in the time of Galileo, that much of nature could be described using mathematics. In the latter part of the seventeenth century this thinking was dramatically reinforced when Newton used his law of gravitation to derive Kepler's three laws of celestial motion (see Section 4.1) to explain the tides, and to show that the earth was flattened at the poles. The impact of this philosophy on mathematics was substantial, and many mathematicians sought to "mathematize" nature. The extent to which mathematics pervades the physical sciences today (and, to an increasing amount, economics and the social and life sciences) is testament to the success of these endeavors. Correspondingly, the attempts to mathematize nature have often led to new mathematical discoveries.

Many of the laws of nature were described in terms of either ordinary differential equations (ODEs, equations involving the derivatives of functions of one variable alone, such as the laws of planetary motion) or partial differential equations (PDEs), that is, equations involving partial derivatives of functions. To give the reader some historical perspective and offer motivation for studying partial derivatives, we present a brief description of three of the most famous partial differential equations: the heat equation, the potential equation (or Laplace's equation), and the wave equation. (Further information on some PDEs is given in Section 8.5.)

The Heat Equation. In the early part of the nineteenth century the French mathematician Joseph Fourier (1768–1830) took up the study of heat. Heat flow had obvious applications to both industrial and scientific problems: A better understanding of it would, for example, make possible more efficient smelting of metals and would enable scientists to determine the temperature of a body given the temperature at its boundary, and to approximate the temperature of the earth's interior.

Let a homogeneous body $\mathcal{B} \subset \mathbb{R}^3$ (Figure 3.1.2) be represented by some region in 3-space. Let $T(x, y, z, t)$ denote the temperature of the body at the

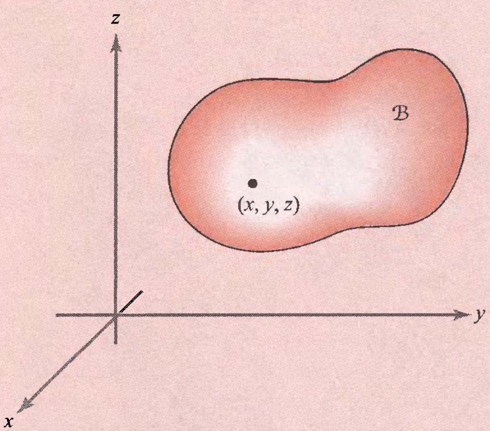

Figure 3.1.2 A homogeneous body in space.

point (x, y, z) at time t. Fourier proved, on the basis of physical principles (described in Section 8.5), that T must satisfy the partial differential equation called the ***heat equation***,

$$k\left(\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}+\frac{\partial^2 T}{\partial z^2}\right)=\frac{\partial T}{\partial t}, \tag{1}$$

where k is a constant whose value depends on the conductivity of the material comprising the body.

Fourier used this equation to solve problems in heat conduction. In fact, his investigations into the solutions of equation (1) led him to the discovery of *Fourier series*.

The Potential Equation. Consider the gravitational potential V (often called Newton's potential) of a mass m at a point (x, y, z) caused by a point mass M situated at the origin. This potential is given by $V = -GmM/r$, where $r = \sqrt{x^2 + y^2 + z^2}$. The potential V satisfies the equation

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+\frac{\partial^2 V}{\partial z^2}=0 \tag{2}$$

everywhere except at the origin, as we will check in the next chapter (see also Exercise 23). This equation is known as ***Laplace's equation***. Pierre-Simon de Laplace (1749–1827) had worked on the gravitational attraction of nonpoint masses and was the first to consider equation (2) with regard to gravitational attraction. He gave arguments (later shown to be incorrect) that equation (2) held for any body and any point whether inside or outside that body. However, Laplace was not the first person to write down equation (2). The potential equation appeared for the first time in one of Euler's major papers in 1752, "Principles of the Motions of Fluids," in which he derived the potential equation with regard to the motion of (incompressible) fluids. Euler remarked that he had no idea how to solve equation (2). Poisson later showed that if (x, y, z) lies inside an attracting body, then V satisfies the equation

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+\frac{\partial^2 V}{\partial z^2}=-4\pi\rho, \tag{3}$$

where ρ is the mass density of the attracting body. Equation (3) is now called ***Poisson's equation***. Poisson was also the first to point out the importance of this equation for problems involving electric fields. Notice that if the temperature T is constant in time, then the heat equation (1) reduces to Laplace's equation (2).

Laplace's and Poisson's equations are fundamental to many fields besides fluid mechanics, gravitational fields, and electrostatic fields. For example, they are useful for studying soap films and liquid crystals (see *The Parsimonious Universe: Shape and Form in the Natural World* by S. Hildebrandt and A. Tromba, Springer-Verlag, New York/Berlin, 1995).

THE WAVE EQUATION. The linear wave equation in space has the form

$$\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}+\frac{\partial^2 f}{\partial z^2}=c^2\frac{\partial^2 f}{\partial t^2}. \tag{4}$$

The one-dimensional wave equation

$$\frac{\partial^2 f}{\partial x^2}=c^2\frac{\partial^2 f}{\partial t^2} \tag{4'}$$

was derived in about 1727 by Johann II Bernoulli and several years later by Jean Le Rond d'Alembert in the study of how to determine the motion of a vibrating string (such as a violin string). Equation (4) became useful in the study of both vibrating bodies and elasticity. As we shall see when we consider Maxwell's equations for electromagnetism in Section 8.5, this equation also arises in the study of the propagation of electromagnetic radiation and sound waves.

EXAMPLE 6 The partial differential equation $u_t + u_{xxx} + uu_x = 0$, called the ***Korteweg–de Vries equation*** (or KdV equation, for short), describes the motion of water waves in a shallow channel.

(a) Show that for any positive constant c, the function

$$u(x,t) = 3c\,\mathrm{sech}^2\left[\tfrac{1}{2}(x-ct)\sqrt{c}\right]$$

is a solution of the Korteweg–de Vries equation. (This solution represents a traveling "hump" of water in the channel and is called a ***soliton***.)[1]

(b) How do the shape and speed of the soliton depend on c?

SOLUTION (a) We compute u_t, u_x, u_{xx}, and u_{xxx} using the chain rule and the differentiation formula $(d/dx)\,\mathrm{sech}\,x = -\,\mathrm{sech}\,x\tanh x$ from one-variable calculus.

---

[1] Solitons were first observed by J. Scott Russell around 1840 in barge canals near Edinburgh. He reported his results in *Trans. R. Soc. Edinburgh* **14** (1840): 47–109.

Letting $\alpha = (x - ct)\sqrt{c}/2$,

$$\begin{aligned} u_t &= 6c \operatorname{sech}\alpha \frac{\partial}{\partial t} \operatorname{sech}\alpha = -6c \operatorname{sech}^2\alpha \tanh\alpha \frac{\partial\alpha}{\partial t} \\ &= 3c^{5/2}\operatorname{sech}^2\alpha\tanh\alpha = c^{3/2}u\tanh\alpha. \end{aligned}$$

Also,

$$\begin{aligned} u_x &= -6c \operatorname{sech}^2\alpha\ \tanh\alpha \frac{\partial\alpha}{\partial x} \\ &= -3c^{3/2}\operatorname{sech}^2\alpha\tanh\alpha = -\sqrt{c}\, u\tanh\alpha, \end{aligned}$$

and so $u_t + cu_x = 0$ and

$$\begin{aligned} u_{xx} &= -\sqrt{c}\left[u_x \ \tanh\ \alpha + u(\operatorname{sech}^2\alpha)\frac{\sqrt{c}}{2}\right] = -\sqrt{c}(\tanh\alpha)u_x - \frac{u^2}{6} \\ &= c(\tanh^2\alpha)u - \frac{u^2}{6} = c(1-\operatorname{sech}^2\alpha)u - \frac{u^2}{6} \\ &= cu - \frac{u^2}{3} - \frac{u^2}{6} = cu - \frac{u^2}{2}. \end{aligned}$$

Thus,

$$u_{xxx} = cu_x - uu_x; \qquad \text{that is,} \qquad u_{xxx} + uu_x = cu_x.$$

Hence,

$$u_t + u_{xxx} + uu_x = u_t + cu_x = 0.$$

(b) The speed of the soliton is c, because $u(x + ct, t) = u(x, 0)$. The soliton is higher and thinner when c is larger. Its shape at time $t = 10$ is shown in Figure 3.1.3. ▲

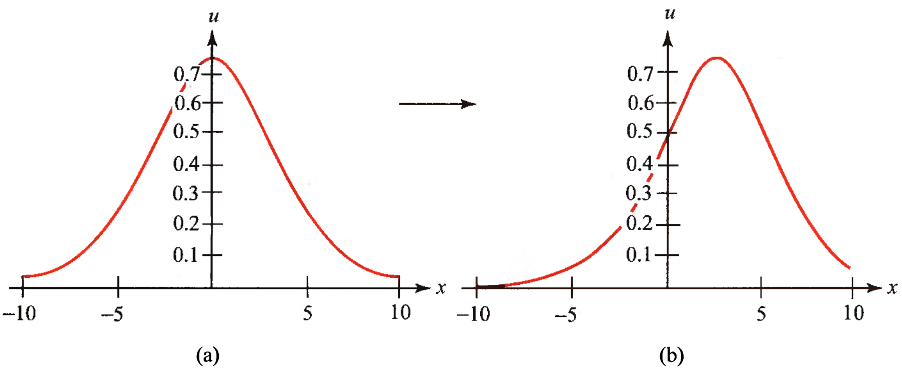

Figure 3.1.3 The graph of $u(x, t) = 3\ \operatorname{sech}^2(\sqrt{c}(x - ct)/2)$ for $c = \frac{1}{4}$ at times (a) $t = 0$ and (b) $t = 10$.

EXERCISES

In Exercises 1 to 6, compute the second partial derivatives $\partial^2 f/\partial x^2$, $\partial^2 f/\partial x\,\partial y$, $\partial^2 f/\partial y\,\partial x$, $\partial^2 f/\partial y^2$ *for each of the following functions. Verify Theorem 1 in each case.*

1. $f(x, y) = 2xy/(x^2 + y^2)^2$, on the region where $(x, y) \neq (0, 0)$

2. $f(x, y, z) = e^z + (1/x) + xe^{-y}$, on the region where $x \neq 0$

3. $f(x, y) = \cos(xy^2)$

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

5. $f(x, y) = 1/(\cos^2 x + e^{-y})$

6. $f(x, y) = \log(x - y)$

7. Find $\partial^2 z/\partial x^2$, $\partial^2 z/\partial x\,\partial y$, $\partial^2 z/\partial y\,\partial x$, and $\partial^2 z/\partial y^2$ for

(a) $z = 3x^2 + 2y^2$
(b) $z = (2x^2 + 7x^2 y)/3xy$, on the region where $x \neq 0$ and $y \neq 0$

8. Find all the second partial derivatives of

(a) $z = \sin(x^2 - 3xy)$
(b) $z = x^2 y^2 e^{2xy}$

9. Find f_{xy}, f_{yz}, f_{zx}, and f_{xyz} for

$$f(x, y, z) = x^2 y + xy^2 + yz^2.$$

10. Let $z = x^4 y^3 - x^8 + y^4$.

(a) Compute $\partial^3 z/\partial y\,\partial x\,\partial x$, $\partial^3 z/\partial x\,\partial y\,\partial x$, and $\partial^3 z/\partial x\,\partial x\,\partial y$ (also denoted $\partial^3 z/\partial x^2 \partial y$).
(b) Compute $\partial^3 z/\partial x\,\partial y\,\partial y$, $\partial^3 z/\partial y\,\partial x\,\partial y$, and $\partial^3 z/\partial y\,\partial y\,\partial x$ (also denoted $\partial^3 z/\partial y^2 \partial x$).

11. Use Theorem 1 to show that if $f(x, y, z)$ is of class C^3, then

$$\frac{\partial^3 f}{\partial x\,\partial y\,\partial z} = \frac{\partial^3 f}{\partial y\,\partial z\,\partial x}.$$

12. Verify that

$$\frac{\partial^3 f}{\partial x\,\partial y\,\partial z} = \frac{\partial^3 f}{\partial z\,\partial y\,\partial x}$$

for $f(x, y, z) = ze^{xy} + yz^3 x^2$.

13. Verify that $f_{xzw} = f_{zwx}$ for $f(x, y, z, w) = e^{xyz}\sin(xw)$.

14. If $f(x, y, z, w)$ is of class C^3, show that $f_{xzw} = f_{zwx}$.

15. Evaluate all first and second partial derivatives of the following functions:

(a) $f(x, y) = x \arctan(x/y)$
(b) $f(x, y) = \cos\sqrt{x^2 + y^2}$
(c) $f(x, y) = \exp(-x^2 - y^2)$

16. Let $w = f(x, y)$ be a function of two variables and let $x = u + v$, $y = u - v$. Show that

$$\frac{\partial^2 w}{\partial u\, \partial v} = \frac{\partial^2 w}{\partial x^2} - \frac{\partial^2 w}{\partial y^2}.$$

17. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a C^2 function and let $\mathbf{c}(t)$ be a C^2 curve in $\mathbb{R}^2$. Write a formula for the second derivative $(d^2/dt^2)((f \circ \mathbf{c})(t))$ using the chain rule twice.

18. Let $f(x, y, z) = e^{xz} \tan(yz)$ and let $x = g(s, t)$, $y = h(s, t)$, $z = k(s, t)$, and define the function $m(s, t) = f(g(s, t), h(s, t), k(s, t))$. Find a formula for m_{st} using the chain rule and verify that your answer is symmetric in s and t.

19. A function $u = f(x, y)$ with continuous second partial derivatives satisfying Laplace's equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

is called a ***harmonic function***. Show that the function $u(x, y) = x^3 - 3xy^2$ is harmonic.

20. Which of the following functions are harmonic? (See Exercise 19.)

(a) $f(x, y) = x^2 - y^2$
(b) $f(x, y) = x^2 + y^2$
(c) $f(x, y) = xy$
(d) $f(x, y) = y^3 + 3x^2 y$
(e) $f(x, y) = \sin x \cosh y$
(f) $f(x, y) = e^x \sin y$

21. Let f and g be C^2 functions of one variable. Set $\phi = f(x - t) + g(x + t)$.

(a) Prove that ϕ satisfies the wave equation: $\partial^2\phi/\partial t^2 = \partial^2\phi/\partial x^2$.
(b) Sketch the graph of ϕ against t and x if $f(x) = x^2$ and $g(x) = 0$.

22. (a) Show that function $g(x, t) = 2 + e^{-t} \sin x$ satisfies the heat equation: $g_t = g_{xx}$. [Here $g(x, t)$ represents the temperature in a metal rod at position x and time t.]

(b) Sketch the graph of g for $t \geq 0$. (HINT: Look at sections by the planes $t = 0$, $t = 1$, and $t = 2$.)

(c) What happens to $g(x, t)$ as $t \to \infty$? Interpret this limit in terms of the behavior of heat in the rod.

23. Show that Newton's potential $V = -GmM/r$ satisfies Laplace's equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0 \qquad \text{for} \qquad (x, y, z) \neq (0, 0, 0).$$

24. Let

$$f(x, y) = \begin{cases} xy(x^2 - y^2)/(x^2 + y^2), & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0) \end{cases}$$

(see Figure 3.1.4).

(a) If $(x, y) \neq (0, 0)$, calculate $\partial f/\partial x$ and $\partial f/\partial y$.
(b) Show that $(\partial f/\partial x)(0, 0) = 0 = (\partial f/\partial y)(0, 0)$.
(c) Show that $(\partial^2 f/\partial x\,\partial y)(0, 0) = 1$, $(\partial^2 f/\partial y\,\partial x)(0, 0) = -1$.
(d) What went wrong? Why are the mixed partials not equal?

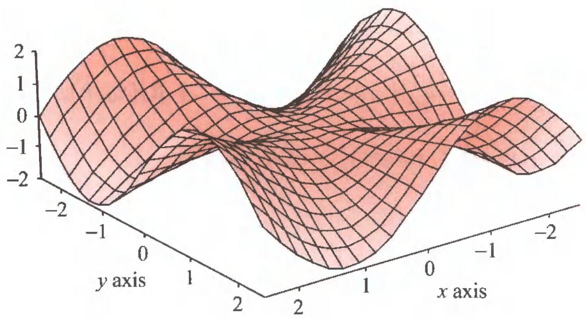

Figure 3.1.4 The graph of the function in Exercise 24.

3.2 Taylor's Theorem

When we introduced the derivative in Chapter 2, we saw that the *linear approximation* of a function played an essential role for a geometric reason—finding the equation of a tangent plane—as well as an analytic reason—finding approximate values of functions. Taylor's theorem deals with the important issue of finding *quadratic and higher-order approximations*.

Taylor's theorem is a central tool for finding accurate numerical approximations of functions, and as such plays an important role in many areas of applied and computational mathematics. We shall use it in the next section to develop the second derivative test for maxima and minima of functions of several variables.

The strategy used to prove Taylor's theorem is to reduce it to the one-variable case by probing a function of many variables along lines of the form $\mathbf{l}(t) = \mathbf{x}_0 + t\mathbf{h}$

emanating from a point $\mathbf{x}_0$ and heading in the direction $\mathbf{h}$. Thus, it will be useful for us to begin by reviewing Taylor's theorem from one-variable calculus.

Single-Variable Taylor Theorem

When recalling a theorem from an earlier course, it is helpful to ask these basic questions: What is the main point of the theorem? What are the key ideas in the proof? Can I understand the result better the second time around?

The main point of the single-variable Taylor theorem is to find approximations of a function near a given point that are accurate to a higher order than the linear approximation. The key idea in the proof is to use the *fundamental theorem of calculus*, followed by *integration by parts*. In fact, just by recalling these basic ideas, one can reconstruct the entire proof. Thinking in this way will help organize all the pieces that need to come together to develop a mastery of Taylor approximations of functions of one and several variables.

For a smooth function $f\colon \mathbb{R} \to \mathbb{R}$ of one variable, Taylor's theorem asserts that:

$$f(x_0+h) = f(x_0) + f'(x_0)\cdot h + \frac{f''(x_0)}{2}h^2 + \cdots + \frac{f^{(k)}(x_0)}{k!}h^k + R_k(x_0, h), \quad (1)$$

where

$$R_k(x_0, h) = \int_{x_0}^{x_0+h} \frac{(x_0 + h - \tau)^k}{k!} f^{k+1}(\tau)\, d\tau$$

is the remainder. For small h, this remainder is small to order k in the sense that

$$\lim_{h\to 0} \frac{R_k(x_0, h)}{h^k} = 0. \quad (2)$$

In other words, $R_k(x_0, h)$ is small compared to the already small quantity h^k.

The preceding is the formal statement of Taylor's theorem. What about the proof? As promised, we begin with the fundamental theorem of calculus, written in the form:

$$f(x_0 + h) = f(x_0) + \int_{x_0}^{x_0+h} f'(\tau)\, d\tau.$$

Next, we write $d\tau = -d(x_0 + h - \tau)$ and integrate parts[2] to give:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \int_{x_0}^{x_0+h} f''(\tau)(x_0 + h - \tau)\, d\tau,$$

which is the first-order Taylor formula. Integrating by parts again:

$$\begin{aligned}
&\int_{x_0}^{x_0+h} f''(\tau)(x_0 + h - \tau)\, d\tau \\
&\quad = -\frac{1}{2}\int_{x_0}^{x_0+h} f''(\tau)\, d(x_0 + h - \tau)^2 \\
&\quad = \frac{1}{2} f''(x_0)h^2 + \frac{1}{2}\int_{x_0}^{x_0+h} f'''(\tau)(x_0 + h - \tau)^2\, d\tau,
\end{aligned}$$

which, when substituted into the preceding formula, gives the ***second-order Taylor formula***:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2} f''(x_0)h^2 + \frac{1}{2}\int_{x_0}^{x_0+h} f'''(\tau)(x_0 + h - \tau)^2\, d\tau.$$

This is Taylor's theorem for $k = 2$.

Taylor's theorem for general k proceeds by repeated integration by parts. The statement (2) that $R_k(x_0, h)/h^k \to 0$ as $h \to 0$ is seen as follows. For τ in the interval $[x_0, x_0 + h]$, we have $|x_0 + h - \tau| \le |h|$ and $f^{k+1}(\tau)$, being continuous, is bounded; say, $|f^{k+1}(\tau)| \le M$. Then:

$$|R_k(x_0, h)| = \left| \int_{x_0}^{x_0+h} \frac{(x_0 + h - \tau)^k}{k!} f^{k+1}(\tau)\, d\tau \right| \le \frac{|h|^{k+1}}{k!} M$$

and, in particular, $|R_k(x_0, h)/h^k| \le |h|\, M/k! \to 0$ as $h \to 0$.

---

[2]Recall that integration by parts (the product rule for the derivative read backwards) reads as:

$$\int_a^b u\, dv = uv|_a^b - \int_a^b v\, du.$$

Here we choose $u = f'(\tau)$ and $v = x_0 + h - \tau$.

Taylor's Theorem for Many Variables

Our next goal in this section is to prove an analogous theorem that is valid for functions of several variables. We already know a first-order version, that is, when $k = 1$. Indeed, if $f\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$ and we define

$$R_1(\mathbf{x}_0, \mathbf{h}) = f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - [\mathbf{D}f(\mathbf{x}_0)](\mathbf{h}),$$

so that

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + [\mathbf{D}f(\mathbf{x}_0)](\mathbf{h}) + R_1(\mathbf{x}_0, \mathbf{h}),$$

then by the definition of differentiability,

$$\frac{|R_1(\mathbf{x}_0, \mathbf{h})|}{\|\mathbf{h}\|} \to 0 \qquad \text{as} \qquad \mathbf{h} \to 0,$$

that is, $R_1(\mathbf{x}_0, \mathbf{h})$ vanishes to first order at $\mathbf{x}_0$. In summary, we have:

THEOREM 2: First-Order Taylor Formula Let $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{x}_0 \in U$. Then

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(\mathbf{x}_0) + R_1(\mathbf{x}_0, \mathbf{h}),$$

where $R_1(\mathbf{x}_0, \mathbf{h})/\|\mathbf{h}\| \to 0$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbb{R}^n$.

The second-order version is as follows:

THEOREM 3: Second-Order Taylor Formula Let $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ have continuous partial derivatives of third order.[3] Then we may write

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(\mathbf{x}_0) + \frac{1}{2} \sum_{i,j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i\, \partial x_j}(\mathbf{x}_0) + R_2(\mathbf{x}_0, \mathbf{h}),$$

where $R_2(\mathbf{x}_0, \mathbf{h})/\|\mathbf{h}\|^2 \to 0$ as $\mathbf{h} \to \mathbf{0}$ and the second sum is over all i's and j's between 1 and n (so there are n^2 terms).

[3] For the statement of the theorem as given here, f actually needs only to be of class C^2, but for a convenient form of the remainder we assume f is of class C^3.

Notice that this result can be written in matrix form as

$$f(\mathbf{x}_0+\mathbf{h}) = f(\mathbf{x}_0) + \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right)\begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

$$+\frac{1}{2}(h_1, \dots, h_n)\begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & & & \\ \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n\,\partial x_n} \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix},$$

$$+ R_2(\mathbf{x}_0, \mathbf{h}),$$

where the derivatives of f are evaluated at $\mathbf{x}_0$.

In the course of the proof of the Theorem 3, we shall obtain a useful explicit formula for the remainder, as in the single-variable theorem.

PROOF OF THEOREM 3 Let $g(t) = f(\mathbf{x}_0 + t\mathbf{h})$ with $\mathbf{x}_0$ and $\mathbf{h}$ fixed, which is a C^3 function of t. Now apply the single-variable Taylor theorem (1) to g, with $k = 2$ to obtain

$$\left.\begin{aligned} g(1) &= g(0) + g'(0) + \frac{g''(0)}{2!} + R_2, \\ \text{where}\qquad & \\ R_2 &= \int_0^1 \frac{(t-1)^2}{2!} g'''(t)\,dt. \end{aligned}\right\} \tag{3}$$

By the chain rule,

$$g'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + t\mathbf{h})\mathbf{h}_i; \qquad g''(t) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i\,\partial x_j}(\mathbf{x}_0 + t\mathbf{h})h_i h_j,$$

and

$$g'''(t) = \sum_{i,j,k=1}^{n} \frac{\partial^3 f}{\partial x_i\,\partial x_j\,\partial x_k}(\mathbf{x}_0 + t\mathbf{h})h_i h_j h_k.$$

Writing $R_2 = R_2(\mathbf{x}_0, \mathbf{h})$ we have thus proved:

$$\left.\begin{aligned} &f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(\mathbf{x}_0) + \frac{1}{2} \sum_{i,j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i\, \partial x_j}(\mathbf{x}_0) + R_2(\mathbf{x}_0, \mathbf{h}), \\ &\text{where} \\ &\qquad R_2(\mathbf{x}_0, \mathbf{h}) = \sum_{i,j,k=1}^{n} \int_0^1 \frac{(t-1)^2}{2} \frac{\partial^3 f}{\partial x_i\, \partial x_j\, \partial x_k}(\mathbf{x}_0 + t\mathbf{h}) h_i h_j h_k \, dt. \end{aligned}\right\} \quad (4)$$

The integrand is a continuous function of t and is therefore bounded by a positive constant C on a small neighborhood of $\mathbf{x}_0$ (because it has to be close to its value at $\mathbf{x}_0$). Also note that $|h_i| \le \|\mathbf{h}\|$, for $\|\mathbf{h}\|$ small, and so

$$|R_2(\mathbf{x}_0, \mathbf{h})| \le \|\mathbf{h}\|^3 C.$$

In particular,

$$\frac{|R_2(\mathbf{x}_0, \mathbf{h})|}{\|\mathbf{h}\|^2} \le \|\mathbf{h}\| C \to 0 \qquad \text{as} \qquad \mathbf{h} \to \mathbf{0},$$

as required by the theorem.

The proof of Theorem 2 follows analogously from the Taylor formula (1) with $k = 1$. A similar argument for R_1 shows that $|R_1(\mathbf{x}_0, \mathbf{h})|/\|\mathbf{h}\| \to 0$ as $\mathbf{h} \to \mathbf{0}$, although this also follows directly from the definition of differentiability. ■

Forms of the Remainder In Theorem 2,

$$R_1(\mathbf{x}_0, \mathbf{h}) = \sum_{i,j=1}^{n} \int_0^1 (1-t) \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0 + t\mathbf{h}) h_i h_j \, dt = \sum_{i,j=1}^{n} \frac{1}{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{c}_{ij}) h_i h_j, \tag{5}$$

where $\mathbf{c}_{ij}$ lies somewhere on the line joining $\mathbf{x}_0$ to $\mathbf{x}_0 + \mathbf{h}$.

In Theorem 3,

$$\begin{aligned} R_2(\mathbf{x}_0, \mathbf{h}) &= \sum_{i,j,k=1}^{n} \int_0^1 \frac{(t-1)^2}{2} \frac{\partial^3 f}{\partial x_i\, \partial x_j\, \partial x_k}(\mathbf{x}_0 + t\mathbf{h}) h_i h_j h_k \, dt \\ &= \sum_{i,j,k=1}^{n} \frac{1}{3!} \frac{\partial^3 f}{\partial x_i\, \partial x_j\, \partial x_k}(\mathbf{c}_{ijk}) h_i h_j h_k, \end{aligned} \tag{5$'$}$$

where $\mathbf{c}_{ijk}$ lies somewhere on the line joining $\mathbf{x}_0$ to $\mathbf{x}_0 + \mathbf{h}$.

The formulas involving $\mathbf{c}_{ij}$ and $\mathbf{c}_{ijk}$ (called Lagrange's form of the remainder) are obtained by making use of the *second mean-value theorem for integrals*. This states that

$$\int_a^b h(t)g(t)\,dt = h(c)\int_a^b g(t)\,dt,$$

provided h and g are continuous and $g \geq 0$ on $[a, b]$; *here c is some number between* a and b.[4] This is applied in formula (5) for the explicit form of the remainder with $h(t) = (\partial^2 f/\partial x_i \partial x_j)(\mathbf{x}_0 + t\mathbf{h})$ and $g(t) = 1 - t$.

The third-order Taylor formula is

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(\mathbf{x}_0) + \frac{1}{2}\sum_{i,j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i\,\partial x_j}(\mathbf{x}_0)$$
$$+ \frac{1}{3!}\sum_{i,j,k=1}^{n} h_i h_j h_k \frac{\partial^3 f}{\partial x_i\,\partial x_j\,\partial x_k}(\mathbf{x}_0) + R_3(\mathbf{x}_0, \mathbf{h}),$$

where $R_3(\mathbf{x}_0, \mathbf{h})/\|\mathbf{h}\|^3 \to 0$ as $\mathbf{h} \to \mathbf{0}$, and so on. The general formula can be proved by induction, using the method of proof already given.

EXAMPLE 1 Compute the second-order Taylor formula for the function $f(x, y) = \sin(x + 2y)$, about the point $\mathbf{x}_0 = (0, 0)$.

SOLUTION Notice that

$$f(0, 0) = 0,$$

$$\frac{\partial f}{\partial x}(0, 0) = \cos(0 + 2\cdot 0) = 1, \qquad \frac{\partial f}{\partial y}(0, 0) = 2\cos(0 + 2\cdot 0) = 2,$$

$$\frac{\partial^2 f}{\partial x^2}(0, 0) = 0, \qquad \frac{\partial^2 f}{\partial y^2}(0, 0) = 0, \qquad \frac{\partial^2 f}{\partial x\,\partial y}(0, 0) = 0.$$

Thus,

$$f(\mathbf{h}) = f(h_1, h_2) = h_1 + 2h_2 + R_2(\mathbf{0}, \mathbf{h}),$$

---

[4]*Proof* If $g = 0$, the result is clear, so we can suppose $g \neq 0$; thus, we can assume $\int_a^b g(t)\,dt > 0$. Let M and m be the maximum and minimum values of h, achieved at t_M and t_m, respectively. Because $g(t) \geq 0$,

$$m\int_a^b g(t)\,dt \leq \int_a^b h(t)g(t)\,dt \leq M\int_a^b g(t)\,dt.$$

Thus, $\left(\int_a^b h(t)g(t)\,dt\right)\Big/\left(\int_a^b g(t)\,dt\right)$ lies between $m = h(t_m)$ and $M = h(t_M)$ and therefore, by the intermediate-value theorem, equals $h(c)$ for some intermediate c. ■

where

$$\frac{R_2(\mathbf{0}, \mathbf{h})}{\|\mathbf{h}\|^2} \to 0 \quad \text{as} \quad \mathbf{h} \to \mathbf{0}. \quad \blacktriangle$$

EXAMPLE 2 Compute the second-order Taylor formula for $f(x, y) = e^x \cos y$ about the point $x_0 = 0$, $y_0 = 0$.

SOLUTION Here

$$f(0,0) = 1, \qquad \frac{\partial f}{\partial x}(0,0) = 1, \qquad \frac{\partial f}{\partial y}(0,0) = 0,$$

$$\frac{\partial^2 f}{\partial x^2}(0,0) = 1, \qquad \frac{\partial^2 f}{\partial y^2}(0,0) = -1, \qquad \frac{\partial^2 f}{\partial x\, \partial y}(0,0) = 0,$$

and so

$$f(\mathbf{h}) = f(h_1, h_2) = 1 + h_1 + \tfrac{1}{2}h_1^2 - \tfrac{1}{2}h_2^2 + R_2(\mathbf{0}, \mathbf{h}),$$

where

$$\frac{R_2(\mathbf{0}, \mathbf{h})}{\|\mathbf{h}\|^2} \to 0 \quad \text{as} \quad \mathbf{h} \to \mathbf{0}. \quad \blacktriangle$$

In the case of functions of one variable, one can expand $f(x)$ in an infinite power series, called the ***Taylor series***:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{f''(x_0)h^2}{2} + \cdots + \frac{f^{(k)}(x_0)h^k}{k!} + \cdots,$$

provided one can show that $R_k(x_0, h) \to 0$ as $k \to \infty$. Similarly, for functions of several variables the preceding terms are replaced by the corresponding ones involving partial derivatives, as we have seen in Theorem 3. Again, one can represent such a function by its Taylor series provided one can show that $R_k \to 0$ as $k \to \infty$. This point is examined further in Exercise 7.

EXAMPLE 3 Find the first- and second-order Taylor approximations to $f(x, y) = \sin(xy)$ at the point $(x_0, y_0) = (1, \pi/2)$.

SOLUTION Here

$$f(x_0, y_0) = \sin(x_0 y_0) = \sin(\pi/2) = 1$$

$$f_x(x_0, y_0) = y_0 \cos(x_0 y_0) = \frac{\pi}{2}\cos(\pi/2) = 0$$

$$f_y(x_0, y_0) = x_0 \cos(x_0 y_0) = \cos(\pi/2) = 0$$

$$f_{xx}(x_0, y_0) = -y_0^2 \sin(x_0 y_0) = -\frac{\pi^2}{4} \sin(\pi/2) = -\pi^2/4$$

$$f_{xy}(x_0, y_0) = \cos(x_0 y_0) - x_0 y_0 \sin(x_0 y_0) = -\frac{\pi}{2} \sin(\pi/2) = -\pi/2$$

$$f_{yy}(x_0, y_0) = -x_0^2 \sin(x_0 y_0) = -\sin(\pi/2) = -1.$$

Thus, the linear (first-order) approximation is

$$\begin{aligned} l(x, y) &= f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ &= 1 + 0 + 0 = 1, \end{aligned}$$

and the second-order (or quadratic) approximation is

$$\begin{aligned} g(x, y) &= 1 + 0 + 0 + \frac{1}{2}\left(-\frac{\pi^2}{4}\right)(x-1)^2 + \left(-\frac{\pi}{2}\right)(x-1)\left(y - \frac{\pi}{2}\right) \\ &\quad + \frac{1}{2}(-1)\left(y - \frac{\pi}{2}\right)^2 \\ &= 1 - \frac{\pi^2}{8}(x-1)^2 - \frac{\pi}{2}(x-1)\left(y - \frac{\pi}{2}\right) - \frac{1}{2}\left(y - \frac{\pi}{2}\right)^2. \end{aligned}$$

See Figure 3.2.1. ▲

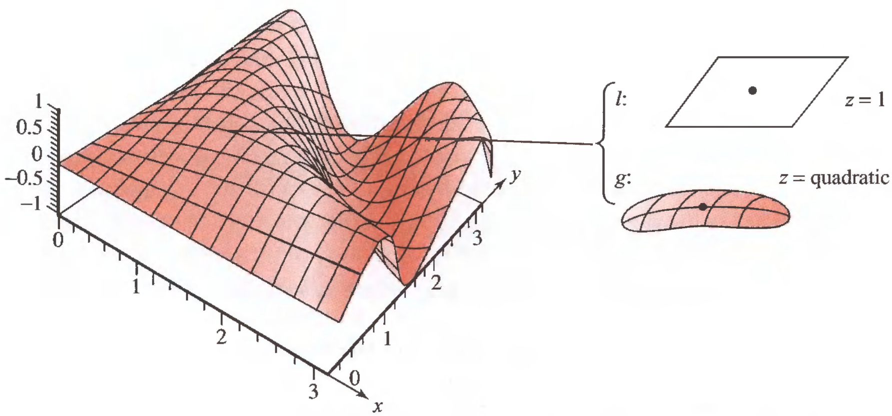

Figure 3.2.1 The linear and quadratic approximations to $z = \sin(xy)$ near $(1, \pi/2)$.

EXAMPLE 4 Find linear and quadratic approximations to the expression $(3.98 - 1)^2/(5.97 - 3)^2$. Compare with the exact value.

SOLUTION Let $f(x, y) = (x - 1)^2/(y - 3)^2$. The desired expression is close to $f(4, 6) = 1$. To find the approximations, we differentiate:

$$f_x = \frac{2(x-1)}{(y-3)^2}, \qquad f_y = \frac{-2(x-1)^2}{(y-3)^3}$$

$$f_{xy} = f_{yx} = \frac{-4(x-1)}{(y-3)^3}, \qquad f_{xx} = \frac{2}{(y-3)^2}, \qquad f_{yy} = \frac{6(x-1)^2}{(y-3)^4}.$$

At the point of approximation, we have

$$f_x(4, 6) = \frac{2}{3}, \qquad f_y = -\frac{2}{3}, \qquad f_{xy} = f_{yx} = -\frac{4}{9}, \qquad f_{xx} = \frac{2}{9}, \qquad f_{yy} = \frac{2}{3}.$$

The linear approximation is then

$$1 + \frac{2}{3}(-0.02) - \frac{2}{3}(-0.03) = 1.00666.$$

The quadratic approximation is

$$1 + \frac{2}{3}(-0.02) - \frac{2}{3}(-0.03) + \frac{2}{9}\frac{(-0.02)^2}{2} - \frac{4}{9}(-0.02)(-0.03) + \frac{2}{3}\frac{(-0.03)^2}{2}$$
$$= 1.00674.$$

The "exact" value using a calculator is 1.00675. ▲

EXERCISES

In each of Exercises 1 to 6, determine the second-order Taylor formula for the given function about the given point (x_0, y_0).

1. $f(x, y) = (x + y)^2$, where $x_0 = 0$, $y_0 = 0$

2. $f(x, y) = 1/(x^2 + y^2 + 1)$, where $x_0 = 0$, $y_0 = 0$

3. $f(x, y) = e^{x+y}$, where $x_0 = 0$, $y_0 = 0$

4. $f(x, y) = e^{-x^2-y^2} \cos(xy)$, where $x_0 = 0$, $y_0 = 0$

5. $f(x, y) = \sin(xy) + \cos(xy)$, where $x_0 = 0$, $y_0 = 0$

6. $f(x, y) = e^{(x-1)^2} \cos y$, where $x_0 = 1$, $y_0 = 0$

7. (Challenging) A function $f\colon \mathbb{R} \to \mathbb{R}$ is called an ***analytic*** function provided

$$f(x + h) = f(x) + f'(x)h + \cdots + \frac{f^{(k)}(x)}{k!}h^k + \cdots$$

[i.e., the series on the right-hand side converges and equals $f(x + h)$].

(a) Suppose f satisfies the following condition: On any closed interval $[a, b]$, there is a constant M such that for all $k = 1, 2, 3, \ldots$, $|f^{(k)}(x)| \leq M^k$ for all $x \in [a, b]$. Prove that f is analytic.

(b) Let $f(x) = \begin{cases} e^{-1/x} & x > 0 \\ 0 & x \leq 0. \end{cases}$

Show that f is a C^∞ function, but f is not analytic.

(c) Give a definition of analytic functions from $\mathbb{R}^n$ to $\mathbb{R}$. Generalize the proof of part (a) to this class of functions.

(d) Develop $f(x, y) = e^{x+y}$ in a power series about $x_0 = 0$, $y_0 = 0$.

3.3 Extrema of Real-Valued Functions

Historical Note

As we saw in the book's Historical introduction, the early Greeks sought to mathematize nature and to find, as in the geometric Ptolemaic model of planetary motion, mathematical laws governing the universe. With the revival of Greek learning during the Renaissance, this point of view again took hold and the search for these laws recommenced. In particular, the question was raised as to whether there was *one* law, one mathematical principle that governed and superseded all others, a principle that the Creator used in His Grand Design of the Universe.

Maupertuis's Principle. In 1744, the French scientist Pierre-Louis de Maupertuis (see Figure 3.3.1) put forth his grand scheme of the world. The "metaphysical principle" of Maupertuis is the assumption that nature always operates with the greatest possible economy. In short, physical laws are a consequence of a principle of "economy of means"; nature always acts in such a way as to minimize some quantity that Maupertuis called the *action.* Action was nothing more than the expenditure of energy over time, or energy × time. In applications, the type of energy changes with each case. For example, physical systems often try to "rearrange themselves" to have a minimum energy—such as a ball rolling from a mountain peak to a valley, or the primordial irregular Earth assuming a more nearly spherical shape. As another example, the spherical shape of soap bubbles is connected with the fact that spheres are the surfaces of least area containing a fixed volume.

We state Maupertuis's principle formally as: *Nature always minimizes action.* Maupertuis saw in this principle an expression of the wisdom of the Supreme Being, of God, according to which everything in nature is

Figure 3.3.1 Pierre-Louis de Maupertuis (1698–1759).

performed in the most economical way. He wrote:

> What satisfaction for the human spirit that, in contemplating these laws which contain the principle of motion and of rest for all bodies in the universe, he finds the proof of existence of Him who governs the world.

Maupertuis indeed believed that he had discovered God's fundamental law, the very secret of Creation itself, but he was actually not the first person to pose this principle.

In 1707, Leibniz wrote down the principle of least action in a letter to Johann Bernoulli, which became lost until 1913, when it was discovered in Germany's Gotha library. For Leibniz, this principle was a natural outgrowth of his great philosophical treatise *The Theodicy*, in which he argues that God may indeed think of all possible worlds, but would want to create only the best among them; and hence our world is necessarily the *best of all possible worlds*.

Action, as defined by Leibniz, was motivated by the following reasoning, used in his letter. Think of a hiker walking along a road, and consider how to describe his action. If he travels 2 kilometers in 1 hour, you would say that he has carried out twice as much action as he would if he traveled 2 kilometers in 2 hours. However, you would also say that he carries out twice as much action in traveling 2 kilometers in 2 hours as he would in traveling 1 kilometer in 1 hour. Altogether then, our hiker, by walking 2 kilometers in 1 hour, carries out 4 times as much action as he would in traveling 1 kilometer in 1 hour.

Using this intuitive idea, Maupertuis defined action as the product of distance, velocity, and mass:

$$\text{Action} = \text{Mass} \times \text{Distance} \times \text{Velocity}.$$

Mass is included in this definition to account for the hiker's backpack. Moreover, according to Leibniz, the kinetic energy E is given by the formula:

$$E = \frac{1}{2} \times \text{Mass} \times (\text{Velocity})^2.$$

So action has the same physical dimension as

$$\text{Energy} \times \text{Time},$$

because velocity is distance divided by time.

Principle of Least Action. In the 250 years after Maupertuis formulated his principle, this *principle of least action* has been found to be a "theoretical basis" for Newton's law of gravity, Maxwell's equations for electromagnetism, Schrödinger's equation of quantum mechanics, and Einstein's field equation in general relativity.

Max Planck (see Figure 3.3.2), one of the greatest scientists of the modern era and the discoverer of the "quantization" of nature, was also a profound believer in the mathematical design of the universe. On June 29, 1922, on "Leibniz Day" in Berlin, Germany, just a few years after World War I, with all its terrible carnage, Planck delivered an address honoring this great scholar.

What follows are some excerpts from Planck's remarks:

> *The Theodicy* culminates with the statement that whatever occurs in our world, in the large as in the small, in nature as in spiritual life, is once and for all regulated by divine reason, and in such a way that our world is the best among possible worlds.
>
> Would Leibniz reaffirm this statement even today, in view of the misery of the present time, in view of the bitter failure of many efforts not immediately aimed at material gain, in view of the

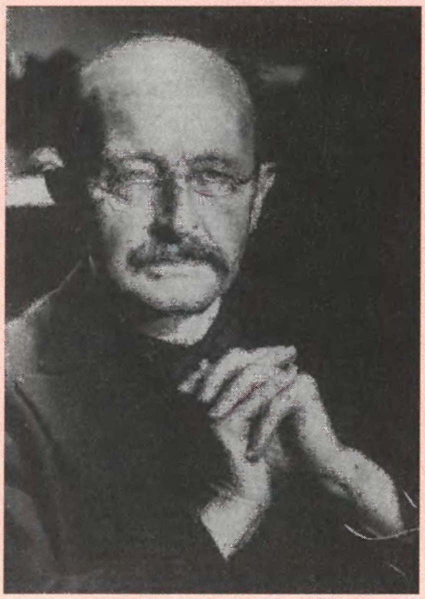

Figure 3.3.2 Max Planck (1858–1947).

undeniable fact that the imagined general harmony of people today seems to be further away from its realization than ever? No doubt, we should have to answer this question in the affirmative, even if we did not know that Leibniz never ceased to earnestly occupy himself until his last years despite an adverse fate and disappointments of all kinds, and we shall hardly err in assuming that it was exactly the *Theodicy* that gave him support and comfort in the most sorrowful days of his life. This once again is a touching example of the old truth that our most profound and most sacred principles are firmly rooted in our innermost soul, independent of experiences in the outer world.

Modern science, in particular under the influence of the development of the notion of causality, has moved far away from Leibniz's teleological point of view. Science has abandoned the assumption of a special, anticipating reason, and it considers each event in the natural and spiritual world, at least in principle, as reducible to prior states. But still we notice a fact, particularly in the most exact science, which, at least in this context, is most surprising. Present-day physics, as far as it is theoretically organized, is completely governed by a system of space–time differential equations which state that each process in nature is totally determined by the events which occur in its immediate temporal and spatial neighborhood. This entire rich system of differential equations, though they differ in detail, since they refer to mechanical, electric, magnetic, and thermal processes, is now completely contained in a single dictum—*the principle of least action*. This, in short, states that, of all possible processes, the only ones that actually occur are those that involve minimum expenditure of action. As we can see, only a short step is required to recognize in the preference for the smallest quantity of action the ruling of divine reason, and thus to discover a part of Leibniz's teleological ordering of the universe.[5]

In present-day physics the principle of least action plays a relatively minor role. It does not quite fit into the framework of present theories. Of course, admittedly it is a correct statement; yet usually it serves not as the foundation of the theory, but as a true but dispensable appendix, because present theoretical physics is entirely tailored to the principle of infinitesimal local effects, and sees extensions to larger spaces and times as an unnecessary and uneconomical complication of the method of treatment. Hence, physics is inclined to view the principle of least action more as a formal and accidental curiosity than as a pillar of physical knowledge.

There is much more to the story of the least action principle, which we will revisit in Section 4.1.

[5]For more information and history, consult S. Hildebrandt and A. J. Tromba, *The Parsimonious Universe: Shape and Form in the Natural World*, Springer-Verlag, New York/Berlin, 1995.

Maxima and Minima for Functions of n-Variables

As the previous remarks show, for Leibniz, Euler, and Maupertuis, and for much of modern science as well, all in nature is a consequence of some maximum or minimum principle. To make such grand schemes—as well as some that are more down to earth—effective, one must first learn the techniques of how to find maxima and minima of functions of n variables.

Extreme Points

Among the most basic geometric features of the graph of a function are its extreme points, at which the function attains its greatest and least values. In this section, we derive a method for determining these points. In fact, the method locates local extrema as well. These are points at which the function attains a maximum or minimum value relative only to nearby points. Let us begin by defining our terms.

DEFINITION If $f: U \subset \mathbb{R}^n \to \mathbb{R}$ is a given scalar function, a point $\mathbf{x}_0 \in U$ is called a ***local minimum*** of f if there is a neighborhood V of $\mathbf{x}_0$ such that for all points $\mathbf{x}$ in V, $f(\mathbf{x}) \geq f(\mathbf{x}_0)$. (See Figure 3.3.3.) Similarly, $\mathbf{x}_0 \in U$ is a ***local maximum*** if there is a neighborhood V of $\mathbf{x}_0$ such that $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ for all $\mathbf{x} \in V$. The point

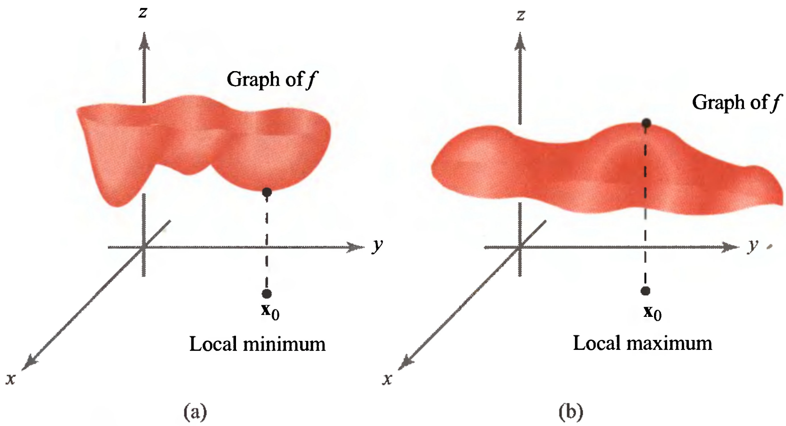

Figure 3.3.3 (a) Local minimum and (b) local maximum points for a function of two variables.

$\mathbf{x}_0 \in U$ is said to be a ***local***, or ***relative***, ***extremum*** if it is either a local minimum or a local maximum. A point $\mathbf{x}_0$ is a ***critical point*** of f if either f is *not differentiable at* $\mathbf{x}_0$, or if it is, $\mathbf{D}f(\mathbf{x}_0) = \mathbf{0}$. A critical point that is not a local extremum is called a ***saddle point***.[6]

First Derivative Test for Local Extrema

The location of extrema is based on the following fact, which should be familiar from one-variable calculus (the case $n = 1$): *Every extremum is a critical point.*

THEOREM 4: First Derivative Test for Local Extrema If $U \subset \mathbb{R}^n$ is open, the function $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable, and $\mathbf{x}_0 \in U$ is a local extremum, then $\mathbf{D}f(\mathbf{x}_0) = \mathbf{0}$; that is, $\mathbf{x}_0$ is a critical point of f.

PROOF Suppose that f achieves a local maximum at $\mathbf{x}_0$. Then for any $\mathbf{h} \in \mathbb{R}^n$, the function $g(t) = f(\mathbf{x}_0 + t\mathbf{h})$ has a local maximum at $t = 0$. Thus, from one-variable calculus $g'(0) = 0$.[7] On the other hand, by the chain rule,

$$g'(0) = [\mathbf{D}f(\mathbf{x}_0)]\mathbf{h}.$$

Thus, $[\mathbf{D}f(\mathbf{x}_0)]\mathbf{h} = 0$ for every $\mathbf{h}$, and so $\mathbf{D}f(\mathbf{x}_0) = \mathbf{0}$. The case in which f achieves a local minimum at $\mathbf{x}_0$ is entirely analogous. ■

---

[6]The term "saddle point" is sometimes not used this generally; we shall discuss saddle points further in the subsequent development.

[7]Recall the proof from one-variable calculus: Because $g(0)$ is a local maximum, $g(t) \le g(0)$ for small $t > 0$, so $g(t) - g(0) \le 0$, and hence $g'(0) = \lim_{t\to 0+} (g(t) - g(0))/t \le 0$, where $\lim_{t\to 0+}$ means the limit as $t \to 0$, $t > 0$. For small $t < 0$, we similarly have $g'(0) = \lim_{t\to 0-} (g(t) - g(0))/t \ge 0$. Therefore, $g'(0) = 0$.

If we remember that $\mathbf{D}f(\mathbf{x}_0) = \mathbf{0}$ means that all the components of $\mathbf{D}f(\mathbf{x}_0)$ are zero, we can rephrase the result of Theorem 4: If $\mathbf{x}_0$ is a local extremum, then

$$\frac{\partial f}{\partial x_i}(\mathbf{x}_0) = 0, \qquad i = 1, \dots, n;$$

that is, each partial derivative is zero at $\mathbf{x}_0$. In other words, $\nabla f(\mathbf{x}_0) = \mathbf{0}$, where ∇f is the gradient of f.

If we seek to find the extrema or local extrema of a function, then Theorem 4 states that we should look among the critical points. Sometimes these can be tested by inspection, but usually we use tests (to be developed below) analogous to the second-derivative test in one-variable calculus.

EXAMPLE 1 Find the maxima and minima of the function $f: \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x, y) = x^2 + y^2$. (Ignore the fact that this example can be done by inspection.)

SOLUTION We first identify the critical points of f by solving the two equations $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$, for x and y. But

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

so the only critical point is the origin $(0, 0)$, where the value of the function is zero. Because $f(x, y) \geq 0$, this point is a relative minimum—in fact, an absolute, or global, minimum—of f. Because $(0, 0)$ is the only critical point, there are no maxima. ▲

EXAMPLE 2 Consider the function $f: \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x^2 - y^2$. Ignoring for the moment that this function has a saddle and no extrema, apply the method of Theorem 4 for the location of extrema.

SOLUTION As in Example 1, we find that f has only one critical point, at the origin, and the value of f there is zero. Examining values of f directly for points near the origin, we see that $f(x, 0) \geq f(0, 0)$ and $f(0, y) \leq f(0, 0)$, with strict inequalities when $x \neq 0$ and $y \neq 0$. Because x or y can be taken arbitrarily small, the origin cannot be either a relative minimum or a relative maximum (so it is a saddle point). Therefore, this function can have no relative extrema (see Figure 3.3.4). ▲

EXAMPLE 3 Find all the critical points of $z = x^2y + y^2x$.

SOLUTION Differentiating, we obtain

$$\frac{\partial z}{\partial x} = 2xy + y^2, \qquad \frac{\partial z}{\partial y} = 2xy + x^2.$$

Equating the partial derivatives to zero yields

$$2xy + y^2 = 0, \qquad 2xy + x^2 = 0.$$

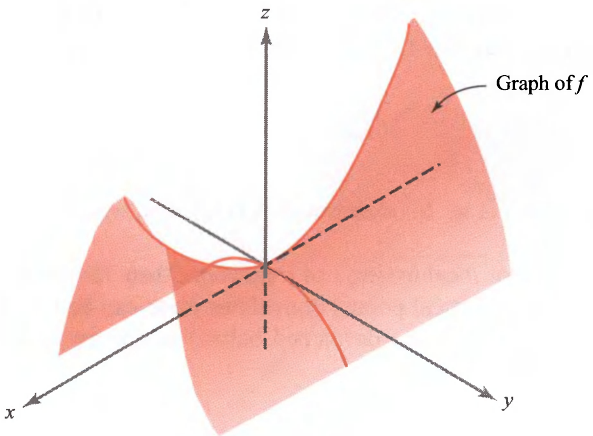

Figure 3.3.4 A function of two variables with a saddle point.

Subtracting, we obtain $x^2 = y^2$. Thus, $x = \pm y$. Substituting $x = +y$ in the first of the two preceding equations, we find that

$$2y^2 + y^2 = 3y^2 = 0,$$

so that $y = 0$ and thus $x = 0$. If $x = -y$, then

$$-2y^2 + y^2 = -y^2 = 0,$$

so $y = 0$ and therefore $x = 0$. Hence, the only critical point is $(0, 0)$. For $x = y$, $z = 2x^3$, which is both positive and negative for x near zero. Thus, $(0, 0)$ is not a relative extremum. ▲

EXAMPLE 4 Refer to Figure 3.3.5, a computer-drawn graph of the function $z = 2(x^2 + y^2)\, e^{-x^2-y^2}$. Where are the critical points?

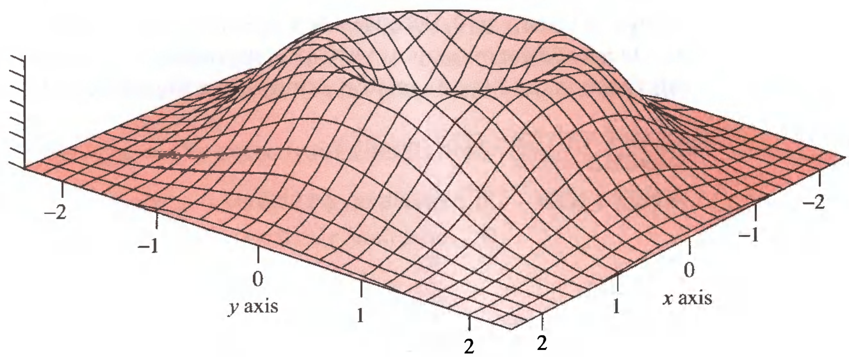

Figure 3.3.5 The volcano: $z = 2(x^2 + y^2)\exp(-x^2 - y^2)$.

SOLUTION Because $z = 2(x^2 + y^2)e^{-x^2-y^2}$, we have

$$\begin{aligned}\frac{\partial z}{\partial x} &= 4x(e^{-x^2-y^2}) + 2(x^2 + y^2)e^{-x^2-y^2}(-2x)\\ &= e^{-x^2-y^2}[4x - 4x(x^2 + y^2)]\\ &= 4x(e^{-x^2-y^2})(1 - x^2 - y^2)\end{aligned}$$

and

$$\frac{\partial z}{\partial y} = 4y(e^{-x^2-y^2})(1 - x^2 - y^2).$$

These both vanish when $x = y = 0$ or when $x^2 + y^2 = 1$. This is consistent with the figure: Points on the crater's rim are maxima and the origin is a minimum. ▲

Second Derivative Test for Local Extrema

The remainder of this section is devoted to deriving a criterion, depending on the second derivative, for a critical point to be a relative extremum. In the special case $n = 1$, our criterion will reduce to the familiar condition from one-variable calculus: $f''(x_0) > 0$ for a minimum and $f''(x_0) < 0$ for a maximum. But in the general context, the second derivative is a fairly complicated mathematical object. To state our criterion, we will introduce a version of the second derivative called the Hessian, which in turn is related to quadratic functions. ***Quadratic functions*** are functions $g: \mathbb{R}^n \to \mathbb{R}$ that have the form

$$g(h_1, \ldots, h_n) = \sum_{i,j=1}^{n} a_{ij}h_ih_j$$

for a $n \times n$ matrix $[a_{ij}]$. In terms of matrix multiplication, we can write

$$g(h_1, \ldots, h_n) = [h_1 \cdots h_n]\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}\begin{bmatrix} h_1\\ \vdots\\ h_n\end{bmatrix}.$$

For example, if $n = 3$,

$$g(h_1, h_2, h_3) = h_1^2 - 2h_1h_2 + h_3^2 = [h_1\, h_2\, h_3]\begin{bmatrix} 1 & -1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix} h_1\\ h_2\\ h_3\end{bmatrix}$$

is a quadratic function.

We can, if we wish, assume that $[a_{ij}]$ is symmetric; in fact, g is unchanged if we replace $[a_{ij}]$ by the symmetric matrix $[b_{ij}]$, where $b_{ij} = \frac{1}{2}(a_{ij} + a_{ji})$, because

$h_i h_j = h_j h_i$ and the sum is over all i and j. The *quadratic* nature of g is reflected in the identity

$$g(\lambda h_1, \dots, \lambda h_n) = \lambda^2 g(h_1, \dots, h_n),$$

which follows from the definition.

Now we are ready to define Hessian functions (named after Ludwig Otto Hesse, who introduced them in 1844).

DEFINITION Suppose that $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ has second-order continuous derivatives $(\partial^2 f/\partial x_i\,\partial x_j)(\mathbf{x}_0)$, for $i, j = 1, \dots, n$, at a point $\mathbf{x}_0 \in U$. The ***Hessian of f at $\mathbf{x}_0$*** is the quadratic function defined by

$$\begin{aligned} Hf(\mathbf{x}_0)(\mathbf{h}) &= \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i\,\partial x_j}(\mathbf{x}_0) h_i h_j \\ &= \frac{1}{2}(h_1, \dots, h_n)\begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1\,\partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \vdots & & \\ \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n\,\partial x_n} \end{pmatrix}\begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}. \end{aligned}$$

Notice that, by equality of mixed partials, the second derivative matrix is symmetric.

This function is usually used at critical points $\mathbf{x}_0 \in U$. In this case, $\mathbf{D}f(\mathbf{x}_0) = \mathbf{0}$, so the Taylor formula (see Theorem 2, Section 3.2) may be written in the form

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + Hf(\mathbf{x}_0)(\mathbf{h}) + R_2(\mathbf{x}_0, \mathbf{h}).$$

Thus, *at a critical point the Hessian equals the first nonconstant term in the Taylor series of f.*

A quadratic function $g\colon \mathbb{R}^n \to \mathbb{R}$ is called ***positive-definite*** if $g(\mathbf{h}) \geq 0$ for all $\mathbf{h} \in \mathbb{R}^n$ and $g(\mathbf{h}) = 0$ only for $\mathbf{h} = \mathbf{0}$. Similarly, g is ***negative-definite*** if $g(\mathbf{h}) \leq 0$ and $g(\mathbf{h}) = 0$ for $\mathbf{h} = \mathbf{0}$ only. Note that if $n = 1$, $Hf(x_0)(h) = \frac{1}{2}f''(x_0)h^2$, which is positive-definite if and only if $f''(x_0) > 0$.

THEOREM 5: Second Derivative Test for Local Extrema If $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ is of class C^3, $\mathbf{x}_0 \in U$ is a critical point of f, and the Hessian $Hf(\mathbf{x}_0)$ is positive-definite, then $\mathbf{x}_0$ is a relative minimum of f. Similarly, if $Hf(\mathbf{x}_0)$ is negative-definite, then $\mathbf{x}_0$ is a relative maximum.

Actually, we shall prove that the extrema given by this criterion are *strict*. A relative maximum $\mathbf{x}_0$ is said to be ***strict*** if $f(\mathbf{x}) < f(\mathbf{x}_0)$ for nearby $\mathbf{x} \neq \mathbf{x}_0$. A strict relative minimum is defined similarly. Also, the theorem is valid even if f is only C^2 but we have assumed C^3 for simplicity.

The proof of Theorem 5 requires Taylor's theorem and the following result from linear algebra.

LEMMA 1 If $B = [b_{ij}]$ is an $n \times n$ real matrix, and if the associated quadratic function

$$H\colon \mathbb{R}^n \to \mathbb{R}, (h_1, \dots, h_n) \mapsto \frac{1}{2}\sum_{i,j=1}^{n} b_{ij}h_ih_j$$

is positive-definite, then there is a constant $M > 0$ such that for all $\mathbf{h} \in \mathbb{R}^n$,

$$H(\mathbf{h}) \geq M\|\mathbf{h}\|^2.$$

PROOF For $\|\mathbf{h}\| = 1$, set $g(\mathbf{h}) = H(\mathbf{h})$. Then g is a continuous function of $\mathbf{h}$ for $\|\mathbf{h}\| = 1$ and so achieves a minimum value, say M.[8] Because H is quadratic, we have

$$H(\mathbf{h}) = H\left(\frac{\mathbf{h}}{\|\mathbf{h}\|}\|\mathbf{h}\|\right) = H\left(\frac{\mathbf{h}}{\|\mathbf{h}\|}\right)\|\mathbf{h}\|^2 = g\left(\frac{\mathbf{h}}{\|\mathbf{h}\|}\right)\|\mathbf{h}\|^2 \geq M\|\mathbf{h}\|^2$$

for any $\mathbf{h} \neq \mathbf{0}$. (The result is obviously valid if $\mathbf{h} = \mathbf{0}$). ■

Note that the quadratic function associated with the symmetric matrix $\frac{1}{2}(\partial^2 f/\partial x_i\, \partial x_j)$ is exactly the Hessian.

PROOF OF THEOREM 5 Recall that if $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ is of class C^3 and $\mathbf{x}_0 \in U$ is a critical point, Taylor's theorem may be expressed in the form

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = Hf(\mathbf{x}_0)(\mathbf{h}) + R_2(\mathbf{x}_0, \mathbf{h}),$$

where $(R_2(\mathbf{x}_0, \mathbf{h}))/\|\mathbf{h}\|^2 \to 0$ as $\mathbf{h} \to \mathbf{0}$.

Because $Hf(\mathbf{x}_0)$ is positive-definite, lemma 1 assures us of a constant $M > 0$ such that for all $\mathbf{h} \in \mathbb{R}^n$

$$Hf(\mathbf{x}_0)(\mathbf{h}) \geq M\|\mathbf{h}\|^2.$$

[8]Here we are using, without proof, a theorem analogous to a theorem in calculus that states that every continuous function on an interval $[a, b]$ achieves a maximum and a minimum; see Theorem 7.

Because $R_2(\mathbf{x}_0, \mathbf{h})/\|\mathbf{h}\|^2 \to 0$ as $\mathbf{h} \to \mathbf{0}$, there is a $\delta > 0$ such that for $0 < \|\mathbf{h}\| < \delta$

$$|R_2(\mathbf{x}_0, \mathbf{h})| < M\|\mathbf{h}\|^2.$$

Thus, $0 < Hf(\mathbf{x}_0)(\mathbf{h}) + R_2(\mathbf{x}_0, \mathbf{h}) = f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)$ for $0 < \|\mathbf{h}\| < \delta$, so that $\mathbf{x}_0$ is a relative minimum, in fact, a strict relative minimum.

The proof in the negative-definite case is similar, or else follows by applying the preceding to $-f$, and is left as an exercise. ■

EXAMPLE 5 Consider again the function $f\colon \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x^2 + y^2$. Then $(0, 0)$ is a critical point, and f is already in the form of Taylor's theorem:

$$f((0, 0) + (h_1, h_2)) = f(0, 0) + (h_1^2 + h_2^2) + 0.$$

We can see directly that the Hessian at $(0, 0)$ is

$$Hf(\mathbf{0})(\mathbf{h}) = h_1^2 + h_2^2,$$

which is clearly positive-definite. Thus, $(0, 0)$ is a relative minimum. This simple case can, of course, be done without calculus. Indeed, it is clear that $f(x, y) > 0$ for all $(x, y) \neq (0, 0)$. ▲

For functions of two variables $f(x, y)$, the Hessian may be written as follows:

$$Hf(x, y)(\mathbf{h}) = \tfrac{1}{2}[h_1, h_2] \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial y\, \partial x} \\[2ex] \dfrac{\partial^2 f}{\partial x\, \partial y} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

Now we shall give a useful criterion for when a quadratic function defined by such a 2×2 matrix is positive-definite. This will be useful in conjunction with Theorem 5.

LEMMA 2 Let

$$B = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \qquad \text{and} \qquad H(\mathbf{h}) = \tfrac{1}{2}[h_1, h_2] B \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

Then $H(\mathbf{h})$ is positive-definite if and only if $a > 0$ and $\det B = ac - b^2 > 0$.

PROOF We have

$$H(\mathbf{h}) = \tfrac{1}{2}[h_1, h_2] \begin{bmatrix} ah_1 + bh_2 \\ bh_1 + ch_2 \end{bmatrix} = \tfrac{1}{2}(ah_1^2 + 2bh_1h_2 + ch_2^2).$$

Let us complete the square, writing

$$H(\mathbf{h}) = \tfrac{1}{2}a\left(h_1 + \frac{b}{a}h_2\right)^2 + \frac{1}{2}\left(c - \frac{b^2}{a}\right)h_2^2.$$

Suppose H is positive-definite. Setting $h_2 = 0$, we see that $a > 0$. Setting $h_1 = -(b/a)h_2$, we get $c - b^2/a > 0$ or $ac - b^2 > 0$. Conversely, if $a > 0$ and $c - b^2/a > 0$, $H(\mathbf{h})$ is a sum of squares, so that $H(\mathbf{h}) \geq 0$. If $H(\mathbf{h}) = 0$, then each square must be zero. This implies that both h_1 and h_2 must be zero, so that $H(\mathbf{h})$ is positive-definite. ■

Similarly, one can see that $H(\mathbf{h})$ is negative-definite if and only if $a < 0$ and $ac - b^2 > 0$. We note that an alternative formulation is that $H(\mathbf{h})$ is positive- (respectively, negative-) definite if $a + c = \text{trace } B > 0$ (respectively, < 0) and $\det B > 0$.

Determinant Test for Positive Definiteness

There are similar criteria for an $n \times n$ symmetric matrix B. Consider the n square submatrices along the diagonal (see Figure 3.3.6). B is positive-definite (that is, the quadratic function associated with B is positive-definite) if and only if the determinants of these diagonal submatrices are all greater than zero. For negative-definite B, the signs should be alternately < 0 and > 0. We shall not prove this general case here.[9] In case the determinants of the diagonal submatrices are all nonzero, but the matrix is not positive- or negative-definite, the critical point is of ***saddle type***; in this case, one can show that the point is neither a maximum nor a minimum in the manner of Example 2.

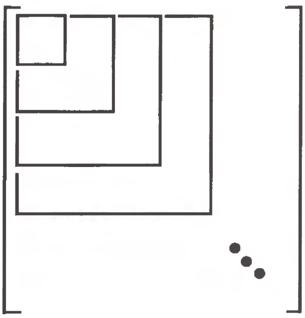

Figure 3.3.6 "Diagonal" submatrices are used in the criterion for positive definiteness; they must all have determinant > 0.

[9]This is proved in, for example, K. Hoffman and R. Kunze, *Linear Algebra*, Prentice Hall, Englewood Cliffs, N.J., 1961, pp. 249–251. For students with sufficient background in linear algebra, it should be noted that B is positive-definite when all of its eigenvalues (which are necessarily real, because B is symmetric) are positive.

Second Derivative Test

Lemma 2 and Theorem 5 imply the following result:

THEOREM 6: Second Derivative Maximum-Minimum Test for Functions of Two Variables Let $f(x, y)$ be of class C^3 on an open set U in $\mathbb{R}^2$. A point (x_0, y_0) is a (strict) local minimum of f provided the following three conditions hold:

(i) $\dfrac{\partial f}{\partial x}(x_0, y_0) = \dfrac{\partial f}{\partial y}(x_0, y_0) = 0$

(ii) $\dfrac{\partial^2 f}{\partial x^2}(x_0, y_0) > 0$

(iii) $D = \left(\dfrac{\partial^2 f}{\partial x^2}\right)\left(\dfrac{\partial^2 f}{\partial y^2}\right) - \left(\dfrac{\partial^2 f}{\partial x\, \partial y}\right)^2 > 0$ at (x_0, y_0)

(D is called the ***discriminant*** of the Hessian.) If in (ii) we have <0 instead of >0 and condition (iii) is unchanged, then we have a (strict) local maximum.

EXAMPLE 6 Classify the critical points of the function $f\colon \mathbb{R}^2 \to \mathbb{R}$, defined by $(x, y) \mapsto x^2 - 2xy + 2y^2$.

SOLUTION As in Example 5, we find that $f(0, 0) = 0$, the origin is the only critical point, and the Hessian is

$$Hf(\mathbf{0})(\mathbf{h}) = h_1^2 - 2h_1h_2 + 2h_2^2 = (h_1 - h_2)^2 + h_2^2,$$

which is clearly positive-definite. Thus, f has a relative minimum at $(0, 0)$. Alternatively, we can apply Theorem 6. At $(0, 0)$, $\partial^2 f/\partial x^2 = 2$, $\partial^2 f/\partial y^2 = 4$, and $\partial^2 f/\partial x\, \partial y = -2$. Conditions (i), (ii), and (iii) hold, so f has a relative minimum at $(0, 0)$ ▲

If $D < 0$ in Theorem 6, then we have a saddle point. In fact, one can prove that $f(x, y)$ is larger than $f(x_0, y_0)$ as we move away from (x_0, y_0) in some direction and smaller in the orthogonal direction (see Exercise 26). The general appearance is thus similar to that shown in Figure 3.3.4. The appearance of the graph near (x_0, y_0) in the case $D = 0$ must be determined by further analysis.

We summarize the procedure for dealing with functions of two variables: After all critical points have been found and their associated Hessians computed, some of these Hessians may be positive-definite, indicating relative minima; some may be negative-definite, indicating relative maxima; and some may be neither positive- nor negative-definite, indicating saddle points. The shape of the graph at a saddle point where $D < 0$ is like that in Figure 3.3.4. Critical points for which $D \neq 0$ are called ***nondegenerate critical points***. Such points are maxima, minima, or saddle points. The remaining critical points, where $D = 0$, may be tested directly, with level sets and sections or by some other method. Such critical points are said to be ***degenerate***; the methods developed in this chapter fail to provide a picture of the behavior of a function near such points, so we examine them case by case.

EXAMPLE 7 Locate the relative maxima, minima, and saddle points of the function

$$f(x, y) = \log(x^2 + y^2 + 1).$$

SOLUTION We must first locate the critical points of this function; therefore, according to Theorem 3, we calculate

$$\nabla f(x, y) = \frac{2x}{x^2 + y^2 + 1}\mathbf{i} + \frac{2y}{x^2 + y^2 + 1}\mathbf{j}.$$

Thus, $\nabla f(x, y) = \mathbf{0}$ if and only if $(x, y) = (0, 0)$, and so the only critical point of f is $(0, 0)$. Now we must determine whether this is a maximum, a minimum, or a saddle point. The second partial derivatives are

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

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

and

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

Therefore,

$$\frac{\partial^2 f}{\partial x^2}(0, 0) = 2 = \frac{\partial^2 f}{\partial y^2}(0, 0) \quad \text{and} \quad \frac{\partial^2 f}{\partial x\, \partial y}(0, 0) = 0,$$

which yields

$$D = 2 \cdot 2 = 4 > 0.$$

Because $(\partial^2 f/\partial x^2)(0, 0) > 0$, we conclude by Theorem 6 that $(0, 0)$ is a local minimum. (Can you show this just from the fact that $\log t$ is an increasing function of $t > 0$?) ▲

EXAMPLE 8 The graph of the function $g(x, y) = 1/xy$ is a surface S in $\mathbb{R}^3$. Find the points on S that are closest to the origin (0, 0, 0). (See Figure 3.3.7.)

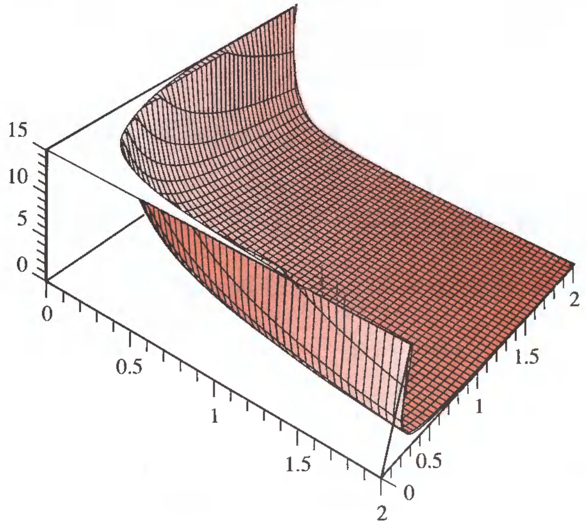

Figure 3.3.7 The surface $z = 1/xy$ in the first quadrant. (There are similar figures in the other quadrants, but notice that $z < 0$ in the second and fourth quadrants.)

SOLUTION Each point on S is of the form $(x, y, 1/xy)$. The distance from this point to the origin is

$$d(x, y) = \sqrt{x^2 + y^2 + \frac{1}{x^2y^2}}.$$

It is easier to work with the square of d, so let $f(x, y) = x^2 + y^2 + (1/x^2y^2)$, which will have the same minimum point. Notice that $f(x, y)$ becomes very large as x and y get larger and larger; $f(x, y)$ also becomes very large as (x, y) approaches the x or y axis where f is not defined, so f must attain a minimum at some critical point. The critical points are determined by:

$$\frac{\partial f}{\partial x} = 2x - \frac{2}{x^3y^2} = 0,$$

$$\frac{\partial f}{\partial y} = 2y - \frac{2}{y^3x^2} = 0,$$

that is, $x^4y^2 - 1 = 0$, and $x^2y^4 - 1 = 0$. From the first equation we get $y^2 = 1/x^4$, and, substituting this into the second equation, we obtain

$$\frac{x^2}{x^8} = 1 = \frac{1}{x^6}.$$

Thus, $x = \pm 1$ and $y = \pm 1$, and it therefore follows that f has four critical points, namely, $(1, 1), (1, -1), (-1, 1)$, and $(-1, -1)$. Note that f has the value 3 for all these points, so they are all minima. Therefore, the points on the surface closest to the point $(0, 0, 0)$ are $(1, 1, 1), (1, -1, -1), (-1, 1, -1)$, and $(-1, -1, 1)$ and the minimum distance is $\sqrt{3}$. Is this consistent with the graph in Figure 3.3.7? ▲

EXAMPLE 9 Analyze the behavior of $z = x^5y + xy^5 + xy$ at its critical points.

SOLUTION The first partial derivatives are

$$\frac{\partial z}{\partial x} = 5x^4y + y^5 + y = y(5x^4 + y^4 + 1)$$

and

$$\frac{\partial z}{\partial y} = x(5y^4 + x^4 + 1).$$

The terms $5x^4 + y^4 + 1$ and $5y^4 + x^4 + 1$ are always greater than or equal to 1, and so it follows that the only critical point is $(0, 0)$.

The second partial derivatives are

$$\frac{\partial^2 z}{\partial x^2} = 20x^3y, \qquad \frac{\partial^2 z}{\partial y^2} = 20xy^3,$$

and

$$\frac{\partial^2 z}{\partial x\, \partial y} = 5x^4 + 5y^4 + 1.$$

Thus, at $(0, 0)$, $D = -1$, and so $(0, 0)$ is a nondegenerate saddle point and the graph of z near $(0, 0)$ looks like the graph in Figure 3.3.4. ▲

Global Maxima and Minima

We end this section with a discussion of the theory of *absolute*, or *global*, maxima and minima of functions of several variables. Unfortunately, the location of absolute maxima and minima for functions on $\mathbb{R}^n$ is, in general, a more difficult problem than for functions of one variable.

DEFINITION Suppose $f\colon A \to \mathbb{R}$ is a function defined on a set A in $\mathbb{R}^2$ or $\mathbb{R}^3$. A point $\mathbf{x}_0 \in A$ is said to be an ***absolute maximum*** (or ***absolute minimum***) point of f if $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ [or $f(\mathbf{x}) \geq f(\mathbf{x}_0)$] for all $\mathbf{x} \in A$.

In one-variable calculus, one learns—but often does not prove—that every continuous function on a closed interval I assumes its absolute maximum (or minimum) value at some point $\mathbf{x}_0$ in I. A generalization of this theoretical fact also holds in $\mathbb{R}^n$.

Such theorems guarantee that the maxima or minima one is seeking actually exist; therefore, the search for them is not in vain.

DEFINITION A set $D \in \mathbb{R}^n$ is said to be ***bounded*** if there is a number $M > 0$ such that $\|\mathbf{x}\| < M$ for all $\mathbf{x} \in D$. A set is ***closed*** if it contains all its boundary points.

Thus, a set is bounded if it can be strictly contained in some (large) ball. The appropriate generalization of the one-variable theorem on maxima and minima is the following result, stated without proof.

THEOREM 7: Global Existence Theorem for Maxima and Minima

Let D be closed and bounded in $\mathbb{R}^n$ and let $f\colon D \to \mathbb{R}$ be continuous. Then f assumes its absolute maximum and minimum values at some points $\mathbf{x}_0$ and $\mathbf{x}_1$ of D.

Simply stated, $\mathbf{x}_0$ and $\mathbf{x}_1$ are points where f assumes its largest and smallest values. As in one-variable calculus, these points need not be uniquely determined.

Suppose now that $D = U \cup \partial U$, where U is open and ∂U is its boundary. If $D \subset \mathbb{R}^2$, we suppose that ∂U is a piecewise smooth curve; that is, D is a region bounded by a collection of smooth curves—for example, a square or the sets depicted in Figure 3.3.8.

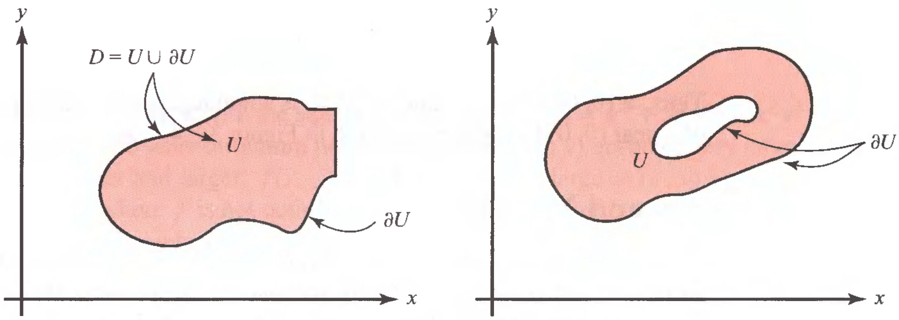

Figure 3.3.8 $D = U \cup \partial U$: Two examples of regions whose boundary is a piecewise smooth curve.

If $\mathbf{x}_0$ and $\mathbf{x}_1$ are in U, we know from Theorem 4 that they are critical points of f. If they are in ∂U, and ∂U is a smooth curve (i.e., the image of a smooth path $\mathbf{c}$ with $\mathbf{c}' \neq 0$), then they are maximum or minimum points of f viewed as a function on ∂U. These observations provide a method of finding the absolute maximum and minimum values of f on a region D.

Strategy for Finding the Absolute Maxima and Minima on a Region with Boundary Let f be a continuous function of two variables defined on a closed and bounded region D in $\mathbb{R}^2$, which is bounded by a smooth closed curve. To find the absolute maximum and minimum of f on D:

(i) Locate all critical points for f in U.

(ii) Find the maximum and minimum of f viewed as a function only on ∂U.

(iii) Compute the value of f at all of these critical points.

(iv) Compare all these values and select the largest and the smallest.

If D is a region bounded by a collection of smooth curves (such as a square), then one follows a similar procedure, but including in step (iii) the points where the curves meet (such as the corners of the square).

All the steps except step (ii) should now be familiar to the student. To carry out step (ii) in the plane, one way is to find a smooth parametrization of ∂U; that is, we find a path $\mathbf{c}\colon I \to \partial U$, where I is some interval, which is onto ∂U. Second, we consider the function of one variable $t \mapsto f(\mathbf{c}(t))$, where $t \in I$, and locate the maximum and minimum points $t_0, t_1 \in I$ (remember to check the endpoints!). Then $\mathbf{c}(t_0)$, $\mathbf{c}(t_1)$ will be maximum and minimum *points* for f as a function on ∂U. Another method for dealing with step (ii) is the Lagrange multiplier method, to be presented in the next section.

EXAMPLE 10 Find the maximum and minimum values of the function $f(x, y) = x^2 + y^2 - x - y + 1$ in the disk D defined by $x^2 + y^2 \leq 1$.

SOLUTION (i) To find the critical points we set $\partial f/\partial x = \partial f/\partial y = 0$. Thus, $2x - 1 = 0$, $2y - 1 = 0$, and hence $(x, y) = (\frac{1}{2}, \frac{1}{2})$ is the only critical point in the open disk $U = \{(x, y) \mid x^2 + y^2 < 1\}$.

(ii) The boundary ∂U can be parametrized by $\mathbf{c}(t) = (\sin t, \cos t)$, $0 \leq t \leq 2\pi$. Thus,

$$f(\mathbf{c}(t)) = \sin^2 t + \cos^2 t - \sin t - \cos t + 1 = 2 - \sin t - \cos t = g(t).$$

To find the maximum and minimum of f on ∂U, it suffices to locate the maximum and minimum of g. Now $g'(t) = 0$ only when

$$\sin t = \cos t, \qquad \text{that is, when} \qquad t = \frac{\pi}{4}, \frac{5\pi}{4}.$$

Thus, the candidates for the maximum and minimum for f on ∂U are the points $\mathbf{c}(\pi/4)$, $\mathbf{c}(5\pi/4)$ and the endpoints $\mathbf{c}(0) = \mathbf{c}(2\pi)$.

(iii) The values of f at the critical points are: $f(\frac{1}{2}, \frac{1}{2}) = \frac{1}{2}$ from step (i) and, from step (ii),

$$f\left(\mathbf{c}\left(\frac{\pi}{4}\right)\right) = f\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) = 2 - \sqrt{2},$$

$$f\left(\mathbf{c}\left(\frac{5\pi}{4}\right)\right) = f\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right) = 2 + \sqrt{2},$$

and

$$f(\mathbf{c}(0)) = f(\mathbf{c}(2\pi)) = f(0, 1) = 1.$$

(iv) Comparing all the values $\frac{1}{2}, 2 - \sqrt{2}, 2 + \sqrt{2}, 1$, it is clear that the absolute minimum occurs at $(1/2, 1/2)$ and the absolute maximum occurs at $(-\sqrt{2}/2, -\sqrt{2}/2)$. ▲

In Section 3.4, we shall consider a generalization of the strategy for finding the absolute maximum and minimum to regions D in $\mathbb{R}^n$.

EXERCISES

In Exercises 1 to 16, find the critical points of the given function and then determine whether they are local maxima, local minima, or saddle points.

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

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

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

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

5. $f(x, y) = e^{1+x^2-y^2}$

6. $f(x, y) = x^2 - 3xy + 5x - 2y + 6y^2 + 8$

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

8. $f(x, y) = \sin(x^2 + y^2)$ [consider only the critical point $(0, 0)$]

9. $f(x, y) = \cos(x^2 + y^2)$ [consider only the three critical points $(0, 0)$, $(\sqrt{\pi/2}, \sqrt{\pi/2})$, and $(0, \sqrt{\pi})$]

10. $f(x, y) = y + x \sin y$

11. $f(x, y) = e^x \cos y$

12. $f(x, y) = (x - y)(xy - 1)$

13. $f(x, y) = xy + \frac{1}{x} + \frac{1}{y}$

14. $f(x, y) = \log(2 + \sin xy)$

15. $f(x, y) = x \sin y$

16. $f(x, y) = (x + y)(xy + 1)$

17. Find the local maxima and minima for $z = (x^2 + 3y^2)\, e^{1-x^2-y^2}$. (See Figure 2.1.15.)

18. Let $f(x, y) = x^2 + y^2 + kxy$. If you imagine the graph changing as k increases, at what values of k does the shape of the graph change qualitatively?

19. An examination of the function $f\colon \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto (y - 3x^2)(y - x^2)$ will give an idea of the difficulty of finding conditions that guarantee that a critical point is a relative extremum when Theorem 6 fails.[10] Show that

(a) the origin is a critical point of f;

(b) f has a relative minimum at $(0, 0)$ on every straight line through $(0, 0)$; that is, if $g(t) = (at, bt)$, then $f \circ g\colon \mathbb{R} \to \mathbb{R}$ has a relative minimum at 0, for every choice of a and b;

(c) the origin is not a relative minimum of f.

20. Let $f(x, y) = Ax^2 + E$ where A and E are constants. What are the critical points of f? Are they local maxima or local minima?

21. Let $f(x, y) = x^2 - 2xy + y^2$. Here $D = 0$. Can you say whether the critical points are local minima, local maxima, or saddle points?

22. Find the point on the plane $2x - y + 2z = 20$ nearest the origin.

23. Show that a rectangular box of given volume has minimum surface area when the box is a cube.

24. Show that the rectangular parallelepiped with fixed surface area and maximum volume is a cube.

25. Write the number 120 as a sum of three numbers so that the sum of the products taken two at a time is a maximum.

26. Show that if (x_0, y_0) is a critical point of a quadratic function $f(x, y)$ and $D < 0$, then there are points (x, y) near (x_0, y_0) at which $f(x, y) > f(x_0, y_0)$ and, similarly, points for which $f(x, y) < f(x_0, y_0)$.

27. Determine the nature of the critical points of the function

$$f(x, y, z) = x^2 + y^2 + z^2 + xy.$$

28. Let n be an integer greater than 2 and set $f(x, y) = ax^n + cy^n$, where $ac \neq 0$. Determine the nature of the critical points of f.

---

[10]This interesting phenomenon was first pointed out by the famous mathematician Giuseppe Peano (1858–1932). Another curious "pathology" is given in Exercise 41.

29. Determine the nature of the critical points of $f(x, y) = x^3 + y^2 - 6xy + 6x + 3y$.

30. Find the absolute maximum and minimum values of the function $f(x, y) = (x^2 + y^2)^4$ on the disk $x^2 + y^2 \leq 1$. (You do not *have* to use calculus.)

31. Repeat Exercise 30 for the function $f(x, y) = x^2 + xy + y^2$.

32. A curve C in space is defined *implicitly* on the cylinder $x^2 + y^2 = 1$ by the additional equation $x^2 - xy + y^2 - z^2 = 1$. Find the point or points on C closest to the origin.

33. Find the absolute maximum and minimum values for $f(x, y) = \sin x + \cos y$ on the rectangle R defined by $0 \leq x \leq 2\pi, 0 \leq y \leq 2\pi$.

34. Find the absolute maximum and minimum values for the function $f(x, y) = xy$ on the rectangle R defined by $-1 \leq x \leq 1, -1 \leq y \leq 1$.

35. Determine the nature of the critical points of $f(x, y) = xy + 1/x + 8/y$.

In Exercises 36 through 40, D denotes the unit disk.

36. Let u be a C^2 function on D which is "strictly subharmonic"; that is, the following inequality holds: $\nabla^2 u = (\partial^2 u/\partial x^2) + (\partial^2 u/\partial y^2) > 0$. Show that u cannot have a maximum point in $D\backslash\partial D$ (the set of points in D, but not in ∂D).

37. Let u be a harmonic function on D; that is, $\nabla^2 u = 0$ on $D\backslash\partial D$ and be continuous on D. Show that if u achieves its maximum value in $D\backslash\partial D$, it also achieves it on ∂D. This is sometimes called the "weak maximum principle" for harmonic functions. [Hint: Consider $\nabla^2(u + \varepsilon e^x)$, $\varepsilon > 0$. You can use the following fact, which is proved in more advanced texts: Given a sequence $\{\mathbf{p}_n\}$, $n = 1, 2, \ldots$, of points in a closed bounded set A in $\mathbb{R}^2$ or $\mathbb{R}^3$, there exists a point $\mathbf{q}$ such that every neighborhood of $\mathbf{q}$ contains at least one member of $\{\mathbf{p}_n\}$.]

38. Define the notion of a strict superharmonic function u on D by mimicking Exercise 36. Show that u cannot have a minimum in $D\backslash\partial D$.

39. Let u be harmonic in D as in Exercise 37. Show that if u achieves its minimum value in $D\backslash\partial D$, it also achieves it on ∂D. This is sometimes called the "weak minimum principle" for harmonic functions.

40. Let $\phi\colon \partial D \to \mathbb{R}$ be continuous and let T be a solution on D to $\nabla^2 T = 0$, continuous on D and $T = \phi$ on ∂D.

(a) Use Exercises 36 to 39 to show that such a solution, if it exists, must be unique.

(b) Suppose that $T(x, y)$ represents a temperature function that is independent of time, with ϕ representing the temperature of a circular plate at its boundary. Can you give a physical interpretation of the principle stated in part (a)?

41. (a) Let f be a C^1 function on the real line $\mathbb{R}$. Suppose that f has exactly one critical point x_0 that is a strict local minimum for f. Show that x_0 is also an absolute minimum for f, that is, that $f(x) \geq f(x_0)$ for all x.

(b) The next example shows that the conclusion of part (a) does not hold for functions of more than one variable. Let $f: \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x, y) = -y^4 - e^{-x^2} + 2y^2\sqrt{e^x + e^{-x^2}}.$$

(i) Show that (0, 0) is the only critical point for f and that it is a local minimum.
(ii) Argue informally that f has no absolute minimum.

42. Suppose that a pentagon is composed of a rectangle topped by an isosceles triangle (see Figure 3.3.9). If the length of the perimeter is fixed, find the maximum possible area.

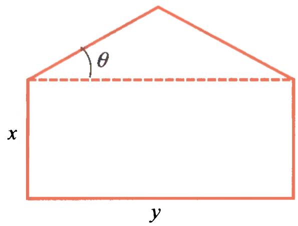

Figure 3.3.9 Maximize the area for fixed perimeter.

3.4 Constrained Extrema and Lagrange Multipliers

Often one is required to maximize or minimize a function subject to certain *constraints* or *side conditions*. For example, we might need to maximize $f(x, y)$ subject to the condition that $x^2 + y^2 = 1$; that is, that (x, y) lie on the unit circle. More generally, we might need to maximize or minimize $f(x, y)$ subject to the side condition that (x, y) also satisfies an equation $g(x, y) = c$ where g is some function and c equals a constant [in the preceding example, $g(x, y) = x^2 + y^2$, and $c = 1$]. The set of such (x, y) is a level curve for g.

The purpose of this section is to develop some methods for handling this sort of problem. In Figure 3.4.1 we picture a graph of a function $f(x, y)$. In this picture,

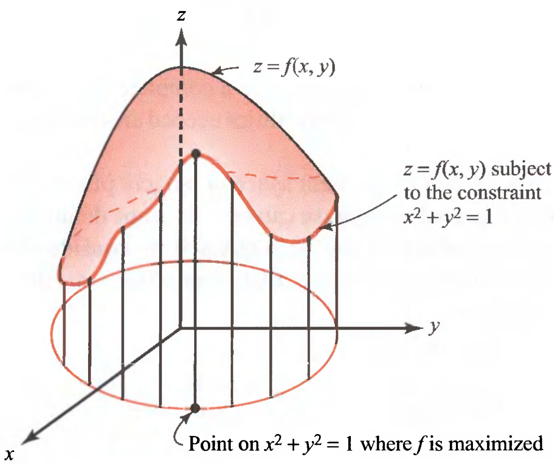

Figure 3.4.1 The geometric meaning of maximizing f subject to the constraint $x^2 + y^2 = 1$.

the maximum of f might be at (0, 0). However, suppose we are not interested in this maximum but only the maximum of $f(x, y)$ when (x, y) belongs to the unit circle; that is, when $x^2 + y^2 = 1$. The cylinder over $x^2 + y^2 = 1$ intersects the graph of $z = f(x, y)$ in a curve that lies on this graph. The problem of maximizing or minimizing $f(x, y)$ subject to the constraint $x^2 + y^2 = 1$ amounts to finding the point on this curve where z is the greatest or the least.

The Lagrange Multiplier Method

In general, let $f: U \subset \mathbb{R}^n \to \mathbb{R}$ and $g: U \subset \mathbb{R}^n \to \mathbb{R}$ be given C^1 functions, and let S be the level set for g with value c [recall that this is the set of points $\mathbf{x} \in \mathbb{R}^n$ with $g(\mathbf{x}) = c$].

When f is restricted to S we again have the notion of local maxima or local minima of f (local extrema), and an absolute maximum (largest value) or absolute minimum (smallest value) must be a local extremum. The following method provides a necessary condition for a constrained extremum:

THEOREM 8: The Method of Lagrange Multipliers Suppose that $f: U \subset \mathbb{R}^n \to \mathbb{R}$ and $g: U \subset \mathbb{R}^n \to \mathbb{R}$ are given C^1 real-valued functions. Let $\mathbf{x}_0 \in U$ and $g(\mathbf{x}_0) = c$, and let S be the level set for g with value c (recall that this is the set of points $\mathbf{x} \in \mathbb{R}^n$ satisfying $g(\mathbf{x}) = c$). Assume $\nabla g(\mathbf{x}_0) \neq \mathbf{0}$.

If $f|S$, which denotes "f restricted to S," has a local maximum or minimum on S at $\mathbf{x}_0$, then there is a real number λ such that

$$\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0). \tag{1}$$

In general, a point $\mathbf{x}_0$ where equation (1) holds is said to be a ***critical point*** of $f|S$.

PROOF We have not developed enough techniques to give a complete proof, but we can provide the essential points. (The additional technicalities needed are discussed in Section 3.5 and in the Internet supplement.)

In Section 2.6 we learned that for $n = 3$ the tangent space or tangent plane of S at $\mathbf{x}_0$ is the space orthogonal to $\nabla g(\mathbf{x}_0)$. For arbitrary n we can give the same definition for the tangent space of S at $\mathbf{x}_0$. This definition can be motivated by considering tangents to paths $\mathbf{c}(t)$ that lie in S, as follows: If $\mathbf{c}(t)$ is a path in S and $\mathbf{c}(0) = \mathbf{x}_0$, then $\mathbf{c}'(0)$ is a tangent vector to S at $\mathbf{x}_0$, but

$$\frac{d}{dt} g(\mathbf{c}(t)) = \frac{d}{dt} c = 0,$$

and on the other hand, by the chain rule,

$$\frac{d}{dt}g(\mathbf{c}(t))\bigg|_{t=0} = \nabla g(\mathbf{x}_0)\cdot\mathbf{c}'(0),$$

so that $\nabla g(\mathbf{x}_0)\cdot\mathbf{c}'(0) = 0$; that is, $\mathbf{c}'(0)$ is orthogonal to $\nabla g(\mathbf{x}_0)$.

If $f|S$ has a maximum at $\mathbf{x}_0$, then $f(\mathbf{c}(t))$ has a maximum at $t = 0$. By one-variable calculus, $df(\mathbf{c}(t))/dt|_{t=0} = 0$. Hence, by the chain rule,

$$0 = \frac{d}{dt}f(\mathbf{c}(t))\bigg|_{t=0} = \nabla f(\mathbf{x}_0)\cdot\mathbf{c}'(0).$$

Thus, $\nabla f(\mathbf{x}_0)$ is perpendicular to the tangent of every curve in S and so is perpendicular to the whole tangent space to S at $\mathbf{x}_0$. Because the space perpendicular to this tangent space is a line, $\nabla f(\mathbf{x}_0)$ and $\nabla g(\mathbf{x}_0)$ are parallel. Because $\nabla g(\mathbf{x}_0) \neq \mathbf{0}$, it follows that $\nabla f(\mathbf{x}_0)$ is a multiple of $\nabla g(\mathbf{x}_0)$, which is the conclusion of the theorem. ∎

Let us extract some geometry from this proof.

THEOREM 9 If f, when constrained to a surface S, has a maximum or minimum at $\mathbf{x}_0$, then $\nabla f(\mathbf{x}_0)$ is perpendicular to S at $\mathbf{x}_0$ (see Figure 3.4.2).

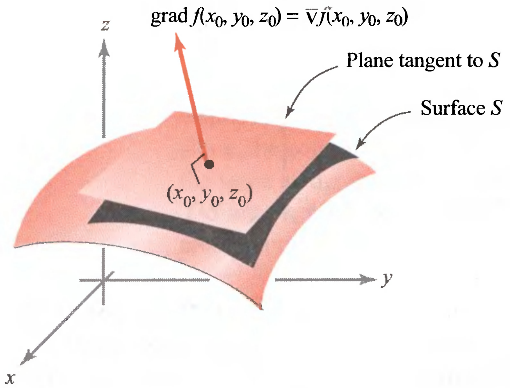

Figure 3.4.2 The geometry of constrained extrema.

These results tell us that in order to find the constrained extrema of f, we must look among those points $\mathbf{x}_0$ satisfying the conclusions of these two theorems. We shall give several illustrations of how to use each.

When the method of Theorem 8 is used, we look for a point $\mathbf{x}_0$ and a constant λ, called a ***Lagrange multiplier***, such that $\nabla f(\mathbf{x}_0) = \lambda\nabla g(\mathbf{x}_0)$. This method is more analytic in nature than the geometric method of Theorem 9. Surprisingly, Euler introduced these multipliers in 1744, some 40 years before Lagrange!

Equation (1) says that the partial derivatives of f are proportional to those of g. Finding such points $\mathbf{x}_0$ at which this occurs means solving the simultaneous equations

$$\left.\begin{aligned}
\frac{\partial f}{\partial x_1}(x_1,\ldots,x_n) &= \lambda\frac{\partial g}{\partial x_1}(x_1,\ldots,x_n)\\
\frac{\partial f}{\partial x_2}(x_1,\ldots,x_n) &= \lambda\frac{\partial g}{\partial x_2}(x_1,\ldots,x_n)\\
&\vdots\\
\frac{\partial f}{\partial x_n}(x_1,\ldots,x_n) &= \lambda\frac{\partial g}{\partial x_n}(x_1,\ldots,x_n)\\
g(x_1,\ldots,x_n) &= c
\end{aligned}\right\} \tag{2}$$

for $x_1, \ldots, x_n$ and λ.

Another way of looking at these equations is as follows: Think of λ as an additional variable and form the auxiliary function

$$h(x_1,\ldots,x_n,\lambda) = f(x_1,\ldots,x_n) - \lambda[g(x_1,\ldots,x_n) - c].$$

The Lagrange multiplier theorem says that to find the extreme points of $f|S$, we should examine the critical points of h. These are found by solving the equations

$$\left.\begin{aligned}
0 &= \frac{\partial h}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda\frac{\partial g}{\partial x_1}\\
&\vdots\\
0 &= \frac{\partial h}{\partial x_n} = \frac{\partial f}{\partial x_n} - \lambda\frac{\partial g}{\partial x_n}\\
0 &= \frac{\partial h}{\partial \lambda} = g(x_1,\ldots,x_n) - c
\end{aligned}\right\}, \tag{3}$$

which are the same as equations (2) above.

Second derivative tests for maxima and minima analogous to those in Section 3.3 will be given in Theorem 10 later in this section. However, in many problems it is possible to distinguish between maxima and minima by direct observation or by geometric means. Because this is often simpler, we consider examples of the latter type first.

EXAMPLE 1 Let $S \subset \mathbb{R}^2$ be the line through $(-1, 0)$ inclined at 45°, and let $f\colon \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x^2 + y^2$. Find the extrema of $f|S$.

SOLUTION Here $S = \{(x, y) \mid y - x - 1 = 0\}$, and therefore we set $g(x, y) = y - x - 1$ and $c = 0$. We have $\nabla g(x, y) = -\mathbf{i} + \mathbf{j} \neq \mathbf{0}$. The relative extrema of $f|S$ must be found among the points at which ∇f is orthogonal to S, that is, inclined at $-45°$. But $\nabla f(x, y) = (2x, 2y)$, which has the desired slope only when $x = -y$, or when (x, y) lies on the line L through the origin inclined at $-45°$. This can occur in the set S only for the single point at which L and S intersect (see Figure 3.4.3). Reference to the level curves of f indicates that this point, $(-1/2, 1/2)$, is a relative minimum of $f|S$ (but not of f).

Notice that in this problem, f on S has a minimum but no maximum. ▲

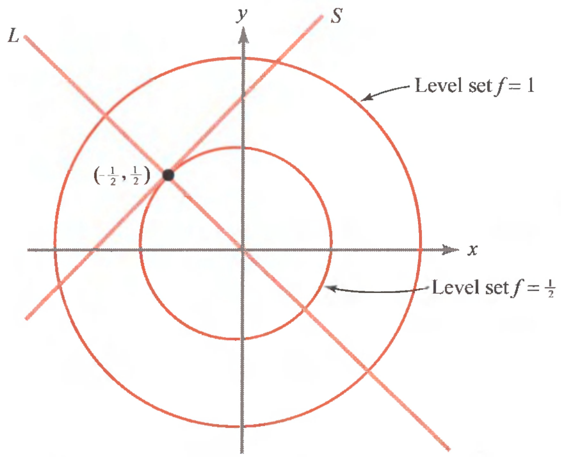

Figure 3.4.3 The geometry associated with finding the extrema of $f(x, y) = x^2 + y^2$ restricted to $S = \{(x, y) \mid y - x - 1 = 0\}$.

EXAMPLE 2 Let $f: \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x^2 - y^2$, and let S be the circle of radius 1 around the origin. Find the extrema of $f|S$.

SOLUTION The set S is the level curve for g with value 1, where $g: \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x^2 + y^2$. Because both of these functions have been studied in previous examples, we know their level curves; these are shown in Figure 3.4.4. In two dimensions, the condition that $\nabla f = \lambda \nabla g$ at $\mathbf{x}_0$, that is, that ∇f and ∇g are parallel at $\mathbf{x}_0$ is the same as the level curves being tangent at $\mathbf{x}_0$ (why?). Thus, the extreme points of $f|S$ are $(0, \pm 1)$ and $(\pm 1, 0)$. Evaluating f, we find $(0, \pm 1)$ are minima and $(\pm 1, 0)$ are maxima.

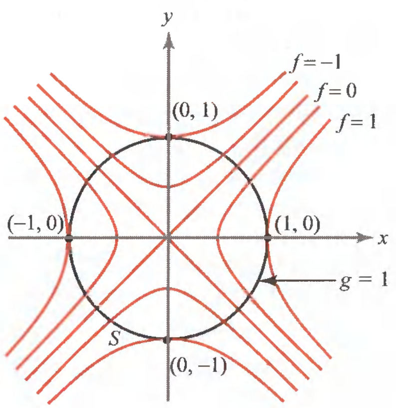

Figure 3.4.4 The geometry associated with the problem of finding the extrema of $x^2 - y^2$ on $S = \{(x, y) \mid x^2 + y^2 = 1\}$.

Let us also do this problem analytically by the method of Lagrange multipliers. Clearly,

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2x, -2y) \qquad \text{and} \qquad \nabla g(x, y) = (2x, 2y).$$

Note that $\nabla g(x, y) \neq \mathbf{0}$ if $x^2 + y^2 = 1$. Thus, according to the Lagrange multiplier theorem, we must find a λ such that

$$(2x, -2y) = \lambda(2x, 2y) \qquad \text{and} \qquad (x, y) \in S, \qquad \text{i.e., } x^2 + y^2 = 1.$$

These conditions yield three equations, which can be solved for the three unknowns x, y, and λ. From $2x = \lambda 2x$ we conclude that either $x = 0$ or $\lambda = 1$. If $x = 0$, then $y = \pm 1$ and $-2y = \lambda 2y$ implies $\lambda = -1$. If $\lambda = 1$, then $y = 0$ and $x = \pm 1$. Thus, we get the points $(0, \pm 1)$ and $(\pm 1, 0)$, as before. As we have mentioned, this method only locates potential extrema; whether they are maxima, minima, or neither must be determined by other means, such as geometric arguments or the second derivative test given below.[11] ▲

EXAMPLE 3 Maximize the function $f(x, y, z) = x + z$ subject to the constraint $x^2 + y^2 + z^2 = 1$.

SOLUTION By Theorem 7 we know that the function f restricted to the unit sphere $x^2 + y^2 + z^2 = 1$ has a maximum (and also a minimum). To find the maximum, we again use the Lagrange multiplier theorem. We seek λ and (x, y, z) such that

$$1 = 2x\lambda, \qquad 0 = 2y\lambda, \qquad \text{and} \qquad 1 = 2z\lambda,$$

and

$$x^2 + y^2 + z^2 = 1.$$

From the first or the third equation, we see that $\lambda \neq 0$. Thus, from the second equation, we get $y = 0$. From the first and third equations, $x = z$, and so from the fourth, $x = \pm 1/\sqrt{2} = z$. Hence, our points are $(1/\sqrt{2}, 0, 1/\sqrt{2})$ and $(-1/\sqrt{2}, 0, -1/\sqrt{2})$. Comparing the values of f at these points, we can see that the first point yields the maximum of f (restricted to the constraint) and the second the minimum. ▲

EXAMPLE 4 Assume that among all rectangular boxes with fixed surface area of 10 square meters there is a box of largest possible volume. Find its dimensions.

SOLUTION If x, y, and z are the lengths of the sides, $x \geq 0$, $y \geq 0$, $z \geq 0$, respectively, and the volume is $f(x, y, z) = xyz$. The constraint is $2(xy + xz + yz) = 10$;

---

[11] In these examples, $\nabla g(\mathbf{x}_0) \neq \mathbf{0}$ on the surface S, as required by the Lagrange multiplier theorem. If $\nabla g(\mathbf{x}_0)$ were zero for some $\mathbf{x}_0$ on S, then it would have to be included among the possible extrema.

that is, $xy + xz + yz = 5$. Thus, the Lagrange multiplier conditions are

$$\begin{aligned} yz &= \lambda(y+z) \\ xz &= \lambda(x+z) \\ xy &= \lambda(y+x) \\ xy + xz + yz &= 5. \end{aligned}$$

First of all, $x \neq 0$, because $x = 0$ implies $yz = 5$ and $0 = \lambda z$, so that $\lambda = 0$ and we get the contradictory equation $yz = 0$. Similarly, $y \neq 0, z \neq 0, x + y \neq 0$. Elimination of λ from the first two equations gives $yz/(y+z) = xz/(x+z)$, which gives $x = y$; similarly, $y = z$. Substituting these values into the last equation, we obtain $3x^2 = 5$, or $x = \sqrt{5/3}$. Thus, we get the solution $x = y = z = \sqrt{5/3}$, and $xyz = (5/3)^{3/2}$. This (cubical) shape must therefore maximize the volume, assuming there is a box of maximum volume. ▲

Existence of Solutions

We should note that the solution to Example 4 does *not* demonstrate that the cube is the rectangular box of largest volume with a given fixed surface area; it proves that the cube is the only possible candidate for a maximum. We shall sketch a proof that it really is the maximum later. The distinction between showing that there is *only one possible solution* to a problem and that, in fact, *a solution exists* is a subtle one that many (even great) mathematicians have overlooked.

Queen Dido (ca. 900 B.C.) realized that among all planar regions with fixed circumference, the disc is the region of maximum area. It is not difficult to prove this fact under the assumption that there is a region of maximum area; however, proving that such a region of maximum area exists is quite another (difficult) matter. A complete proof was not given until the second half of the nineteenth century by the German mathematician Weierstrass.

Let us consider a nonmathematical parallel to this situation. Put yourself in the place of Lord Peter Wimsey, Dorothy Sayers' famous detective:

> "Undoubtedly," said Wimsey, "but if you think that this identification is going to make life one grand, sweet song for you, you are mistaken. . . . Since we have devoted a great deal of time and thought to the case on the assumption that it was murder, it's a convenience to know that the assumption is correct."

Wimsey has found the body of a dead man, and after some time has located ten suspects. He is sure that no one else other than one of the suspects could be the murderer. By collecting all the evidence and checking alibis, he then reduces the number of suspects one by one, until, finally, only the butler remains; hence he is the murderer! But wait, Peter is a very cautious man. By checking everything once again, he discovers that the man died by suicide; so there is no murder. You see the point: It does not suffice to find a clear and uniquely determined suspect in a criminal case where murder is suspected; you must prove that a murder actually took place.

The same goes for our cube; the fact that it is the only possible candidate for a maximum does not prove that it is maximum. (For more information see *The Parsimonious Universe: Shape and Form in the Natural World*, by S. Hildebrandt and A. Tromba, Springer-Verlag, New York/Berlin, 1995.)

The key to showing that $f(x, y, z) = xyz$ really has a maximum lies in the fact that f is a continuous function that is defined on the unbounded surface $S: xy + xz + yz = 5$, and not on a bounded set, which includes its boundary, where Theorem 7 of Section 3.3 would apply. We have already seen problems of this sort for functions of one and two variables.

The way to show that $f(x, y, z) = xyz \geq 0$ does indeed have a maximum on $xy + yz + xz = 5$ is to show that if either x, y, or z tend to ∞, then $f(x, y, z) \to 0$. We may then conclude that the maximum of f on S must exist by appealing to Theorem 7 (the student should supply the details). So, suppose (x, y, z) lies in S and $x \to \infty$, then $y \to 0$ and $z \to 0$ (why?). Multiplying the equation defining S by z we obtain the equation $xyz + xz^2 + yz^2 = 5z \to 0$ as $x \to \infty$. Because $x, y, z \geq 0$, $xyz = f(x, y, z) \to 0$. Similarly, $f(x, y, z) \to 0$ if either y or z tend to ∞. Thus, a box of maximum volume must exist.

Some general guidelines may be useful for maximum and minimum problems with constraints. First of all, if the surface S is bounded (as an ellipsoid is, for example), then f must have a maximum and a minimum on S. (See Theorem 7 in the preceding section.) In particular, if f has only two points satisfying the conditions of the Lagrange multiplier theorems or Theorem 9, then one must be a maximum and one must be a minimum. Evaluating f at each point will tell the maximum from the minimum. However, if there are more than two such points, some can be saddle points. Also, if S is not bounded (for example, if it is a hyperboloid), then f need not have any maxima or minima.

Several Constraints

If a surface S is defined by a number of constraints, namely,

$$\left.\begin{aligned} g_1(x_1, \dots, x_n) &= c_1 \\ g_2(x_1, \dots, x_n) &= c_2 \\ &\vdots \\ g_k(x_1, \dots, x_n) &= c_k \end{aligned}\right\}, \tag{4}$$

then the Lagrange multiplier theorem may be generalized as follows: *If f has a maximum or a minimum at $\mathbf{x}_0$ on S, there must exist constants $\lambda_1, \dots, \lambda_k$ such that*[12]

$$\nabla f(\mathbf{x}_0) = \lambda_1 \nabla g_1(\mathbf{x}_0) + \cdots + \lambda_k \nabla g_k(\mathbf{x}_0). \tag{5}$$

[12]As with the hypothesis $\nabla g(\mathbf{x}_0) \neq \mathbf{0}$ in the Lagrange multiplier theorem, here one must assume that the vectors $\nabla g_1(\mathbf{x}_0), \dots, \nabla g_k(\mathbf{x}_0)$ are linearly independent; that is, each $\nabla g_i(\mathbf{x}_0)$ is not a linear combination of the other $\nabla g_j(\mathbf{x}_0)$, $j \neq i$.

This case may be proved by generalizing the method used to prove the Lagrange multiplier theorem. Let us give an example of how this more general formulation is used.

EXAMPLE 5 Find the extreme points of $f(x, y, z) = x + y + z$ subject to the two conditions $x^2 + y^2 = 2$ and $x + z = 1$.

SOLUTION Here there are two constraints:

$$g_1(x, y, z) = x^2 + y^2 - 2 = 0 \qquad \text{and} \qquad g_2(x, y, z) = x + z - 1 = 0.$$

Thus, we must find x, y, z, λ_1, and λ_2 such that

$$\nabla f(x, y, z) = \lambda_1 \nabla g_1(x, y, z) + \lambda_2 \nabla g_2(x, y, z),$$
$$g_1(x, y, z) = 0, \qquad \text{and} \qquad g_2(x, y, z) = 0.$$

Computing the gradients and equating components, we get

$$\begin{aligned} 1 &= \lambda_1 \cdot 2x + \lambda_2 \cdot 1, \\ 1 &= \lambda_1 \cdot 2y + \lambda_2 \cdot 0, \\ 1 &= \lambda_1 \cdot 0 + \lambda_2 \cdot 1, \\ x^2 + y^2 &= 2, \qquad \text{and} \qquad x + z = 1. \end{aligned}$$

These are five equations for x, y, z, λ_1, and λ_2. From the third equation, $\lambda_2 = 1$, and so $2x\lambda_1 = 0$, $2y\lambda_1 = 1$. Because the second implies $\lambda_1 \neq 0$, we have $x = 0$. Thus, $y = \pm\sqrt{2}$ and $z = 1$. Hence, the possible extrema are $(0, \pm\sqrt{2}, 1)$. By inspection, $(0, \sqrt{2}, 1)$ gives a relative maximum, and $(0, -\sqrt{2}, 1)$ a relative minimum.

The condition $x^2 + y^2 = 2$ implies that x and y must be bounded. The condition $x + z = 1$ implies that z is also bounded. If follows that the constraint set S is closed and bounded. By Theorem 7 it follows that f has a maximum and minimum on S that must therefore occur at $(0, \sqrt{2}, 1)$ and $(0, -\sqrt{2}, 1)$, respectively. ▲

The method of Lagrange multipliers provides us with another tool to locate the absolute maxima and minima of differentiable functions on bounded regions in $\mathbb{R}^2$ (see the strategy for finding absolute maximum and minimum in Section 3.3).

EXAMPLE 6 Find the absolute maximum of $f(x, y) = xy$ on the unit disk D, where D is the set of points (x, y) with $x^2 + y^2 \leq 1$.

SOLUTION By Theorem 7 of Section 3.3, we know the absolute maximum exists. First, we find all the critical points of f in U, the set of points (x, y) with

$x^2 + y^2 < 1$. Because

$$\frac{\partial f}{\partial x} = y \quad \text{and} \quad \frac{\partial f}{\partial y} = x,$$

$(0, 0)$ is the only critical point of f in U. Now consider f on the unit circle, the level curve $g(x, y) = 1$, where $g(x, y) = x^2 + y^2$. To locate the maximum and minimum of f on C, we write down the Lagrange multiplier equations: $\nabla f(x, y) = (y, x) = \lambda \nabla g(x, y) = \lambda(2x, 2y)$ and $x^2 + y^2 = 1$. Rewriting these in component form, we get

$$\begin{aligned} y &= 2\lambda x, \\ x &= 2\lambda y, \\ x^2 + y^2 &= 1. \end{aligned}$$

Thus,

$$y = 4\lambda^2 y,$$

or $\lambda = \pm 1/2$ and $y = \pm x$, which means that $x^2 + x^2 = 2x^2 = 1$ or $x = \pm 1/\sqrt{2}$, $y = \pm 1/\sqrt{2}$. On C we compute four candidates for the absolute maximum and minimum, namely,

$$\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right), \quad \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \quad \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \quad \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right).$$

The value of f at both $(-1/\sqrt{2}, -1/\sqrt{2})$ and $(1/\sqrt{2}, 1/\sqrt{2})$ is $1/2$. The value of f at $(-1/\sqrt{2}, 1/\sqrt{2})$ and $(1/\sqrt{2}, -1/\sqrt{2})$ is $-1/2$, and the value of f at $(0, 0)$ is 0. Therefore, the absolute maximum of f is $1/2$ and the absolute minimum is $-1/2$, both occurring on C. At $(0, 0)$, $\partial^2 f/\partial x^2 = 0$, $\partial^2 f/\partial y^2 = 0$ and $\partial^2 f/\partial x\, \partial y = 1$, so the discriminant is -1 and thus $(0, 0)$ is a saddle point. ▲

EXAMPLE 7 Find the absolute maximum and minimum of $f(x, y) = \frac{1}{2}x^2 + \frac{1}{2}y^2$ in the elliptical region D defined by $\frac{1}{2}x^2 + y^2 \leq 1$.

SOLUTION Again by Theorem 7, Section 3.3, the absolute maximum exists. We first locate the critical points of f in U, the set of points (x, y) with $\frac{1}{2}x^2 + y^2 < 1$. Because

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

the only critical point is the origin $(0, 0)$.

We now find the maximum and minimum of f on C, the boundary of U, which is the level curve $g(x, y) = 1$, where $g(x, y) = \frac{1}{2}x^2 + y^2$. The Lagrange multiplier equations are

$$\nabla f(x, y) = (x, y) = \lambda \nabla g(x, y) = \lambda(x, 2y)$$

and $(x^2/2) + y^2 = 1$. In other words,

$$x = \lambda x$$
$$y = 2\lambda y$$
$$\frac{x^2}{2} + y^2 = 1.$$

If $x = 0$, then $y = \pm 1$ and $\lambda = \frac{1}{2}$. If $y = 0$, then $x = \pm\sqrt{2}$ and $\lambda = 1$. If $x \neq 0$ and $y \neq 0$, we get both $\lambda = 1$ and $1/2$, which is impossible. Thus, the candidates for the maxima and minima of f on C are $(0, \pm 1)$, $(\pm\sqrt{2}, 0)$ and for f inside D, the candidate is $(0, 0)$. The value of f at $(0, \pm 1)$ is $1/2$, at $(\pm\sqrt{2}, 0)$ it is 1, and at $(0, 0)$ it is 0. Thus, the absolute minimum of f occurs at $(0, 0)$ and is 0. The absolute maximum of f on D is thus 1 and occurs at the points $(\pm\sqrt{2}, 0)$. ▲

Global Maxima and Minima

The method of Lagrange multipliers enhances our techniques for finding global maxima and minima. In this respect, the following is useful.

DEFINITION Let U be an open region in $\mathbb{R}^n$ with boundary ∂U. We say that ∂U is ***smooth*** if ∂U is the level set of a smooth function g whose gradient ∇g never vanishes (i.e., $\nabla g \neq \mathbf{0}$). Then we have the following strategy.

Lagrange Multiplier Strategy for Finding Absolute Maxima and Minima on Regions with Boundary Let f be a differentiable function on a closed and bounded region $D = U \cup \partial U$, U open in $\mathbb{R}^n$, with smooth boundary ∂U.

To find the absolute maximum and minimum of f on D:

(i) Locate all critical points of f in U.

(ii) Use the method of Lagrange multiplier to locate all the critical points of $f|\partial U$.

(iii) Compute the values of f at all these critical points.

(iv) Select the largest and the smallest.

EXAMPLE 8 Find the absolute maximum and minimum of the function $f(x, y, z) = x + y + z$ on the set $D = \{(x, y, z) \mid x^2 + y^2 + z^2 \leq 1\}$.

SOLUTION As in the previous examples, we know the absolute maximum and minimum exists. Now $D = U \cup \partial U$, where

$$U = \{(x, y, z) \mid x^2 + y^2 + z^2 < 1\}$$

and

$$\partial U = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}.$$

The gradient of f is $\nabla f = (1, 1, 1)$, and so f has no critical points in U. Therefore, the maximum and minimum values of f must occur on ∂U.

Let $g(x, y, z) = x^2 + y^2 + z^2$. Then ∂U is the level set $g(x, y, z) = 1$. By the method of Lagrange multipliers, the maximum and minimum must occur at a critical point of $f|\partial U$, that is, at a point $\mathbf{x}_0$ where $\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0)$ for some scalar λ.

Thus,

$$(1, 1, 1) = \lambda(2x, 2y, 2z); \qquad \text{that is,} \qquad x = \frac{1}{2\lambda}, \qquad y = \frac{1}{2\lambda}, \qquad z = \frac{1}{2\lambda}.$$

Because $x^2 + y^2 + z^2 = 1$, we obtain $\lambda = \pm\sqrt{3}/2$ and so $\mathbf{x}_0 = \pm(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$. Clearly, $-(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ is the point where f assumes its absolute minimum (namely, $-\sqrt{3}$) and $(1/\sqrt{3}, 1/\sqrt{3}, 1\sqrt{3})$, the point where f assumes its maximum value $\sqrt{3}$. ▲

Two Additional Applications

We now present two further applications of the mathematical techniques developed in this section to geometry and to economics. We shall begin wth a geometric example.

EXAMPLE 9 Suppose we have a curve defined by the equation

$$\phi(x, y) = Ax^2 + 2Bxy + Cy^2 - 1 = 0.$$

Find the maximum and minimum distance of the curve to the origin. (These are the lengths of the ***semimajor*** and the ***semiminor*** axis of this quadric.)

SOLUTION The problem is equivalent to finding the extreme values of $f(x, y) = x^2 + y^2$ subject to the constraining condition $\phi(x, y) = 0$. Using the Lagrange multiplier method, we have the following equations:

$$2x + \lambda(2Ax + 2By) = 0 \tag{6}$$

$$2y + \lambda(2Bx + 2Cy) = 0 \tag{7}$$

$$Ax^2 + 2Bxy + Cy^2 = 1. \tag{8}$$

Adding x times equation (6) to y times equation (7), we obtain

$$2(x^2+y^2)+2\lambda(Ax^2+2Bxy+Cy^2)=0.$$

By equation (8), it follows that $x^2+y^2+\lambda=0$. Let $t=-1/\lambda=1/(x^2+y^2)$ [the case $\lambda=0$ is impossible, because (0, 0) is not on the curve $\phi(x,y)=0$]. Then equations (6) and (7) can be written as follows:

$$\begin{aligned} 2(A-t)x+2By&=0\\ 2Bx+2(C-t)y&=0. \end{aligned} \tag{9}$$

If these two equations are to have a nontrivial solution [remember that $(x,y)=(0,0)$ is not on our curve and so is not a solution], it follows from a theorem of linear algebra that their determinant vanishes:[13]

$$\begin{vmatrix} A-t & B \\ B & C-t \end{vmatrix}=0.$$

Because this equation is quadratic in t, there are two solutions, which we shall call t_1 and t_2. Because $-\lambda=x^2+y^2$, we have $\sqrt{x^2+y^2}=\sqrt{-\lambda}$. Now $\sqrt{x^2+y^2}$ is the distance from the point (x,y) to the origin. Therefore, if (x_1,y_1) and (x_2,y_2) denote the nontrivial solutions to equation (9) corresponding to t_1 and t_2, and if t_1 and t_2 are positive, we get $\sqrt{x_2^2+y_2^2}=1/\sqrt{t_2}$ and $\sqrt{x_1^2+y_1^2}=1/\sqrt{t_1}$. Consequently, if $t_1>t_2$, the lengths of the semiminor and semimajor axes are $1/\sqrt{t_1}$ and $1/\sqrt{t_2}$, respectively. If the curve is an ellipse, both t_1 and t_2 are, in fact, real and positive. What happens with a hyperbola or a parabola? ▲

Finally, we discuss an application to economics.

EXAMPLE 10 Suppose that the output of a manufacturing firm is a quantity Q of a certain product, where Q is a function $f(K,L)$, where K is the amount of capital equipment (or investment) and L is the amount of labor used. If the price of labor is p, the price of capital is q, and the firm can spend no more than B dollars, how can we find the amount of capital and labor to maximize the output Q?

SOLUTION We would expect that if the amount of capital or labor is increased, then the output Q should also increase; that is,

$$\frac{\partial Q}{\partial K}\geq 0 \qquad \text{and} \qquad \frac{\partial Q}{\partial L}\geq 0.$$

[13]The matrix of coefficients of the equations cannot have an inverse, because this would imply that the solution is zero. Recall that a matrix that does not have an inverse has determinant zero.

We also expect that as more labor is added to a given amount of capital equipment, we get less additional output for our effort; that is,

$$\frac{\partial^2 Q}{\partial L^2} < 0.$$

Similarly,

$$\frac{\partial^2 Q}{\partial K^2} < 0.$$

With these assumptions on Q, it is reasonable to expect the level curves of output (called ***isoquants***) $Q(K, L) = c$ to look something like the curves sketched in Figure 3.4.5, with $c_1 < c_2 < c_3$.

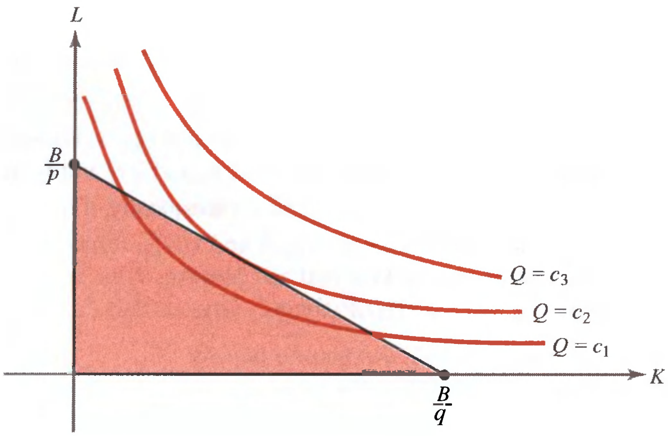

Figure 3.4.5 What is the largest value of Q in the shaded triangle?

We can interpret the convexity of the isoquants as follows: As one moves to the right along a given isoquant, it takes more and more capital to replace a unit of labor and still produce the same output. The budget constraint means that we must stay inside the triangle bounded by the axes and the line $pL + qK = B$. Geometrically, it is clear that we produce the most by spending all our money in such a way as to pick the isoquant that just touches, but does not cross, the budget line.

Because the maximum point lies on the boundary of our domain, we apply the method of Lagrange multipliers to find the maximum. To maximize $Q = f(K, L)$ subject to the constraint $pL + qK = B$, we look for critical points of the auxiliary function,

$$h(K, L, \lambda) = f(K, L) - \lambda(pL + qK - B).$$

Thus, we want

$$\frac{\partial Q}{\partial K} = \lambda q, \qquad \frac{\partial Q}{\partial L} = \lambda p, \qquad \text{and} \qquad pL + qK = B.$$

These are the conditions we must meet in order to maximize output. (The reader is asked to work out a specific case in Exercise 31.) ▲

In the preceding example, λ represents something interesting. Let $k = qK$ and $l = pL$, so that k is the dollar value of the capital used and l is the dollar value of the labor used. Then the first two equations become

$$\frac{\partial Q}{\partial k} = \frac{1}{q}\frac{\partial Q}{\partial K} = \lambda = \frac{1}{p}\frac{\partial Q}{\partial L} = \frac{\partial Q}{\partial l}.$$

Thus, at the optimum production point the marginal change in output per dollar's worth of additional capital investment is equal to the marginal change of output per dollar's worth of additional labor, and λ is this common value. At the optimum point, the exchange of a dollar's worth of capital for a dollar's worth of labor does not change the output. Away from the optimum point the marginal outputs are different, and one exchange or the other will increase the output.

A Second Derivative Test for Constrained Extrema

In Section 3.3, we developed a second derivative test for extrema of functions of several variables by looking at the second-degree term in the Taylor series of f. If the Hessian matrix of second partial derivatives is either positive-definite or negative-definite at a critical point of f, this point is a relative minimum or maximum, respectively.

The question naturally arises as to whether there is a second derivative test for maximum and minimum problems *in the presence of constraints*. The answer is yes and the test involves a matrix called a *bordered Hessian*. We will first discuss the test and how to apply it for the case of a function $f(x, y)$ of two variables subject to the constraint $g(x, y) = c$.

THEOREM 10 Let $f\colon U \subset \mathbb{R}^2 \to \mathbb{R}$ and $g\colon U \subset \mathbb{R}^2 \to \mathbb{R}$ be smooth (at least C^2) functions. Let $\mathbf{v}_0 \in U$, $g(\mathbf{v}_0) = c$, and S be the level curve for g with value c. Assume that $\nabla g(\mathbf{v}_0) \neq \mathbf{0}$ and that there is a real number λ such that $\nabla f(\mathbf{v}_0) = \lambda \nabla g(\mathbf{v}_0)$. Form the auxiliary function $h = f - \lambda g$ and the ***bordered Hessian*** determinant

$$|\bar{H}| = \begin{vmatrix} 0 & -\dfrac{\partial g}{\partial x} & -\dfrac{\partial g}{\partial y} \\[2mm] -\dfrac{\partial g}{\partial x} & \dfrac{\partial^2 h}{\partial x^2} & \dfrac{\partial^2 h}{\partial x\,\partial y} \\[2mm] -\dfrac{\partial g}{\partial y} & \dfrac{\partial^2 h}{\partial x\,\partial y} & \dfrac{\partial^2 h}{\partial y^2} \end{vmatrix} \quad \text{evaluated at } \mathbf{v}_0.$$

(i) If $|\bar{H}| > 0$, then $\mathbf{v}_0$ is a local maximum point for $f|S$.

(ii) If $|\bar{H}| < 0$, then $\mathbf{v}_0$ is a local minimum point for $f|S$.

(iii) If $|\bar{H}| = 0$, the test is inconclusive and $\mathbf{v}_0$ may be a minimum, a maximum, or neither.

This theorem is proved in the Internet supplement for this section.

EXAMPLE 11 Find extreme points of $f(x, y) = (x - y)^n$ subject to the constraint $x^2 + y^2 = 1$, where $n \geq 1$.

SOLUTION We set the first derivatives of the auxiliary function h defined by $h(x, y, \lambda) = (x - y)^n - \lambda(x^2 + y^2 - 1)$ equal to 0:

$$\begin{aligned} n(x-y)^{n-1} - 2\lambda x &= 0 \\ -n(x-y)^{n-1} - 2\lambda y &= 0 \\ -(x^2 + y^2 - 1) &= 0. \end{aligned}$$

From the first two equations we see that $\lambda(x + y) = 0$. If $\lambda = 0$, then $x = y = \pm\sqrt{2}/2$. If $\lambda \neq 0$, then $x = -y$. The four critical points are represented in Figure 3.4.6 and the corresponding values of $f(x, y)$ are listed below:

(A)	$x = \sqrt{2}/2$	$y = \sqrt{2}/2$	$\lambda = 0$	$f(x, y) = 0$
(B)	$x = \sqrt{2}/2$	$y = -\sqrt{2}/2$	$\lambda = n(\sqrt{2})^{n-2}$	$f(x, y) = (\sqrt{2})^n$
(C)	$x = -\sqrt{2}/2$	$y = -\sqrt{2}/2$	$\lambda = 0$	$f(x, y) = 0$
(D)	$x = -\sqrt{2}/2$	$y = \sqrt{2}/2$	$\lambda = (-1)^{n-2} n(\sqrt{2})^{n-2}$	$f(x, y) = (-\sqrt{2})^n$.

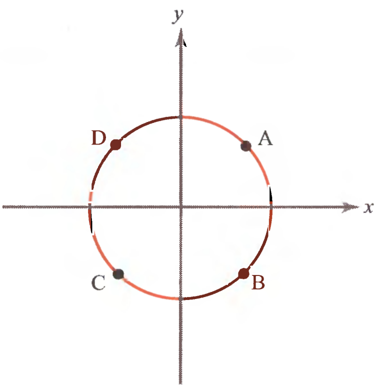

Figure 3.4.6 The four critical points in Example 11.

By inspection, we see that if n is even, then A and C are minimum points and B and D are maxima. If n is odd, then B is a maximum point, D is a minimum, and A and C are neither. Let us see whether Theorem 10 is consistent with these observations.

The bordered Hessian determinant is

$$|\bar{H}| = \begin{vmatrix} 0 & -2x & -2y \\ -2x & n(n-1)(x-y)^{n-2} - 2\lambda & -n(n-1)(x-y)^{n-2} \\ -2y & -n(n-1)(x-y)^{n-2} & n(n-1)(x-y)^{n-2} - 2\lambda \end{vmatrix}$$

$$= -4n(n-1)(x-y)^{n-2}(x+y)^2 + 8\lambda(x^2 - y^2).$$

If $n = 1$ or if $n \geq 3$, $|\bar{H}| = 0$ at A, B, C, and D. If $n = 2$, then $|\bar{H}| = 0$ at B and D and -16 at A and C. Thus, the second-derivative test picks up the minima at A and C, but is inconclusive in testing the maxima at B and D for $n = 2$. It is also inconclusive for all other values of n. ▲

Just as in the unconstrained case, there is also a second-derivative test for functions of more than two variables. If we are to find extreme points for $f(x_1, \ldots, x_n)$ subject to a single constraint $g(x_1, \ldots, x_n) = c$, we first form the bordered Hessian for the auxiliary function $h(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - \lambda(g(x_1, \ldots, x_n) - c)$ as follows:

$$\begin{vmatrix} 0 & \dfrac{-\partial g}{\partial x_1} & \dfrac{-\partial g}{\partial x_2} & \cdots & \dfrac{-\partial g}{\partial x_n} \\ \dfrac{-\partial g}{\partial x_1} & \dfrac{\partial^2 h}{\partial x_1^2} & \dfrac{\partial^2 h}{\partial x_1\, \partial x_2} & \cdots & \dfrac{\partial^2 h}{\partial x_1\, \partial x_n} \\ \dfrac{-\partial g}{\partial x_2} & \dfrac{\partial^2 h}{\partial x_1\, \partial x_2} & \dfrac{\partial^2 h}{\partial x_2^2} & \cdots & \dfrac{\partial^2 h}{\partial x_2\, \partial x_n} \\ \vdots & \vdots & \vdots & & \vdots \\ \dfrac{-\partial g}{\partial x_n} & \dfrac{\partial^2 h}{\partial x_1\, \partial x_n} & \dfrac{\partial^2 h}{\partial x_2\, \partial x_n} & \cdots & \dfrac{\partial^2 h}{\partial x_n^2} \end{vmatrix}.$$

Second, we examine the determinants of the diagonal submatrices of order ≥ 3 at the critical points of h. If they are all negative, that is, if

$$\begin{vmatrix} 0 & -\dfrac{\partial g}{\partial x_1} & -\dfrac{\partial g}{\partial x_2} \\ -\dfrac{\partial g}{\partial x_1} & \dfrac{\partial^2 h}{\partial x_1^2} & \dfrac{\partial^2 h}{\partial x_1\, \partial x_2} \\ -\dfrac{\partial g}{\partial x_2} & \dfrac{\partial^2 h}{\partial x_1\, \partial x_2} & \dfrac{\partial^2 h}{\partial x_2^2} \end{vmatrix} < 0, \quad \begin{vmatrix} 0 & -\dfrac{\partial g}{\partial x_1} & -\dfrac{\partial g}{\partial x_2} & -\dfrac{\partial g}{\partial x_3} \\ -\dfrac{\partial g}{\partial x_1} & \dfrac{\partial^2 h}{\partial x_1^2} & \dfrac{\partial^2 h}{\partial x_1\, \partial x_2} & \dfrac{\partial^2 h}{\partial x_1\, \partial x_3} \\ -\dfrac{\partial g}{\partial x_2} & \dfrac{\partial^2 h}{\partial x_1\, \partial x_2} & \dfrac{\partial^2 h}{\partial x_2^2} & \dfrac{\partial^2 h}{\partial x_2\, \partial x_3} \\ -\dfrac{\partial g}{\partial x_3} & \dfrac{\partial^2 h}{\partial x_1\, \partial x_3} & \dfrac{\partial^2 h}{\partial x_2\, \partial x_3} & \dfrac{\partial^2 h}{\partial x_3^2} \end{vmatrix} < 0, \ldots,$$

then we are at a local minimum of $f|S$. If they start out with a positive 3×3 subdeterminant and alternate in sign (that is, >0, <0, >0, <0, $\ldots$), then we are at a local maximum. If they are all nonzero and do not fit one of these patterns, then the point is neither a maximum nor a minimum (it is said to be of the saddle type).[14]

EXAMPLE 12 Study the local extreme points of $f(x, y, z) = xyz$ on the surface of the unit sphere $x^2 + y^2 + z^2 = 1$ using the second-derivative test.

SOLUTION Setting the partial derivatives of the auxiliary function $h(x, y, z, \lambda) = xyz - \lambda(x^2 + y^2 + z^2 - 1)$ equal to zero gives

$$\begin{aligned} yz &= 2\lambda x \\ xz &= 2\lambda y \\ xy &= 2\lambda z \\ x^2 + y^2 + z^2 &= 1. \end{aligned}$$

Thus, $3xyz = 2\lambda(x^2 + y^2 + z^2) = 2\lambda$. If $\lambda = 0$, the solutions are $(x, y, z, \lambda) = (\pm 1, 0, 0, 0)$, $(0, \pm 1, 0, 0)$, and $(0, 0, \pm 1, 0)$. If $\lambda \neq 0$, then we have $2\lambda = 3xyz = 6\lambda z^2$ and so $z^2 = \frac{1}{3}$. Similarly, $x^2 = y^2 = \frac{1}{3}$. Thus, the solutions are given by $\lambda = \frac{3}{2}xyz = \pm\sqrt{3}/6$. The critical points of h and the corresponding values of f are given in Table 3.1. From it, we see that points E, F, G, and K are minima. Points D, H, I, and J are maxima. To see whether this is in accord with the second-derivative

Table 3.1 The critical points A, B, ..., J, K of h and corresponding values of f

	x	y	z	λ	$f(x, y, z)$
±A	± 1	0	0	0	0
±B	0	± 1	0	0	0
±C	0	0	± 1	0	0
D	$\sqrt{3}/3$	$\sqrt{3}/3$	$\sqrt{3}/3$	$\sqrt{3}/6$	$\sqrt{3}/9$
E	$-\sqrt{3}/3$	$\sqrt{3}/3$	$\sqrt{3}/3$	$-\sqrt{3}/6$	$-\sqrt{3}/9$
F	$\sqrt{3}/3$	$-\sqrt{3}/3$	$\sqrt{3}/3$	$-\sqrt{3}/6$	$-\sqrt{3}/9$
G	$\sqrt{3}/3$	$\sqrt{3}/3$	$-\sqrt{3}/3$	$-\sqrt{3}/6$	$-\sqrt{3}/9$
H	$\sqrt{3}/3$	$-\sqrt{3}/3$	$-\sqrt{3}/3$	$\sqrt{3}/6$	$\sqrt{3}/9$
I	$-\sqrt{3}/3$	$\sqrt{3}/3$	$-\sqrt{3}/3$	$\sqrt{3}/6$	$\sqrt{3}/9$
J	$-\sqrt{3}/3$	$-\sqrt{3}/3$	$\sqrt{3}/3$	$\sqrt{3}/6$	$\sqrt{3}/9$
K	$-\sqrt{3}/3$	$-\sqrt{3}/3$	$-\sqrt{3}/3$	$-\sqrt{3}/6$	$-\sqrt{3}/9$

[14]For a detailed discussion, see C. Caratheodory, *Calculus of Variations and Partial Differential Equations*, Holden-Day, San Francisco, 1965; Y. Murata, *Mathematics for Stability and Optimization of Economic Systems*, Academic Press, New York, 1977, pp. 263–271; or D. Spring, *Am. Math. Mon.* 92 (1985): 631–643.

test, we need to consider two determinants. First, we look at the following:

$$|\bar{H}_2| = \begin{vmatrix} 0 & -\partial g/\partial x & -\partial g/\partial y \\ -\partial g/\partial x & \partial^2 h/\partial x^2 & \partial^2/\partial x\,\partial y \\ -\partial g/\partial y & \partial^2 h/\partial x\,\partial y & \partial^2 h/\partial y^2 \end{vmatrix} = \begin{vmatrix} 0 & -2x & -2y \\ -2x & -2\lambda & z \\ -2y & z & -2\lambda \end{vmatrix}$$
$$= 8\lambda x^2 + 8\lambda y^2 + 8xyz = 8\lambda(x^2 + y^2 + 2z^2).$$

Observe that sign $(|\bar{H}_2|) =$ sign $\lambda =$ sign (xyz), where the sign of a number is 1 if that number is positive, or is -1 if that number is negative. Second, we consider

$$|\bar{H}_3| = \begin{vmatrix} 0 & -\partial g/\partial x & -\partial g/\partial y & -\partial g/\partial z \\ -\partial g/\partial x & \partial^2 h/\partial x^2 & \partial^2 h/\partial x\,\partial y & \partial^2 h/\partial x\,\partial z \\ -\partial g/\partial y & \partial^2 h/\partial x\,\partial y & \partial^2 h/\partial y^2 & \partial^2 h/\partial y\,\partial z \\ -\partial g/\partial z & \partial^2 h/\partial x\,\partial z & \partial^2 h/\partial y\,\partial z & \partial^2 h/\partial z^2 \end{vmatrix}$$
$$= \begin{vmatrix} 0 & -2x & -2y & -2z \\ -2x & -2\lambda & z & y \\ -2y & z & -2\lambda & x \\ -2z & y & x & -2\lambda \end{vmatrix},$$

which works out to be $+4$ at points $\pm$A, $\pm$B, and $\pm$C and $-\frac{16}{3}$ at the other eight points. At E, F, G, and K, we have $|\bar{H}_2| < 0$ *and* $|\bar{H}_3| < 0$, and so the test indicates these are local minima. At D, H, I, and J we have $|\bar{H}_2| > 0$ and $|\bar{H}_3| < 0$, and so the test says these are local maxima. Finally, the second-derivative test shows that $\pm$A, $\pm$B, and $\pm$C are saddle points. ▲

EXERCISES

In Exercises 1 to 5 find the extrema of f subject to the stated constraints.

1. $f(x, y, z) = x - y + z$, subject to $x^2 + y^2 + z^2 = 2$

2. $f(x, y) = x - y$, subject to $x^2 - y^2 = 2$

3. $f(x, y) = x$, subject to $x^2 + 2y^2 = 3$

4. $f(x, y, z) = x + y + z$, subject to $x^2 - y^2 = 1, 2x + z = 1$

5. $f(x, y) = 3x + 2y$, subject to $2x^2 + 3y^2 = 3$

Find the relative extrema of $f|S$ in Exercises 6 to 9.

6. $f\colon \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x^2 + y^2, S = \{(x, 2) \mid x \in \mathbb{R}\}$

7. $f\colon \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x^2 + y^2, S = \{(x, y) \mid y \geq 2\}$

8. $f\colon \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x^2 - y^2, S = \{(x, \cos x) \mid x \in \mathbb{R}\}$

9. $f\colon \mathbb{R}^3 \to \mathbb{R}, (x, y, z) \mapsto x^2 + y^2 + z^2, S = \{(x, y, z) \mid z \geq 2 + x^2 + y^2\}$

10. Use the method of Lagrange multipliers to find the absolute maximum and minimum values of $f(x, y) = x^2 + y^2 - x - y + 1$ on the unit disk (see Example 10 of Section 3.3).

11. Consider the function $f(x, y) = x^2 + xy + y^2$ defined on the unit disk, namely, $D = \{(x, y) \mid x^2 + y^2 \leq 1\}$. Use the method of Lagrange multipliers to locate the maximum and minimum points for f on the unit circle. Use this to determine the absolute maximum and minimum values for f on D.

12. A rectangular box with no top is to have a surface area of 16 m^2. Find the dimensions that maximize its volume.

13. Design a cylindrical can (with a lid) to contain 1 liter (= 1000 cm^3) of water, using the minimum amount of metal.

14. Show that solutions of equations (4) and (5) are in one-to-one correspondence with the critical points of

$$\begin{aligned} h(x_1, \dots, x_n, \lambda_1, \dots, \lambda_k) = {} & f(x_1, \dots, x_n) - \lambda_1[g_1(x_1, \dots, x_n) - c_1] \\ & - \cdots - \lambda_k[g_k(x_1, \dots, x_n) - c_k]. \end{aligned}$$

15. Find the absolute maximum and minimum for the function $f(x, y, z) = x + y - z$ on the ball $B = \{(x, y, z) \mid x^2 + y^2 + z^2 \leq 1\}$.

16. Repeat Exercise 15 for $f(x, y, z) = x + yz$.

17. A rectangular mirror with area A square feet is to have trim along the edges. If the trim along the horizontal edges costs p cents per foot and that for the vertical edges costs q cents per foot, find the dimensions that will minimize the total cost.

18. An irrigation canal in Arizona has concrete sides and bottom with trapezoidal cross section of area $A = y(x + y \tan\theta)$ and wetted perimeter $P = x + 2y/\cos\theta$, where $x =$ bottom width, $y =$ water depth, $\theta =$ side inclination, measured from vertical. The best design for a fixed inclination θ is found by solving $P =$ minimum subject to the condition $A =$ constant. Show that $y^2 = (A\cos\theta)/(2 - \sin\theta)$.

19. Apply the second-derivative test to study the nature of the extrema in Exercises 1 and 5.

20. A light ray travels from point A to point B crossing a boundary between two media (see Figure 3.4.7). In the first medium its speed is v_1, and in the second it is v_2. Show that the trip is made in minimum time when *Snell's law* holds:

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2}.$$

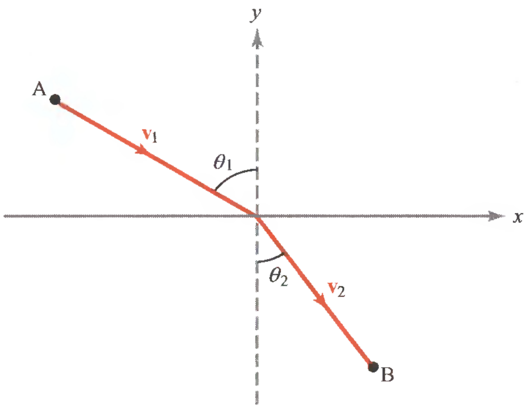

Figure 3.4.7 Snell's law of refraction.

21. A parcel delivery service requires that the dimensions of a rectangular box be such that the length plus twice the width plus twice the height be no more than 108 inches ($l + 2w + 2h \leq 108$). What is the volume of the largest-volume box the company will deliver?

22. Let P be a point on a surface S in $\mathbb{R}^3$ defined by the equation $f(x, y, z) = 1$, where f is of class C^1. Suppose that P is a point where the distance from the origin to S is maximized. Show that the vector emanating from the origin and ending at P is perpendicular to S.

23. Let A be a nonzero symmetric 3×3 matrix. Thus, its entries satisfy $a_{ij} = a_{ji}$. Consider the function $f(\mathbf{x}) = \frac{1}{2}(A\mathbf{x}) \cdot \mathbf{x}$.

(a) What is ∇f?

(b) Consider the restriction of f to the unit sphere $S = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}$ in $\mathbb{R}^3$. By Theorem 7 we know that f must have a maximum and a minimum on S. Show that there must be an $\mathbf{x} \in S$ and a $\lambda \neq 0$ such that $A\mathbf{x} = \lambda\mathbf{x}$. (The vector $\mathbf{x}$ is called an ***eigenvector***, while the scalar λ is called an ***eigenvalue***.)

(c) What are the maxima and minima for f on $B = \{(x, y, z) \mid x^2 + y^2 + z^2 \leq 1\}$?

24. Suppose that A in the function f defined in Exercise 23 is not necessarily symmetric.

(a) What is ∇f?

(b) Can one conclude the existence of an eigenvector and eigenvalues as in Exercise 23?

25. (a) Find the critical points of $x + y^2$ subject to the constraint $2x^2 + y^2 = 1$.

(b) Use the bordered Hessian to classify the critical points.

26. Answer the question posed in the last line of Example 9.

27. Try to find the extrema of $xy + yz$ among points satisfying $xz = 1$.

28. A company's production function is $Q(x, y) = xy$. The cost of production is $C(x, y) = 2x + 3y$. If this company can spend $C(x, y) = 10$, what is the maximum quantity that can be produced?

29. Find the point on the curve $(\cos t, \sin t, \sin(t/2))$ that is farthest from the origin.

30. A firm uses wool and cotton fiber to produce cloth. The amount of cloth produced is given by $Q(x, y) = xy - x - y + 1$, where x is the number of pounds of wool, y the number of pounds of cotton, $x > 1$, and $y > 1$. If wool costs p dollars per pound, and cotton q dollars per pound and the firm can spend B dollars on material, what should the ratio of cotton and wool be to produce the most cloth?

31. Carry out the analysis of Example 10 for the production function $Q(K, L) = AK^{\alpha}L^{1-\alpha}$, where A and α are positive constants and $0 < \alpha < 1$. This is called a ***Cobb–Douglas production function*** and is sometimes used as a simple model for the national economy. Q is then the aggregate output of the economy for a given input of capital and labor.

3.5 The Implicit Function Theorem

In this section, we state two versions of the *implicit function theorem*, arguably the most important theorem in all of mathematical analysis. The entire theoretical basis of the idea of a surface as well as the method of Lagrange multipliers depends on it. Moreover, it is a cornerstone of several fields of mathematics, such as differential topology and geometry.

The One-Variable Implicit Function Theorem

In one-variable calculus, we learn the importance of the inversion process. For example, $x = \ln y$ is the inverse of $y = e^x$, and $x = \sin^{-1} y$ is the inverse of $y = \sin x$. The inversion process is also important for functions of several variables; for example, the switch between Cartesian and polar coordinates in the plane involves inverting two functions of two variables.

Recall from one-variable calculus that if $y = f(x)$ is a C^1 function and $f'(x_0) \neq 0$, then locally near x_0 we can solve for x to give the inverse function: $x = f^{-1}(y)$. We learn that $(f^{-1})'(y) = 1/f'(x)$; that is, $dx/dy = 1/(dy/dx)$. That $y = f(x)$ can be inverted is plausible because $f'(x_0) \neq 0$ means that the slope of $y = f(x)$ is nonzero, so that the graph is rising or falling near x_0. Thus, if we reflect the graph across the line $y = x$, it is still a graph *near* (x_0, y_0) where $y_0 = f(x_0)$. For example, in Figure 3.5.1, we can invert $y = f(x)$ in the shaded box, so in this range, $x = f^{-1}(y)$ is defined.

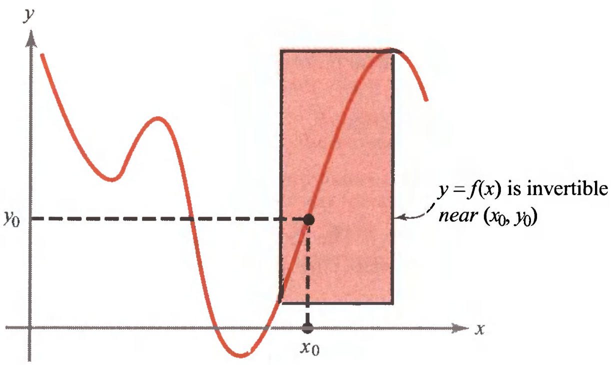

Figure 3.5.1 If $f'(x_0) \neq 0$, then $y = f(x)$ is locally invertible.

A Special Result

We next turn to the situation for real-valued functions of variables $x_1, \ldots, x_n$ and z.

THEOREM 11: Special Implicit Function Theorem Suppose that $F\colon \mathbb{R}^{n+1} \to \mathbb{R}$ has continuous partial derivatives. Denoting points in $\mathbb{R}^{n+1}$ by $(\mathbf{x}, z)$, where $\mathbf{x} \in \mathbb{R}^n$ and $z \in \mathbb{R}$, assume that $(\mathbf{x}_0, z_0)$ satisfies

$$F(\mathbf{x}_0, z_0) = 0 \qquad \text{and} \qquad \frac{\partial F}{\partial z}(\mathbf{x}_0, z_0) \neq 0.$$

Then there is a ball U containing $\mathbf{x}_0$ in $\mathbb{R}^n$ and a neighborhood V of z_0 in $\mathbb{R}$ such that there is a unique function $z = g(\mathbf{x})$ defined for $\mathbf{x}$ in U and z in V that satisfies

$$F(\mathbf{x}, g(\mathbf{x})) = 0.$$

Moreover, if $\mathbf{x}$ in U and z in V satisfy $F(\mathbf{x}, z) = 0$, then $z = g(\mathbf{x})$. Finally, $z = g(\mathbf{x})$ is continuously differentiable, with the derivative given by

$$\mathbf{D}g(\mathbf{x}) = -\frac{1}{\dfrac{\partial F}{\partial z}(\mathbf{x}, z)} \mathbf{D}_{\mathbf{x}} F(\mathbf{x}, z)\bigg|_{z=g(\mathbf{x})},$$

where $\mathbf{D}_{\mathbf{x}} F$ denotes the (partial) derivative of F with respect to the variable $\mathbf{x}$, that is, we have $\mathbf{D}_{\mathbf{x}} F = [\partial F/\partial x_1, \ldots, \partial F/\partial x_n]$; in other words,

$$\frac{\partial g}{\partial x_i} = -\frac{\partial F/\partial x_i}{\partial F/\partial z}, \qquad i = 1, \ldots, n. \tag{1}$$

A proof of this theorem is given in the Internet supplement.

Once it is known that $z = g(\mathbf{x})$ exists and is differentiable, formula (1) may be checked by implicit differentiation; to see this, note that the chain rule applied to $F(\mathbf{x}, g(\mathbf{x})) = 0$ gives

$$\mathbf{D}_{\mathbf{x}} F(\mathbf{x}, g(\mathbf{x})) + \left[\frac{\partial F}{\partial z}(\mathbf{x}, g(\mathbf{x}))\right][\mathbf{D}g(\mathbf{x})] = 0,$$

which is equivalent to formula (1).

EXAMPLE 1 In the special implicit function theorem, it is important to recognize the necessity of taking sufficiently small neighborhoods U and V. For example, consider the equation

$$x^2 + z^2 - 1 = 0,$$

that is, $F(x, z) = x^2 + z^2 - 1$, with $n = 1$. Here $(\partial F/\partial z)(x, z) = 2z$, and so the special implicit function theorem applies to a point (x_0, z_0) satisfying $x_0^2 + z_0^2 - 1 = 0$ and $z_0 \neq 0$. Thus, near such points, z is a unique function of x. This function is

$z = \sqrt{1 - x^2}$ if $z_0 > 0$ and $z = -\sqrt{1 - x^2}$ if $z_0 < 0$. Note that z is defined for $|x| < 1$ only (U must not be too big) and z is unique only if it is near z_0 (V must not be too big). These facts and the nonexistence of $\partial z/\partial x$ at $z_0 = 0$ are, of course, clear from the fact that $x^2 + z^2 = 1$ defines a circle in the xz plane (Figure 3.5.2). ▲

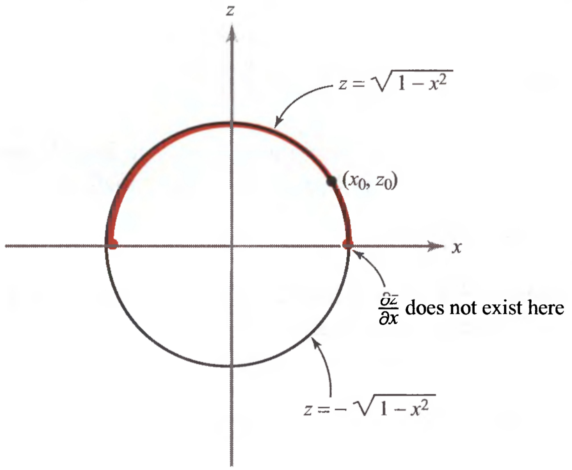

Figure 3.5.2 It is necessary to take small neighborhoods in the implicit function theorem.

The Implicit Function Theorem and Surfaces

Let us apply Theorem 11 to the study of surfaces. We are concerned with the level set of a function $g: U \subset \mathbb{R}^n \to \mathbb{R}$, that is, with the surface S consisting of the set of $\mathbf{x}$ satisfying $g(\mathbf{x}) = c_0$, where $c_0 = g(\mathbf{x}_0)$ and where $\mathbf{x}_0$ is given. Let us take $n = 3$ for concreteness. Thus, we are dealing with the level surface of a function $g(x, y, z)$ through a given point (x_0, y_0, z_0). As in the Lagrange multiplier theorem, assume that $\nabla g(x_0, y_0, z_0) \neq \mathbf{0}$. This means that at least one of the partial derivatives of g is nonzero. For definiteness, suppose that $(\partial g/\partial z)(x_0, y_0, z_0) \neq 0$. By applying Theorem 11 to the function $(x, y, z) \mapsto g(x, y, z) - c_0$, we know there is a unique function $z = k(x, y)$ satisfying $g(x, y, k(x, y)) = c_0$ for (x, y) near (x_0, y_0) and z near z_0. Thus, near z_0 the surface S is the graph of the function k. Because k is continuously differentiable, this surface has a tangent plane at (x_0, y_0, z_0) given by

$$z = z_0 + \left[\frac{\partial k}{\partial x}(x_0, y_0)\right](x - x_0) + \left[\frac{\partial k}{\partial y}(x_0, y_0)\right](y - y_0). \tag{2}$$

But by formula (1),

$$\frac{\partial k}{\partial x}(x_0, y_0) = -\frac{\dfrac{\partial g}{\partial x}(x_0, y_0, z_0)}{\dfrac{\partial g}{\partial z}(x_0, y_0, z_0)} \quad \text{and} \quad \frac{\partial k}{\partial y}(x_0, y_0) = -\frac{\dfrac{\partial g}{\partial y}(x_0, y_0, z_0)}{\dfrac{\partial g}{\partial z}(x_0, y_0, z_0)}.$$

Substituting these two equations into the equation for the tangent plane gives this equivalent description:

$$0 = (z - z_0)\frac{\partial g}{\partial z}(x_0, y_0, z_0) + (x - x_0)\frac{\partial g}{\partial x}(x_0, y_0, z_0) + (y - y_0)\frac{\partial g}{\partial y}(x_0, y_0, z_0);$$

that is,

$$(x - x_0, y - y_0, z - z_0) \cdot \nabla g(x_0, y_0, z_0) = 0.$$

Thus, the tangent plane to the level surface of g is the orthogonal complement to $\nabla g(x_0, y_0, z_0)$ through the point (x_0, y_0, z_0). This agrees with our characterization of tangent planes to level sets from Chapter 2.

We are now ready to complete the proof of the Lagrange multiplier theorem. To do this, we must show that every vector tangent to S at (x_0, y_0, z_0) is tangent to a curve in S. By Theorem 11, we need only show this for a graph of the form $z = k(x, y)$. However, if $\mathbf{v} = (x - x_0, y - y_0, z - z_0)$ is tangent to the graph [that is, if it satisfies equation (2)], then $\mathbf{v}$ is tangent to the path in S given by

$$\mathbf{c}(t) = (x_0 + t(x - x_0), y_0 + t(y - y_0), k(x_0 + t(x - x_0), y_0 + t(y - y_0)))$$

at $t = 0$. This can be checked by using the chain rule. (See Figure 3.5.3.)

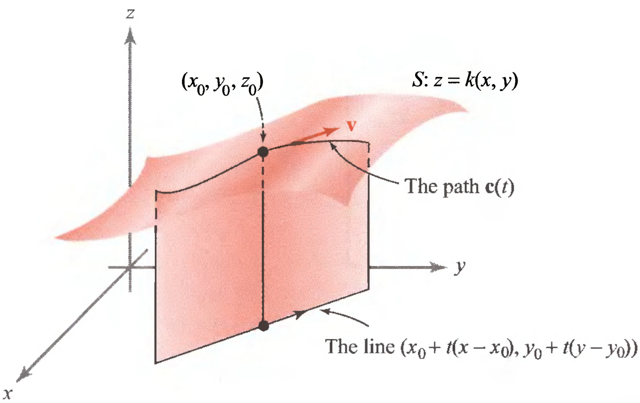

Figure 3.5.3 The construction of a path $\mathbf{c}(t)$ in the surface S whose tangent vector is $\mathbf{v}$.

EXAMPLE 2 Near what points may the surface

$$x^3 + 3y^2 + 8xz^2 - 3z^3y = 1$$

be represented as a graph of a differentiable function $z = k(x, y)$?

SOLUTION Here we take $F(x, y, z) = x^3 + 3y^2 + 8xz^2 - 3z^3y - 1$ and attempt to solve $F(x, y, z) = 0$ for z as a function of (x, y). By Theorem 11, this may be done near a point (x_0, y_0, z_0) if $(\partial F/\partial z)(x_0, y_0, z_0) \neq 0$, that is, if

$$z_0(16x_0 - 9z_0y_0) \neq 0,$$

which means, in turn,

$$z_0 \neq 0 \qquad \text{and} \qquad 16x_0 \neq 9z_0y_0. \quad \blacktriangle$$

General Implicit Function Theorem

Next we shall state, without proof, the *general implicit function theorem*.[15] Instead of attempting to solve one equation for one variable, we attempt to solve m equations for m variables $z_1, \dots, z_m$:

$$\begin{aligned} F_1(x_1, \dots, x_n, z_1, \dots, z_m) &= 0 \\ F_2(x_1, \dots, x_n, z_1, \dots, z_m) &= 0 \\ \vdots \qquad\qquad \vdots \qquad & \quad \vdots \\ F_m(x_1, \dots, x_n, z_1, \dots, z_m) &= 0. \end{aligned} \tag{3}$$

In Theorem 11 we had the condition $\partial F/\partial z \neq 0$. The condition appropriate to the general implicit function theorem is that $\Delta \neq 0$,[16] where Δ is the determinant of the $m \times m$ matrix

$$\begin{bmatrix} \dfrac{\partial F_1}{\partial z_1} & \cdots & \dfrac{\partial F_1}{\partial z_m} \\ \vdots & & \vdots \\ \dfrac{\partial F_m}{\partial z_1} & \cdots & \dfrac{\partial F_m}{\partial z_m} \end{bmatrix}$$

---

[15] For three different proofs of the general case, consult:

(a) E. Goursat, *A Course in Mathematical Analysis*, I, Dover, New York, 1959, p. 45. (This proof derives the general theorem by successive application of Theorem 11.)

(b) T. M. Apostol, *Mathematical Analysis*, 2d ed., Addison-Wesley, Reading, Mass., 1974.

(c) J. E. Marsden and M. Hoffman, *Elementary Classical Analysis*, 2d ed., Freeman, New York, 1993.

Of these sources, the last two use more sophisticated ideas that are usually not covered until a junior-level course in analysis. The first, however, is easily understood by the reader who has some knowledge of linear algebra.

[16] For students who have had linear algebra: The condition $\Delta \neq 0$ has a simple interpretation in the case that F *is linear*; namely, $\Delta \neq 0$ is equivalent to the rank of F being equal to m, which in turn is equivalent to the fact that the solution space of $F = 0$ is m-dimensional.

evaluated at the point $(\mathbf{x}_0, \mathbf{z}_0)$; in the neighborhood of such a point, we can uniquely solve for $\mathbf{z}$ in terms of $\mathbf{x}$.

THEOREM 12: General Implicit Function Theorem If $\Delta \neq 0$, then near the point $(\mathbf{x}_0, \mathbf{z}_0)$, equation (3) defines unique (smooth) functions

$$z_i = k_i(x_1, \dots, x_n) \qquad (i = 1, \dots, m).$$

Their derivatives may be computed by implicit differentiation.

EXAMPLE 3 Show that near the point $(x, y, u, v) = (1, 1, 1, 1)$, we can solve

$$xu + yvu^2 = 2$$
$$xu^3 + y^2v^4 = 2$$

uniquely for u and v as functions of x and y. Compute $\partial u/\partial x$ at the point $(1, 1)$.

SOLUTION To check solvability, we form the equations

$$F_1(x, y, u, v) = xu + yvu^2 - 2$$
$$F_2(x, y, u, v) = xu^3 + y^2v^4 - 2$$

and the determinant

$$\Delta = \begin{vmatrix} \dfrac{\partial F_1}{\partial u} & \dfrac{\partial F_1}{\partial v} \\ \dfrac{\partial F_2}{\partial u} & \dfrac{\partial F_2}{\partial v} \end{vmatrix} \quad \text{at} \quad (1, 1, 1, 1)$$

$$= \begin{vmatrix} x + 2yuv & yu^2 \\ 3u^2x & 4y^2v^3 \end{vmatrix} \quad \text{at} \quad (1, 1, 1, 1)$$

$$= \begin{vmatrix} 3 & 1 \\ 3 & 4 \end{vmatrix} = 9.$$

Because $\Delta \neq 0$, solvability is assured by the general implicit function theorem. To find $\partial u/\partial x$, we implicitly differentiate the given equations in x using the chain rule:

$$x\frac{\partial u}{\partial x} + u + y\frac{\partial v}{\partial x}u^2 + 2yvu\frac{\partial u}{\partial x} = 0$$

$$3xu^2\frac{\partial u}{\partial x} + u^3 + 4y^2v^3\frac{\partial v}{\partial x} = 0.$$

Setting $(x, y, u, v) = (1, 1, 1, 1)$ gives

$$3\frac{\partial u}{\partial x} + \frac{\partial v}{\partial x} = -1$$

$$3\frac{\partial u}{\partial x} + 4\frac{\partial v}{\partial x} = -1.$$

Solving for $\partial u/\partial x$ by multiplying the first equation by 4 and subtracting gives $\partial u/\partial x = -\frac{1}{3}$. ▲

Inverse Function Theorem

A special case of the general implicit function theorem is the *inverse function theorem*. Here we attempt to solve the n equations

$$\left.\begin{aligned} f_1(x_1, \dots, x_n) &= y_1 \\ &\cdots \\ f_n(x_1, \dots, x_n) &= y_n \end{aligned}\right\} \tag{4}$$

for $x_1, \dots, x_n$ as functions of $y_1, \dots, y_n$; that is, we are trying to invert the equations of system (4). This is analogous to forming the inverses of functions like $\sin x = y$ and $e^x = y$, with which the reader should be familiar from elementary calculus. Now, however, we are concerned with functions of several variables. The question of solvability is answered by the general implicit function theorem applied to the functions $y_i - f_i(x_1, \dots, x_n)$ with the unknowns $x_1, \dots, x_n$ (called $z_1, \dots, z_n$ earlier). The condition for solvability in a neighborhood of a point $\mathbf{x}_0$ is $\Delta \neq 0$, where Δ is the determinant of the matrix $\mathbf{D}f(\mathbf{x}_0)$, and $f = (f_1, \dots, f_n)$. The quantity Δ is denoted by $\partial(f_1, \dots, f_n)/\partial(x_1, \dots, x_n)$, or $\partial(y_1, \dots, y_n)/\partial(x_1, \dots, x_n)$ or $J(f)(\mathbf{x}_0)$ and is called the ***Jacobian determinant*** of f. Explicitly,

$$\left.\frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)}\right|_{\mathbf{x}=\mathbf{x}_0} = J(f)(\mathbf{x}_0) = \begin{vmatrix} \dfrac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \cdots & \dfrac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \dfrac{\partial f_n}{\partial x_1}(\mathbf{x}_0) & \cdots & \dfrac{\partial f_n}{\partial x_n}(\mathbf{x}_0) \end{vmatrix}. \tag{5}$$

The reader should note that in the case when f is linear, for example $f(x) = Ax$, where A is an $n \times n$ matrix, the condition $\Delta \neq 0$ is equivalent to the fact that the determinant of A, $\det A \neq 0$, and from Section 1.5 we know that A, and therefore f, has an inverse.

The Jacobian determinant will play an important role in our work on integration (see Chapter 5). The following theorem summarizes this discussion:

THEOREM 13: Inverse Function Theorem Let $U \subset \mathbb{R}^n$ be open and let $f_1: U \to \mathbb{R}, \ldots, f_n: U \to \mathbb{R}$ have continuous partial derivatives. Consider the equations (4) near a given solution $\mathbf{x}_0, \mathbf{y}_0$. If $J(f)(\mathbf{x}_0)$ [defined by equation (5)] is nonzero, then equation (4) can be solved uniquely as $\mathbf{x} = g(\mathbf{y})$ for $\mathbf{x}$ near $\mathbf{x}_0$ and $\mathbf{y}$ near $\mathbf{y}_0$. Moreover, the function g has continuous partial derivatives.

EXAMPLE 4 Consider the equations

$$\frac{x^4 + y^4}{x} = u, \qquad \sin x + \cos y = v.$$

Near which points (x, y) can we solve for x, y in terms of u, v?

SOLUTION Here the functions are $u = f_1(x, y) = (x^4 + y^4)/x$ and $v = f_2(x, y) = \sin x + \cos y$. We want to know the points near which we can solve for x, y as functions of u and v. According to the inverse function theorem, we must first compute the Jacobian determinant $\partial(f_1, f_2)/\partial(x, y)$. We take the domain of $f = (f_1, f_2)$ to be $U = \{(x, y) \in \mathbb{R}^2 \mid x \neq 0\}$. Now

$$\frac{\partial(f_1, f_2)}{\partial(x, y)} = \begin{vmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} \end{vmatrix} = \begin{vmatrix} \dfrac{3x^4 - y^4}{x^2} & \dfrac{4y^3}{x} \\ \cos x & -\sin y \end{vmatrix} = \frac{\sin y}{x^2}(y^4 - 3x^4) - \frac{4y^3}{x} \cos x.$$

Therefore, at points where this does not vanish we can solve for x, y in terms of u and v. In other words, we can solve for x, y near those x, y for which $x \neq 0$ and $(\sin y)(y^4 - 3x^4) \neq 4xy^3 \cos x$. Such conditions generally cannot be solved explicitly. For example, if $x_0 = \pi/2$, $y_0 = \pi/2$, we can solve for x, y near (x_0, y_0) because there, $\partial(f_1, f_2)/\partial(x, y) \neq 0$. ▲

EXERCISES

1. Let $F(x, y) = 0$ define a curve in the xy plane through the point (x_0, y_0), where F is C^1. Assume that $(\partial F/\partial y)(x_0, y_0) \neq 0$. Show that this curve can be locally represented by the graph of a function $y = g(x)$. Show that (i) the line orthogonal to $\nabla F(x_0, y_0)$ agrees with (ii) the tangent line to the graph of $y = g(x)$.

2. Show that $xy + z + 3xz^5 = 4$ is solvable for z as a function of (x, y) near $(1, 0, 1)$. Compute $\partial z/\partial x$ and $\partial z/\partial y$ at $(1, 0)$.

3. (a) Check directly (i.e., without using Theorem 11) where we can solve the equation $F(x, y) = y^2 + y + 3x + 1 = 0$ for y in terms of x.

(b) Check that your answer in part (a) agrees with the answer you expect from the implicit function theorem. Compute dy/dx.

4. Repeat Exercise 3 with $F(x, y) = xy^2 - 2y + x^2 + 2 = 0$.

5. Show that $x^3z^2 - z^3yx = 0$ is solvable for z as a function of (x, y) near $(1, 1, 1)$, but not near the origin. Compute $\partial z/\partial x$ and $\partial z/\partial y$ at $(1, 1)$.

6. Discuss the solvability in the system

$$\begin{aligned} 3x + 2y + z^2 + u + v^2 &= 0 \\ 4x + 3y + z + u^2 + v + w + 2 &= 0 \\ x + z + w + u^2 + 2 &= 0 \end{aligned}$$

for u, v, w in terms of x, y, z near $x = y = z = 0$, $u = v = 0$, and $w = -2$.

7. Discuss the solvability of

$$\begin{aligned} y + x + uv &= 0 \\ uxy + v &= 0 \end{aligned}$$

for u, v in terms of x, y near $x = y = u = v = 0$ and check directly.

8. Investigate whether or not the system

$$\begin{aligned} u(x, y, z) &= x + xyz \\ v(x, y, z) &= y + xy \\ w(x, y, z) &= z + 2x + 3z^2 \end{aligned}$$

can be solved for x, y, z in terms of u, v, w near $(x, y, z) = (0, 0, 0)$.

9. Consider $f(x, y) = ((x^2 - y^2)/(x^2 + y^2), xy/(x^2 + y^2))$. Does this map of $\mathbb{R}^2\backslash(0, 0)$ to $\mathbb{R}^2$ have a local inverse near $(x, y) = (0, 1)$?

10. (a) Define $x\colon \mathbb{R}^2 \to \mathbb{R}$ by $x(r, \theta) = r\cos\theta$ and define $y\colon \mathbb{R}^2 \to \mathbb{R}$ by $y(r, \theta) = r\sin\theta$. Show that

$$\left.\frac{\partial(x, y)}{\partial(r, \theta)}\right|_{(r_0, \theta_0)} = r_0.$$

(b) When can we form a smooth inverse function $(r(x, y), \theta(x, y))$? Check directly and with the inverse function theorem.

(c) Consider the following transformations for spherical coordinates (see Section 1.4):

$$\begin{aligned} x(\rho, \phi, \theta) &= \rho\sin\phi\cos\theta \\ x(\rho, \phi, \theta) &= \rho\sin\phi\sin\theta \\ z(\rho, \phi, \theta) &= \rho\cos\phi. \end{aligned}$$

Show that the Jacobian determinant is given by

$$\frac{\partial(x, y, z)}{\partial(\rho, \phi, \theta)} = \rho^2 \sin\phi.$$

(d) When can we solve for (ρ, ϕ, θ) in terms of (x, y, z)?

11. Let (x_0, y_0, z_0) be a point of the locus defined by $z^2 + xy - a = 0$, $z^2 + x^2 - y^2 - b = 0$, where a and b are constants.

(a) Under what conditions may the part of the locus near (x_0, y_0, z_0) be represented in the form $x = f(z)$, $y = g(z)$?
(b) Compute $f'(z)$ and $g'(z)$.

12. Is it possible to solve the system of equations

$$xy^2 + xzu + yv^2 = 3$$
$$u^3yz + 2xv - u^2v^2 = 2$$

for $u(x, y, z)$, $v(x, y, z)$ near $(x, y, z) = (1, 1, 1)$, $(u, v) = (1, 1)$? Compute $\partial v/\partial y$ at $(x, y, z) = (1, 1, 1)$.

13. The problem of factoring a polynomial $x^n + a_{n-1}x^{n-1} + \cdots + a_0$ into linear factors is, in a sense, an "inverse function" problem. The coefficients a_i may be thought of as functions of the n roots r_j. We would like to find the roots as functions of the coefficients in some region. With $n = 3$, apply the inverse function theorem to this problem and state what it tells you about the possibility of doing this.

REVIEW EXERCISES FOR CHAPTER 3

1. Analyze the behavior of the following functions at the indicated points. [Your answer in part (b) may depend on the constant C.]

(a) $z = x^2 - y^2 + 3xy$, $\quad (x, y) = (0, 0)$
(b) $z = x^2 - y^2 + Cxy$, $\quad (x, y) = (0, 0)$

2. Find and classify the extreme values (if any) of the functions on $\mathbb{R}^2$ defined by the following expressions:

(a) $y^2 - x^3$ (b) $(x - 1)^2 + (x - y)^2$ (c) $x^2 + xy^2 + y^4$

3. (a) Find the minimum distance from the origin in $\mathbb{R}^3$ to the surface $z = \sqrt{x^2 - 1}$.
(b) Repeat part (a) for the surface $z = 6xy + 7$.

4. Find the first few terms in the Taylor expansion of $f(x, y) = e^{xy}\cos x$ about $x = 0$, $y = 0$.

5. Prove that

$$z = \frac{3x^4 - 4x^3 - 12x^2 + 18}{12(1 + 4y^2)}$$

has one local maximum, one local minimum, and one saddle point. (The graph is shown in Figure 3.R.1.)

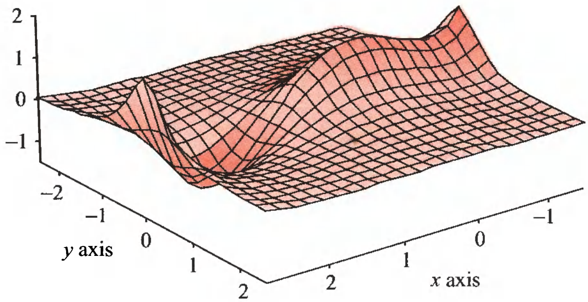

Figure 3.R.1 Graph of $z = (3x^4 - 4x^3 - 12x^2 + 18)/12(1 + 4y^2)$.

6. Find the maxima, minima, and saddles of the function $z = (2 + \cos \pi x)(\sin \pi y)$, which is graphed in Figure 3.R.2.

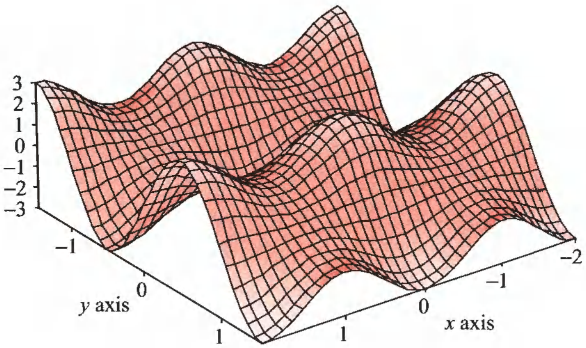

Figure 3.R.2 Graph of $z = (2 + \cos \pi x)(\sin \pi y)$.

7. Find and describe the critical points of $f(x, y) = y \sin(\pi x)$. (See Figure 3.R.3.)

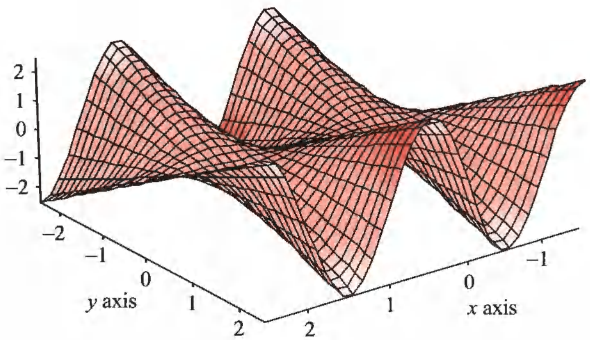

Figure 3.R.3 Graph of $z = y \sin(\pi x)$.

8. A graph of the function $z = \sin(\pi x)/(1 + y^2)$ is shown in Figure 3.R.4. Verify that this function has alternating maxima and minima on the x axis, with no other critical points.

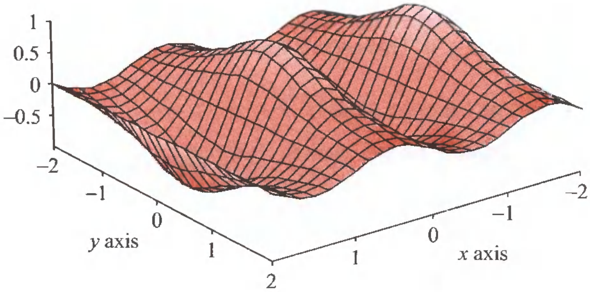

Figure 3.R.4 Graph of $z = \sin(\pi x)/(1 + y^2)$.

In Exercises 9 to 14 find the extrema of the given functions subject to the given constraints.

9. $f(x, y) = x^2 - 2xy + 2y^2$, subject to $x^2 + y^2 = 1$

10. $f(x, y) = xy - y^2$, subject to $x^2 + y^2 = 1$

11. $f(x, y) = \cos(x^2 - y^2)$, subject to $x^2 + y^2 = 1$

12. $f(x, y) = \dfrac{x^2 - y^2}{x^2 + y^2}$, subject to $x + y = 1$

13. $z = xy$, subject to the condition $x + y = 1$.

14. $z = \cos^2 x + \cos^2 y$, subject to the condition $x + y = \pi/4$.

15. Find the points on the surface $z^2 - xy = 1$ nearest to the origin.

16. Use the implicit function theorem to compute dy/dx for

(a) $x/y = 10$ (b) $x^3 - \sin y + y^4 = 4$ (c) $e^{x+y^2} + y^3 = 0$

17. Find the shortest distance from the point $(0, b)$ to the parabola $x^2 - 4y = 0$. Solve this problem using the Lagrange multiplier method and also without using Lagrange's method.

18. Solve the following geometric problems by Lagrange's method.

(a) Find the shortest distance from the point (a_1, a_2, a_3) in $\mathbb{R}^3$ to the plane whose equation is given by $b_1x_1 + b_2x_2 + b_3x_3 + b_0 = 0$, where $(b_1, b_2, b_3) \neq (0, 0, 0)$.

(b) Find the point on the line of intersection of the two planes $a_1x_1 + a_2x_2 + a_3x_3 = 0$ and $b_1x_1 + b_2x_2 + b_3x_3 + b_0 = 0$ that is nearest to the origin.

(c) Show that the volume of the largest rectangular parallelepiped that can be inscribed in the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

is $8abc/3\sqrt{3}$.

19. A particle moves in a potential $V(x, y) = x^3 - y^2 + x^2 + 3xy$. Determine whether $(0, 0)$ is a stable equilibrium point—that is, whether or not $(0, 0)$ is a strict local minimum of V.

20. Study the nature of the function $f(x, y) = x^3 - 3xy^2$ near $(0, 0)$. Show that the point $(0, 0)$ is a degenerate critical point, that is, $D = 0$. This surface is called a *monkey saddle*.

21. Find the maximum of $f(x, y) = xy$ on the curve $(x + 1)^2 + y^2 = 1$.

22. Find the maximum and minimum of $f(x, y) = xy - y + x - 1$ on the set $x^2 + y^2 \leq 2$.

23. The Baraboo, Wisconsin, plant of International Widget Co., Inc., uses aluminium, iron, and magnesium to produce high-quality widgets. The quantity of widgets that may be produced using x tons of aluminum, y tons of iron, and z tons of magnesium is $Q(x, y, z) = xyz$. The cost of raw materials is aluminum, \$6 per ton; iron, \$4 per ton; and magnesium, \$8 per ton. How many tons each of aluminum, iron, and magnesium should be used to manufacture 1000 widgets at the lowest possible cost? (HINT: Find an extreme value for what function subject to what constraint?)

24. Let $f: \mathbb{R} \to \mathbb{R}$ be of class C^1 and let

$$\begin{aligned} u &= f(x) \\ v &= -y + xf(x). \end{aligned}$$

If $f'(x_0) \neq 0$, show that this transformation of $\mathbb{R}^2$ to $\mathbb{R}^2$ is invertible near (x_0, y_0) and its inverse is given by

$$\begin{aligned} x &= f^{-1}(u) \\ y &= -v + uf^{-1}(u). \end{aligned}$$

25. Show that the pair of equations

$$\begin{aligned} x^2 - y^2 - u^3 + v^2 + 4 &= 0 \\ 2xy + y^2 - 2u^2 + 3v^4 + 8 &= 0 \end{aligned}$$

determine functions $u(x, y)$ and $v(x, y)$ defined for (x, y) near $x = 2$ and $y = -1$ such that $u(2, -1) = 2$ and $v(2, -1) = 1$. Compute $\partial u/\partial x$ at $(2, -1)$.

26. Show that there are positive numbers p and q and unique functions u and v from the interval $(-1 - p, -1 + p)$ into the interval $(1 - q, 1 + q)$ satisfying

$$xe^{u(x)} + u(x)e^{v(x)} = 0 = xe^{v(x)} + v(x)e^{u(x)}$$

for all x in the interval $(-1 - p, -1 + p)$ with $u(-1) = 1 = v(-1)$.

27. To work this exercise, the reader should be familiar with the technique of diagonalizing a 2×2 matrix. Let $a(x)$, $b(x)$, and $c(x)$ be three continuous functions defined on $U \cup \partial U$, where U is an open set and ∂U denotes its set of boundary points (see Section 2.2). Use the notation of Lemma 2 in Section 3.3, and assume that for each $x \in U \cup \partial U$ the quadratic form defined by the matrix

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

is positive-definite. For a C^2 function v on $U \cup \partial U$, we define a differential operator L by $Lv = a(\partial^2 v/\partial x^2) + 2b(\partial^2 v/\partial x \partial y) + c(\partial^2 v/\partial y^2)$. With this positive-definite condition, such an operator is said to be ***elliptic***. A function v is said to be ***strictly subharmonic relative*** to L if $Lv > 0$. Show that a strictly subharmonic function cannot have a maximum point in U.

28. A function v is said to be in the ***kernel*** of the operator L described in Exercise 27 if $Lv = 0$ on $U \cup \partial U$. Arguing as in Exercise 37 of Section 3.3, show that if v achieves its maximum on U it also achieves it on ∂U. This is called the weak maximum principle for elliptic operators.

29. Let L be an elliptic differential operator as in Exercises 27 and 28.

(a) Define the notion of a strict superharmonic function.
(b) Show that such functions cannot achieve a minimum on U.
(c) If v is as in Exercise 28, show that if v achieves its minimum on U it also achieves it on ∂U.

The following ***method of least squares*** *should be applied to Exercises 30 to 35.*

It sometimes happens that the theory behind an experiment indicates that the experimental data should lie approximately along a straight line of the form $y = mx + b$. The actual results, of course, never match the theory exactly. We are then faced with the problem of finding the straight line that *best fits* some set of experimental data $(x_1, y_1), \ldots, (x_n, y_n)$ as in Figure 3.R.5. If we guess at a straight line $y = mx + b$ to fit the data, each point will deviate vertically from the line by an amount $d_i = y_i - (mx_i + b)$.

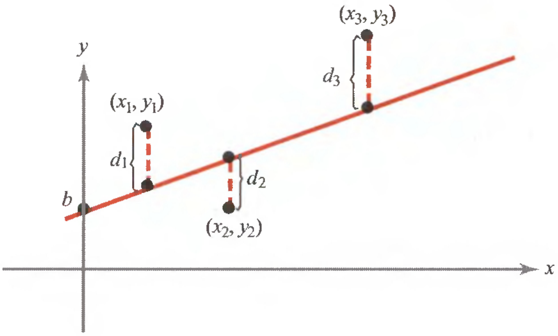

Figure 3.R.5 The method of least squares tries to find a straight line that best approximates a set of data.

We would like to choose m and b in such a way as to make the total effect of these deviations as small as possible. However, because some are negative and some positive, we could get a lot of cancellations and still have a pretty bad fit. This leads us to suspect that a better measure of the total error might be the sum of the *squares* of these deviations. Thus, we are led to the problem of finding the m and b that minimize the function

$$s = f(m, b) = d_1^2 + d_2^2 + \cdots + d_n^2 = \sum_{i=1}^{n} (y_i - mx_i - b)^2,$$

where $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are the given data.

30. For each set of three data points, plot the points, write down the function $f(m, b)$ from the preceding equation, find m and b to give the best straight-line fit according to the method of least squares, and plot the straight line.

(a) $(x_1, y_1) = (1, 1)$
$(x_2, y_2) = (2, 3)$
$(x_3, y_3) = (4, 3)$

(b) $(x_1, y_1) = (0, 0)$
$(x_2, y_2) = (1, 2)$
$(x_3, y_3) = (2, 3)$

31. Show that if only two data points (x_1, y_1) and (x_2, y_2) are given, this method produces the line through (x_1, y_1) and (x_2, y_2).

32. Show that the equations for a critical point, $\partial s/\partial b = 0$ and $\partial s/\partial m = 0$, are equivalent to

$$m\left(\sum x_i\right) + nb = \left(\sum y_i\right) \quad \text{and} \quad m\left(\sum x_i^2\right) + b\left(\sum x_i\right) = \left(\sum x_i y_i\right),$$

where all the sums run from $i = 1$ to $i = n$.

33. If $y = mx + b$ is the best-fitting straight line to the data points $(x_1, y_1), \ldots, (x_n, y_n)$ according to the least-square method, show that

$$\sum_{i=1}^{n} (y_i - mx_i - b) = 0;$$

that is, the positive and negative deviations cancel (see Exercise 32).

34. Use the second derivative test to show that the critical point of f is a minimum.

35. Use the method of least squares to find the straight line that best fits the points $(0, 1)$, $(1, 3)$, $(2, 2)$, $(3, 4)$, and $(4, 5)$. Plot the points and line.[17]

---

[17]The method of least squares may be varied and generalized in a number of ways. The basic idea can be applied to equations of more complicated curves than the straight line. For example, this might be done to find the parabola that best fits a given set of data points. These ideas also formed part of the basis for the development of the science of cybernetics by Norbert Wiener. Another version of the data is the following problem of least-square approximation: Given a function f defined and integrable on an interval $[a, b]$, find a polynomial P of degree $\leq n$ such that the mean square error

$$\int_a^b |f(x) - P(x)|^2 \, dx$$

is as small as possible.