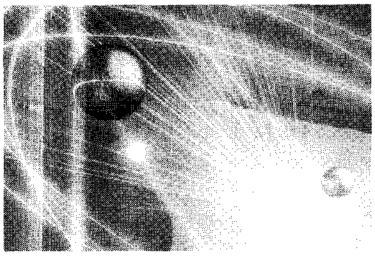

CHAPTER 13

Applications of Partial Derivatives

Introduction In this chapter we will discuss some of the ways partial derivatives contribute to the understanding and solution of problems in applied mathematics. Many such problems can be put in the context of determining maximum or minimum values for functions of several variables, and the first four sections of this chapter deal with that subject. The remaining sections discuss some miscellaneous problems involving the differentiation of functions with respect to parameters, and also Newton's Method for approximating solutions of systems of nonlinear equations. Much of the material in this chapter may be considered *optional*. Only Sections 13.1–13.3 contain *core material*, and even parts of those sections can be omitted (e.g., the discussion of linear programming in Section 13.2).

13.1 Extreme Values

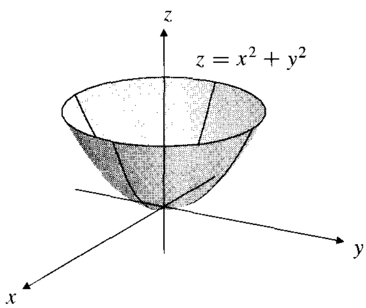

Figure 13.1 $x^2 + y^2$ has minimum value 0 at the origin

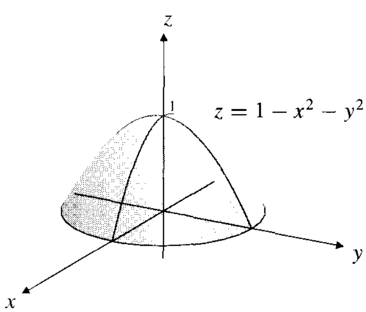

Figure 13.2 $1 - x^2 - y^2$ has maximum value 1 at the origin

The function $f(x, y) = x^2 + y^2$, part of whose graph is shown in Figure 13.1, has a minimum value of 0; this value occurs at the origin $(0, 0)$ where the graph has a horizontal tangent plane. Similarly, the function $g(x, y) = 1 - x^2 - y^2$, part of whose graph appears in Figure 13.2, has a maximum value of 1 at $(0, 0)$. What techniques could be used to discover these facts if they were not evident from a diagram? Finding maximum and minimum values of functions of several variables is, like its single-variable counterpart, the crux of many applications of advanced calculus to problems that arise in other disciplines. Unfortunately, this problem is often much more complicated than in the single-variable case. Our discussion will begin by developing the techniques for functions of two variables. Some of the techniques extend to functions of more variables in obvious ways. The extension of those that do not will be discussed later in this section.

Let us begin by reviewing what we know about the single-variable case. Recall that a function $f(x)$ has a *local maximum value* (or a *local minimum value*) at a point a in its domain if $f(x) \le f(a)$ (or $f(x) \ge f(a)$) for all x in the domain of f that are *sufficiently close* to a. If the appropriate inequality holds *for all* x in the domain of f, then we say that f has an *absolute maximum* (or *absolute minimum*) value at a. Moreover, such local or absolute extreme values can occur only at points of one of the following three types:

(a) critical points, where $f'(x) = 0$,

(b) singular points, where $f'(x)$ does not exist, or

(c) endpoints of the domain of f.

A similar situation exists for functions of several variables. For example, we say that a function of two variables has a **local maximum** or **relative maximum** value at the point (a, b) in its domain if $f(x, y) \le f(a, b)$ for all points (x, y) in the domain of f that are *sufficiently close* to the point (a, b). If the inequality holds *for all* (x, y) in the domain of f, then we say that f has a **global maximum** or **absolute maximum** value at (a, b). Similar definitions obtain for local (relative) and absolute (global) minimum values. In practice, the word *absolute* or *global* is usually omitted, and we refer simply to *the maximum* or *the minimum* value of f.

The following theorem shows that there are three possibilities for points where extreme values can occur, analogous to those for the single-variable case.

THEOREM 1

Necessary conditions for extreme values

A function $f(x, y)$ can have a local or absolute extreme value at a point (a, b) in its domain only if (a, b) is one of the following:

(a) a **critical point** of f, that is, a point satisfying $\nabla f(a, b) = \mathbf{0}$,

(b) a **singular point** of f, that is, a point where $\nabla f(a, b)$ does not exist, or

(c) a **boundary point** of the domain of f.

PROOF Suppose that (a, b) belongs to the domain of f. If (a, b) is not on the boundary of the domain of f, then it must belong to the interior of that domain, and if (a, b) is not a singular point of f then $\nabla f(a, b)$ exists. Finally, if (a, b) is not a critical point of f, then $\nabla f(a, b) \neq \mathbf{0}$, so f has a positive directional derivative in the direction of $\nabla f(a, b)$ and a negative directional derivative in the direction of $-\nabla f(a, b)$; that is, f is increasing as we move from (a, b) in one direction and decreasing as we move in the opposite direction. Hence, f cannot have either a maximum or a minimum value at (a, b). Therefore, any point where an extreme value occurs must be either a critical point or a singular point of f, or a boundary point of the domain of f.

Note that Theorem 1 remains valid with unchanged proof for functions of any number of variables. Of course, Theorem 1 does not guarantee that a given function will have any extreme values. It only tells us where to look to find any that may exist. Theorem 2, below, provides conditions that guarantee the existence of absolute maximum and minimum values for a continuous function. It is analogous to the Max-Min Theorem for functions of one variable. The proof is beyond the scope of this book; an interested student should consult an elementary text on mathematical analysis. A set in $\mathbb{R}^n$ is **bounded** if it is contained inside some *ball* $x_1^2 + x_2^2 + \cdots + x_n^2 \leq R^2$ of finite radius R. A set on the real line is bounded if it is contained in an interval of finite length.

THEOREM 2

Sufficient conditions for extreme values

If f is a *continuous* function of n variables whose domain is a *closed* and *bounded* set in $\mathbb{R}^n$, then the range of f is a bounded set of real numbers, and there are points in its domain where f takes on absolute maximum and minimum values.

Example 1 The function $f(x, y) = x^2 + y^2$ (see Figure 13.1) has a critical point at $(0, 0)$ since $\nabla f = 2x\mathbf{i} + 2y\mathbf{j}$ and both components of ∇f vanish at $(0, 0)$. Since

$$f(x, y) > 0 = f(0, 0) \quad \text{if} \quad (x, y) \neq (0, 0),$$

f must have (absolute) minimum value 0 at that point. If the domain of f is not restricted, f has no maximum value. Similarly, $g(x, y) = 1 - x^2 - y^2$ has (absolute) maximum value 1 at its critical point $(0, 0)$. (See Figure 13.2.)

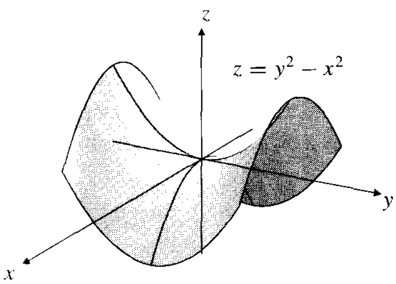

Figure 13.3 $y^2 - x^2$ has a saddle point at $(0, 0)$

Example 2 The function $h(x, y) = y^2 - x^2$ also has a critical point at $(0, 0)$ but has neither a local maximum nor a local minimum value at that point. Observe that $h(0, 0) = 0$ but $h(x, 0) < 0$ and $h(0, y) > 0$ for all nonzero values of x and y. (See Figure 13.3.) The graph of h is a hyperbolic paraboloid. In view of its shape we call the critical point $(0, 0)$ a **saddle point** of h.

In general, we will somewhat loosely call any *interior critical* point of the domain of a function f of several variables a **saddle point** if f does not have a local maximum or minimum value there. Even for functions of two variables, the graph will not always look like a saddle near a saddle point. For instance, the function $f(x, y) = -x^3$ has a whole line of *saddle points* along the y-axis (see Figure 13.4), although its graph does not resemble a saddle anywhere. These points resemble inflection points of a function of one variable. Saddle points are higher-dimensional analogues of such horizontal inflection points.

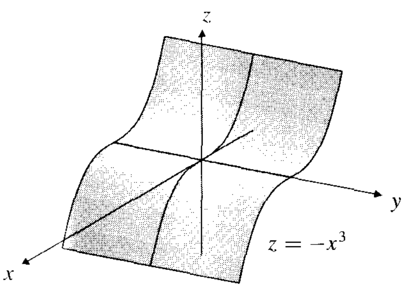

Figure 13.4 A line of saddle points

Example 3 The function $f(x, y) = \sqrt{x^2 + y^2}$ has no critical points but does have a singular point at $(0, 0)$ where it has a local (and absolute) minimum value, zero. The graph of f is a circular cone. (See Figure 13.5(a).)

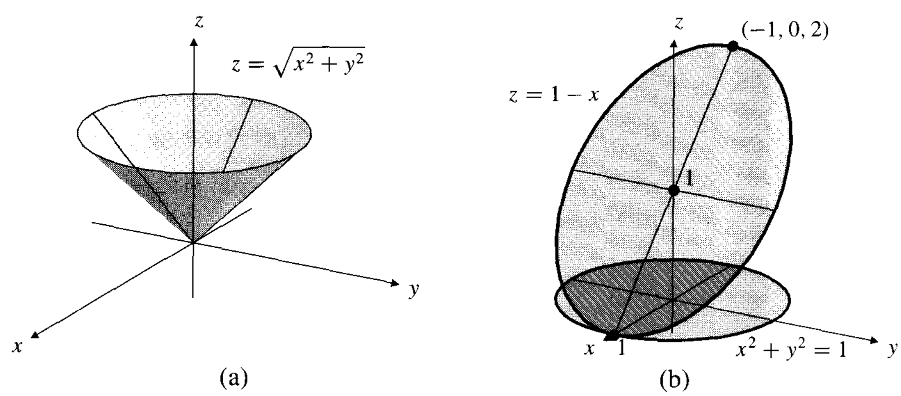

Figure 13.5
(a) $\sqrt{x^2 + y^2}$ has a minimum value at the singular point $(0, 0)$
(b) When restricted to the disk $x^2 + y^2 \leq 1$, the function $1 - x$ has maximum and minimum values at boundary points

Example 4 The function $f(x, y) = 1 - x$ is defined everywhere in the xy-plane and has no critical or singular points. ($\nabla f(x, y) = -\mathbf{i}$ at every point (x, y).) Therefore f has no extreme values. However, if we restrict the domain of f to the points in the disk $x^2 + y^2 \leq 1$ (a closed bounded set in the xy-plane), then f does have absolute maximum and minimum values, as it must by Theorem 2. The maximum value is 2 at the boundary point $(-1, 0)$ and the minimum value is 0 at $(1, 0)$. (See Figure 13.5(b).)

Classifying Critical Points

The above examples were very simple ones; it was immediately obvious in each case whether the function had a local maximum, local minimum, or a saddle point at the critical or singular point. For more complicated functions, it may be harder to classify the interior critical points. In theory such a classification can always be made by considering the difference

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

for small values of h and k, where (a, b) is the critical point in question. If the difference is always nonnegative (or nonpositive) for small h and k, then f must have a local minimum (or maximum) at (a, b); if the difference is negative for some points (h, k) arbitrarily near $(0, 0)$ and positive for others, then f must have a saddle point at (a, b).

Example 5 Find and classify the critical points of $f(x, y) = 2x^3 - 6xy + 3y^2$.

Solution The critical points must satisfy the system of equations:

$$\begin{aligned} 0 &= f_1(x, y) = 6x^2 - 6y \quad &\iff \quad x^2 &= y \\ 0 &= f_2(x, y) = -6x + 6y \quad &\iff \quad x &= y. \end{aligned}$$

Together, these equations imply that $x^2 = x$ so that $x = 0$ or $x = 1$. Therefore, the critical points are $(0, 0)$ and $(1, 1)$.

Consider $(0, 0)$. Here Δf is given by

$$\Delta f = f(h, k) - f(0, 0) = 2h^3 - 6hk + 3k^2.$$

Since $f(h, 0) - f(0, 0) = 2h^3$ is positive for small positive h and negative for small negative h, f cannot have a maximum or minimum value at $(0, 0)$. Therefore $(0, 0)$ is a saddle point.

Now consider $(1, 1)$. Here Δf is given by

$$\begin{aligned} \Delta f &= f(1 + h, 1 + k) - f(1, 1) \\ &= 2(1 + h)^3 - 6(1 + h)(1 + k) + 3(1 + k)^2 - (-1) \\ &= 2 + 6h + 6h^2 + 2h^3 - 6 - 6h - 6k - 6hk + 3 + 6k + 3k^2 + 1 \\ &= 6h^2 - 6hk + 3k^2 + 2h^3 \\ &= 3(h - k)^2 + h^2(3 + 2h). \end{aligned}$$

Both terms in the latter expression are nonnegative if $|h| < 3/2$, and they are not both zero unless $h = k = 0$. Hence, $\Delta f > 0$ for small h and k, and f has a local minimum value -1 at $(1, 1)$.

The method used to classify critical points in the above example takes on a "brute force" aspect if the function involved is more complicated. However, there is a *second derivative test* similar to that for functions of one variable. The n-variable version is the subject of the following theorem, the proof of which is based on properties of quadratic forms presented in Section 10.6.

THEOREM 3

A second derivative test

Suppose that $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ is a critical point of $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$ and is interior to the domain of f. Also, suppose that all the second partial derivatives of f are continuous throughout a neighbourhood of $\mathbf{a}$, so that the **Hessian matrix**

$$\mathcal{H}(\mathbf{x}) = \begin{pmatrix} f_{11}(\mathbf{x}) & f_{12}(\mathbf{x}) & \cdots & f_{1n}(\mathbf{x}) \\ f_{21}(\mathbf{x}) & f_{22}(\mathbf{x}) & \cdots & f_{2n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1}(\mathbf{x}) & f_{n2}(\mathbf{x}) & \cdots & f_{nn}(\mathbf{x}) \end{pmatrix}$$

is also continuous in that neighbourhood. Note that the continuity of the partials guarantees that $\mathcal{H}$ is a symmetric matrix.

(a) If $\mathcal{H}(\mathbf{a})$ is positive definite, then f has a local minimum at $\mathbf{a}$.

(b) If $\mathcal{H}(\mathbf{a})$ is negative definite, then f has a local maximum at $\mathbf{a}$.

(c) If $\mathcal{H}(\mathbf{a})$ is indefinite, then f has a saddle point at $\mathbf{a}$.

(d) If $\mathcal{H}(\mathbf{a})$ is neither positive nor negative definite nor indefinite, this test gives no information.

PROOF Let $g(t) = f(\mathbf{a} + t\mathbf{h})$ for $0 \le t \le 1$, where $\mathbf{h}$ is an n-vector. Then

$$g'(t) = \sum_{i=1}^{n} f_i(\mathbf{a} + t\mathbf{h})\, h_i$$

$$g''(t) = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}(\mathbf{a} + t\mathbf{h})\, h_i\, h_j = \mathbf{h}^T \mathcal{H}(\mathbf{a} + t\mathbf{h})\mathbf{h}.$$

(In the latter expression, $\mathbf{h}$ is being treated as a column vector.) We apply Taylor's Formula with Lagrange remainder to g to write

$$g(1) = g(0) + g'(0) + \frac{1}{2} g''(\theta)$$

for some θ between 0 and 1. Thus,

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \sum_{i=1}^{n} f_i(\mathbf{a})\, h_i + \frac{1}{2}\mathbf{h}^T \mathcal{H}(\mathbf{a} + \theta\mathbf{h})\mathbf{h}.$$

Since $\mathbf{a}$ is a critical point of f, $f_i(\mathbf{a}) = 0$ for $1 \le i \le n$, so

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = \frac{1}{2}\mathbf{h}^T \mathcal{H}(\mathbf{a} + \theta\mathbf{h})\mathbf{h}.$$

If $\mathcal{H}(\mathbf{a})$ is positive definite, then, by the continuity of $\mathcal{H}$, so is $\mathcal{H}(\mathbf{a} + \theta\mathbf{h})$ for $|\mathbf{h}|$ sufficiently small. Therefore, $f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) > 0$ for nonzero $\mathbf{h}$, proving (a).

Parts (b) and (c) are proved similarly. The functions $f(x, y) = x^4 + y^4$, $g(x, y) = -x^4 - y^4$, and $h(x, y) = x^4 - y^4$ all fall under part (d) and show that in this case a function can have a minimum, a maximum, or a saddle point.

Example 6 Find and classify the critical points of the function $f(x, y, z) = x^2y + y^2z + z^2 - 2x$.

Solution The equations that determine the critical points are

$$\begin{aligned} 0 &= f_1(x, y, z) = 2xy - 2, \\ 0 &= f_2(x, y, z) = x^2 + 2yz, \\ 0 &= f_3(x, y, z) = y^2 + 2z. \end{aligned}$$

The third equation implies $z = -y^2/2$, and the second then implies $y^3 = x^2$. From the first equation we get $y^{5/2} = 1$. Thus $y = 1$ and $z = -\frac{1}{2}$. Since $xy = 1$, we
\0 2 2/

Since

$$2 > 0, \quad \begin{vmatrix} 2 & 2 \\ 2 & -1 \end{vmatrix} = -6 < 0, \quad \begin{vmatrix} 2 & 2 & 0 \\ 2 & -1 & 2 \\ 0 & 2 & 2 \end{vmatrix} = -20 < 0,$$

$\mathcal{H}$ is indefinite by Theorem 8 of Section 10.6, so P is a saddle point of f. ■

Remark Applying the test for positive or negative definiteness or indefiniteness given in Theorem 8 of Section 10.6, we can paraphrase the second derivative test for a function of two variables as follows:

Suppose that (a, b) is a critical point of the function $f(x, y)$ that is interior to the domain of f. Suppose also that the second partial derivatives of f are continuous in a neighbourhood of (a, b) and have at that point the values

$$A = f_{11}(a, b), \quad B = f_{12}(a, b) = f_{21}(a, b), \quad \text{and} \quad C = f_{22}(a, b).$$

(a) If $AC > B^2$ and $A > 0$, then f has a local minimum value at (a, b).
(b) If $AC > B^2$ and $A < 0$, then f has a local maximum value at (a, b).
(c) If $AC < B^2$, then f has a saddle point at (a, b).
(d) If $AC = B^2$, this test provides no information; f may have a local maximum or a local minimum value or a saddle point at (a, b).

Example 7 Reconsider Example 5 and use the second derivative test to classify the two critical points $(0, 0)$ and $(1, 1)$ of $f(x, y) = 2x^3 - 6xy + 3y^2$.

Solution We have

$$f_{11}(x, y) = 12x, \quad f_{12}(x, y) = -6, \quad \text{and} \quad f_{22}(x, y) = 6.$$

At (0, 0) we therefore have

$$A = 0, \quad B = -6, \quad C = 6, \text{ and } \quad B^2 - AC = 36 > 0,$$

so (0, 0) is a saddle point. At (1, 1) we have

$$A = 12 > 0, \quad B = -6, \quad C = 6, \text{ and } \quad B^2 - AC = -36 < 0,$$

so f must have a local minimum at (1, 1). ■

Example 8 Find and classify the critical points of

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

Does f have absolute maximum and minimum values? Why?

Solution We begin by calculating the first- and second-order partial derivatives of f:

$$\begin{aligned} f_1(x, y) &= y(1 - x^2)\, e^{-(x^2+y^2)/2}, \\ f_2(x, y) &= x(1 - y^2)\, e^{-(x^2+y^2)/2}, \\ f_{11}(x, y) &= xy(x^2 - 3)\, e^{-(x^2+y^2)/2}, \\ f_{12}(x, y) &= (1 - x^2)(1 - y^2)\, e^{-(x^2+y^2)/2}, \\ f_{22}(x, y) &= xy(y^2 - 3)\, e^{-(x^2+y^2)/2}. \end{aligned}$$

At any critical point $f_1 = 0$ and $f_2 = 0$, so the critical points are the solutions of the system of equations

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

The first of these equations says that $y = 0$ or $x = \pm 1$. The second equation says that $x = 0$ or $y = \pm 1$. There are five points satisfying both conditions: $(0, 0)$, $(1, 1)$, $(1, -1)$, $(-1, 1)$, and $(-1, -1)$. We classify them using the second derivative test.

At $(0, 0)$ we have $A = C = 0$, $B = 1$, so that $B^2 - AC = 1 > 0$. Thus f has a saddle point at $(0, 0)$.

At $(1, 1)$ and $(-1, -1)$ we have $A = C = -2/e < 0$, $B = 0$. It follows that $B^2 - AC = -4/e^2 < 0$. Thus, f has local maximum values at these points. The value of f is $1/e$ at each point.

At $(1, -1)$ and $(-1, 1)$ we have $A = C = 2/e > 0$, $B = 0$. If follows that $B^2 - AC = -4/e^2 < 0$. Thus, f has local minimum values at these points. The value of f at each of them is $-1/e$.

Indeed, f has absolute maximum and minimum values, namely, the values obtained above as local extrema. To see why, observe that $f(x, y)$ approaches 0 as the point (x, y) recedes to infinity in any direction because the negative exponential dominates the power factor xy for large $x^2 + y^2$. Pick a number between 0 and the local maximum value $1/e$ found above, say, the number $1/(2e)$. For some R, we must have $|f(x, y)| \leq 1/(2e)$ whenever $x^2 + y^2 \geq R^2$. On the closed disk $x^2 + y^2 \leq R^2$, f must have absolute maximum and minimum values by Theorem 2. These cannot occur on the boundary circle $x^2 + y^2 = R^2$ because $|f|$ is smaller there ($\leq 1/(2e)$) than it is at the critical points considered above. Since f has no singular points, the absolute maximum and minimum values for the disk, and therefore for the whole plane, must occur at those critical points. ■

Example 9 Find the shape of a rectangular box with no top having given volume V and the least possible total surface area of its five faces.

Solution If the horizontal dimensions of the box are x, y, and its height is z (see Figure 13.6), then we want to minimize

$$S = xy + 2yz + 2xz$$

subject to the restriction that $xyz = V$, the required volume. We can use this restriction to reduce the number of variables on which S depends, for instance, by substituting

$$z = \frac{V}{xy}.$$

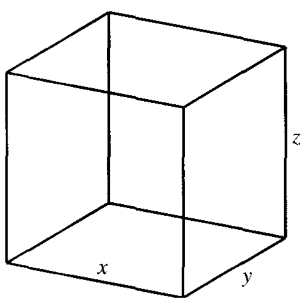

Figure 13.6

Then S becomes a function of the two variables x and y:

$$S = S(x, y) = xy + \frac{2V}{x} + \frac{2V}{y}.$$

A real box has positive dimensions, so the domain of S should consist of only those points (x, y) that satisfy $x > 0$ and $y > 0$. If either x or y approaches 0 or ∞, then $S \to \infty$, so the minimum value of S must occur at a critical point. (S has no singular points.) For critical points we solve the equations

$$0 = \frac{\partial S}{\partial x} = y - \frac{2V}{x^2} \iff x^2 y = 2V,$$
$$0 = \frac{\partial S}{\partial y} = x - \frac{2V}{y^2} \iff xy^2 = 2V.$$

Thus $x^2y - xy^2 = 0$, or $xy(x - y) = 0$. Since $x > 0$ and $y > 0$, this implies that $x = y$. Therefore, $x^3 = 2V$, $x = y = (2V)^{1/3}$, and $z = V/(xy) = 2^{-2/3}V^{1/3} = x/2$. Since there is only one critical point it must minimize S. (Why?) The box having minimal surface area has a square base but is only half as high as its horizontal dimensions. ■

Remark The preceding problem is a *constrained* extreme value problem in three variables; the equation $xyz = V$ is a *constraint* limiting the freedom of x, y, and z. We used the constraint to eliminate one variable, z, and so to reduce the problem to a *free* (i.e., *unconstrained*) problem in two variables. In Section 13.3 we will develop a more powerful method for solving constrained extreme value problems.

Exercises 13.1

In Exercises 1–15, find and classify the critical points of the given functions.

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

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

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

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

5. $f(x, y) = \frac{x}{y} + \frac{8}{x} - y$

6. $f(x, y) = \cos(x + y)$

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

8. $f(x, y) = \cos x + \cos y$

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

10. $f(x, y) = \frac{xy}{2 + x^4 + y^4}$

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

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

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

* 14. $f(x, y, z) = xyz - x^2 - y^2 - z^2$

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

* 16. Show that $f(x, y, z) = 4xyz - x^4 - y^4 - z^4$ has a local maximum value at the point $(1, 1, 1)$.

17. Find the maximum and minimum values of $f(x, y) = xy e^{-x^2-y^4}$.

18. Find the maximum and minimum values of $f(x, y) = x/(1 + x^2 + y^2)$.

* 19. Find the maximum and minimum values of $f(x, y, z) = xyz e^{-x^2-y^2-z^2}$. How do you know that such extreme values exist?

20. Find the minimum value of $f(x, y) = x + 8y + \frac{1}{xy}$ in the first quadrant $x > 0$, $y > 0$. How do you know that a minimum exists?

21. Postal regulations require that the sum of the height and girth (horizontal perimeter) of a package should not exceed L units. Find the largest volume of a rectangular box that can satisfy this requirement.

22. The material used to make the bottom of a rectangular box is twice as expensive per unit area as the material used to make the top or side walls. Find the dimensions of the box of given volume V for which the cost of materials is minimum.

23. Find the volume of the largest rectangular box (with faces parallel to the coordinate planes) that can be inscribed inside the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

24. Find the three positive numbers a, b, and c, whose sum is 30 and for which the expression ab^2c^3 is maximum.

25. Find the critical points of the function $z = g(x, y)$ that satisfies the equation $e^{2zx-x^2} - 3e^{2zy+y^2} = 2$.

* 26. Classify the critical points of the function g in the previous exercise.

* 27. Let $f(x, y) = (y - x^2)(y - 3x^2)$. Show that the origin is a critical point of f and that the restriction of f to every straight line through the origin has a local minimum value at the origin. (That is, show that $f(x, kx)$ has a local minimum value at $x = 0$ for every k and that $f(0, y)$ has a local minimum value at $y = 0$.) Does $f(x, y)$ have a local minimum value at the origin? What happens to f on the curve $y = 2x^2$? What does the second derivative test say about this situation?

28. Verify by completing the square (that is, without appealing to Theorem 8 of Section 10.6) that the quadratic form

$$Q(u, v) = (x, y)\begin{pmatrix} A & B \\ B & C \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = Au^2 + 2Buv + Cv^2$$

is positive definite if $A > 0$ and $\begin{vmatrix} A & B \\ B & C \end{vmatrix} > 0$, negative definite if $A < 0$ and $\begin{vmatrix} A & B \\ B & C \end{vmatrix} > 0$, and indefinite if $\begin{vmatrix} A & B \\ B & C \end{vmatrix} < 0$. This gives independent confirmation of the assertion in the remark preceding Example 7.

* 29. State and prove (using square completion arguments rather than appealing to Theorem 8 of Section 10.6) a result analogous to that of the previous exercise for a Quadratic form $Q(u, v, w)$ involving three variables. What are the implications of this for a critical point (a, b, c) of a function $f(x, y, z)$ all of whose second partial derivatives are known at (a, b, c)?

13.2 Extreme Values of Functions Defined on Restricted Domains

Much of the previous section was concerned with techniques for determining whether a critical point of a function provides a local maximum or minimum value or is a saddle point. In this section we address the problem of determining absolute maximum and minimum values for functions that have them—usually functions whose domains are restricted to subsets of $\mathbb{R}^2$ (or $\mathbb{R}^n$) having nonempty interiors. In Example 8 of Section 13.1 we had to *prove* that the given function had absolute extreme values. If, however, we are dealing with a continuous function on a domain that is closed and bounded, then we can rely on Theorem 2 to guarantee the existence of such extreme values, but we will always have to check boundary points as well as any interior critical or singular points to find them. The following examples illustrate the technique.

Example 1 Find the maximum and minimum values of $f(x, y) = 2xy$ on the closed disk $x^2 + y^2 \le 4$. (See Figure 13.7.)

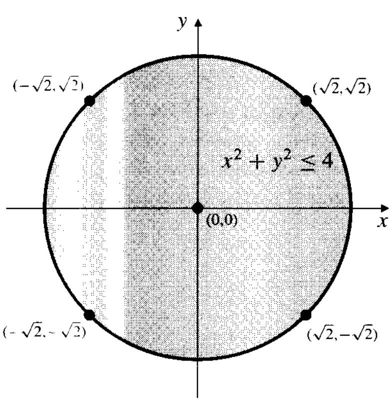

Figure 13.7 Points that are candidates for extreme values in Example 1

Solution Since f is continuous and the disk is closed, f must have absolute maximum and minimum values at some points of the disk. The first partial derivatives of f are

$$f_1(x, y) = 2y \qquad \text{and} \qquad f_2(x, y) = 2x,$$

so there are no singular points, and the only critical point is $(0, 0)$, where f has the value 0.

We must still consider values of f on the boundary circle $x^2 + y^2 = 4$. We can express f as a function of a single variable on this circle by using a convenient parametrization of the circle, say,

$$x = 2\cos t, \qquad y = 2\sin t, \qquad (-\pi \le t \le \pi).$$

We have

$$f\Big(2\cos t, 2\sin t\Big) = 8\cos t \sin t = g(t).$$

We must find any extreme values of $g(t)$. We can do this in either of two ways. If we rewrite $g(t) = 4\sin 2t$, it is clear that $g(t)$ has maximum value 4 (at $t = \frac{\pi}{4}$ and $-\frac{3\pi}{4}$) and minimum value -4 (at $t = -\frac{\pi}{4}$ and $\frac{3\pi}{4}$). Alternatively, we can differentiate g to find its critical points:

$$\begin{aligned} 0 = g'(t) = -8\sin^2 t + 8\cos^2 t \quad &\Longleftrightarrow \quad \tan^2 t = 1 \\ &\Longleftrightarrow \quad t = \pm\frac{\pi}{4} \text{ or } \pm\frac{3\pi}{4}, \end{aligned}$$

which again yield the maximum value 4 and the minimum value -4. (It is not necessary to check the endpoints $t = -\pi$ and $t = \pi$; since g is everywhere differentiable and is periodic with period π, any absolute maximum or minimum will occur at a critical point.)

In any event, f has maximum value 4 at the boundary points $(\sqrt{2}, \sqrt{2})$ and $(-\sqrt{2}, -\sqrt{2})$ and minimum value -4 at the boundary points $(\sqrt{2}, -\sqrt{2})$ and $(-\sqrt{2}, \sqrt{2})$. It is easily shown by the Second Derivative Test (or otherwise) that the interior critical point $(0, 0)$ is a saddle point. (See Figure 13.7.) ■

Example 2 Find the extreme values of the function $f(x, y) = x^2 y e^{-(x+y)}$ on the triangular region T given by $x \ge 0$, $y \ge 0$, and $x + y \le 4$.

Solution First, we look for critical points:

$$0 = f_1(x, y) = xy(2 - x)e^{-(x+y)} \iff x = 0, \ y = 0, \ \text{or } x = 2,$$
$$0 = f_2(x, y) = x^2(1 - y)e^{-(x+y)} \iff x = 0 \text{ or } y = 1.$$

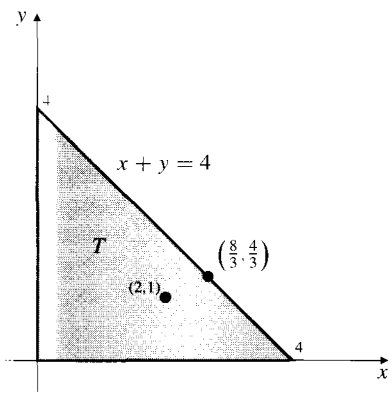

Figure 13.8 Points of interest in Example 2

The critical points are $(0, y)$ for any y and $(2, 1)$. Only $(2, 1)$ is an interior point of T. (See Figure 13.8.) $f(2, 1) = 4/e^3 \approx 0.199$. The boundary of T consists of three straight line segments. On two of these, the coordinate axes, f is identically zero. The third segment is given by

$$y = 4 - x, \qquad 0 \le x \le 4,$$

so the values of f on this segment can be expressed as a function of x alone:

$$g(x) = f(x, 4 - x) = x^2(4 - x)e^{-4}, \qquad 0 \le x \le 4.$$

Note that $g(0) = g(4) = 0$ and $g(x) > 0$ if $0 < x < 4$. The critical points of g are given by $0 = g'(x) = (8x - 3x^2)e^{-4}$, so they are $x = 0$ and $x = 8/3$. We have

$$g\left(\frac{8}{3}\right) = f\left(\frac{8}{3}, \frac{4}{3}\right) = \frac{256}{27}e^{-4} \approx 0.174 < f(2, 1).$$

We conclude that the maximum value of f over the region T is $4/e^3$ and that it occurs at the interior critical point $(2, 1)$. The minimum value of f is zero and occurs at all points of the two perpendicular boundary segments. Note that f has neither a local maximum nor a local minimum at the boundary point $(8/3, 4/3)$, although g has a local maximum there. Of course, that point is not a saddle point of f either. It is not a critical point of f. ■

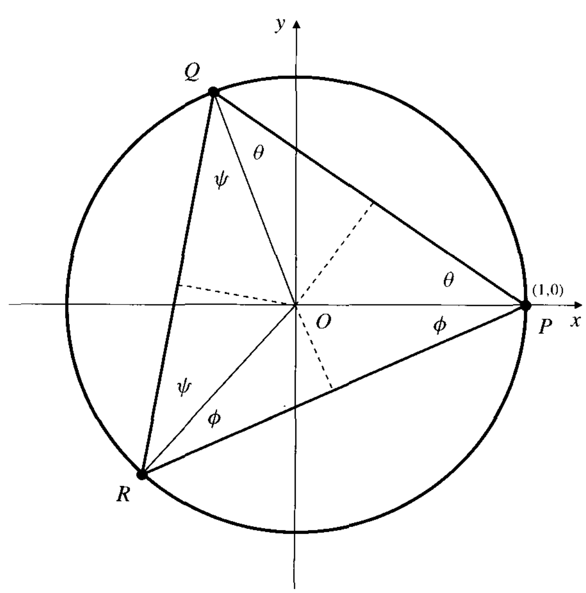

Figure 13.9 Where should Q and R be to ensure that triangle PQR has maximum area?

Example 3 Among all triangles with vertices on a given circle, find those that have the largest area.

Solution Intuition tells us that the equilateral triangles must have the largest area. However, proving this can be quite difficult unless a good choice of variables in which to set up the problem analytically is made. With a suitable choice of units and axes we can assume the circle is $x^2 + y^2 = 1$ and that one vertex of the triangle is the point P with coordinates $(1, 0)$. Let the other two vertices, Q and R, be as shown in Figure 13.9. There is no harm in assuming that Q lies on the upper semicircle and R on the lower, and that the origin O is inside triangle PQR. Let PQ and PR make angles θ and ϕ, respectively, with the negative direction of the x-axis. Clearly $0 \le \theta \le \pi/2$ and $0 \le \phi \le \pi/2$. The lines from O to Q and R make equal angles ψ with the line QR, where $2\theta + 2\phi + 2\psi = \pi$. Dropping perpendiculars from O to the three sides of the triangle PQR, we can write the area A of the triangle as the sum of the areas of six small, right-angled triangles:

$$\begin{aligned} A &= 2 \times \frac{1}{2} \sin\theta\cos\theta + 2 \times \frac{1}{2} \sin\phi\cos\phi + 2 \times \frac{1}{2} \sin\psi\cos\psi \\ &= \frac{1}{2}\big(\sin 2\theta + \sin 2\phi + \sin 2\psi\big). \end{aligned}$$

Since $2\psi = \pi - 2(\theta + \phi)$, we can express A as a function of the two variables θ and ϕ:

$$A = A(\theta, \phi) = \frac{1}{2}\big(\sin 2\theta + \sin 2\phi + \sin 2(\theta + \phi)\big).$$

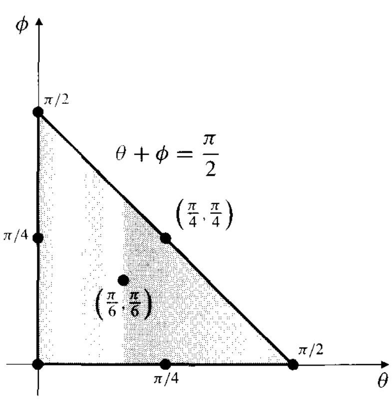

Figure 13.10 The domain of $A(\theta, \phi)$

The domain of A is the triangle $\theta \ge 0$, $\phi \ge 0$, $\theta + \phi \le \pi/2$. $A = 0$ at the vertices of the triangle and is positive elsewhere. (See Figure 13.10.) We show that the maximum value of $A(\theta, \phi)$ on any edge of the triangle is 1 and occurs at the midpoint of that edge. On the edge $\theta = 0$ we have

$$A(0, \phi) = \frac{1}{2}\big(\sin 2\phi + \sin 2\phi\big) = \sin 2\phi \le 1 = A(0, \pi/4).$$

Similarly, on $\phi = 0$, $A(\theta, 0) \le 1 = A(\pi/4, 0)$. On the edge $\theta + \phi = \pi/2$ we have

$$\begin{aligned} A\Big(\theta, \frac{\pi}{2} - \theta\Big) &= \frac{1}{2}\big(\sin 2\theta + \sin(\pi - 2\theta)\big) \\ &= \sin 2\theta \le 1 = A\Big(\frac{\pi}{4}, \frac{\pi}{4}\Big). \end{aligned}$$

We must now check for any interior critical points of $A(\theta, \phi)$. (There are no singular points.) For critical points we have

$$\begin{aligned} 0 &= \frac{\partial A}{\partial \theta} = \cos 2\theta + \cos(2\theta + 2\phi), \\ 0 &= \frac{\partial A}{\partial \phi} = \cos 2\phi + \cos(2\theta + 2\phi), \end{aligned}$$

so the critical points satisfy $\cos 2\theta = \cos 2\phi$ and, hence, $\theta = \phi$. We now substitute this equation into either of the above equations to determine θ:

$$\cos 2\theta + \cos 4\theta = 0$$
$$2\cos^2 2\theta + \cos 2\theta - 1 = 0$$
$$(2\cos 2\theta - 1)(\cos 2\theta + 1) = 0$$
$$\cos 2\theta = \frac{1}{2} \quad \text{or} \quad \cos 2\theta = -1.$$

The only solution leading to an interior point of the domain of A is $\theta = \phi = \pi/6$. Note that

$$A\left(\frac{\pi}{6}, \frac{\pi}{6}\right) = \frac{1}{2}\left(\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2}\right) = \frac{3\sqrt{3}}{4} > 1;$$

this interior critical point maximizes the area of the inscribed triangle. Finally, observe that for $\theta = \phi = \pi/6$, we also have $\psi = \pi/6$, so the largest triangle is indeed equilateral. ■

Remark Since the area A of the inscribed triangle must have a maximum value (A is continuous and its domain is closed and bounded), a strictly geometric argument can be used to show that the largest triangle is equilateral. If an inscribed triangle has two unequal sides, its area can be made larger by moving the common vertex of these two sides along the circle to increase its perpendicular distance from the opposite side of the triangle.

Linear Programming

Linear programming is a branch of linear algebra that develops systematic techniques for finding maximum or minimum values of a *linear function* subject to several *linear inequality constraints.* Such problems arise frequently in management science and operations research. Because of their linear nature they do not usually involve calculus in their solution; linear programming is frequently presented in courses on *finite mathematics.* We will not attempt any formal study of linear programming here, but we will make a few observations for comparison with the more general nonlinear extreme value problems considered above that involve calculus in their solution.

The inequality $ax + by \le c$ is an example of a linear inequality in two variables. The *solution set* of this inequality consists of a half-plane lying on one side of the straight line $ax + by = c$. The solution set of a system of several two-variable linear inequalities is an intersection of such half-planes, so it is a *convex* region of the plane bounded by a *polygonal line.* If it is a bounded set, then it is a convex polygon together with its interior. (A set is called **convex** if it contains the entire line segment between any two of its points. On the real line the convex sets are intervals.)

Let us examine a simple concrete example that involves only two variables and a few constraints.

Example 4 Find the maximum value of $F(x, y) = 2x + 7y$ subject to the constraints $x + 2y \le 6$, $2x + y \le 6$, $x \ge 0$, and $y \ge 0$.

Solution The solution set $\mathcal{S}$ of the system of four constraint equations is shown in Figure 13.11. It is the quadrilateral region with vertices $(0, 0)$, $(3, 0)$, $(2, 2)$, and $(0, 3)$. Several level curves of the linear function F are also shown in the figure. They are parallel straight lines with slope $-\frac{2}{7}$. We want the line that gives F the greatest value and that still intersects $\mathcal{S}$. Evidently this is the line $F = 21$ that passes through the vertex $(0, 3)$ of $\mathcal{S}$. The maximum value of F subject to the constraints is 21. ■

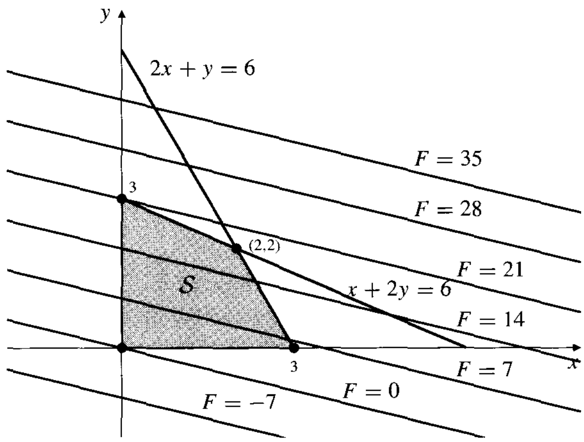

Figure 13.11 The shaded region is the solution set for the constraint inequalities in Example 4

As this simple example illustrates, a *linear* function with domain restricted by *linear inequalities* does not achieve maximum or minimum values at points in the interior of its domain (if that domain has an interior). Any such extreme value occurs at a boundary point of the domain or a set of such boundary points. Where an extreme value occurs at a set of boundary points, that set will *always* contain at least one vertex. This phenomenon holds in general for extreme value problems for linear functions in any number of variables with domains restricted by any number of linear inequalities. For problems involving three variables the domain will be a convex region of $\mathbb{R}^3$ bounded by planes. For a problem involving n variables the domain will be a convex region in $\mathbb{R}^n$ bounded by $(n - 1)$-dimensional hyperplanes. Such *polyhedral* regions still have vertices (where n hyperplanes intersect), and maximum or minimum values of linear functions subject to the constraints will still occur at subsets of the boundary containing such vertices. These problems can therefore be solved by evaluating the linear function to be extremized (it is called the **objective function**) at all the vertices and selecting the greatest or least value.

In practice, linear programming problems can involve hundreds or even thousands of variables and even more constraints. Such problems need to be solved with computers, but even then it is extremely inefficient, if not impossible, to calculate all the vertices of the constraint solution set and the values of the objective function at them. Much of the study of linear programming therefore centres on devising techniques for getting to (or at least near) the optimizing vertex in as few steps as possible. Usually this involves criteria whereby large numbers of vertices can be rejected on geometric grounds. We will not delve into such techniques here but will content ourselves with one more example to illustrate, in a very simple case, how the underlying geometry of a problem can be used to reduce the number of vertices that must be considered.

Example 5 A tailor has 230 m of a certain fabric and has orders for up to 20 suits, up to 30 jackets, and up to 40 pairs of slacks to be made from the fabric. Each suit requires 6 m, each jacket 3 m, and each pair of slacks 2 m of the fabric. If the tailor's profit is \$20 per suit, \$14 per jacket, and \$12 per pair of slacks, how many of each should he make to realize the maximum profit from his supply of the fabric?

Solution Suppose he makes x suits, y jackets, and z pairs of slacks. Then his profit will be

$$P = 20x + 14y + 12z.$$

The constraints posed in the problem are

$$\begin{aligned} x &\geq 0, \qquad & x &\leq 20, \\ y &\geq 0, \qquad & y &\leq 30, \\ z &\geq 0, \qquad & z &\leq 40, \end{aligned}$$

$$6x + 3y + 2z \leq 230.$$

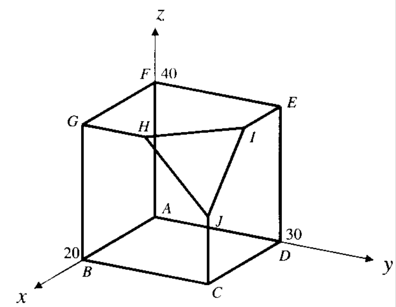

Figure 13.12 The convex set of points satisfying the constraints in Example 5

The last inequality is due to the limited supply of fabric. The solution set is shown in Figure 13.12. It has 10 vertices, $A, B, \ldots, J$. Since P increases in the direction of the vector $\nabla P = 20\mathbf{i} + 14\mathbf{j} + 12\mathbf{k}$, which points into the first octant, its maximum value cannot occur at any of the vertices $A, B, \ldots, G$. (Think about why.) Thus we need look only at the vertices H, I, and J.

$$\begin{aligned} H &= (20, 10, 40), \qquad & P &= 1{,}020 \text{ at } H. \\ I &= (10, 30, 40), \qquad & P &= 1{,}100 \text{ at } I. \\ J &= (20, 30, 10), \qquad & P &= 940 \text{ at } J. \end{aligned}$$

Thus, the tailor should make 10 suits, 30 jackets, and 40 pairs of slacks to realize the maximum profit, \$1,100, from the fabric. ■

Exercises 13.2

1. Find the maximum and minimum values of $f(x, y) = x - x^2 + y^2$ on the rectangle $0 \leq x \leq 2$, $0 \leq y \leq 1$.
2. Find the maximum and minimum values of $f(x, y) = xy - 2x$ on the rectangle $-1 \leq x \leq 1$, $0 \leq y \leq 1$.
3. Find the maximum and minimum values of $f(x, y) = xy - y^2$ on the disk $x^2 + y^2 \leq 1$.
4. Find the maximum and minimum values of $f(x, y) = x + 2y$ on the disk $x^2 + y^2 \leq 1$.
5. Find the maximum value of $f(x, y) = xy - x^3y^2$ over the square $0 \leq x \leq 1, 0 \leq y \leq 1$.
6. Find the maximum and minimum values of $f(x, y) = xy(1 - x - y)$ over the triangle with vertices $(0, 0)$, $(1, 0)$, and $(0, 1)$.
7. Find the maximum and minimum values of $f(x, y) = \sin x \cos y$ on the closed triangular region bounded by the coordinate axes and the line $x + y = 2\pi$.
8. Find the maximum value of
$$f(x, y) = \sin x \sin y \sin(x + y)$$
over the triangle bounded by the coordinate axes and the line $x + y = \pi$.
9. The temperature at all points in the disk $x^2 + y^2 \leq 1$ is given by
$$T = (x + y)\, e^{-x^2-y^2}.$$
Find the maximum and minimum temperatures at points of the disk.

10. Find the maximum and minimum values of

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

on the upper half-plane $y \geq 0$.

first quadrant: $x \geq 0$, $y \geq 0$. Show that $\lim_{x \to \infty} f(x, kx) = 0$. Does f have a limit as (x, y) recedes arbitrarily far from the origin in the first quadrant? Does f have a maximum value in the first quadrant?

14. Repeat Exercise 13 for the function $f(x, y) = xy^2 e^{-xy}$.

15. In a certain community there are two breweries in competition, so that sales of each negatively affect the profits of the other. If brewery A produces x L of beer per month and brewery B produces y L per month, then brewery A's monthly profit $\$P$ and brewery B's monthly profit $\$Q$ are assumed to be

$$P = 2x - \frac{2x^2 + y^2}{10^6},$$
$$Q = 2y - \frac{4y^2 + x^2}{2 \times 10^6}.$$

Find the sum of the profits of the two breweries if each brewery independently sets its own production level to maximize its own profit and assumes its competitor does likewise. Find the sum of the profits if the two breweries cooperate to determine their respective productions to maximize that sum.

16. Equal angle bends are made at equal distances from the two ends of a 100 m long straight length of fence so the resulting

18. Minimize $F(x, y, z) = 2x + 3y + 4z$ subject to $x \geq 0$, $y \geq 0$, $z \geq 0$, $x + y \geq 2$, $y + z \geq 2$, and $x + z \geq 2$.

19. A textile manufacturer produces two grades of wool-cotton-polyester fabric. The deluxe grade has composition (by weight) 20% wool, 50% cotton, and 30% polyester, and it sells for \$3 per kilogram. The standard grade has composition 10% wool, 40% cotton, and 50% polyester, and sells for \$2 per kilogram. If he has in stock 2,000 kg of wool and 6,000 kg each of cotton and polyester, how many kilograms of fabric of each grade should he manufacture to maximize his revenue?

20. A 10 hectare parcel of land is zoned for building densities of 6 detached houses per hectare, 8 duplex units per hectare, or 12 apartments per hectare. The developer who owns the land can make a profit of \$40,000 per house, \$20,000 per duplex unit, and \$16,000 per apartment that he builds. Municipal bylaws require him to build at least as many apartments as houses or duplex units. How many of each type of dwelling should he build to maximize his profit?

13.3 Lagrange Multipliers

A constrained extreme-value problem is one in which the variables of the function to be maximized or minimized are not completely independent of one another, but must satisfy one or more constraint equations or inequalities. For instance, the problems

$$\textbf{maximize} \quad f(x, y) \quad \textbf{subject to} \quad g(x, y) = C$$

and

$$\textbf{minimize} \quad f(x, y, z, w) \quad \textbf{subject to} \quad g(x, y, z, w) = C_1, \quad \textbf{and} \quad h(x, y, z, w) = C_2$$

have, respectively, one and two constraint equations, while the problem

$$\textbf{maximize} \quad f(x, y, z) \quad \textbf{subject to} \quad g(x, y, z) \leq C$$

has a single constraint inequality.

Generally, inequality constraints can be regarded as restricting the domain of the function to be extremized to a smaller set that still has interior points. Section 13.2 was devoted to such problems. In each of the first three examples of that section we looked for *free* (i.e., *unconstrained*) extreme values in the interior of the domain, and we also examined the boundary of the domain, which was specified by one or more *constraint equations*. In Example 1 we parametrized the boundary and expressed the function to be extremized as a function of the parameter, thus reducing the boundary case to a free problem in one variable instead of a constrained problem in two variables. In Example 2 the boundary consisted of three line segments, on two of which the function was obviously zero. We solved the equation for the third boundary segment for y in terms of x, again in order to express the values of $f(x, y)$ on that segment as a function of one free variable. A similar approach was used in Example 3 to deal with the triangular boundary of the domain of the area function $A(\theta, \phi)$.

The reduction of extremization problems with equation constraints to free problems with fewer independent variables is only feasible when the constraint equations can be solved either explicitly for some variables in terms of others or parametrically for all variables in terms of some parameters. It is often very difficult or impossible to solve the constraint equations, so we need another technique.

The Method of Lagrange Multipliers

A technique for finding extreme values of $f(x, y)$ subject to the equality constraint $g(x, y) = 0$ is based on the following theorem:

THEOREM 4

Suppose that f and g have continuous first partial derivatives near the point $P_0 = (x_0, y_0)$ on the curve $\mathcal{C}$ with equation $g(x, y) = 0$. Suppose also that, when restricted to points on $\mathcal{C}$, the function $f(x, y)$ has a local maximum or minimum value at P_0. Finally, suppose that

(i) P_0 is not an endpoint of $\mathcal{C}$, and

(ii) $\nabla g(P_0) \neq \mathbf{0}$.

Then there exists a number λ_0 such that (x_0, y_0, λ_0) is a critical point of the **Lagrangian function**

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y).$$

PROOF Together, (i) and (ii) imply that $\mathcal{C}$ is smooth enough to have a tangent line at P_0 and that $\nabla g(P_0)$ is normal to that tangent line. If $\nabla f(P_0)$ is not parallel to $\nabla g(P_0)$, then $\nabla f(P_0)$ has a nonzero vector projection $\mathbf{v}$ along the tangent line to $\mathcal{C}$ at P_0. (See Figure 13.13.) Therefore, f has a positive directional derivative at P_0 in the direction of $\mathbf{v}$ and a negative directional derivative in the opposite direction. Thus, $f(x, y)$ increases or decreases as we move away from P_0 along $\mathcal{C}$ in the direction of $\mathbf{v}$ or $-\mathbf{v}$, and f cannot have a maximum or minimum value at P_0. Since we are assuming that f *does* have an extreme value at P_0, it must be that $\nabla f(P_0)$ is parallel to $\nabla g(P_0)$. Since $\nabla g(P_0) \neq \mathbf{0}$, there must exist a real number λ_0 such that $\nabla f(P_0) = -\lambda_0 \nabla g(P_0)$, or

$$\nabla(f + \lambda_0 g)(P_0) = \mathbf{0}.$$

The two components of the above vector equation assert that $\partial L/\partial x = 0$ and $\partial L/\partial y = 0$ at (x_0, y_0, λ_0). The third equation that must be satisfied by a critical point of L is $\partial L/\partial \lambda = g(x, y) = 0$. This is satisfied at (x_0, y_0, λ_0) because P_0 lies on $\mathcal{C}$. Thus (x_0, y_0, λ_0) is a critical point of $L(x, y, \lambda)$.

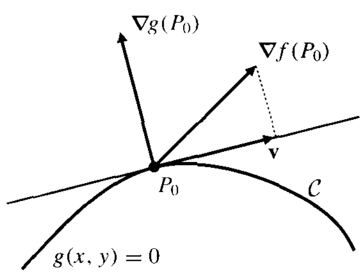

Figure 13.13 If $\nabla f(P_0)$ is not a multiple of $\nabla g(P_0)$, then $\nabla f(P_0)$ has a nonzero projection $\mathbf{v}$ tangent to the level curve of g through P_0

Theorem 4 suggests that to find candidates for points on the curve $g(x, y) = 0$ at which $f(x, y)$ is maximum or minimum, we should look for critical points of the Lagrangian function

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y).$$

At any critical point of L we must have

$$\left.\begin{aligned} 0 &= \frac{\partial L}{\partial x} = f_1(x, y) + \lambda g_1(x, y), \\ 0 &= \frac{\partial L}{\partial y} = f_2(x, y) + \lambda g_2(x, y), \end{aligned}\right\} \text{ i.e., } \nabla f \text{ is parallel to } \nabla g,$$

$$\text{and} \quad 0 = \frac{\partial L}{\partial \lambda} = g(x, y) \qquad \text{the constraint equation.}$$

Note, however, that it is *assumed* that the constrained problem *has a solution*. Theorem 4 does not guarantee that a solution exists; it only provides a means for finding a solution already known to exist. It is usually necessary to satisfy yourself that the problem you are trying to solve has a solution before using this method to find the solution.

Let us put the method to a concrete test:

Example 1 Find the shortest distance from the origin to the curve $x^2y = 16$.

Solution The graph of $x^2y = 16$ is shown in Figure 13.14. There appear to be two points on the curve that are closest to the origin, and no points that are farthest from the origin. (The curve is unbounded.) To find the closest points it is sufficient to minimize the *square* of the distance from the point (x, y) on the curve to the origin. (It is easier to work with the square of the distance rather than the distance itself, which involves a square root and so is harder to differentiate.) Thus we want to solve the problem

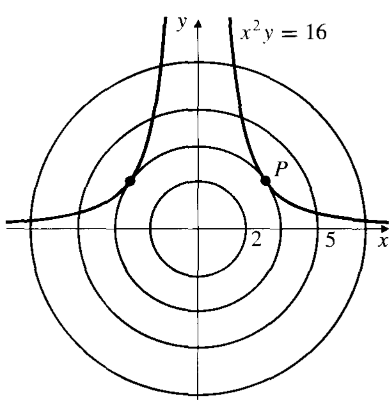

Figure 13.14 The level curve of the function representing the square of distance from the origin is tangent to the curve $x^2y = 16$ at the two points on that curve that are closest to the origin

$$\textbf{minimize} \quad f(x, y) = x^2 + y^2 \quad \textbf{subject to} \quad g(x, y) = x^2y - 16 = 0.$$

Let $L(x, y, \lambda) = x^2 + y^2 + \lambda(x^2y - 16)$. For critical points of L we want

$$0 = \frac{\partial L}{\partial x} = 2x + 2\lambda xy = 2x(1 + \lambda y) \qquad \text{(A)}$$

$$0 = \frac{\partial L}{\partial y} = 2y + \lambda x^2 \qquad \text{(B)}$$

$$0 = \frac{\partial L}{\partial \lambda} = x^2y - 16. \qquad \text{(C)}$$

Equation (A) requires that either $x = 0$ or $\lambda y = -1$. However, $x = 0$ is inconsistent with equation (C). Therefore $\lambda y = -1$. From equation (B) we now have

$$0 = 2y^2 + \lambda y x^2 = 2y^2 - x^2.$$

Thus, $x = \pm\sqrt{2}y$, and (C) now gives $2y^3 = 16$, so $y = 2$. There are, therefore, two candidates for points on $x^2 y = 16$ closest to the origin, $(\pm 2\sqrt{2}, 2)$. Both of these points are at distance $\sqrt{8+4} = 2\sqrt{3}$ units from the origin, so this must be the minimum distance from the origin to the curve. Some level curves of $x^2 + y^2$ are shown, along with the constraint curve $x^2 y = 16$, in Figure 13.14. Observe how the constraint curve is tangent to the level curve passing through the minimizing points $(\pm 2\sqrt{2}, 2)$, reflecting the fact that the two curves have parallel normals there. ■

Remark In the above example we could, of course, have solved the constraint equation for $y = 16/x^2$, substituted into f, and thus reduced the problem to one of finding the (unconstrained) minimum value of

$$F(x) = f\left(x, \frac{16}{x^2}\right) = x^2 + \frac{256}{x^4}.$$

The reader is invited to verify that this gives the same result.

The number λ that occurs in the Lagrangian function is called a **Lagrange multiplier**. The technique for solving an extreme-value problem with equation constraints by looking for critical points of an unconstrained problem in more variables (the original variables plus a Lagrange multiplier corresponding to each constraint equation) is called **the method of Lagrange multipliers.** It can be expected to give results as long as the function to be maximized or minimized (called the **objective function**) and the constraint equations have *smooth* graphs in a neighbourhood of the points where the extreme values occur, and these points are not on *edges* of those graphs. See Example 3 and Exercise 26 below.

Example 2 Find the points on the curve $17x^2 + 12xy + 8y^2 = 100$ that are closest to and farthest away from the origin.

Solution The quadratic form on the left side of the equation above is positive definite, as can be seen by completing a square. Hence the curve is bounded and must have points closest to and farthest from the origin. (In fact, the curve is an ellipse with centre at the origin and oblique principal axes. The problem asks us to find the ends of the major and minor axes.)

Again, we want to extremize $x^2 + y^2$ subject to an equation constraint. The Lagrangian in this case is

$$L(x, y, \lambda) = x^2 + y^2 + \lambda(17x^2 + 12xy + 8y^2 - 100),$$

and its critical points are given by

$$0 = \frac{\partial L}{\partial x} = 2x + \lambda(34x + 12y) \qquad \text{(A)}$$

$$0 = \frac{\partial L}{\partial y} = 2y + \lambda(12x + 16y) \qquad \text{(B)}$$

$$0 = \frac{\partial L}{\partial \lambda} = 17x^2 + 12xy + 8y^2 - 100. \qquad \text{(C)}$$

Solving each of equations (A) and (B) for λ and equating the two expressions for λ obtained, we get

$$\frac{-2x}{34x+12y} = \frac{-2y}{12x+16y} \qquad \text{or} \qquad 12x^2 + 16xy = 34xy + 12y^2.$$

This equation simplifies to

$$2x^2 - 3xy - 2y^2 = 0. \tag{D}$$

We multiply equation (D) by 4 and add the result to equation (C) to get $25x^2 = 100$, so that $x = \pm 2$. Finally, we substitute each of these values of x into (D) and obtain (for each) two values of y from the resulting quadratics:

$$\text{For } x = 2: \quad y^2 + 3y - 4 = 0, \quad (y-1)(y+4) = 0. \qquad \text{For } x = -2: \quad y^2 - 3y - 4 = 0, \quad (y+1)(y-4) = 0.$$

We therefore obtain four candidate points: (2, 1), (−2, −1), (2, −4), and (−2, 4). The first two points are closest to the origin (they are the ends of the minor axis of the ellipse); the second pair are farthest from the origin (the ends of the major axis). (See Figure 13.15.)

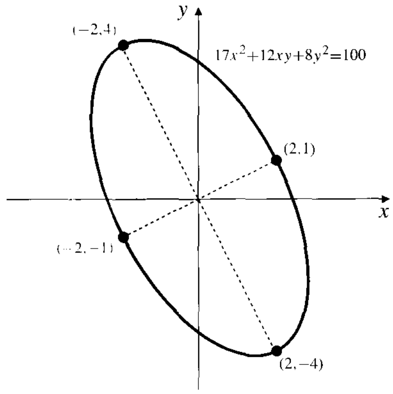

Figure 13.15 The points on the ellipse that are closest to and farthest from the origin

Considering the geometric underpinnings of the method of Lagrange multipliers, we would not expect the method to work if the level curves of the functions involved are not smooth or if the maximum or minimum occurs at an endpoint of the constraint curve. One of the pitfalls of the method is that the level curves of functions may not be smooth, even though the functions themselves have partial derivatives. Problems can occur where a gradient vanishes, as the following example shows.

Example 3 Find the minimum value of $f(x, y) = y$ subject to the constraint equation $g(x, y) = y^3 - x^2 = 0$.

Solution The semicubical parabola $y^3 = x^2$ has a cusp at the origin. (See Figure 13.16.) Clearly, $f(x, y) = y$ has minimum value 0 at that point. Suppose, however, that we try to solve the problem using the method of Lagrange multipliers. The Lagrangian here is

$$L(x, y, \lambda) = y + \lambda(y^3 - x^2),$$

which has critical points given by

$$\begin{aligned} -2\lambda x &= 0, \\ 1 + 3\lambda y^2 &= 0, \\ y^3 - x^2 &= 0. \end{aligned}$$

Observe that $y = 0$ cannot satisfy the second equation, and, in fact, the three equations have *no solution* (x, y, λ). (The first equation implies either $\lambda = 0$ or $x = 0$, but neither of these is consistent with the other two equations.)

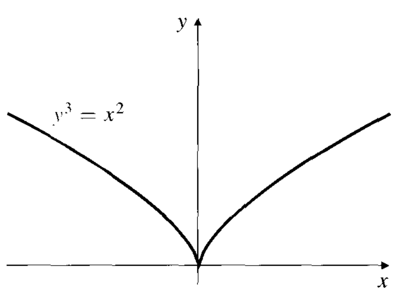

Figure 13.16 The minimum of y occurs at a point on the curve where the curve has no tangent line

Remark The method of Lagrange multipliers breaks down in the above example because $\nabla g = \mathbf{0}$ at the solution point, and therefore the curve $g(x, y) = 0$ need not be smooth there. (In this case, it isn't smooth!) The geometric condition that ∇f should be parallel to ∇g at the solution point is meaningless in this case. When applying the method of Lagrange multipliers, be aware that an extreme value may occur at

(i) a critical point of the Lagrangian,
(ii) a point where $\nabla g = \mathbf{0}$,
(iii) a point where ∇f or ∇g does not exist, or
(iv) an "endpoint" of the constraint set.

This situation is similar to that for extreme values of a function f of one variable, which can occur at a critical point of f, a singular point of f, or an endpoint of the domain of f.

Problems with More than One Constraint

Next consider a three-dimensional problem requiring us to find a maximum or minimum value of a function of three variables subject to two equation constraints:

$$\textbf{extremize} \quad f(x, y, z) \quad \textbf{subject to} \quad g(x, y, z) = 0 \text{ and } h(x, y, z) = 0.$$

Again, we assume that the problem has a solution, say at the point $P_0 = (x_0, y_0, z_0)$, and that the functions f, g, and h have continuous first partial derivatives near P_0. Also, we assume that $\mathbf{T} = \nabla g(P_0) \times \nabla h(P_0) \neq \mathbf{0}$. These conditions imply that the surfaces $g(x, y, z) = 0$ and $h(x, y, z) = 0$ are smooth near P_0 and are not tangent to each other there, so they must intersect in a curve $\mathcal{C}$ that is smooth near P_0. The curve $\mathcal{C}$ has tangent vector $\mathbf{T}$ at P_0. The same geometric argument used in the proof of Theorem 4 again shows that $\nabla f(P_0)$ must be perpendicular to $\mathbf{T}$. (If not, then it would have a nonzero vector projection along $\mathbf{T}$, and f would have nonzero directional derivatives in the directions $\pm\mathbf{T}$ and would therefore increase and decrease as we moved away from P_0 along $\mathcal{C}$ in opposite directions.)

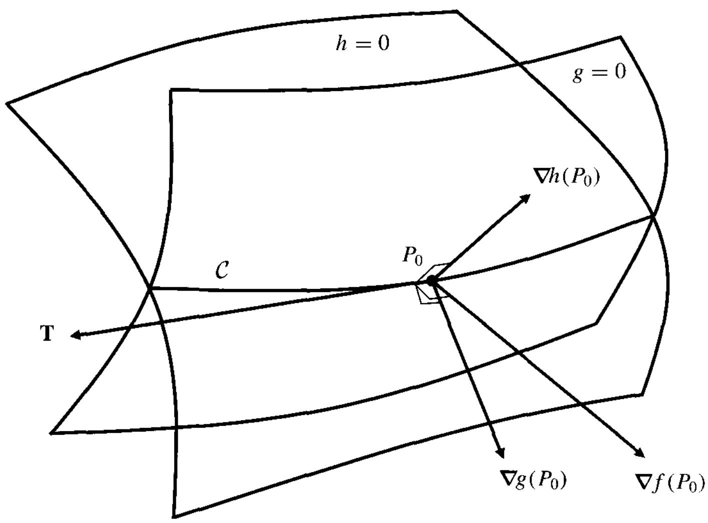

Figure 13.17 At P_0, ∇f, ∇g, and ∇h are all perpendicular to $\mathbf{T}$. Thus, ∇f is in the plane spanned by ∇g and ∇h

Since $\nabla g(P_0)$ and $\nabla h(P_0)$ are nonzero and both are perpendicular to $\mathbf{T}$ (see Figure 13.17), $\nabla f(P_0)$ must lie in the plane spanned by these two vectors and hence must be a linear combination of them:

$$\nabla f(x_0, y_0, z_0) = -\lambda_0 \nabla g(x_0, y_0, z_0) - \mu_0 \nabla h(x_0, y_0, z_0)$$

for some constants λ_0 and μ_0. It follows that $(x_0, y_0, z_0, \lambda_0, \mu_0)$ is a critical point of the Lagrangian function

$$L(x, y, z, \lambda, \mu) = f(x, y, z) + \lambda g(x, y, z) + \mu h(x, y, z).$$

We look for triples (x, y, z) that extremize $f(x, y, z)$ subject to the two constraints $g(x, y, z) = 0$ and $h(x, y, z) = 0$ among the points (x, y, z, λ, μ) that are critical points of the above Lagrangian function, and we therefore solve the system of equations

$$\begin{aligned} f_1(x, y, z) + \lambda g_1(x, y, z) + \mu h_1(x, y, z) &= 0, \\ f_2(x, y, z) + \lambda g_2(x, y, z) + \mu h_2(x, y, z) &= 0, \\ f_3(x, y, z) + \lambda g_3(x, y, z) + \mu h_3(x, y, z) &= 0, \\ g(x, y, z) &= 0, \\ h(x, y, z) &= 0. \end{aligned}$$

Solving such a system can be very difficult. It should be noted that, in using the method of Lagrange multipliers instead of solving the constraint equations, we have traded the problem of having to solve two equations for two variables as *functions* of a third one for a problem of having to solve five equations for *numerical* values of five unknowns.

Example 4 Find the maximum and minimum values of $f(x, y, z) = xy + 2z$ on the circle that is the intersection of the plane $x + y + z = 0$ and the sphere $x^2 + y^2 + z^2 = 24$.

Solution The function f is continuous, and the circle is a closed bounded set in 3-space. Therefore, maximum and minimum values must exist. We look for critical points of the Lagrangian

$$L = xy + 2z + \lambda(x + y + z) + \mu(x^2 + y^2 + z^2 - 24).$$

Setting the first partial derivatives of L equal to zero, we obtain

$$\begin{aligned} y + \lambda + 2\mu x &= 0, & \text{(A)} \\ x + \lambda + 2\mu y &= 0, & \text{(B)} \\ 2 + \lambda + 2\mu z &= 0, & \text{(C)} \\ x + y + z &= 0, & \text{(D)} \\ x^2 + y^2 + z^2 - 24 &= 0. & \text{(E)} \end{aligned}$$

When none of the equations factors, try to combine two or more of them to produce an equation that does factor.

Subtracting (A) from (B) we get $(x - y)(1 - 2\mu) = 0$. Therefore either $\mu = \frac{1}{2}$ or $x = y$. We analyze both possibilities.

CASE I If $\mu = \frac{1}{2}$, we obtain from (B) and (C)

$$x + \lambda + y = 0 \quad \text{and} \quad 2 + \lambda + z = 0.$$

Thus $x + y = 2 + z$. Combining this with (D), we get $z = -1$ and $x + y = 1$. Now, by (E), $x^2 + y^2 = 24 - z^2 = 23$. Since $x^2 + y^2 + 2xy = (x + y)^2 = 1$, we have $2xy = 1 - 23 = -22$ and $xy = -11$. Now $(x - y)^2 = x^2 + y^2 - 2xy = 23 + 22 = 45$, so $x - y = \pm 3\sqrt{5}$. Combining this with $x + y = 1$, we obtain two critical points arising from $\mu = \frac{1}{2}$, namely, $\big((1 + 3\sqrt{5})/2, (1 - 3\sqrt{5})/2, -1\big)$ and $\big((1 - 3\sqrt{5})/2, (1 + 3\sqrt{5})/2, -1\big)$. At both of these points we find that $f(x, y, z) = xy + 2z = -11 - 2 = -13$.

CASE II If $x = y$, then (D) implies that $z = -2x$, and (E) then gives $6x^2 = 24$, so $x = \pm 2$. Therefore, points $(2, 2, -4)$ and $(-2, -2, 4)$ must be considered. We have $f(2, 2, -4) = 4 - 8 = -4$ and $f(-2, -2, 4) = 4 + 8 = 12$.

We conclude that the maximum value of f on the circle is 12, and the minimum value is -13.

■

The method of Lagrange multipliers can be applied to find extreme values of a function of n variables, that is, of a vector variable $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ subject to $m \le n - 1$ constraints:

$$\textbf{extremize} \quad f(\mathbf{x}) \quad \textbf{subject to} \quad g_{(1)}(\mathbf{x}) = 0, \quad \ldots \quad g_{(m)}(\mathbf{x}) = 0.$$

Assuming that the problem has a solution at the point P_0, that f and all of the functions $g_{(j)}$ have continuous first partial derivatives in a neighbourhood of P_0, and that the intersection of the constraint (hyper)surfaces is smooth near P_0, then we should look for P_0 among the critical points of the $(n + m)$-variable Lagrangian function

$$L(\mathbf{x}, \lambda_1, \ldots, \lambda_m) = f(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j g_{(j)}(\mathbf{x}).$$

We will not attempt to prove this general assertion. (A proof can be based on the Implicit Function Theorem.) Any critical points must satisfy the $n + m$ equations

$$\frac{\partial L}{\partial x_i} = 0, \quad (1 \le i \le n), \qquad \frac{\partial L}{\partial \lambda_j} = g_{(j)}(\mathbf{x}) = 0, \quad (1 \le j \le m).$$

Nonlinear Programming

When we looked for extreme values of functions f on restricted domains R in the previous section, we had to look separately for critical points of f in the interior of R and then for critical points of the restriction of f to the boundary of R. The interior of R is typically specified by one or more inequality constraints of the form $g < 0$, while the boundary corresponds to equation constraints of the form $g = 0$ (for which Lagrange multipliers can be used).

It is possible to unify these approaches into a single method for finding extreme values of functions defined on regions specified by inequalities of the form $g \le 0$.

Consider, for example, the problem of finding extreme values of $f(x, y)$ over the region R specified by $g(x, y) \le 0$. We can proceed by trying to find critical points of the four-variable function

$$L(x, y, \lambda, u) = f(x, y) + \lambda\big(g(x, y) + u^2\big).$$

Such critical points must satisfy the four equations

$$0 = \frac{\partial L}{\partial x} = f_1(x, y) + \lambda g_1(x, y), \tag{A}$$
$$0 = \frac{\partial L}{\partial y} = f_2(x, y) + \lambda g_2(x, y), \tag{B}$$
$$0 = \frac{\partial L}{\partial \lambda} = g(x, y) + u^2, \tag{C}$$
$$0 = \frac{\partial L}{\partial u} = 2\lambda u. \tag{D}$$

Suppose that (x, y, λ, u) satisfies these equations. We consider two cases:

CASE I $u \neq 0$. Then (D) implies that $\lambda = 0$, (C) implies that $g(x, y) = -u^2 < 0$, and (A) and (B) imply that $f_1(x, y) = 0$ and $f_2(x, y) = 0$. Thus, (x, y) is an interior critical point of f.

CASE II $u = 0$. Then (C) implies that $g(x, y) = 0$, and (A) and (B) imply that $\nabla f(x, y) = -\lambda \nabla g(x, y)$, so that (x, y) is a boundary point candidate for the location of the extreme value.

This technique can be extended to the problem of finding extreme values of a function of n variables, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, over the intersection R of m regions R_j defined by inequality constraints of the form $g_{(j)}(\mathbf{x}) \leq 0$.

$$\textbf{extremize} \quad f(\mathbf{x}) \quad \textbf{subject to} \quad g_{(1)}(\mathbf{x}) \leq 0, \quad \ldots \quad g_{(m)}(\mathbf{x}) \leq 0.$$

In this case we look for critical points of the $(n + 2m)$-variable Lagrangian

$$L(\mathbf{x}, \lambda_1, \ldots, \lambda_m, u_1, \ldots, u_m) = f(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j \big(g_{(j)}(\mathbf{x}) + u_j^2\big).$$

The critical points will satisfy $n + 2m$ equations

$$\nabla f(\mathbf{x}) = -\sum_{j=1}^{m} \lambda_j \nabla g_{(j)}(\mathbf{x}), \qquad (n \text{ equations})$$
$$g_{(j)}(\mathbf{x}) = -u_j^2, \qquad (1 \leq j \leq m), \qquad (m \text{ equations})$$
$$2\lambda_j u_j = 0, \qquad (1 \leq j \leq m). \qquad (m \text{ equations})$$

The last m equations show that $\lambda_j = 0$ for any j for which $u_j \neq 0$. If all $u_j \neq 0$ then $\mathbf{x}$ is a critical point of f interior to R. Otherwise some of the u_j will be zero, say those corresponding to j in a subset J of $\{1, 2, \ldots, m\}$. In this case $\mathbf{x}$ will lie on the part of the boundary of R consisting of points lying on the boundaries of each of the regions R_j for which $j \in J$, and ∇f will be a linear combination of the corresponding gradients $\nabla g_{(j)}$:

$$\nabla f(\mathbf{x}) = -\sum_{j \in J} \lambda_j \nabla g_{(j)}(\mathbf{x}).$$

These are known as **Kuhn-Tucker conditions**, and this technique for solving extreme-value problems on restricted domains is called **nonlinear programming**.

Exercises 13.3

1. Use the method of Lagrange multipliers to maximize x^3y^5 subject to the constraint $x + y = 8$.

2. Find the shortest distance from the point $(3, 0)$ to the parabola $y = x^2$,
   (a) by reducing to an unconstrained problem in one variable, and
   (b) by using the method of Lagrange multipliers.

3. Find the distance from the origin to the plane $x + 2y + 2z = 3$,
   (a) using a geometric argument (no calculus),
   (b) by reducing the problem to an unconstrained problem in two variables, and
   (c) using the method of Lagrange multipliers.

4. Find the maximum and minimum values of the function $f(x, y, z) = x + y - z$ over the sphere $x^2 + y^2 + z^2 = 1$.

5. Use the Lagrange multiplier method to find the greatest and least distances from the point $(2, 1, -2)$ to the sphere with equation $x^2 + y^2 + z^2 = 1$. (Of course, the answer could be obtained more easily using a simple geometric argument.)

6. Find the shortest distance from the origin to the surface $xyz^2 = 2$.

7. Find a, b, and c so that the volume $V = 4\pi abc/3$ of an ellipsoid $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$ passing through the point $(1, 2, 1)$ is as small as possible.

8. Find the ends of the major and minor axes of the ellipse $3x^2 + 2xy + 3y^2 = 16$.

9. Find the maximum and minimum values of $f(x, y, z) = xyz$ on the sphere $x^2 + y^2 + z^2 = 12$.

10. Find the maximum and minimum values of $x + 2y - 3z$ over the ellipsoid $x^2 + 4y^2 + 9z^2 \le 108$.

11. Find the maximum and minimum values of the function $f(x, y, z) = x$ over the curve of intersection of the plane $z = x + y$ and the ellipsoid $x^2 + 2y^2 + 2z^2 = 8$.

12. Find the maximum and minimum values of $f(x, y, z) = x^2 + y^2 + z^2$ on the ellipse formed by the intersection of the cone $z^2 = x^2 + y^2$ and the plane $x - 2z = 3$.

13. Find the maximum and minimum values of $f(x, y, z) = 4 - z$ on the ellipse formed by the intersection of the cylinder $x^2 + y^2 = 8$ and the plane $x + y + z = 1$.

14. Find the maximum and minimum values of $f(x, y, z) = x + y^2z$ subject to the constraints $y^2 + z^2 = 2$ and $z = x$.

* 15. Use the method of Lagrange multipliers to find the shortest distance between the straight lines $x = y = z$ and $x = -y$, $z = 2$. (There are, of course, much easier ways to get the answer. This is an object lesson in the folly of shooting sparrows with cannons.)

16. Find the maximum and minimum values of the n-variable function $x_1 + x_2 + \cdots + x_n$ subject to the constraint $x_1^2 + x_2^2 + \cdots + x_n^2 = 1$.

17. Repeat Exercise 16 for the function $x_1 + 2x_2 + 3x_3 + \cdots + nx_n$ with the same constraint.

18. Find the most economical shape of a rectangular box with no top.

19. Find the maximum volume of a rectangular box with faces parallel to the coordinate planes if one corner is at the origin and the diagonally opposite corner lies on the plane $4x + 2y + z = 2$.

20. Find the maximum volume of a rectangular box with faces parallel to the coordinate planes if one corner is at the origin and the diagonally opposite corner is on the first octant part of the surface $xy + 2yz + 3xz = 18$.

21. A rectangular box having no top and having a prescribed volume $V\,\text{m}^3$ is to be constructed using two different materials. The material used for the bottom and front of the box is five times as costly (per square metre) as the material used for the back and the other two sides. What should be the dimensions of the box to minimize the cost of materials?

* 22. Find the maximum and minimum values of $xy + z^2$ on the ball $x^2 + y^2 + z^2 \le 1$. Use Lagrange multipliers to treat the boundary case.

* 23. Repeat Exercise 22 but handle the boundary case by parametrizing the sphere $x^2 + y^2 + z^2 = 1$ using

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

where $0 \le \phi \le \pi$ and $0 \le \theta \le 2\pi$.

* 24. If α, β, and γ are the angles of a triangle, show that

$$\sin\frac{\alpha}{2}\sin\frac{\beta}{2}\sin\frac{\gamma}{2} \le \frac{1}{8}.$$

For what triangles does equality occur?

* 25. Suppose that f and g have continuous first partial derivatives throughout the xy-plane, and suppose that $g_2(a, b) \ne 0$. This implies that the equation $g(x, y) = g(a, b)$ defines y implicitly as a function of x near the point (a, b). Use the Chain Rule to show that if $f(x, y)$ has a local extreme value at (a, b) subject to the constraint $g(x, y) = g(a, b)$, then for some number λ the point (a, b, λ) is a critical point of the function

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y).$$

This constitutes a more formal justification of the method of Lagrange multipliers in this case.

26. What is the shortest distance from the point $(0, -1)$ to the curve $y = \sqrt{1 - x^2}$? Can this problem be solved by the Lagrange multiplier method? Why?

27. Example 3 showed that the method of Lagrange multipliers might fail to find a point that extremizes $f(x, y)$ subject to the constraint $g(x, y) = 0$ if $\nabla g = \mathbf{0}$ at the extremizing point. Can the method also fail if $\nabla f = \mathbf{0}$ at the extremizing point? Why?

13.4 The Method of Least Squares

Important optimization problems arise in the statistical analysis of experimental data. Frequently experiments are designed to measure the values of one or more quantities supposed to be constant or to demonstrate a supposed functional relationship among variable quantities. Experimental error is usually present in the measurements, and experiments need to be repeated several times in order to arrive at *mean* or *average* values of the quantities being measured.

Consider a very simple example. An experiment to measure a certain physical constant c is repeated n times, yielding the values $c_1, c_2, \ldots, c_n$. If none of the measurements is suspected of being faulty, intuition tells us that we should use the mean value $\bar{c} = (c_1 + c_2 + \cdots + c_n)/n$ as the value of c determined by the experiments. Let us see how this intuition can be justified.

Various methods for determining c from the data values are possible. We could, for instance, choose c to minimize the sum T of its distances from the data points:

$$T = |c - c_1| + |c - c_2| + \cdots + |c - c_n|.$$

This is unsatisfactory for a number of reasons. Since absolute values have singular points, it is difficult to determine the minimizing value of c. More importantly, c may not be determined uniquely. If $n = 2$, any point in the interval between c_1 and c_2 will give the same minimum value to T. (See Exercise 24 below for a generalization of this phenomenon.)

A more promising approach is to minimize the sum S of *squares* of the distances from c to the data points:

$$S = (c - c_1)^2 + (c - c_2)^2 + \cdots + (c - c_n)^2 = \sum_{i=1}^{n} (c - c_i)^2.$$

This function of c is smooth, and its (unconstrained) minimum value will occur at a critical point $\bar{c}$ given by

$$0 = \left.\frac{dS}{dc}\right|_{c=\bar{c}} = \sum_{i=1}^{n} 2(\bar{c} - c_i) = 2n\bar{c} - 2\sum_{i=1}^{n} c_i.$$

Thus $\bar{c}$ is the *mean* of the data values:

$$\bar{c} = \frac{1}{n}\sum_{i=1}^{n} c_i = \frac{c_1 + c_2 + \cdots + c_n}{n}.$$

The technique used to obtain $\bar{c}$ above is an example of what is called **the method of least squares.** It has the following geometric interpretation. If the data values $c_1, c_2, \ldots, c_n$ are regarded as components of a vector $\mathbf{c}$ in $\mathbb{R}^n$, and $\mathbf{w}$ is the vector with components $1, 1, \ldots, 1$, then the vector projection of $\mathbf{c}$ in the direction of $\mathbf{w}$,

$$\mathbf{c_w} = \frac{\mathbf{c} \bullet \mathbf{w}}{|\mathbf{w}|^2}\mathbf{w} = \frac{c_1 + c_2 + \cdots + c_n}{n}\mathbf{w},$$

has all its components equal to the average of the data values. Thus, determining c from the data by the method of least squares corresponds to finding the vector projection of the data vector onto the one-dimensional subspace of $\mathbb{R}^n$ spanned by $\mathbf{w}$. Had there been no error in the measurements c_i, then $\mathbf{c}$ would have been equal to $c\mathbf{w}$.

Linear Regression

In scientific investigations it is often believed that the response of a system is a certain kind of function of one or more input variables. An investigator can set up an experiment to measure the response of the system for various values of those variables in order to determine the parameters of the function.

For example, suppose that the response y of a system is suspected to depend on the input x according to the linear relationship

$$y = ax + b,$$

where the values of a and b are unknown. An experiment set up to measure values of y corresponding to several values of x yields n data points, (x_i, y_i), $i = 1, 2, \ldots, n$. If the supposed linear relationship is valid, these data points should lie *approximately* along a straight line, but not exactly on one because of experimental error. Suppose the points are as shown in Figure 13.18. The linear relationship seems reasonable in this case. We want to find values of a and b so that the straight line $y = ax + b$ "best" fits the data.

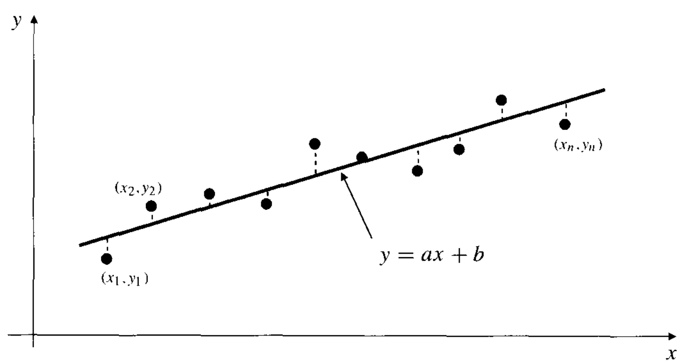

Figure 13.18 Fitting a straight line through experimental data

In this situation the method of least squares requires that a and b be chosen to minimize the sum S of the squares of the vertical displacements of the data points from the line:

$$S = \sum_{i=1}^{n} (y_i - ax_i - b)^2.$$

This is an unconstrained minimum problem in two variables, a and b. The minimum will occur at a critical point of S that satisfies

$$0 = \frac{\partial S}{\partial a} = -2 \sum_{i=1}^{n} x_i (y_i - ax_i - b),$$

$$0 = \frac{\partial S}{\partial b} = -2 \sum_{i=1}^{n} (y_i - ax_i - b).$$

These equations can be rewritten

$$\begin{aligned}\left(\sum_{i=1}^{n} x_i^2\right)a \quad + \quad \left(\sum_{i=1}^{n} x_i\right)b \quad &= \quad \sum_{i=1}^{n} x_i y_i,\\ \left(\sum_{i=1}^{n} x_i\right)a \quad + \quad nb \quad &= \quad \sum_{i=1}^{n} y_i.\end{aligned}$$

Solving this pair of linear equations, we obtain the desired parameters:

$$a = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2},$$

$$b = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\overline{x^2}\bar{y} - \bar{x}\overline{xy}}{\overline{x^2} - (\bar{x})^2}.$$

In these formulas we have used a bar to indicate the mean value of a quantity; thus $\overline{xy} = (1/n)\sum_{i=1}^{n} x_i y_i$, and so on.

This procedure for fitting the "best" straight line through data points by the method of least squares is called **linear regression**, and the line $y = ax + b$ obtained in this way is called the **empirical regression line** corresponding to the data. Some scientific calculators with statistical features provide for linear regression by accumulating the sums of x_i, y_i, x_i^2, and $x_i y_i$ in various registers and keeping track of the number n of data points entered in another register. At any time it has available the information necessary to calculate a and b and the value of y corresponding to any given x.

Example 1 Find the empirical regression line for the data $(x, y) = (0, 2.10)$, $(1, 1.92)$, $(2, 1.84)$, and $(3, 1.71)$, $(4, 1.64)$. What is the predicted value of y at $x = 5$?

Solution We have

$$\begin{aligned}\bar{x} &= \frac{0 + 1 + 2 + 3 + 4}{5} = 2,\\ \bar{y} &= \frac{2.10 + 1.92 + 1.84 + 1.71 + 1.64}{5} = 1.842,\end{aligned}$$

$$\overline{xy} = \frac{(0)(2.10) + (1)(1.92) + (2)(1.84) + (3)(1.71) + (4)(1.64)}{5} = 3.458,$$
$$\overline{x^2} = \frac{0^2 + 1^2 + 2^2 + 3^2 + 4^2}{5} = 6.$$

Therefore,

$$a = \frac{3.458 - (2)(1.842)}{6 - 2^2} = -0.113,$$
$$b = \frac{(6)(1.842) - (2)(3.458)}{6 - 2^2} = 2.068,$$

and the empirical regression line is

$$y = 2.068 - 0.113x.$$

The predicted value of y at $x = 5$ is $2.068 - 0.113 \times 5 = 1.503$. ■

Linear regression can also be interpreted in terms of vector projection. The data points define two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ with components $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$, respectively. Let $\mathbf{w}$ be the vector with components $1, 1, \ldots, 1$. Finding the coefficients a and b for the regression line corresponds to finding the orthogonal projection of $\mathbf{y}$ onto the two-dimensional subspace (plane) in $\mathbb{R}^n$ spanned by $\mathbf{x}$ and $\mathbf{w}$. (See Figure 13.19.) This projection is $\mathbf{p} = a\mathbf{x} + b\mathbf{w}$. In fact, the two equations obtained above by setting the partial derivatives of S equal to zero are just the two conditions

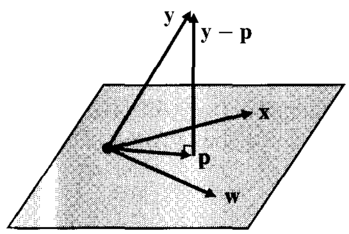

Figure 13.19 $\mathbf{p} = a\mathbf{x} + b\mathbf{w}$ is the projection of $\mathbf{y}$ onto the plane spanned by $\mathbf{x}$ and $\mathbf{w}$

$$(\mathbf{y} - \mathbf{p}) \bullet \mathbf{x} = 0,$$
$$(\mathbf{y} - \mathbf{p}) \bullet \mathbf{w} = 0,$$

stating that $\mathbf{y}$ minus its projection onto the subspace is perpendicular to the subspace. The angle between $\mathbf{y}$ and this $\mathbf{p}$ provides a measure of how well the empirical regression line fits the data; the smaller the angle, the better the fit.

Linear regression can be used to find specific functional relationships of types other than linear if suitable transformations are applied to the data.

Example 2 Find the values of constants K and s for which the curve

$$y = Kx^s$$

best fits the experimental data points (x_i, y_i), $i = 1, 2, \ldots, n$. (Assume all data values are positive.)

Solution Observe that the required functional form corresponds to a linear relationship between $\ln y$ and $\ln x$:

$$\ln y = \ln K + s \ln x.$$

If we determine the parameters a and b of the empirical regression line $\eta = a\xi + b$ corresponding to the transformed data $(\xi_i, \eta_i) = (\ln x_i, \ln y_i)$, then $s = a$ and $K = e^b$ are the required values. ■

Remark It should be stressed that the constants K and s obtained by the method used in the solution above are not the same as those that would be obtained by direct application of the least squares method to the untransformed problem, that is, by minimizing $\sum_{i=1}^{n}(y_i - Kx_i^s)^2$. This latter problem cannot readily be solved. (Try it!)

Generally, the method of least squares is applied to fit an equation in which the response is expressed as a sum of constants times functions of one or more input variables. The constants are determined as critical points of the sum of squared deviations of the actual response values from the values predicted by the equation.

Applications of the Least Squares Method to Integrals

The method of least squares can be used to find approximations to reasonably well-behaved (say, piecewise continuous) functions as sums of constants times specified functions. The idea is to choose the constants to minimize the *integral* of the square of the difference.

For example, suppose we want to approximate the continuous function $f(x)$ over the interval $[0, 1]$ by a linear function $g(x) = px + q$. The method of least squares would require that p and q be chosen to minimize the integral

$$I(p, q) = \int_0^1 \Big(f(x) - px - q\Big)^2 dx.$$

Assuming that we can "differentiate through the integral" (we will investigate this issue in Section 13.5), the critical point of $I(p, q)$ can be found from

$$0 = \frac{\partial I}{\partial p} = -2\int_0^1 x\Big(f(x) - px - q\Big)\,dx,$$
$$0 = \frac{\partial I}{\partial q} = -2\int_0^1 \Big(f(x) - px - q\Big)\,dx.$$

Thus,

$$\frac{p}{3} + \frac{q}{2} = \int_0^1 xf(x)\,dx,$$
$$\frac{p}{2} + q = \int_0^1 f(x)\,dx,$$

and solving this linear system for p and q we get

$$p = \int_0^1 (12x - 6)f(x)\,dx,$$
$$q = \int_0^1 (4 - 6x)f(x)\,dx.$$

The following example concerns the approximation of a function by a **trigonometric polynomial**. Such approximations form the basis for the study of **Fourier series**, which are of fundamental importance in the solution of boundary-value problems for the Laplace, heat, and wave equations and other partial differential equations that arise in applied mathematics.

Example 3 Use a least squares integral to approximate $f(x)$ by the sum

$$\sum_{k=1}^{n} b_k \sin kx$$

on the interval $0 \le x \le \pi$.

Solution We want to choose the constants to minimize

$$I = \int_0^\pi \left(f(x) - \sum_{k=1}^{n} b_k \sin kx \right)^2 dx.$$

For each $1 \le j \le n$ we have

$$0 = \frac{\partial I}{\partial b_j} = -2 \int_0^\pi \left(f(x) - \sum_{k=1}^{n} b_k \sin kx \right) \sin jx\, dx.$$

Thus,

$$\sum_{k=1}^{n} b_k \int_0^\pi \sin kx \sin jx\, dx = \int_0^\pi f(x) \sin jx\, dx.$$

However, if $j \neq k$, then $\sin kx \sin jx$ is an even function, so that

$$\begin{aligned} \int_0^\pi \sin kx \sin jx\, dx &= \frac{1}{2} \int_{-\pi}^{\pi} \sin kx \sin jx\, dx \\ &= \frac{1}{4} \int_{-\pi}^{\pi} \big(\cos(k-j)x - \cos(k+j)x \big)\, dx = 0. \end{aligned}$$

If $j = k$, then we have

$$\int_0^\pi \sin^2 jx\, dx = \frac{1}{2} \int_0^\pi (1 - \cos 2jx)\, dx = \frac{\pi}{2},$$

so that

$$b_j = \frac{2}{\pi} \int_0^\pi f(x) \sin jx\, dx.$$

■

Remark The series

$$\sum_{k=1}^{\infty} b_k \sin kx, \quad \text{where} \quad b_k = \frac{2}{\pi} \int_0^\pi f(x) \sin kx\, dx, \quad k = 1, 2, \ldots,$$

is called the **Fourier sine series** representation of $f(x)$ on the interval $]0, \pi[$. If f is continuous on $[0, \pi]$, it can be shown that

$$\lim_{n \to \infty} \int_0^\pi \left(f(x) - \sum_{k=1}^{n} b_k \sin kx \right)^2 dx = 0,$$

but more than just continuity is required of f to ensure that this Fourier sine series converges to $f(x)$ at each point of $]0, \pi[$. Such questions are studied in *harmonic analysis*. Similarly, the series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx, \quad \text{where} \quad a_k = \frac{2}{\pi}\int_0^{\pi} f(x) \cos kx\, dx, \quad k = 0, 1, 2, \ldots,$$

is called the **Fourier cosine series** representation of $f(x)$ on the interval $]0, \pi[$.

$$(f + g)(x) = f(x) + g(x), \qquad (cf)(x) = cf(x),$$

and with the "dot product" defined as

$$f \bullet g = \int_0^{\pi} f(x)g(x)\, dx,$$

then the functions $e_k(x) = \sqrt{2/\pi}\, \sin kx$ form a "basis." As shown in the example above, $e_j \bullet e_j = 1$, and if $k \neq j$, then $e_k \bullet e_j = 0$. Thus these "basis vectors" are "mutually perpendicular unit vectors." The Fourier sine coefficients b_j of a function f are the components of f with respect to that basis.

Exercises 13.4

1. A generator is to be installed in a factory to supply power to n machines located at positions (x_i, y_i), $i = 1, 2, \ldots, n$. Where should the generator be located to minimize the sum of the squares of its distances from the machines?
2. The relationship $y = ax^2$ is known to hold between certain variables. Given the experimental data (x_i, y_i), $i = 1, 2, \ldots, n$, determine a value for a by the method of least squares.
3. Repeat Exercise 2 but with the relationship $y = ae^x$.
4. Use the method of least squares to find the plane $z = ax + by + c$ that best fits the data (x_i, y_i, z_i), $i = 1, 2, \ldots, n$.
5. Repeat Exercise 4 using a vector projection argument instead of the method of least squares.

In Exercises 6–11, show how to adapt linear regression to determine the two parameters p and q so that the given relationship fits the experimental data (x_i, y_i), $i = 1, 2, \ldots, n$. In which of these situations are the values of p and q obtained identical to those obtained by direct application of the method of least squares with no change of variable?

6. $y = p + qx^2$
7. $y = pe^{qx}$
8. $y = \ln(p + qx)$
9. $y = px + qx^2$
10. $y = \sqrt{px + q}$
11. $y = pe^x + qe^{-x}$
12. Find the parabola of the form $y = p + qx^2$ that best fits the data $(x, y) = (1, 0.11)$, $(2, 1.62)$, $(3, 4.07)$, $(4, 7.55)$, $(6, 17.63)$, and $(7, 24.20)$. No value of y was measured at $x = 5$. What value would you predict at this point?
13. Use the method of least squares to find constants a, b, and c so that the relationship $y = ax^2 + bx + c$ best describes the experimental data (x_i, y_i), $i = 1, 2, \ldots, n$, $(n \geq 3)$. How is this situation interpreted in terms of vector projection?
14. How can the result of Exercise 13 be used to fit a curve of the form $y = pe^x + q + re^{-x}$ through the same data points?
15. Find the value of the constant a for which the function $f(x) = ax^2$ best approximates the function $g(x) = x^3$ on the interval $[0, 1]$, in the sense that the integral

$$I = \int_0^1 \big(f(x) - g(x)\big)^2\, dx$$

is minimized. What is the minimum value of I?

16. Find a to minimize $I = \int_0^{\pi} \big(ax(\pi - x) - \sin x\big)^2\, dx$. What is the minimum value of the integral?

17. Repeat Exercise 15 with the function $f(x) = ax^2 + b$ and the same g. Find a and b.

$*$ **18.** Find a, b, and c to minimize $\int_0^1 (x^3 - ax^2 - bx - c)^2\,dx$. What is the minimum value of the integral?

$*$ **19.** Find a and b to minimize $\int_0^{\pi} (\sin x - ax^2 - bx)^2\,dx$.

20. Find a, b, and c to minimize the integral

$$J = \int_{-1}^{1} \left(x - a\sin \pi x - b\sin 2\pi x - c\sin 3\pi x\right)^2 dx.$$

$*$ **21.** Find constants a_j, $j = 0, 1, \ldots, n$, to minimize

$$\int_0^{\pi} \left(f(x) - \frac{a_0}{2} - \sum_{k=1}^{n} a_k \cos kx\right)^2 dx.$$

22. Find the Fourier sine series for the function $f(x) = x$ on $0 < x < \pi$. Assuming the series does converge to x on the interval $]0, \pi[$, to what function would you expect the series to converge on $]-\pi, 0[$?

23. Repeat Exercise 22 but obtaining instead a Fourier cosine series.

24. Suppose $x_1, x_2, \ldots, x_n$ satisfy $x_i \le x_j$ whenever $i < j$. Find x that minimizes $\sum_{i=1}^{n} |x - x_i|$. Treat the cases n odd and n even separately. For what values of n is x unique? *Hint:* use no calculus in this problem.

13.5 Parametric Problems

In this section we will briefly examine three unrelated situations in which one wants to differentiate a function with respect to a parameter rather than one of the basic variables of the function. Such situations arise frequently in mathematics and its applications.

Differentiating Integrals with Parameters

The Fundamental Theorem of Calculus shows how to differentiate a definite integral with respect to the upper limit of integration:

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x).$$

We are going to look at a different problem about differentiating integrals. If the integrand of a definite integral also depends on variables other than the variable of integration, then the integral will be a function of those other variables. How are we to find the derivative of such a function? For instance, consider the function $F(x)$ defined by

$$F(x) = \int_a^b f(x,t)\,dt.$$

We would like to be able to calculate $F'(x)$ by taking the derivative inside the integral:

$$F'(x) = \frac{d}{dx}\int_a^b f(x,t)\,dt = \int_a^b \frac{\partial}{\partial x} f(x,t)\,dt.$$

Observe that we use "d/dx" outside the integral and "$\partial/\partial x$" inside; this is because the integral is a function of x only, but the integrand f is a function of both x and t. If the integrand depends on more than one parameter, then partial derivatives would be needed inside and outside the integral:

$$\frac{\partial}{\partial x}\int_a^b f(x,y,t)\,dt = \int_a^b \frac{\partial}{\partial x} f(x,y,t)\,dt.$$

The operation of taking a derivative with respect to a parameter inside the integral, or *differentiating through the integral*, as it is usually called, seems plausible. We differentiate sums term by term, and integrals are the limits of sums. However, both the differentiation and integration operations involve the taking of limits (limits of Newton quotients for derivatives, limits of Riemann sums for integrals). Differentiating through the integral requires changing the order in which the two limits are taken and, therefore, requires justification.

We have already seen another example of change of order of limits. When we assert that two mixed partial derivatives with respect to the same variables are equal,

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

we are, in fact, saying that limits corresponding to differentiation with respect to x and y can be taken in either order with the same result. This is not true in general; we proved it under the assumption that both of the mixed partials were *continuous*. (See Theorem 1 and Exercise 16 of Section 12.4.) In general, some assumptions are required to justify the interchange of limits. The following theorem gives one set of conditions that justify the interchange of limits involved in differentiating through the integral.

THEOREM 5

Differentiating through an integral

Suppose that for every x satisfying $c < x < d$ the following conditions hold:

(i) the integrals

$$\int_a^b f(x,t)\,dt \qquad \text{and} \qquad \int_a^b f_1(x,t)\,dt$$

both exist (either as proper or convergent improper integrals).

(ii) $f_{11}(x,t)$ exists and satisfies

$$|f_{11}(x,t)| \le g(t), \qquad a < t < b,$$

where

$$\int_a^b g(t)\,dt = K < \infty.$$

Then for each x satisfying $c < x < d$ we have

$$\frac{d}{dx}\int_a^b f(x,t)\,dt = \int_a^b \frac{\partial}{\partial x} f(x,t)\,dt.$$

PROOF Let

$$F(x) = \int_a^b f(x,t)\,dt.$$

If $c < x < d$, $h \ne 0$, and $|h|$ is sufficiently small that $c < x+h < d$, then, by Taylor's Formula,

$$f(x+h,t) = f(x,t) + hf_1(x,t) + \frac{h^2}{2} f_{11}(x+\theta h, t)$$

for some θ between 0 and 1. Therefore,

$$\begin{aligned}
&\left|\frac{F(x+h)-F(x)}{h} - \int_a^b f_1(x,t)\,dt\right| \\
&\qquad = \left|\int_a^b \frac{f(x+h,t)-f(x,t)}{h}\,dt - \int_a^b f_1(x,t)\,dt\right| \\
&\qquad \le \int_a^b \left|\frac{f(x+h,t)-f(x,t)}{h} - f_1(x,t)\right| dt \\
&\qquad = \int_a^b \left|\frac{h}{2} f_{11}(x+\theta h,t)\right| dt \\
&\qquad \le \frac{h}{2}\int_a^b g(t)\,dt = \frac{Kh}{2} \;\to\; 0 \text{ as } h \to 0.
\end{aligned}$$

Therefore,

$$F'(x) = \lim_{h\to 0} \frac{F(x+h)-F(x)}{h} = \int_a^b f_1(x,t)\,dt,$$

which is the desired result.

Remark It can be shown that the conclusion of Theorem 5 also holds under the sole assumption that $f_1(x,t)$ is continuous on the *closed, bounded* rectangle $c \le x \le d$, $a \le t \le b$. We cannot prove this here; the proof depends on a subtle property called *uniform continuity* possessed by continuous functions on closed bounded sets in $\mathbb{R}^n$. In any event, Theorem 5 is more useful for our purposes because it allows for improper integrals.

Example 1 Evaluate $\displaystyle\int_0^\infty t^n e^{-t}\,dt$.

Solution Starting with the convergent improper integral

$$\int_0^\infty e^{-s}\,ds = \lim_{R\to\infty} \left.\frac{e^{-s}}{-1}\right|_0^R = \lim_{R\to\infty}(1-e^{-R}) = 1,$$

we introduce a parameter by substituting $s = xt$, $ds = x\,dt$ (where $x > 0$) and get

$$\int_0^\infty e^{-xt}\,dt = \frac{1}{x}.$$

Now differentiate n times (each resulting integral converges):

$$\begin{aligned}
\int_0^\infty -t\,e^{-xt}\,dt &= -\frac{1}{x^2}, \\
\int_0^\infty (-t)^2 e^{-xt}\,dt &= (-1)^2\frac{2}{x^3}, \\
&\;\;\vdots \\
\int_0^\infty (-t)^n e^{-xt}\,dt &= (-1)^n\frac{n!}{x^{n+1}}.
\end{aligned}$$

Putting $x = 1$ we get

$$\int_0^\infty t^n e^{-t}\, dt = n!.$$

Note that this result could be obtained by integration by parts (n times) or a reduction formula. This method is a little easier.

Remark The reader should check that the function $f(x, t) = t^k e^{-xt}$ satisfies the conditions of Theorem 5 for $x > 0$ and $k \geq 0$. We will normally not make a point

$$\frac{\partial F}{\partial x} = -\int_0^\infty e^{-xt}\, dt = -\frac{1}{x} \quad \text{and} \quad \frac{\partial F}{\partial y} = \int_0^\infty e^{-yt}\, dt = \frac{1}{y}.$$

It follows that

$$F(x, y) = -\ln x + C_1(y) \quad \text{and} \quad F(x, y) = \ln y + C_2(x).$$

Comparing these two formulas for F, we are forced to conclude that $C_1(y) = \ln y + C$ for some constant C. Therefore,

$$F(x, y) = \ln y - \ln x + C = \ln\frac{y}{x} + C.$$

Since $F(1, 1) = 0$, we must have $C = 0$ and $F(x, y) = \ln(y/x)$.

Remark We can combine Theorem 5 and the Fundamental Theorem of Calculus to differentiate an integral with respect to a parameter that appears in the limits of integration as well as in the integrand. If

$$F(x, b, a) = \int_a^b f(x, t)\, dt,$$

then, by the Chain Rule,

$$\frac{d}{dx} F\big(x, b(x), a(x)\big) = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial b}\frac{db}{dx} + \frac{\partial F}{\partial a}\frac{da}{dx}.$$

Accordingly, we have

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f(x, t)\, dt$$
$$= \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x, t)\, dt + f\big(x, b(x)\big)b'(x) - f\big(x, a(x)\big)a'(x).$$

We require that $a(x)$ and $b(x)$ be differentiable at x, and for the application of Theorem 5, that $a \leq a(x) \leq b$ and $a \leq b(x) \leq b$ for all x satisfying $c < x < d$.

Example 3 Solve the *integral equation*

$$f(x) = a - \int_b^x (x-t)f(t)\,dt.$$

Solution Assume, for the moment, that the equation has a sufficiently well behaved solution to allow for differentiation through the integral. Differentiating twice, we get

$$f'(x) = -(x-x)f(x) - \int_b^x f(t)\,dt = -\int_b^x f(t)\,dt,$$
$$f''(x) = -f(x).$$

The latter equation is the differential equation of simple harmonic motion. Observe that the given equation for f and that for f' imply the initial conditions

$$f(b) = a \qquad \text{and} \qquad f'(b) = 0.$$

Accordingly, we write the general solution of $f''(x) = -f(x)$ in the form

$$f(x) = A\cos(x-b) + B\sin(x-b).$$

The initial conditions then imply $A = a$ and $B = 0$, so the required solution is $f(x) = a\cos(x-b)$. Finally, we note that this function is indeed smooth enough to allow the differentiations through the integral and is, therefore, the solution of the given integral equation. (If you wish, verify it in the integral equation.) ■

Envelopes

An equation $f(x, y, c) = 0$ that involves a parameter c as well as the variables x and y represents a family of curves in the xy-plane. Consider, for instance, the family

$$f(x, y, c) = \frac{x}{c} + cy - 2 = 0.$$

This family consists of straight lines with intercepts $(2c, 2/c)$ on the coordinate axes. Several of these lines are sketched in Figure 13.20. It appears that there is a curve to which all these lines are tangent. This curve is called the *envelope* of the family of lines.

In general, a curve $\mathcal{C}$ is called the **envelope** of the family of curves with equations $f(x, y, c) = 0$ if, for each value of c, the curve $f(x, y, c) = 0$ is tangent to $\mathcal{C}$ at some point depending on c.

For the family of lines in Figure 13.20 it appears that the envelope may be the rectangular hyperbola $xy = 1$. We will verify this after developing a method for determining the equation of the envelope of a family of curves. We assume that the function $f(x, y, c)$ has continuous first partials and that the envelope is a smooth curve.

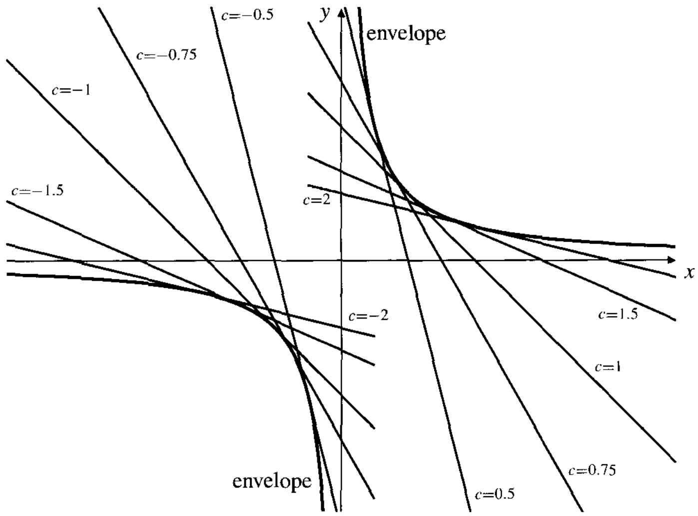

Figure 13.20 A family of straight lines and their envelope

BEWARE
This is a subtle argument. Take your time and try to understand each step in the development.

For each c, the curve $f(x, y, c) = 0$ is tangent to the envelope at a point (x, y) that depends on c. Let us express this dependence in the explicit form $x = g(c)$, $y = h(c)$; these equations are parametric equations of the envelope. Since (x, y) lies on the curve $f(x, y, c) = 0$, we have

$$f\big(g(c), h(c), c\big) = 0.$$

Differentiating this equation with respect to c, we obtain

$$f_1 g'(c) + f_2 h'(c) + f_3 = 0, \qquad (*)$$

where the partials of f are evaluated at $\big(g(c), h(c), c\big)$.

The slope of the curve $f(x, y, c) = 0$ at $\big(g(c), h(c), c\big)$ can be obtained by differentiating its equation implicitly with respect to x:

$$f_1 + f_2 \frac{dy}{dx} = 0.$$

On the other hand, the slope of the envelope $x = g(c)$, $y = h(c)$ at that point is $dy/dx = h'(c)/g'(c)$. Since the curve and the envelope are tangent at $f\big(g(c), h(c), c\big)$, these slopes must be equal. Therefore,

$$f_1 + f_2 \frac{h'(c)}{g'(c)} = 0, \qquad \text{so} \qquad f_1 g'(c) + f_2 h'(c) = 0.$$

Combining this with equation $(*)$ we get $f_3(x, y, c) = 0$ at all points of the envelope.

The equation of the envelope can be found by eliminating c between the two equations

$$f(x, y, c) = 0 \qquad \text{and} \qquad \frac{\partial}{\partial c} f(x, y, c) = 0.$$

Example 4 Find the envelope of the family of straight lines

$$f(x, y, c) = \frac{x}{c} + cy - 2 = 0.$$

Solution We eliminate c between the equations

$$f(x, y, c) = \frac{x}{c} + cy - 2 = 0 \quad \text{and} \quad f_3(x, y, c) = -\frac{x}{c^2} + y = 0.$$

These equations can be easily solved and give $x = c$ and $y = 1/c$. Hence, they imply that the envelope is $xy = 1$, as we conjectured earlier.

Example 5 Find the envelope of the family of circles

$$(x - c)^2 + y^2 = c.$$

Solution Here, $f(x, y, c) = (x - c)^2 + y^2 - c$. The equation of the envelope is obtained by eliminating c from the pair of equations

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

From the second equation, $x = c - \frac{1}{2}$, and then from the first, $y^2 = c - \frac{1}{4}$. Hence, the envelope is the parabola

$$x = y^2 - \frac{1}{4}.$$

This envelope and some of the circles in the family are sketched in Figure 13.21.

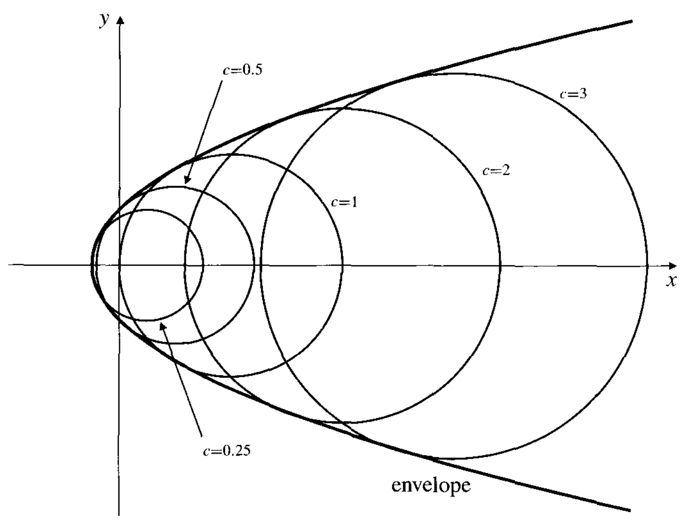

Figure 13.21 Circles $(x - c)^2 + y^2 = c$ and their envelope

A similar technique can be used to find the envelope of a family of surfaces. This will be a surface tangent to each member of the family.

Example 6 **(The Mach cone)** Suppose that sound travels at speed c in still air and that a supersonic aircraft is travelling at speed $v > c$ along the x-axis, so that its position at time t is $(vt, 0, 0)$. Find the envelope at time t of the sound waves created by the aircraft at previous times. See Figure 13.22.

Solution The sound created by the aircraft at time $\tau < t$ spreads out as a spherical wave front at speed c. The centre of this wave front is $(v\tau, 0, 0)$, the position of the aircraft at time τ. At time t the radius of this wave front is $c(t-\tau)$, so its equation is

$$f(x, y, z, \tau) = (x - v\tau)^2 + y^2 + z^2 - c^2(t-\tau)^2 = 0. \qquad (*)$$

At time t the envelope of all these wave fronts created at earlier times τ is obtained by eliminating the parameter τ from the above equation and the equation

$$\frac{\partial}{\partial \tau} f(x, y, z, \tau) = -2v(x - v\tau) + 2c^2(t-\tau) = 0.$$

Solving this latter equation for τ, we get $\tau = \dfrac{vx - c^2 t}{v^2 - c^2}$. Thus,

$$\begin{aligned} x - v\tau &= x - \frac{v^2 x - vc^2 t}{v^2 - c^2} = \frac{c^2}{v^2 - c^2}(vt - x) \\ t - \tau &= t - \frac{vx - c^2 t}{v^2 - c^2} = \frac{v}{v^2 - c^2}(vt - x). \end{aligned}$$

We substitute these two expressions into equation $(*)$ to eliminate τ:

$$\begin{aligned} &\frac{c^4}{(v^2 - c^2)^2}(vt - x)^2 + y^2 + z^2 - \frac{c^2 v^2}{(v^2 - c^2)^2}(vt - x)^2 = 0 \\ &y^2 + z^2 = \frac{c^2}{(v^2 - c^2)^2}(v^2 - c^2)(vt - x)^2 = \frac{c^2}{v^2 - c^2}(vt - x)^2. \end{aligned}$$

The envelope is the cone

$$x = vt - \frac{\sqrt{v^2 - c^2}}{c}\sqrt{y^2 + z^2},$$

which extends backward in the x direction from its vertex at $(vt, 0, 0)$, the position of the aircraft at time t. This is called the **Mach cone**. The sound of the aircraft cannot be heard at any point until the cone reaches that point. ■

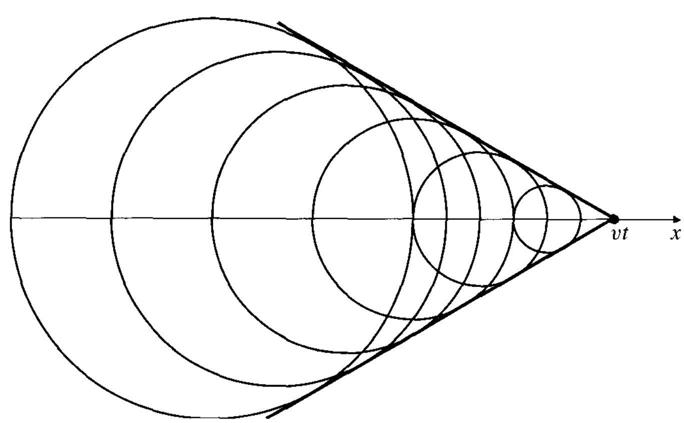

Figure 13.22 The Mach cone

Equations with Perturbations

In applied mathematics one frequently encounters intractable equations for which at least approximate solutions are desired. Sometimes such equations result from adding an extra term to what would otherwise be a simple and easily solved equation. This extra term is called a **perturbation** of the simpler equation. Often the perturbation has coefficient smaller than the other terms in the equation, that is, it is a **small perturbation**. When this is the case you can find approximate solutions to the perturbed equation by replacing the small coefficient by a parameter and calculating Maclaurin polynomials in that parameter. One example should serve to clarify the method.

Example 7 Find an approximate solution of the equation

$$y + \frac{1}{50}\ln(1+y) = x^2.$$

Solution Without the logarithm term the equation would clearly have the solution $y = x^2$. Let us replace the coefficient $1/50$ with the parameter ϵ and look for a solution $y = y(x, \epsilon)$ to the equation

$$y + \epsilon\ln(1+y) = x^2 \qquad (*)$$

in the form

$$y = y(x,\epsilon) = y(x,0) + \epsilon y_\epsilon(x,0) + \frac{\epsilon^2}{2!}y_{\epsilon\epsilon}(x,0) + \cdots,$$

where the subscripts ϵ denote derivatives with respect to ϵ. We shall calculate the terms up to second order in ϵ. Evidently $y(x,0) = x^2$. Differentiating equation (*) twice with respect to ϵ and evaluating the results at $\epsilon = 0$, we obtain

$$\frac{\partial y}{\partial \epsilon} + \ln(1+y) + \frac{\epsilon}{1+y}\frac{\partial y}{\partial \epsilon} = 0,$$

$$\frac{\partial^2 y}{\partial \epsilon^2} + \frac{2}{1+y}\frac{\partial y}{\partial \epsilon} + \epsilon\frac{\partial}{\partial \epsilon}\left(\frac{1}{1+y}\frac{\partial y}{\partial \epsilon}\right) = 0,$$

$$y_\epsilon(x,0) = -\ln(1+x^2),$$

$$y_{\epsilon\epsilon}(x,0) = \frac{2}{1+x^2}\ln(1+x^2).$$

Hence,

$$y(x,\epsilon) = x^2 - \epsilon\ln(1+x^2) + \frac{\epsilon^2}{1+x^2}\ln(1+x^2) + \cdots,$$

and the given equation has the approximate solution

$$y \approx x^2 - \frac{\ln(1+x^2)}{50} + \frac{\ln(1+x^2)}{2{,}500(1+x^2)}.$$

■

Similar perturbation techniques can be used for systems of equations and for differential equations.

Exercises 13.5

1. Let $F(x) = \int_0^1 t^x\,dt = \dfrac{1}{x+1}$ for $x > -1$. By repeated differentiation of F evaluate the integral

$$\int_{-\infty}^{1} e^{[illegible]}\,dt = \sqrt{\pi},$$

and differentiating with respect to x, evaluate

$$\int_{-\infty}^{\infty} t^2 e^{-t^2}\,dt \quad \text{and} \quad \int_{-\infty}^{\infty} t^4 e^{-t^2}\,dt.$$

3. Evaluate $\displaystyle\int_{-\infty}^{\infty} \frac{e^{-xt^2} - e^{-yt^2}}{t^2}\,dt$ for $x > 0$, $y > 0$.

4. Evaluate $\displaystyle\int_0^1 \frac{t^x - t^y}{\ln t}\,dt$ for $x > -1$, $y > -1$.

5. Given that $\displaystyle\int_0^\infty e^{-xt}\sin t\,dt = \frac{1}{1+x^2}$ for $x > 0$ (which can be shown by integration by parts), evaluate

$$\int_0^\infty te^{-xt}\sin t\,dt \quad \text{and} \quad \int_0^\infty t^2 e^{-xt}\sin t\,dt.$$

* **6.** Referring to Exercise 5, for $x > 0$ evaluate

$$F(x) = \int_0^\infty e^{-xt}\frac{\sin t}{t}\,dt.$$

Show that $\lim_{x\to\infty} F(x) = 0$ and hence evaluate the integral

$$\int_0^\infty \frac{\sin t}{t}\,dt = \lim_{x\to 0} F(x).$$

7. Evaluate $\displaystyle\int_0^\infty \frac{dt}{x^2+t^2}$ and use the result to help you evaluate

$$\int_0^\infty \frac{dt}{(x^2+t^2)^2} \quad \text{and} \quad \int_0^\infty \frac{dt}{(x^2+t^2)^3}.$$

* **8.** Evaluate $\displaystyle\int_0^x \frac{dt}{x^2+t^2}$ and use the result to help you evaluate

$$\int_0^x \frac{dt}{(x^2+t^2)^2} \quad \text{and} \quad \int_0^x \frac{dt}{(x^2+t^2)^3}.$$

9. Find $f^{(n+1)}(a)$ if $f(x) = 1 + \displaystyle\int_a^x (x-t)^n f(t)\,dt$.

Solve the integral equations in Exercises 10–12.

[illegible]

Find the envelopes of the families of curves in Exercises 13–18.

13. $y = 2cx - c^2$

14. $y - (x-c)\cos c = \sin c$

15. $x\cos c + y\sin c = 1$

16. $\dfrac{x}{\cos c} + \dfrac{y}{\sin c} = 1$

17. $y = c + (x-c)^2$

18. $(x-c)^2 + (y-c)^2 = 1$

19. Does every one-parameter family of curves in the plane have an envelope? Try to find the envelope of $y = x^2 + c$.

20. For what values of k does the family of curves $x^2 + (y-c)^2 = kc^2$ have an envelope?

21. Try to find the envelope of the family $y^3 = (x+c)^2$. Are the curves of the family tangent to the envelope? What have you actually found in this case? Compare with Example 3 of Section 13.3.

* **22.** Show that if a two-parameter family of surfaces $f(x, y, z, \lambda, \mu) = 0$ has an envelope, then the equation of that envelope can be obtained by eliminating λ and μ from the three equations

$$f(x, y, z, \lambda, \mu) = 0,$$
$$\frac{\partial}{\partial\lambda} f(x, y, z, \lambda, \mu) = 0,$$
$$\frac{\partial}{\partial\mu} f(x, y, z, \lambda, \mu) = 0.$$

23. Find the envelope of the two-parameter family of planes

$$x\sin\lambda\cos\mu + y\sin\lambda\sin\mu + z\cos\lambda = 1.$$

24. Find the envelope of the two-parameter family of spheres

$$(x-\lambda)^2 + (y-\mu)^2 + z^2 = \frac{\lambda^2+\mu^2}{2}.$$

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

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

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

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

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

$*$ **29.** Use perturbation methods to find approximate values for x and y from the system

$$x + 2y + \frac{1}{100}e^{-x} = 3, \quad x - y + \frac{1}{100}e^{-y} = 0.$$

Calculate all terms up to second order in $\epsilon = 1/100$.

13.6 Newton's Method

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

$$f(r) = 0.$$

Such a root is called a **zero** of the function f. In Section 4.6 we introduced Newton's Method, a simple but powerful method for determining roots of functions that are sufficiently smooth. The method involves *guessing* an approximate value x_0 for a root r of the function f, and then calculating successive approximations $x_1, x_2, \dots$, using the formula

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

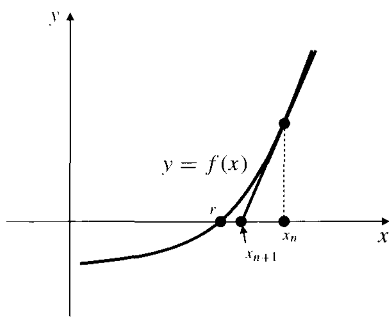

Figure 13.23 x_{n+1} is the x-intercept of the tangent at x_n

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

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

(See Figure 13.23.) The x-intercept, x_{n+1}, of this line is determined by setting $y = 0$, $x = x_{n+1}$ in this equation, so is given by the formula in the shaded box above.

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

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

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

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

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

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

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

Solving these two equations for x_1 and y_1, we obtain

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

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

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

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

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

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

We stop when the desired accuracy has been achieved.

Example 1 Find the root of the system

$$x(1 + y^2) - 1 = 0, \qquad y(1 + x^2) - 2 = 0$$

with sufficient accuracy to ensure that the left sides of the equations vanish to the sixth decimal place.

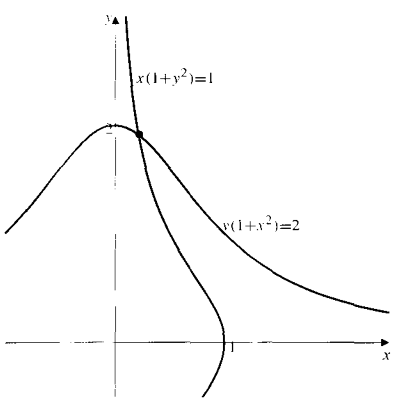

Figure 13.24 The two graphs intersect near (0.2, 1.8)

Solution A sketch of the graphs of the two equations (see Figure 13.24) in the xy-plane indicates that the system has only one root near the point (0.2, 1.8). Application of Newton's Method requires successive computations of the quantities

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

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

Table 1. Root near (0.2, 1.8)

n	x_n	y_n	$f(x_n, y_n)$	$g(x_n, y_n)$
0	0.200000	1.800000	−0.152000	−0.128000
1	0.216941	1.911349	0.009481	0.001303
2	0.214827	1.911779	−0.000003	0.000008
3	0.214829	1.911769	0.000000	0.000000

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

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

Implementing Newton's Method Using a Spreadsheet

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

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

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

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

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

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

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

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

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

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

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

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

Exercises 13.6

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

1. $y - e^x = 0, \quad x - \sin y = 0$
2. $x^2 + y^2 - 1 = 0, \quad y - e^x = 0$ (two solutions)
3. $x^4 + y^2 - 16 = 0, \quad xy - 1 = 0$ (four solutions)
4. $x^2 - xy + 2y^2 = 10, \quad x^3y^2 = 2$ (four solutions)
5. $y - \sin x = 0, \quad x^2 + (y + 1)^2 - 2 = 0$ (two solutions)
6. $\sin x + \sin y - 1 = 0, \quad y^2 - x^3 = 0$ (two solutions)
7. * Write formulas for obtaining successive Newton's Method approximations to a solution of the system

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

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

8. Use the formulas from Exercise 7 to find the first octant intersection point of the surfaces $y^2 + z^2 = 3$, $x^2 + z^2 = 2$, and $x^2 - z = 0$.
9. The equations $y - x^2 = 0$ and $y - x^3 = 0$ evidently have the solutions $x = y = 0$ and $x = y = 1$. Try to obtain these solutions using the two-variable form of Newton's Method with starting values (a) $x_0 = y_0 = 0.1$ and (b) $x_0 = y_0 = 0.9$.
How many iterations are required to obtain 6-decimal-place accuracy for the appropriate solution in each case?
How do you account for the difference in the behaviour of Newton's Method for these equations near $(0, 0)$ and $(1, 1)$?

13.7 Calculations with Maple

The calculations involved in finding extreme values of functions of several variables can be very lengthy even if the number of variables is small. In particular, even locating critical points of a function of n variables involves solving a system of n (usually nonlinear) equations in n unknowns. In such situations the effective use of a computer algebra system like Maple can be very helpful. In this optional section we present a few examples of how Maple can be used to find and classify critical points and solve extreme-value problems.[1]

As observed previously, Maple has many functions for processing vectors and matrices. Many of these are only available in the **linalg** package, so we must remember to include the Maple command `with(linalg):` when using these functions.

[1] The author is grateful to his colleague, Professor Peter Kiernan, for suggesting the procedures **newtroot** and **newtcp** presented in this section.

Solving Systems of Equations

In the previous section we considered the 2-variable (and n-variable) versions of Newton's Method for approximating solutions of systems of equations. The method applies to a system of n equations in n unknowns, that is, to an equation of the form

$$\mathbf{F}(x_1, x_2, \ldots, x_n) = \mathbf{0},$$

where $\mathbf{F}$ is an n-vector-valued function.

Below we investigate a Maple procedure, **newtroot**, which automates the use of Newton's Method to solve such a vector equation, provided a good starting value can be found. This procedure can be found in the file **newton.def**, which is available on the website `www.pearsoned.ca/text/adams_calc`. The file can be read into a Maple session with the command

```
> read "newton.def";
```

The file also loads the linalg package, parts of which are needed by `newtroot` and its companion procedure `newtcp`.

The procedure `newtroot` searches for a solution of $\mathbf{F} = \mathbf{0}$ using $\mathbf{v}$ as an initial approximation. Here, $\mathbf{F}$ is an n-vector function of n real variables, and $\mathbf{v}$ is a list of starting values of those variables. However, if $n = 1$, then $\mathbf{F}$ should just be a scalar function and $\mathbf{v}$ should be a number. The procedure requires four arguments: the name of the function $\mathbf{F}$, the starting value or list $\mathbf{v}$, the maximum number m of iterations to allow before declaring failure to find a root, and the maximum tolerated norm, *tol*, of the difference between two consecutive approximations for success. (The norm of a vector is the maximum of the absolute values of its components.) At each iteration the procedure outputs a line with the iteration number, the current approximation, the value of $\mathbf{F}$ there, and the error (the norm of the difference of the last two approximations). When the error is less than the tolerance, the procedure exits, returning the final approximation to the root as its output. If the error continues to exceed the tolerance, the procedure exits after m iterations, declaring failure and returning the most recent approximation to the root as its output.

A listing of "newtroot" can be found in Appendix V.

Example 1 Solve the system

$$\begin{cases} x^2 + y^4 = 1 \\ z = x^3 y \\ e^x = 2y - z. \end{cases}$$

Solution Assuming that the above procedure has been read into a Maple session as described above, we proceed to define a vector-valued function of 3 variables:

```
> F := (x,y,z) -> vector([x^2+y^4-1,
>  z-x^3*y, exp(x)-2*y+z]);
```

This is the procedure we will feed to `newtroot`. But what should our starting value $\mathbf{v}$ be? The first equation cannot be satisfied by any points outside the square $-1 \le x \le 1$, $-1 \le y \le 1$, so we need only consider starting values for x and y inside this square. The second equation then forces z to lie between -1 and 1 also. We could just try many starting points that satisfy these conditions and see what we get using newtroot. Alternatively, we can make several implicit plots of the three equations for fixed values of z between -1 and 1, looking for cases where the three curves come close to having a common intersection point:

```
> with(plots):
> for z from -1 by .2 to 1 do print('z =' z);
> implicitplot({x^2+y^4-1, z-x^3*y, exp(x)-2*y+z},
> x=-1.5 .. 1.5, y=-1.5 .. 1.5) od;
```

These commands produce 11 graphs of the three equations, considered as depending on x and y for z values ranging from -1 to 1 in steps of 0.2. Two of them are shown in Figure 13.25 and Figure 13.26. They correspond to $z = -0.2$ and $z = 0.2$ and indicate that the three equations likely have solutions near $(-1, 0.2, -0.2)$ and $(0.5, 0.9, 0.2)$. We run `newtroot` with these starting values, allowing up to 10 iterations to try and achieve zero values for the components of F to within a tolerance of 0.000005. To limit the output, we set the Maple variable *Digits* to 6:

```
> Digits := 6; newtroot(F,[-1,.2,-.2],10,.000005);
```

$$1, [-1.00044, .122484, -.122748], [.00111, -.000102, .2\,10^{-5}], .077516$$

$$2, [.999885, .122655, -.122613], [-.4\,10^{-5}, 0, -.1\,10^{-5}], .000555$$

$$3, [-.999887, .122654, -.122613], [0, -.1\,10^{-5}, 0], .2\,10^{-5}$$

$$[-.999887, .122654, -.122613]$$

```
> newtroot(F,[.5,.9,.2],10,.000005);
```

$$1, [.533423, .920740, .137653], [.00324, -.002097, .000933], .062347$$

$$2, [.531838, .920244, .138429], [.00001, -.4\,10^{-5}, -.1\,10^{-5}], .001585$$

$$3, [.531835, .920242, .138430], [-.5\,10^{-5}, -.1\,10^{-5}, 0], .3\,10^{-5}$$

$$[.531835, .920242, .138430]$$

The last line of each output gives the appropriate solution to 6 significant digits. ∎

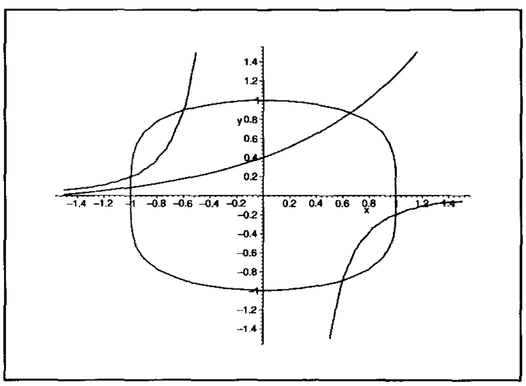

Figure 13.25 $z = -0.2$

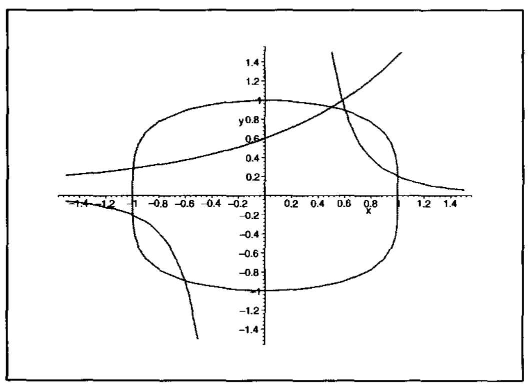

Figure 13.26 $z = 0.2$

Finding and Classifying Critical Points

Finding the critical points of a function of several variables amounts to solving the system of equations obtained by setting the components of the gradient of the function to zero. The file **newton.def** also contains a modification of the `newtroot` procedure, called **newtcp**, that mechanizes this procedure. It searches for a solution of **grad** $F(x_1, x_2, \ldots, x_n) = \mathbf{0}$ using **v** as an initial approximation. Here, F is a scalar-valued function of n real variables, and **v** is an n-vector, with the exception that if $n = 1$, then the routine assumes that **v** is also a scalar. At each iteration of Newton's Method, the procedure prints the number n of the iteration, the coordinates of the nth approximation to the critical point, the value of the function at that approximation, and the error (the norm of the difference between the last two approximations).

The procedure continues iterating until either the error is less than a prescribed tolerance *tol* or the number of iterations exceeds a prescribed maximum number m. In the former case it prints the eigenvalues of the Hessian matrix of F at the final approximation to the critical point and then exits, returning the critical point and the value of F there. Recall that having all eigenvalues negative (or positive) implies that F has a local maximum (or minimum) at the critical point. If some eigenvalues are positive and some are negative, F has a saddle point there. If the procedure fails to find a critical point within the prescribed maximum number of iterations, it reports this failure along with the value of the error at the final step and returns with the final approximation to the critical point.

A listing of "newtcp" is given in Appendix V.

Example 2 Find and classify the critical points of

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

Solution Assuming the procedure **newtcp** has been loaded into a Maple session, say, by reading in the file **newton.def**, we proceed to define f as a function of x and y as follows:

```
> f := (x,y) -> (x^2+x*y+5*y^2+x-y)*exp(-x^2 -y^2);
```

Plotting some level curves of F can suggest likely locations for critical points.

```
> with(plots):
> contourplot(F(x,y),x=-3..3, y=-3..3, grid=[50,50],
>  contours=16);
```

The contour plot (Figure 13.27) suggests that there are five critical points, three local extrema near $(0.3, 1)$, $(0, -1)$, and $(-0.6, 0.1)$ and two saddle points near $(1, 0)$ and $(-1.6, 0.2)$. We home in on each of these using `newtcp`. We limit ourselves to 6 significant digits and set the tolerance to 0.00001.

```
> Digits := 6; newtcp(f, [.3,1], 10, .00001);
```

$$1, [.275095, 1.00131], 1.57773, .024905$$

$$2, [.275058, 1.00132], 1.57773, .000037$$

$$3, [.275057, 1.00132], 1.57773, .1\,10^{-5}$$

$$\mathit{Eigenvalues}, -6.61494, -2.41894$$

$$[.275057, 1.00132], 1.57773$$

The eigenvalues indicate a local maximum value of 1.57773 at the point $(0.275057, 1.00132)$.

```
> newtcp(f, [0,-1], 10, .00001);
```

$$1, [.00456621, -.954338], 2.21553, .045662$$

$$2, [.00492090, -.955505], 2.21553, .001167$$

$$3, [.00492106, -.955505], 2.21553, .16\ 10^{-6}$$

$$\textit{Eigenvalues}, -8.54884, -3.58874$$

$$[.00492106, -.955505], 2.21553$$

The eigenvalues indicate a local maximum value of 2.21553 at the point (0.00492106, −0.955505). This is, in fact, the absolute maximum.

```
> newtcp(f, [-.6,.1], 10, .00001);
```

$$1, [-.341894, .139437], -.274764, .258106$$

$$2, [-.417293, .133079], -.283308, .075399$$

$$3, [-.421350, .132976], -.283327, .004057$$

$$4, [-.421366, .132977], -.283328, .000016$$

$$5, [-.421365, .132977], -.283328, .1\ 10^{-5}$$

$$\textit{Eigenvalues}, 2.32435, 8.90196$$

$$[-.421365, .132977], -.283328$$

The eigenvalues indicate a local minimum value of −0.283328 at the point (−0.421365, 0.132977). This is, in fact, the absolute minimum.

```
> newtcp(f, [1,0], 10, .00001);
```

$$1, [.837838, .0270270270], .762222, .162162$$

$$2, [.858275, .0210235], .762817, .020437$$

$$3, [.858435, .0207853], .762810, .0002382$$

$$4, [.858433, .0207854], .762809, .2\ 10^{-5}$$

$$\textit{Eigenvalues}, -2.84681, 3.28635$$

$$[.858433, .0207854], .76280945$$

The eigenvalues indicate a saddle point at (0.858433, 0.0207854).

```
> newtcp(f, [-1.6, .2], 10, .00001);
```

$$1, [-1.60684, .291329], .0444636, .091329$$

$$2, [-1.57946, .292991], .0445837, .02738$$

$$3, [-1.58082, .292688], .0445834, .00136$$

$$4, [-1.58082, .292686], .0445843, .2\ 10^{-5}$$

$$\textit{Eigenvalues}, -.407585, .673366$$

$$[-1.58082, .292686], .0445843$$

The eigenvalues indicate a saddle point at (−1.58082, 0.292686). ■

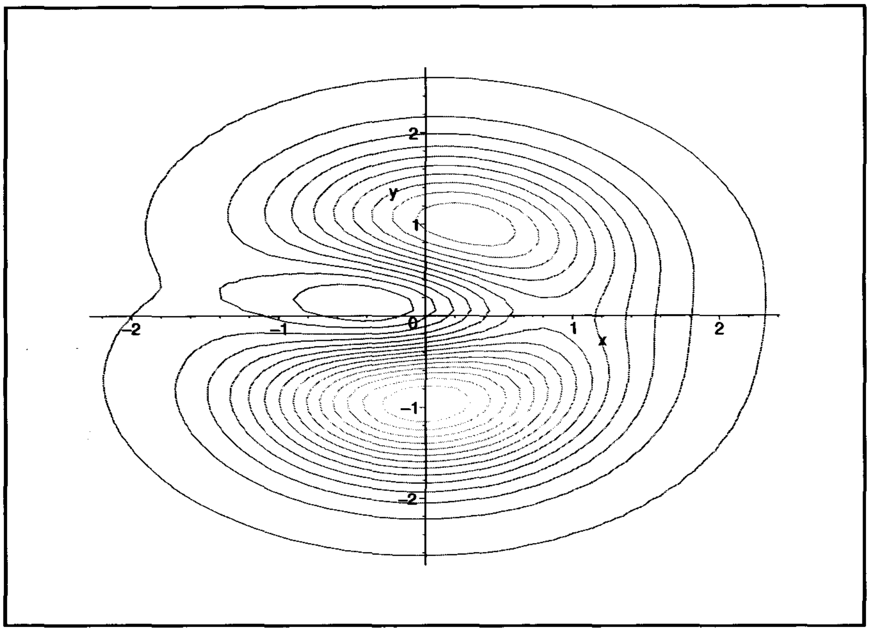

Figure 13.27 Contours of $f(x, y)$.

Remark The hardest part of using Newton's Method for large systems is determining suitable starting values for the roots or critical points. Graphical means are really only suitable for small systems (one, two, or three equations), and even then it is important to analyze the equations or functions involved for clues on where the roots or critical points may be. Here are some possibilities to consider:

1. Sometimes some of the equations will be simple enough that they can be solved for some variables and thus used to reduce the size of the system. We could have used the second equation in Example 1 to eliminate z from the first and third equations and, hence, reduced the system to two equations in two unknowns.
2. The system might result from adding a small extra term to a simpler system, the location of whose roots is known. In this case you can use those known roots as starting points.
3. Always be alert for equations limiting the possible values of some variables. For instance, in Example 1 the equation $x^2 + y^4 = 1$ limited x and y to the interval $[-1, 1]$.

Exercises 13.7

In Exercises 1–2, solve the given systems of equations using `newtonroots`. Quote the roots to 5 significant figures and use a tolerance of 0.00001. Be alert for simple substitutions that can reduce the number of equations that must be fed to `newtroot`.

1. $\begin{cases} x^2 + y^2 + z^2 = 1 \\ z = xy \\ 6xz = 1 \end{cases}$

2. $\begin{cases} x^4 + y^2 + z^2 = 1 \\ y = \sin z \\ z + z^3 + z^4 = x + y \end{cases}$

In Exercises 3–6, use `newtcp` to calculate the requested results. In each case, use a tolerance of 0.00001, and quote the results to 5 significant digits.

3. Find the maximum and minimum values and their locations for $f(x, y) = (xy - x - 2y)/((1 + x^2 + y^2)^2)$. Use a contour plot to help you determine suitable starting points.

4. Evidently $f(x, y, z) = 1 - 10x^4 - 8y^4 - 7z^4$ has maximum value 1 at $(0, 0, 0)$. Find the absolute maximum value of $g(x, y, z) = f(x, y, z) + yz - xyz - x - 2y + z$ by starting at various points near $(0, 0, 0)$.

5. Find the minimum value of

$$f(x, y, z) = x^2 + y^2 + z^2 + 0.2xy - 0.3xz + 4x - y.$$

6. Find the maximum and minimum values of

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

Chapter Review

Key Ideas

- **What is meant by the following terms?**
  - ◇ a critical point of $f(x, y)$
  - ◇ a singular point of $f(x, y)$
  - ◇ an absolute maximum value of $f(x, y)$
  - ◇ a local minimum value of $f(x, y)$
  - ◇ a saddle point of $f(x, y)$
  - ◇ a quadratic form
  - ◇ a constraint
  - ◇ linear programming
  - ◇ an envelope of a family of curves
- **State the second derivative test for a critical point of $f(x, y)$.**
- **Describe the method of Lagrange multipliers.**
- **Describe the method of least squares.**
- **Describe Newton's Method for two equations.**

Review Exercises

In Exercises 1–4, find and classify all the critical points of the given functions.

1. $xy\,e^{-x+y}$

2. $x^2y - 2xy^2 + 2xy$

3. $\dfrac{1}{x} + \dfrac{4}{y} + \dfrac{9}{4 - x - y}$

4. $x^2y(2 - x - y)$

5. Let $f(x, y, z) = x^2 + y^2 + z^2 + 1/(x^2 + y^2 + z^2)$. Does f have a minimum value? If so, what is it and where is it assumed?

6. Show that $x^2 + y^2 + z^2 - xy - xz - yz$ has a local minimum value at $(0, 0, 0)$. Is the minimum value assumed anywhere else?

7. Find the absolute maximum and minimum values of $f(x, y) = xye^{-x^2-4y^2}$. Justify your answer.

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

   (a) Find the maximum and minimum values of $f(x, y)$ on the xy-plane.

   (b) Find the maximum and minimum values of $f(x, y)$ on the wedge-shaped region $0 \le y \le 3x$.

9. A wire of length L cm is cut into at most three pieces, and each piece is bent into a square. What is the (a) minimum and (b) maximum of the sum of the areas of the squares?

10. A delivery service will accept parcels in the shape of rectangular boxes the sum of whose girth and height is at most 120 inches. (The girth is the perimeter of a horizontal cross-section.) What is the largest possible volume of such a box?

11. Find the area of the smallest ellipse $(x/a)^2 + (y/b)^2 = 1$ that contains the rectangle $-1 \le x \le 1$, $-2 \le y \le 2$.

12. Find the volume of the smallest ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

that contains the rectangular box $-1 \le x \le 1$, $-2 \le y \le 2$, $-3 \le z \le 3$.

13. Find the volume of the smallest region of the form

$$0 \le z \le a\left(1 - \frac{x^2}{b^2} - \frac{y^2}{c^2}\right)$$

that contains the box $-1 \le x \le 1, -2 \le y \le 2, 0 \le z \le 2$.

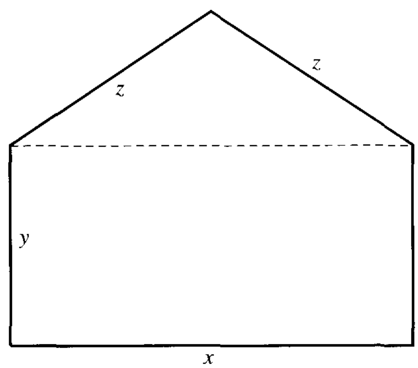

Figure 13.28

14. A window has the shape of a rectangle surmounted by an isosceles triangle. What are the dimensions x, y, and z of the window (see Figure 13.28) if its perimeter is L and its area is maximum?

15. A widget manufacturer determines that if she manufactures x thousands of widgets per month and sells the widgets for y dollars each, then her monthly profit (in thousands of dollars) will be $P = xy - \frac{1}{27}x^2y^3 - x$. If her factory is capable of producing at most 3,000 widgets per month, and government regulations prevent her from charging more than \$2 per widget, how many should she manufacture, and how much should she charge for each, to maximize her monthly profit?

16. Find the envelope of the curves $y = (x - c)^3 + 3c$.

17. Find an approximate solution $y(x, \epsilon)$ of the equation $y + \epsilon x e^y = -2x$ having terms up to second degree in ϵ.

18. (a) Calculate $G'(y)$ if $G(y) = \displaystyle\int_0^\infty \frac{\tan^{-1}(xy)}{x}\,dx$.

(b) Evaluate $\displaystyle\int_0^\infty \frac{\tan^{-1}(\pi x) - \tan^{-1}x}{x}\,dx$. *Hint:* this integral is $G(\pi) - G(1)$.

Challenging Problems

1. (Fourier series)
Show that the constants a_k, $(k = 0, 1, 2, \ldots, n)$, and b_k, $(k = 1, 2, \ldots, n)$, which minimize the integral

$$I_n = \int_{-\pi}^{\pi} \left[f(x) - \frac{a_0}{2} - \sum_{k=0}^{n} \big(a_k \cos kx + b_k \sin kx\big) \right]^2 dx,$$

are given by

$$a_k = \frac{1}{\pi} \int_{\pi}^{\pi} f(x) \cos kx\,dx, \quad b_k = \frac{1}{\pi} \int_{\pi}^{\pi} f(x) \sin kx\,dx.$$

Note that these numbers, called the **Fourier coefficients** of f on $[-\pi, \pi]$, do not depend on n. If they can be calculated for all positive integers k, then the series

$$\frac{a_0}{2} + \sum_{k=0}^{\infty} \big(a_k \cos kx + b_k \sin kx\big)$$

is called the **(full-range) Fourier series** of f on $[-\pi, \pi]$.

2. This is a continuation of Problem 1. Find the (full range) Fourier coefficients a_k and b_k of

$$f(x) = \begin{cases} 0 & \text{if } -\pi \le x < 0 \\ x & \text{if } 0 \le x \le \pi. \end{cases}$$

What is the minimum value of I_n in this case? How does it behave as $n \to \infty$?

* **3.** Evaluate $\displaystyle\int_0^x \frac{\ln(tx+1)}{1+t^2}\,dt$.

* **4. (Steiner's problem)** The problem of finding a point in the plane (or a higher-dimensional space) that minimizes the sum of its distances from n given points is very difficult. The case $n = 3$ is known as Steiner's problem. If $P_1P_2P_3$ is a triangle whose largest angle is less than $120°$, there is a point Q inside the triangle so that the lines QP_1, QP_2, and QP_3 make equal $120°$ angles with one another. Show that the sum of the distances from the vertices of the triangle to a point P is minimum when $P = Q$. *Hint:* first show that if $P = (x, y)$ and $P_i = (x_i, y_i)$, then

$$\frac{d|PP_i|}{dx} = \cos\theta_i \quad \text{and} \quad \frac{d|PP_i|}{dy} = \sin\theta_i,$$

where θ_i is the angle between $\overrightarrow{P_iP}$ and the positive direction of the x-axis. Hence, show that the minimal point P satisfies two trigonometric equations involving θ_1, θ_2, and θ_3. Then try to show that any two of those angles differ by $\pm 2\pi/3$. Where should P be taken if the triangle has an angle of $120°$ or greater?