Dividing by $\|y\|^2$ we obtain the inequality

$$|E_2(a, y)| \le \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} |D_{ij}f(a + cy) - D_{ij}f(a)|$$

for $y \neq 0$. Since each second-order partial derivative $D_{ij}f$ is continuous at a, we have $D_{ij}f(a + cy) \to D_{ij}f(a)$ as $y \to O$, so $E_2(a, y) \to 0$ as $y \to 0$. This completes the proof.

9.11 The nature of a stationary point determined by the eigenvalues of the Hessian matrix

At a stationary point we have $\nabla f(a) = 0$, so the Taylor formula in Equation (9.35) becomes

$$f(a + y) - f(a) = \tfrac{1}{2}yH(a)y^t + \|y\|^2 E_2(a, y).$$

Since the error term $\|y\|^2 E_2(a, y)$ tends to zero faster than $\|y\|^2$, it seems reasonable to expect that for small y the algebraic sign of $f(a + y) - f(a)$ is the same as that of the quadratic form $yH(a)y^t$; hence the nature of the stationary point should be determined by the algebraic sign of the quadratic form. This section is devoted to a proof of this fact.

First we give a connection between the algebraic sign of a quadratic form and its eigenvalues.

THEOREM 9.5. *Let $A = [a_{ij}]$ be an $n \times n$ real symmetric matrix, and let*

$$Q(y) = yAy^t = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}y_iy_j.$$

Then we have:

(a) ***$Q(y) > 0$ for all $y \neq 0$ if and only if all the eigenvalues of A are positive.***
(b) ***$Q(y) < 0$ for all $y \neq 0$ if and only if all the eigenvalues of A are negative.***

Note: In case (a), the quadratic form is called *positive definite;* in case (b) it is called ***negative definite.***

Proof. According to Theorem 5.11 there is an orthogonal matrix C that reduces the quadratic form yAy^t to a diagonal form. That is

$$Q(y) = yAy^t = \sum_{i=1}^{n} \lambda_i x_i^2 \tag{9.38}$$

where $x = (x_1, \ldots, x_n)$ is the row matrix $x = yC$, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. The eigenvalues are real since A is symmetric.

If all the eigenvalues are positive, Equation (9.38) shows that $Q(y) > 0$ whenever $x \neq 0$. But since $x = yC$ we have $y = xC^{-1}$, so $x \neq 0$ if and only if $y \neq 0$. Therefore $Q(y) > 0$ for all $y \neq 0$.

Conversely, if $Q(y) > 0$ for all $y \neq 0$ we can choose y so that $x = yC$ is the kth coordinate vector e_k. For this y, Equation (9.38) gives us $Q(y) = \lambda_k$, so each $\lambda_k > 0$. This proves part (a). The proof of (b) is entirely analogous.

The next theorem describes the nature of a stationary point in terms of the algebraic sign of the quadratic form $yH(a)y^t$.

THEOREM 9.6. *Let f be a scalar field with continuous second-order partial derivatives $D_{ij}f$ in an n-ball B(a), and let H(a) denote the Hessian matrix at a stationary point a. Then we have:*

(a) *If all the eigenvalues of H(a) are positive, f has a relative minimum at a.*

(b) *If all the eigenvalues of H(a) are negative, f has a relative maximum at a.*

(c) *If H(a) has both positive and negative eigenvalues, then f has a saddle point at a.*

Proof. Let $Q(y) = yH(a)y^t$. The Taylor formula gives us

$$f(a + y) - f(a) = \tfrac{1}{2}Q(y) + \|y\|^2 E_2(a, y), \tag{9.39}$$

where $E_2(a, y) \to 0$ as $y \to 0$. We will prove that there is a positive number r such that, if $0 < \|y\| < r$, the algebraic sign of $f(a + y)$ -f(u) is the same as that of Q(y).

Assume first that all the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $H(a)$ are positive. Let h be the smallest eigenvalue. If $u < h$, the n numbers

$$\lambda_1 - u, \ldots, \lambda_n - u$$

are also positive. These numbers are the eigenvalues of the real symmetric matrix $H(a) - uI$, where I is the $n \times n$ identity matrix. By Theorem 9.5, the quadratic form $y[H(a) - uI]y^t$ is positive definite, and hence $y[H(a) - uI]y^t > 0$ for all $y \neq 0$. Therefore

$$yH(a)y^t > y(uI)y^t = u\,\|y\|^2$$

for all real $u < h$. Taking $u = \tfrac{1}{2}h$ we obtain the inequality

$$Q(y) > \tfrac{1}{2}h\,\|y\|^2$$

for all $y \neq 0$. Since $E_2(a, y) \to 0$ as $y \to O$, there is a positive number r such that $|E_2(a, y)| < \tfrac{1}{4}h$ whenever $0 < \|y\| < r$. For such y we have

$$0 \le \|y\|^2\,|E_2(a, y)| < \tfrac{1}{4}h\,\|y\|^2 < \tfrac{1}{2}Q(y),$$

and Taylor's formula (9.39) shows that

$$f(a + y) - f(a) \ge \tfrac{1}{2}Q(y) - \|y\|^2|E_2(a, y)| > 0.$$

Therefore f has a relative minimum at a, which proves part (a). To prove (b) we can use a similar argument, or simply apply part (a) to $-f$.

To prove (c), let I, and λ_2 be two eigenvalues of $H(a)$ of opposite signs. Let $h = \min\{|\lambda_1|, |\lambda_2|\}$. Then for each real u satisfying $-h < u < h$ the numbers

$$\lambda_1 - u \qquad \text{and} \qquad \lambda_2 - u$$

are eigenvalues of opposite sign for the matrix $H(\boldsymbol{a}) - uI$. Therefore, if $u \in (-h, h)$, the quadratic form $\boldsymbol{y}[H(\boldsymbol{a}) - uI]\boldsymbol{y}^t$ takes both positive and negative values in every neighborhood of $\boldsymbol{y} = 0$. Choose $r > 0$ as above so that $|E_2(\boldsymbol{a}, \boldsymbol{y})| < \frac{1}{4}h$ whenever $0 < \|\boldsymbol{y}\| < r$. Then, arguing as above, we see that for such $\boldsymbol{y}$ the algebraic sign of $f(\boldsymbol{a} + \boldsymbol{y})$ -f(a) is the same as that of $Q(\boldsymbol{y})$. Since both positive and negative values occur as $\boldsymbol{y} \to \boldsymbol{O}$, f has a saddle point at $\boldsymbol{a}$. This completes the proof.

> *Note:* If all the eigenvalues of $H(\boldsymbol{a})$ are zero, Theorem 9.6 gives no information concerning the stationary point. Tests involving higher order derivatives can be used to treat such examples, but we shall not discuss them here.

9.12 Second-derivative test for extrema of functions of two variables

In the case $n = 2$ the nature of the stationary point can also be determined by the algebraic sign of the second derivative $D_{1,1}f(\boldsymbol{a})$ and the determinant of the Hessian matrix.

THEOREM 9.7. *Let a be a stationary point of a scalar field f (x,, x_2) with continuous second-order partial derivatives in a 2-ball B(u). Let*

$$A = D_{1,1}f(\boldsymbol{a}), \qquad B = D_{1,2}f(\boldsymbol{a}), \qquad C = D_{2,2}f(\boldsymbol{a}),$$

and let

$$A = \det H(\boldsymbol{a}) = \det \begin{bmatrix} A & B \\ B & C \end{bmatrix} = AC - B^2.$$

Then we have:

(a) *If* $A < 0$, *f has a saddle point at* $\boldsymbol{a}$.
(b) *If* $A > 0$ *and* $A > 0$, *f has a relative minimum at a.*
(c) If $A > 0$ *and* $A < 0$, *f has a relative maximum at a.*
(d) *If* $A = 0$, *the test is inconclusive.*

Proof. In this case the characteristic equation det $[\lambda I -$ H(u)] = 0 is a quadratic equation,

$$\lambda^2 - (A + C)\lambda + \Delta = 0.$$

The eigenvalues λ_1, λ_2 are related to the coefficients by the equations

$$\lambda_1 + \lambda_2 = A + C, \qquad \lambda_1\lambda_2 = A.$$

If $A < 0$ the eigenvalues have opposite signs, so f has a saddle point at a, which proves (a). If $A > 0$, the eigenvalues have the same sign. In this case $AC > B^2 \geq 0$, so A and C have the same sign. This sign must be that of λ_1 and λ_2 since $A + C = \lambda_1 + \lambda_2$. This proves (b) and (c).

To prove (d) we refer to Examples 4 and 5 of Section 9.9. In both these examples we have $A = 0$ at the origin. In Example 4 the origin is a saddle point, and in Example 5 it is a relative minimum.

Even when Theorem 9.7 is applicable it may not be the simplest way to determine the nature of a stationary point. For example, when $f(x, y) = e^{1/g(x,y)}$, where $g(x, y) = x^2 + 2 + \cos^2 y - 2 \cos y$, the test is applicable, but the computations are lengthy. In this case we may express $g(x, y)$ as a sum of squares by writing $g(x, y) = 1 + x^2 + (1 - \cos y)^2$. We see at once that f has relative maxima at the points at which $x^2 = 0$ and $(1 - \cos y)^2 = 0$. These are the points $(0, 2n\pi)$, when n is any integer.

9.13 Exercises

In Exercises 1 through 15, locate and classify the stationary points (if any) of the surfaces having the Cartesian equations given.

1. $z = x^2 + (y - 1)^2$.
2. $z = x^2 - (y - 1)^2$.
3. $z = 1 + x^2 - y^2$.
4. $z = (x - y + 1)^2$.
5. $z = 2x^2 - xy - 3y^2 - 3x + 7y$.
6. $z = x^2 - xy + y^2 - 2x + y$.
7. $z = x^3 - 3xy^2 + y^3$.
8. $z = x^2y^3(6 - x - y)$.
9. $z = x^3 + y^3 - 3xy$.
10. $z = \sin x \cosh y$.
11. $z = e^{2x+3y}(8x^2 - 6xy + 3y^2)$.
12. $z = (5x + 7y - 25)e^{-(x^2+xy+y^2)}$.
13. $z = \sin x \sin y \sin (x + y)$, $0 \le x \le \pi, 0 \le y \le \pi$.
14. $z = x - 2y + \log \sqrt{x^2 + y^2} + 3 \arctan \dfrac{y}{x}$, $x > 0$.
15. $z = (x^2 + y^2)e^{-(x^2+y^2)}$.
16. Let $f(x, y) = 3x^4 - 4x^2y + y^2$. Show that on every line $y = mx$ the function has a minimum at (0, 0), but that there is no relative minimum in any two-dimensional neighborhood of the origin. Make a sketch indicating the set of points (x, y) at which $f(x, y) > 0$ and the set at which $f(x, y) < 0$.
17. Let $f(x, y) = (3 - x)(3 - y)(x + y - 3)$.
    (a) Make a sketch indicating the set of points (x, y) at which $f(x, y) \ge 0$.
    (b) Find all points (x, y) in the plane at which $D_1 f(x, y) = D_2 f(x, y) = 0$. [*Hint:* $D_1 f(x, y)$ has $(3 - y)$ as a factor.]
    (c) Which of the stationary points are relative maxima? Which are relative minima? Which are neither? Give reasons for your answers.
    (d) Does f have an absolute minimum or an absolute maximum on the whole plane? Give reasons for your answers.
18. Determine all the relative and absolute extreme values and the saddle points for the function $f(x, y) = xy(1 - x^2 - y^2)$ on the square $0 \le x \le 1, 0 \le y \le 1$.
19. Determine constants a and b such that the integral

$$\int_0^1 \{ax + b - f(x)\}^2 \, dx$$

will be as small as possible if (a) $f(x) = x^2$; (b) $f(x) = (x^2 + 1)^{-1}$.

20. Let $f(x, y) = Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F$, where $A > 0$ and $B^2 < AC$.
    (a) Prove that a point (x_1, y_1) exists at which f has a minimum. [*Hint:* Transform the quadratic part to a sum of squares.]
    (b) Prove that $f(x_1, y_1) = Dx_1 + Ey_1 + F$ at this minimum.
    (c) Show that

$$f(x_1, y_1) = \frac{1}{AC - B^2} \begin{vmatrix} A & B & D \\ B & C & E \\ D & E & F \end{vmatrix}.$$

21. ***Method of least squares.*** Given n distinct numbers $x_1, \ldots, x_n$ and n further numbers $y_1, \ldots, y_n$ (not necessarily distinct), it is generally impossible to find a straight line $f(x) = ax + b$ which passes through all the points (x_i, y_i), that is, such that $f(x_i) = y_i$ for each i. However, we can try a linear function which makes the "total square error"

$$E(a, b) = \sum_{i=1}^{n} [f(x_i) - y_i]^2$$

a minimum. Determine values of a and b which do this.

22. Extend the method of least squares to 3-space. That is, find a linear function $f(x, y) = ax + by + c$ which minimizes the total square error

$$E(a, b, c) = \sum_{i=1}^{n} [f(x_i, y_i) - z_i]^2,$$

where (x_i, y_i) are n given distinct points and $z_1, \ldots, z_n$ are n given real numbers.

23. Let $z_1, \ldots, z_n$ be n distinct points in m-space. If $x \in \mathbf{R}^m$, define

$$f(x) = \sum_{k=1}^{n} \|x - z_k\|^2 .$$

Prove that f has a minimum at the point $a = \frac{1}{n} \sum_{k=1}^{n} z_k$ (the centroid).

24. Let a be a stationary point of a scalar field f with continuous second-order partial derivatives in an n-ball $B(a)$. Prove that f has a saddle point at a if at least two of the diagonal entries of the Hessian matrix $H(a)$ have opposite signs.

25. Verify that the scalar field $f(x, y, z) = x^4 + y^4 + z^4 - 4xyz$ has a stationary point at $(1, 1, 1)$, and determine the nature of this stationary point by computing the eigenvalues of its Hessian matrix.

9.14 Extrema with constraints. Lagrange's multipliers

We begin this section with two examples of extremum problems with constraints.

EXAMPLE 1. Given a surface S not passing through the origin, determine those points of S which are nearest to the origin.

EXAMPLE 2. If $f(x, y, z)$ denotes the temperature at (x, y, z), determine the maximum and minimum values of the temperature on a given curve C in 3-space.

Both these examples are special cases of the following general problem: ***Determine the extreme values of a scalar field $f(x)$ when x is restricted to lie in a given subset of the domain of f.***

In Example 1 the scalar field to be minimized is the distance function,

$$f(x, y, z) = (x^2 + y^2 + z^2)^{1/2} ;$$

the constraining subset is the given surface S. In Example 2 the constraining subset is the given curve C.

Constrained extremum problems are often very difficult; no general method is known for attacking them in their fullest generality. Special methods are available when the constraining subset has a fairly simple structure, for instance, if it is a surface as in Example 1, or a curve as in Example 2. This section discusses the method of Lagrange's multipliers for solving such problems. First we describe the method in its general form, and then we give geometric arguments to show why it works in the two examples mentioned above.

The method of Lagrange's multipliers. If a scalar field $f(x_1, \ldots, x_n)$ has a relative extremum when it is subject to m constraints, say

$$g_1(x_1, \ldots, x_n) = 0, \qquad \ldots, \qquad g_m(x_1, \ldots, x_n) = 0, \tag{9.40}$$

where $m < n$, then there exist m scalars $\lambda_1, \ldots, \lambda_m$ such that

$$\nabla f = \lambda_1 \nabla g_1 + \cdots + \lambda_m \nabla g_m \tag{9.41}$$

at each extremum point.

To determine the extremum points in practice we consider the system of $n + m$ equations obtained by taking the m constraint equations in (9.40) along with the n scalar equations determined by the vector relation (9.41). These equations are to be solved (if possible) for the $n + m$ unknowns $x_1, \ldots, x_n$ and $\lambda_1, \ldots, \lambda_m$. The points $(x_1, \ldots, x_n)$ at which relative extrema occur are found among the solutions to these equations.

The scalars $\lambda_1, \ldots, \lambda_m$ which are introduced to help us solve this type of problem are called ***Lagrange's multipliers.*** One multiplier is introduced for each constraint. The scalar field f and the constraint functions $g_1, \ldots, g_m$ are assumed to be differentiable. The method is valid if the number of constraints, m, is less than the number of variables, n, and if not all the Jacobian determinants of the constraint functions with respect to m of the variables $x_1, \ldots, x_n$ are zero at the extreme value in question. The proof of the validity of the method is an important result in advanced calculus and will not be discussed here. (See Chapter 7 of the author's ***Mathematical Analysis,*** Adldison-Wesley, Reading, Mass., 1957.) Instead we give geometric arguments to show why the method works in the two examples described at the beginning of this section.

Geometric solution of Example 1. We wish to determine: those points on a given surface S which are nearest to the origin. A point (x, y, z) in 3-space lies at a distance r from the origin if and only if it lies on the sphere

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

This sphere is a level surface of the function $f(x, y, z) = (x^2 + y^2 + z^2)^{1/2}$ which is being minimized. If we start with $r = 0$ and let r increase until the corresponding level surface first touches the given surface S, each point of contact will be a point of S nearest to the origin.

To determine the coordinates of the contact points we assume that S is described by a Cartesian equation $g(x, y, z) = 0$. If S has a tangent plane at a point of contact, this plane must also be tangent to the contacting level surface. Therefore the gradient vector of the surface $g(x, y, z) = 0$ must be parallel to the gradient vector of the contacting level surface $f(x, y, z) = r$. Hence there is a constant λ such that

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

at each contact point. This is the vector equation (9.41) provided by Lagrange's method when there is one constraint.

Geometric solution *to Example 2.* We seek the extreme values of a temperature function $f(x, y, z)$ on a given curve C. If we regard the curve C as the intersection of two surfaces, say

$$g_1(x, y, z) = 0 \qquad \text{and} \qquad g_2(x, y, z) = 0,$$

we have an extremum problem with two constraints. The two gradient vectors ∇g_1 and ∇g_2 are normals to these surfaces, hence they are also normal to C, the curve of intersection. (See Figure 9.8.) We show next that the gradient vector ∇f of the temperature

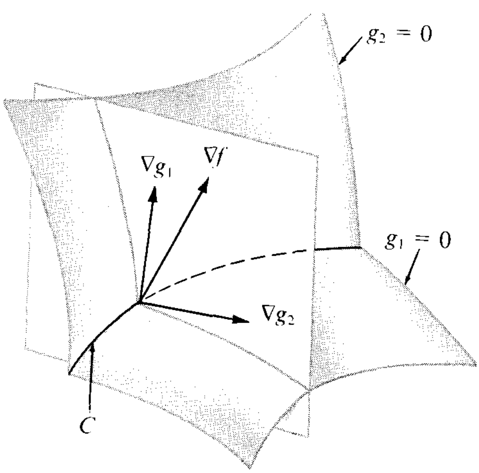

FIGURE 9.8 The vectors ∇g_1, ∇g_2, and ∇f shown lying in the same plane.

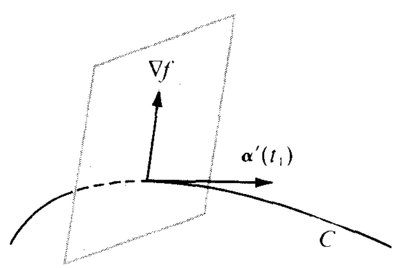

FIGURE 9.9 The gradient vector ∇f lies in a plane normal to C.

function is also normal to C at each relative extremum on C. This implies that ∇f lies in the same plane as ∇g_1 and ∇g_2; hence if ∇g_1 and ∇g_2 are independent we can express ∇f as a linear combination of ∇g_1 and ∇g_2, say

$$\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 .$$

This is the vector equation (9.41) provided by Lagrange's method when there are two constraints.

To show that ∇f is normal to C at an extremum we imagine C as being described by a vector-valued function $\alpha(t)$, where t varies over an interval $[a, b]$. On the curve C the temperature becomes a function of t, say $\varphi(t) = f[\alpha(t)]$. If φ has a relative extremum at an interior point t_1 of $[a, b]$ we must have $\varphi'(t_1) = 0$. On the other hand, the chain rule tells us that $\varphi'(t)$ is given by the dot product

$$\varphi'(t) = \nabla f\,[\alpha(t)] \cdot \alpha'(t).$$

This dot product is zero at t_1, hence ∇f is perpendicular to $\alpha'(t_1)$. But $\alpha'(t_1)$ is tangent to C, so $\nabla f\,[\alpha(t_1)]$ lies in the plane normal to C, as shown in Figure 9.9.

The two gradient vectors ∇g_1 and ∇g_2 are independent if and only if their cross product is nonzero. The cross product is given by

$$\nabla g_1 \times \nabla g_2 = \begin{vmatrix} i & j & k \\ \dfrac{\partial g_1}{\partial x} & \dfrac{\partial g_1}{\partial y} & \dfrac{\partial g_1}{\partial z} \\ \dfrac{\partial g_2}{\partial x} & \dfrac{\partial g_2}{\partial y} & \dfrac{\partial g_2}{\partial z} \end{vmatrix} = \frac{\partial(g_1, g_2)}{\partial(y, z)}\, i + \frac{\partial(g_1, g_2)}{\partial(z, x)}\, j + \frac{\partial(g_1, g_2)}{\partial(x, y)}\, k\,.$$

Therefore, independence of ∇g_1 and ∇g_2 means that not all three of the Jacobian determinants on the right are zero. As remarked earlier, Lagrange's method is applicable whenever this condition is satisfied.

If ∇g_1 and ∇g_2 are dependent the method may fail. For example, suppose we try to apply Lagrange's method to find the extreme values of $f(x, y, z) = x^2 + y^2$ on the curve of intersection of the two surfaces $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 0$, where $g_1(x, y, z) = z$ and $g_2(x, y, z) = z^2 - (y - 1)^3$. The two surfaces, a plane and a cylinder, intersect along the straight line C shown in Figure 9.10. The problem obviously has a solution, because

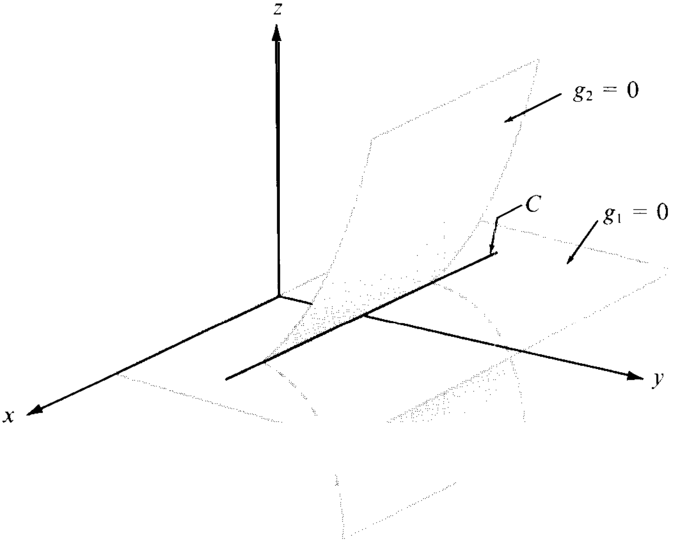

FIGURE 9.10 An example where Lagrange's method is not applicable.

$f(x, y, z)$ represents the distance of the point (x, y, z) from the z-axis and this distance is a minimum on C when the point is at (0, 1, 0). However, at this point the gradient vectors are $\nabla g_1 = \boldsymbol{k}$, $\nabla g_2 = \boldsymbol{O}$, and $\nabla f = 2\boldsymbol{j}$, and it is clear that there are no scalars λ_1 and λ_2 that satisfy Equation (9.41).

9.15 Exercises

1. Find the extreme values of $z = xy$ subject to the condition $x + y = 1$.
2. Find the maximum and minimum distances from the origin to the curve $5x^2 + 6xy + 5y^2 = 8$.
3. Assume a and b are fixed positive numbers.
   (a) Find the extreme values of $z = x/a + y/b$ subject to the condition $x^2 + y^2 = 1$.
   (b) Find the extreme values of $z = x^2 + y^2$ subject to the condition $x/a + y/b = 1$.
   In each case, interpret the problem geometrically.
4. Find the extreme values of $z = \cos^2 x + \cos^2 y$ subject to the side condition $x - y = \pi/4$.
5. Find the extreme values of the scalar field $f(x, y, z) = x - 2y + 2z$ on the sphere $x^2 + y^2 + z^2 = 1$.
6. Find the points of the surface $z^2 - xy = 1$ nearest to the origin.
7. Find the shortest distance from the point $(1, 0)$ to the parabola $y^2 = 4x$.
8. Find the points on the curve of intersection of the two surfaces

$$x^2 - xy + y^2 - z^2 = 1 \qquad \text{and} \qquad x^2 + y^2 = 1$$

   which are nearest to the origin.
9. If a, b, and c are positive numbers, find the maximum value of $f(x, y, z) = x^a y^b z^c$ subject to the side condition $x + y + z = 1$.
10. Find the minimum volume bounded by the planes $x = 0$, $y = 0$, $z = 0$, and a plane which is tangent to the ellipsoid

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

    at a point in the octant $x > 0, y > 0, z > 0$.
11. Find the maximum of $\log x + \log y + 3 \log z$ on that portion of the sphere $x^2 + y^2 + z^2 = 5r^2$ where $x > 0$, $y > 0$, $z > 0$. Use the result to prove that for real positive numbers a, b, c we have

$$abc^3 \le 27\left(\frac{a + b + c}{5}\right)^5.$$

12. Given the conic section $Ax^2 + 2Bxy + Cy^2 = 1$, where $A > 0$ and $B^2 < AC$. Let m and M denote the distances from the origin to the nearest and furthest points of the conic. Show that

$$M^2 = \frac{A + C + \sqrt{(A - C)^2 + 4B^2}}{2(AC - B^2)}$$

    and find a companion formula for m^2.
13. Use the method of Lagrange's multipliers to find the greatest and least distances of a point on the ellipse $x^2 + 4y^2 = 4$ from the straight line $x + y = 4$.
14. The cross section of a trough is an isosceles trapezoid. If the trough is made by bending up the sides of a strip of metal c inches wide, what should be the angle of inclination of the sides and the width across the bottom if the cross-sectional area is to be a maximum?

9.16 The extreme-value theorem for continuous scalar fields

The extreme-value theorem for real-valued functions continuous on a closed and bounded interval can be extended to scalar fields. We consider scalar fields continuous on a closed n-dimensional interval. Such an interval is defined as the Cartesian product of n one-dimensional closed intervals. If a = *(a,,* . . . , *a,)* and 6 = $(b_1, \ldots, b_n)$ we write

$$[a, b] = [a_1, b_1] \times \cdots \times [a_n, b_n] = \{(x_1, \ldots, x_n) \mid x_1 \in [a_1, b_1], \ldots, x_n \in [a_n, b_n]\}.$$

For example, when n = 2 the Cartesian product [a, b] is a rectangle.

The proof of the extreme-value theorem parallels the proof given in Volume I for the l-dimensional case. First we prove that continuity offimplies boundedness, then we prove thatfactually attains its maximum and minimum values somewhere in [a, b].

THEOREM 9.8. BOUNDEDNESS THEOREM FOR CONTINUOUS SCALAR FIELDS. *If fis a scalar field continuous at each point of a closed interval [a, b] in* R", *then f is bounded on [a, b]. That is, there is a number C $\geq$ 0 such that $|f(x)| \leq$ Cfor all x in [a, b].*

Proof. We argue by contradiction, using the method of successive bisection. Figure 9.11 illustrates the method for the case ***n* = *2.***

Assumefis unbounded on [a, ***b].*** Let $I^{(1)}$ = [a, ***b]*** and let $I_k^{(1)}$ = ***[a,,*** $b_k]$, so that

$$I^{(1)} = I_1^{(1)} \times \cdots \times I_n^{(1)}.$$

Bisect each one-dimensional interval $I_k^{(1)}$ to form two subintervals, a left half $I_{k,1}^{(1)}$ and a right half $I_{k,2}^{(1)}$. Now consider all possible Cartesian products of the form

$$I_{1,j_1}^{(1)} \times \cdots \times I_{n,j_n}^{(1)},$$

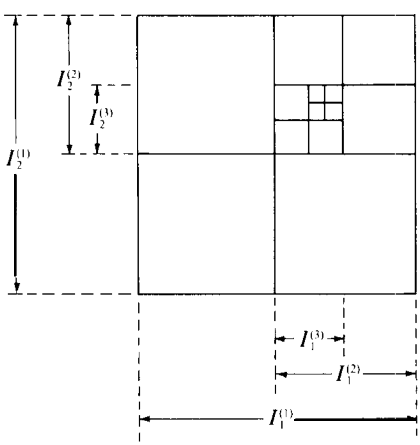

FIGURE 9.11 Illustrating the method of successive bisection in the plane.

where each j_i = I or 2. There are exactly 2” such products. Each product is an n-dimensional subinterval of [a, b], and their union is equal to [a, b]. The functionf’is unbounded in ***at least one*** of these subintervals (if it were bounded in each of them it would also be bounded on [a, b]). One of these we denote by $I^{(2)}$ which we express as

$$I^{(2)} = I_1^{(2)} \times \cdots \times I_n^{(2)},$$

where each $I_k^{(2)}$ is one of the one-dimensional subintervals of $I_k^{(1)}$, of length $\frac{1}{2}(b_k - a_k)$.

We now proceed with $I^{(2)}$ as we did with $I^{(1)}$, bisecting each one-dimensional component interval $I_k^{(2)}$ and arriving at an n-dimensional interval $I^{(3)}$ in which f is unbounded. We continue the process, obtaining an infinite set of n-dimensional intervals

$$I^{(1)}, I^{(2)}, \ldots, \qquad \text{with} \quad I^{(m+1)} \subseteq I^{(m)},$$

in each of which f is unbounded. The mth interval $I^{(m)}$ can be expressed in the form

$$I^{(m)} = I_1^{(m)} \times \cdots \times I_n^{(m)}.$$

Since each one-dimensional interval $I_k^{(m)}$ is obtained by $m - 1$ successive bisections of [a,, b_k], if we write $I_k^{(m)} = [a_k^{(m)}, b_k^{(m)}]$ we have

(9.42) $$b_k^{(m)} - a_k^{(m)} = \frac{b_k - a_k}{2^{m-1}}, \qquad \text{for} \quad k = 1, 2, \ldots, n.$$

For each fixed k, the supremum of all left endpoints $a_k^{(m)}$ $(m = 1, 2, \ldots)$ must therefore be equal to the infimum of all right endpoints $b_k^{(m)}$ $(m = 1, 2, \ldots)$, and their common value we denote by t_k. The point $t = (t_1, \ldots, t_n)$ lies in [a, b]. By continuity off’at t there is an n-ball $B(t; r)$ in which we have

$$|f(x) - f(t)| \leq 1 \qquad \text{for all } x \text{ in } B(t; r) \cap [a, b].$$

This inequality implies

$$|f(x)| < 1 + |f(t)| \qquad \text{for all } x \text{ in } B(t; r) \cap [a, b],$$

so f is bounded on the set $B(t; r) \cap [a, b]$. But this set contains the entire interval $I^{(m)}$ when m is large enough so that each of the n numbers in (9.42) is less than $r/\sqrt{n}$. Therefore for such m the function f is bounded on $I^{(m)}$, contradicting the fact that f is unbounded on $I^{(m)}$. This contradiction completes the proof.

If f is bounded on [a, b], the set of all function values $f(x)$ is a set of real numbers bounded above and below. Therefore this set has a supremum and an infimum which we denote by sup f and inf f, respectively. That is, we write

$$\sup f = \sup \{f(x) \mid x \in [a, b]\}, \qquad \inf f = \inf \{f(x) \mid x \in [a, b]\}.$$

Now we prove that a continuous function takes on both values inff'and supfsomewhere in [a, ***b***].

THEOREM *9.9* EXTREME-VALUE **THEOREM** FOR CONTINUOUS SCALAR **FIELDS.** *Iff is continuous on a closed interval [a, b] in* $\mathbf{R}^n$, *then there exist points c and d in [a, b] such that*

$$f(\boldsymbol{c}) = \sup f \qquad \text{and} \qquad f(\boldsymbol{d}) = \text{inff.}$$

Proof. It suffices to prove that f attains its supremum in [a, ***b].*** The result for the infimum then follows as a consequence because the infimum **off** is the supremum of ***-f.***

Let M = sup ***f.*** We shall assume that there is no $\boldsymbol{x}$ in [$\boldsymbol{a}$, ***b]*** for which $f(\boldsymbol{x}) = M$ and obtain a contradiction. Let g(x) = $M - f(\boldsymbol{x})$. Then g(x) > 0 for all $\boldsymbol{x}$ in [a, *b] so* the reciprocal l/g is continuous on [a, ***b].*** By the boundedness theorem, l/g is bounded on [a, ***b],*** say $1/g(\boldsymbol{x}) < C$ for all $\boldsymbol{x}$ in [a, ***b],*** where C > 0. This implies M -f(x) **>** l/C, so that $f(\boldsymbol{x})$ < A4 **–** l/C for all $\boldsymbol{x}$ in [a, ***b].*** This contradicts the fact that M is the least upper bound **off** on [a, ***b].*** Hence ***f(x)*** = M for at least one x in [a, ***b].***

9.17 The small-span theorem for continuous scalar fields (uniform continuity)

Let ***f*** be continuous on a bounded closed interval [a, ***b]*** in $\mathbf{R}^n$, and let ***M(f)*** and *m(f)* denote, respectively, the maximum and minimum values **off** on [a, ***b].*** The difference

$$M(f) - m(f)$$

is called the *span* off on [a, ***b].*** As in the one-dimensional case we have a small-span theorem for continuous functions which tells us that the interval [a, ***b]*** can be partitioned so that the span **off** in each subinterval is arbitrarily small.

Write [a, ***b] = [a,,*** b_1] x . . . x *[a,,* b_n], and let P_k be a partition of the interval *[a,,* b_k]. That is, P_k is a set of points

$$P_k = \{x_0, x_1, \ldots, x_{r-1}, x_r\}$$

such that $a_k = x_0 \le x_1 \le \cdots \le x_{r-1} \le x_r$ = b,. The Cartesian product

$$P = P_1 \times \cdots \times P_n$$

is called a partition of the interval [a, *b].* The small-span theorem, also called the theorem on uniform continuity, now takes the following form.

THEOREM 9.10. *Let* ***f*** *be a scalar field continuous on a closed interval [a, b] in* $\mathbf{R}^n$. *Then* ***for*** *every* $\epsilon > 0$ *there is a partition* ***of*** *[a, b] into a finite number* ***of*** *subintervals such that the span* ***off*** *in every subintercal* ***is*** *less than* ϵ.

Proof. The proof is entirely analogous to the one-dimensional case so we only outline the principal steps. We argue by contradiction, using the method of successive bisection. We assume the theorem is false; that is, we assume that for some ϵ_0 the interval [a, ***b]*** cannot be partitioned into a finite number of subintervals in each of which the span **off** is less than ϵ_0. By successive bisection we obtain an infinite set of subintervals $I^{(1)}, I^{(2)}, \ldots,$ in each of which the span **off** is at least ϵ_0. By considering the least upper bound of the leftmost endpoints of the component intervals of $I^{(1)}, I^{(2)}, \ldots$ we obtain a point $\boldsymbol{t}$ in [a, $\boldsymbol{b}$] lying in all these intervals. By continuity **off** at $\boldsymbol{t}$ there is an n-ball $B(\boldsymbol{t}; r)$ such that the span **off** is less than $\frac{1}{2}\epsilon_0$ in $B(\boldsymbol{t}; r) \cap$ [a, $\boldsymbol{b}$]. But, when m is sufficiently large, the interval $I^{(m)}$ lies in the set $B(\boldsymbol{t}; r) \cap$ [a, $\boldsymbol{b}$], so the span off is no larger than $\frac{1}{2}\epsilon_0$ in $I^{(m)}$, contradicting the fact that the span **off** is at least ϵ_0 in $I^{(m)}$.

10

LINE INTEGRALS

10.1 Introduction

In Volume I we discussed the integral $\int_a^b f(x)\,dx$, first for real-valued functions defined and bounded on finite intervals, and then for unbounded functions and infinite intervals. The concept was later extended to vector-valued functions and, in Chapter 7 of Volume II, to matrix functions.

This chapter extends the notion of integral in another direction. The interval $[a, b]$ is replaced by a curve in n-space described by a vector-valued function $\boldsymbol{\alpha}$, and the integrand is a vector field $\boldsymbol{f}$ defined and bounded on this curve. The resulting integral is called a *line integral,* a *curvilinear integral,* or a *contour integral,* and is denoted by $\int \boldsymbol{f} \cdot d\boldsymbol{\alpha}$ or by some similar symbol. The dot is used purposely to suggest an inner product of two vectors. The curve is called a *path of integration.*

Line integrals are of fundamental importance in both pure and applied mathematics. They occur in connection with work, potential energy, heat flow, change in entropy, circulation of a fluid, and other physical situations in which the behavior of a vector or scalar field is studied along a curve.

10.2 Paths and line integrals

Before defining line integrals we recall the definition of curve given in Volume I. Let $\boldsymbol{\alpha}$ be a vector-valued function defined on a finite closed interval $J = [a, b]$. As t runs through J, the function values $\boldsymbol{\alpha}(t)$ trace out a set of points in n-space called the *graph* of the function. If $\boldsymbol{\alpha}$ is continuous on J the graph is called a *curve;* more specifically, the curve described by $\boldsymbol{\alpha}$.

In our study of curves in Volume I we found that different functions can trace out the same curve in different ways, for example, in different directions or with different velocities. In the study of line integrals we are concerned not only with the set of points on a curve but with the actual manner in which the curve is traced out, that is, with the function $\boldsymbol{\alpha}$ itself. Such a function will be called a *continuous path.*

DEFINITION. *Let $J = [a, b]$ be a finite closed interval in $\mathbf{R}^1$. A function $\boldsymbol{\alpha}: J \to \mathbf{R}^n$ which is continuous on J is called a continuous path in n-space. The path is called smooth if the derivative $\boldsymbol{\alpha}'$ exists and is continuous in the open interval (a, b). The path is called piecewise*