CHAPTER 3

Differentiation of Multivariable Functions

16. Functions of Several Variables

The concept of a function of several variables can be qualitatively understood from simple examples in everyday life. The temperature in a room may vary from point to point. A point in space can be defined by an ordered triple of numbers that are coordinates of the point in some coordinate system, say, (x, y, z). Measurements of the temperature at every point from a set D in space assign a real number T (the temperature) to every point of D. The dependence of T on coordinates of the point is indicated by writing $T = T(x, y, z)$. Similarly, the concentration of a chemical can depend on a point in space. In addition, if the chemical reacts with other chemicals, its concentration at a point may also change with time. In this case, the concentration $C = C(x, y, z, t)$ depends on four variables, three spatial coordinates and the time t. In general, if the value of a quantity f depends on values of several other quantities, say, x_1, x_2,..., x_m, this dependence is indicated by writing $f = f(x_1, x_2, ..., x_m)$. In other words, $f = f(x_1, x_2, ..., x_m)$ *indicates a rule that assigns a unique real number* f *to each ordered* m*-tuple of real numbers* $(x_1, x_2, ..., x_m)$:

$$f: \quad (x_1, x_2, ..., x_m) \ \rightarrow \ f(x_1, x_2, ..., x_m)$$

Each number in the m-tuple may be of a different nature and measured in different units. In the above example, the concentration depends on ordered quadruples (x, y, z, t), where x, y, and z are the coordinates of a point in space (measured in units of length) and t is time (measured in units of time).

To analyze properties of functions of several variables, a notion of a distance between two ordered $m-$tuples is needed. For example, a rate of change of a function is naturally defined as the difference of values of the function at two points divided by the distance between them. This allows us to determine that one function changes more rapidly than the other. In what follows, *functions on Euclidean spaces will be studied*. In other words, the distance between two ordered $m-$tuples (or two arguments of a function) is assumed to be the Euclidean distance. To simplify notations, the argument of a function of several variables will often be written in the vector form

$$f(x_1, x_2, ..., x_m) = f(\mathbf{r}) \, , \quad \mathbf{r} \in \mathbb{R}^m \, .$$

The value of a function at a particular point P of a Euclidean space will also be denoted by $f(P)$ to emphasize that this value is *independent* of the choice of a coordinate system in which coordinates of P are given. For

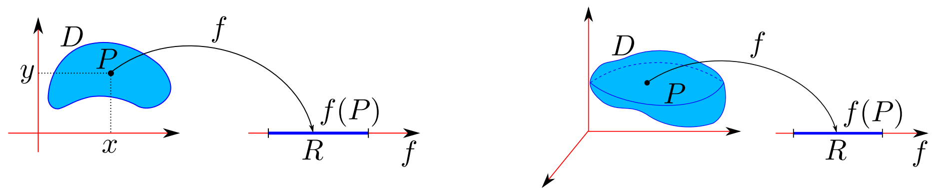


FIGURE **16.1**. **Left**: A function f of two variables is a rule, $f:\ P \to f(P)$, that assigns a number $f(P)$ to every point P of a planar region D. The set R of all numbers $f(P)$ is the range of f. The region D is the domain of f. **Right**: A function f of three variables is a rule that assigns a number $f(P)$ to every point P of a solid region D.

example, a temperature in a room $T = T(P)$ depends on a point P, but the coordinates of P (the ordered triples of numbers) are different in different coordinate systems.

16.1. Real-Valued Functions of Several Variables.

DEFINITION **16.1**. (Real-Valued Function of Several Variables). *Let D be a set of ordered m-tuples of real numbers $(x_1, x_2, ..., x_m)$. A function f of m variables is a rule that assigns to each $m-$tuple in the set D a unique real number denoted by $f(x_1, x_2, ..., x_m)$. The set D is the domain of f, and its range is the set of values that f takes on it, that is, $\{f(x_1, x_2, ..., x_m)\,|\,(x_1, x_2, ..., x_m) \in D\}$.*

This definition is illustrated in Fig. **16.1**. The rule may be defined by different means. For example, if D is a collection of N points P_i, $i = 1, 2, ..., N$, then a function f can be defined by a table $(P_i, f(P_i))$, where $f(P_i)$ is the value of f at P_i. A function f can be defined geometrically. For example, the height of a mountain relative to sea level is a function of its position on the globe. So the height is a function of two variables, the longitude and latitude. A function can be defined by an algebraic rule that prescribes algebraic operations to be carried out with real numbers in any n-tuple to obtain the value of the function. For example,

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

The value of this function at $(1, 2, 3)$ is

$$f(1, 2, 3) = 1^2 - 2 + 3^3 = 26 \,.$$

Unless specified otherwise, *the domain of a function defined by an algebraic rule is the set of m-tuples for which the rule makes sense.*

EXAMPLE **16.1**. *Find the domain and the range of the function of two variables $f(x, y) = \ln(1 - x^2 - y^2)$.*

Solution: The logarithm is defined for any strictly positive number. Therefore, the doublets (x, y) must be such that $1 - x^2 - y^2 > 0$ or $x^2 + y^2 < 1$. Hence,

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

Since any doublet (x, y) can be uniquely associated with a point on a plane, the set D can be given a geometrical description as a disk of radius 1 whose boundary, the circle $x^2 + y^2 = 1$, is not included in D. For any point in the interior of the disk, the argument of the logarithm lies in the interval $0 < 1 - x^2 - y^2 \leq 1$. So the range of f is the set of values of the logarithm in the interval $(0, 1]$, which is $-\infty < f \leq 0$. □

Example **16.2**. *Find the domain and the range of the function of three variables* $f(x, y, z) = x^2\sqrt{z - x^2 - y^2}$.

Solution: The square root is defined only for nonnegative numbers. Therefore, ordered triples (x, y, z) must be such that $z - x^2 - y^2 \geq 0$, that is,

$$D = \{(x, y, z) \mid z \geq x^2 + y^2\}.$$

This set can be given a geometrical description as a point set in space because any triple can be associated with a unique point in space. The equation

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

describes a *circular paraboloid*. So the domain is the spatial (solid) region containing points that lie on or above the paraboloid. The function is nonnegative. By fixing x and y and increasing z, one can see that the value of f can be any positive number. So the range is $0 \leq f(x, y, z) < \infty$. □

The domain of a function of m variables is viewed as a subset of an m-dimensional Euclidean space. For example, the domain of the function

$$f(\mathbf{r}) = (1 - x_1^2 - x_2^2 - \cdots - x_m^2)^{1/2} = (1 - \|\mathbf{r}\|^2)^{1/2},$$

where $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$, is the set of points in the m-dimensional Euclidean space whose distance from the origin (the zero vector) does not exceed 1,

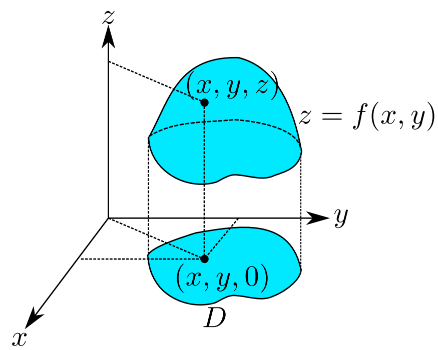
$$D = \{\mathbf{r} \in \mathbb{R}^m \mid \|\mathbf{r}\| \leq 1\};$$

that is, it is a closed m-dimensional ball of radius 1. So the domain of a multivariable function defined by an algebraic rule can be described by conditions on the components of the ordered m-tuple $\mathbf{r}$ under which the rule makes sense.

16.2. The Graph of a Function of Two Variables. The graph of a function of one variable $f(x)$ is the set of points (x, y) of a plane such that

$$y = f(x), \quad x \in D$$

where the domain D is a collection of points on the x axis. The graph is obtained by moving a point of the domain parallel to the y axis by an amount determined by the value of the function $y = f(x)$. The graph provides a



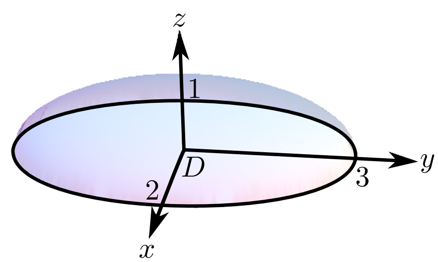


FIGURE **16.2**. **Left**: The graph of a function of two variables is the surface defined by the equation $z = f(x, y)$. It is obtained from the domain D of f by moving each point $(x, y, 0)$ in D along the z axis to the point $(x, y, f(x, y))$. **Right**: The graph of the function studied in Example **16.3**.

useful picture of the behavior of the function. The idea can be extended to functions of two variables.

DEFINITION **16.2**. (Graph of a Function of Two Variables). *The graph of a function $f(x, y)$ with domain D is a collection of points (x, y, z) in space such that*

$$z = f(x, y)\,, \quad (x, y) \in D\,.$$

The domain D is a set of points in the xy plane. The graph is then obtained by moving each point of D parallel to the z axis by an amount equal to the corresponding value of the function $z = f(x, y)$. If D is a portion of the plane, then the graph of f is generally a surface (see Fig. **16.2** (left panel)). One can think of the graph as "mountains" of height $f(x, y)$ on the xy plane.

EXAMPLE **16.3**. *Sketch the graph of the function* $f(x, y) = \sqrt{1 - (x/2)^2 - (y/3)^2}$.

SOLUTION: The domain is the portion of the xy plane

$$D = \{(x, y)\,|\,(x/2)^2 + (y/3)^2 \le 1\}\,.$$

It is bounded by the ellipse with semiaxes 2 and 3. The graph is the surface defined by the equation

$$z = \sqrt{1 - (x/2)^2 - (y/3)^2}\,.$$

By squaring both sides of this equation, one finds

$$(x/2)^2 + (y/3)^2 + z^2 = 1\,,$$

which defines an ellipsoid. The graph is its upper portion with $z \ge 0$ as depicted in the right panel of Fig. **16.2**. $\square$

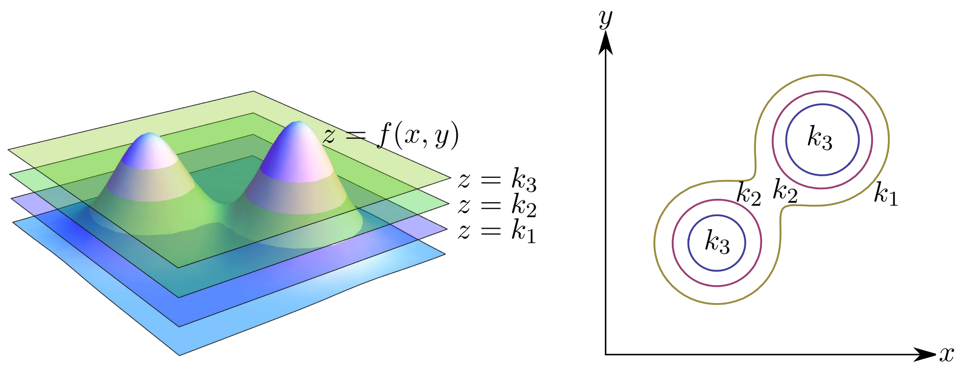


FIGURE **16.3**. **Left**: Cross sections of the graph $z = f(x, y)$ by horizontal planes $z = k_i$, $i = 1, 2, 3$, are level curves $f(x, y) = k_i$ of the function f. **Right**: Contour map of the function f consists of level sets (curves) $f(x, y) = k_i$. The number k_i indicates the value of f along each level curve.

The concept of the graph is obviously hard to extend to functions of more than two variables. The graph of a function of three variables would be a three-dimensional surface in four-dimensional space! So the qualitative behavior of a function of three variables should be studied by different graphical means.

16.3. Level Sets. When visualizing the shape of quadric surfaces, the method of cross sections by coordinate planes has been helpful. It can also be applied to visualize the shape of the graph $z = f(x, y)$. In particular, consider the cross sections of the graph with horizontal planes $z = k$. The curve of intersection is defined by the equation $f(x, y) = k$. Continuing the analogy that $f(x, y)$ defines the height of a mountain, a hiker traveling along the path $f(x, y) = k$ does not have to climb or descend as the elevation along the path remains constant.

DEFINITION **16.3**. (Level Sets).
The level sets of a function f are subsets of the domain of f on which the function has a fixed value; that is, they are determined by the equation $f(\mathbf{r}) = k$, where k is a number.

For functions of two variables, the equation $f(x, y) = k$ generally defines a curve, but not necessarily so. For example, if $f(x, y) = x^2 + y^2$, then the equation $x^2 + y^2 = k$ defines concentric circles of radii $\sqrt{k}$ for any $k > 0$.

However, for $k = 0$, the level set consists of a single point $(x, y) = (0, 0)$. If $k < 0$, then the corresponding level sets are *empty*. Clearly, if k is not from the range of a function, then the corresponding level set is empty. If f is a constant function on D, then it has just one non-empty level set; it coincides with the entire domain D. In general, a level set of a function of two variables may contain curves, isolated points, and even portions of the domain with nonzero area.

Suppose that each level set $f(x, y) = k$ for some k in the range of f is a curve or a collection of curves. These curves are referred to as *level curves* of a function. Recall that a curve in a plane can be described by parametric equations $x = x(t)$, $y = y(t)$ where $x(t)$ and $y(t)$ are continuous functions on an interval $a \le t \le b$. Therefore the equation $f(x, y) = k$ defines a curve (or a collection of curves) if there exist continuous functions $x(t)$ and $y(t)$ (or a collection of pairs of continuous functions) such that $f(x(t), y(t)) = k$ for all values of t from an interval. For example, let

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

Then the level set $f(x, y) = 4$ is an ellipse

$$x^2 + \frac{y^2}{4} = 4 \quad \Leftrightarrow \quad \frac{x^2}{2^2} + \frac{y^2}{4^2} = 1$$

with semi-axes $a = 2$ and $b = 4$. The ellipse is also described by parametric equations

$$x = 2\cos t\,, \quad y = 4\sin t\,, \quad 0 \le t \le 2\pi\,.$$

Indeed, $x^2/2^2 + y^2/4^2 = \cos^2 t + \sin^2 t = 1$ for all t.

EXAMPLE **16.4**. *Determine the level set $f(x, y) = 1$ of the function $f(x, y) = (3 - x^2 - y^2)^2$. If the level set contains curves, find their parameterization.*

SOLUTION: It follows from the equation for the level set

$$(3 - x^2 - y^2)^2 = 1 \quad \Leftrightarrow \quad 3 - x^2 - y^2 = \pm 1 \quad \Leftrightarrow \quad x^2 + y^2 = 3 \pm 1$$

The equations $x^2 + y^2 = 4$ and $x^2 + y^2 = 2$ describe circles of radii 2 and $\sqrt{2}$, respectively. Their parametric equations may chosen in the form

$$x = a\cos t\,, \quad y = a\sin t\,, \quad 0 \le t \le 2\pi\,,$$

where $a = 2$ or $a = \sqrt{2}$. $\square$

DEFINITION **16.4**. (Contour Map).
A collection of level curves of a function of two variables is called a contour map *of the function.*

The concept of level curves and a contour map of a function of two variables are illustrated in Fig. **16.3**. The contour map of the function in

Example **16.3** consists of ellipses. Indeed, the range is the interval $[0, 1]$. For any $0 \le k < 1$, a level curve is an ellipse,

$$1 - \left(\frac{x}{2}\right)^2 - \left(\frac{y}{3}\right)^2 = k^2 \quad \Leftrightarrow \quad \left(\frac{x}{2}\right)^2 + \left(\frac{y}{3}\right)^2 = 1 - k^2$$

or, after dividing both sides of the latter equation by $k^2 - 1 > 0$,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\,, \quad a = 2\sqrt{1-k^2}\,, \quad b = 3\sqrt{1-k^2}\,.$$

The level set for $k = 1$ consists of a single point, the origin. Larger values of the function (larger k) correspond to smaller ellipses (the semi-axes a and b decrease with increasing k). So, the contour map of the function consists of ellipses that have a common center (the origin) and lie inside the ellipse with $a = 2$ and $b = 3$.

A contour map is a useful tool for studying the qualitative behavior of a function. Consider the contour map that consists of level curves C_i, $i = 1, 2, ...,$ $f(x, y) = k_i$, where $k_{i+1} - k_i = \Delta k$ is fixed. The values of the function along the neighboring curves C_i and C_{i+1} differ by Δk. So, in the region where the level curves are dense (close to one another), the function $f(x, y)$ changes rapidly. Indeed, let P be a point of C_i and let Δs be the distance from P to C_{i+1} along the normal to C_i at P. Then the slope of the graph of f or the rate of change of f at P in that direction is $\Delta k/\Delta s$. Thus, the closer the curves C_i are to one another, the faster the function changes. Contour maps are used in topography to indicate the steepness of mountains on maps.

EXAMPLE **16.5**. *Describe the level sets (a contour map) of the function* $f(x, y) = (a^2 - (x^2 + y^2/4))^2$.

SOLUTION: The function depends only on single combination variables x and y:

$$u^2 = x^2 + y^2/4 \quad \Rightarrow \quad f(x, y) = (a^2 - u^2)^2\,.$$

Therefore the level sets $f(x, y) = k \ge 0$ are sets on which u^2 has a constant value, that is, they contain ellipses:

$$f(x, y) = k \quad \Rightarrow \quad u^2 = a^2 \pm \sqrt{k} \quad \Rightarrow \quad x^2 + \frac{y^2}{4} = a^2 \pm \sqrt{k}$$

The level set $k = 0$, is the ellipse $x^2/a^2 + y^2/(2a)^2 = 1$. The level sets with $0 < k < a^4$ contain two ellipses because $a^2 \pm \sqrt{k} > 0$. For $k = a^4$, the level set consists of the ellipse $u^2 = 2a^2$ and the point $(x, y) = (0, 0)$. The level set for $k > a^4$ is the ellipse $u^2 = a^2 + \sqrt{k}$ (the other root becomes negative and the corresponding point set is empty). So the contour map contains the ellipse $x^2/a^2 + y^2/(2a)^2 = 1$ along which the function attains its absolute minimum $f(x, y) = 0$. As the value of k increases, this ellipse splits into two ellipses. The smaller ellipse is shrinking with increasing k, while the larger ellipse is expanding. At $k = a^4$ the smaller ellipse collapses to a point and disappear. This shows that the function f has a local maximum at the origin, $f(0,0) = a^4$. The larger ellipse keeps expanding in size with increasing k. The graph of f looks like a Mexican hat stretched along the y axis (by a factor 2). □

16.4. Level Surfaces. In contrast to the graph, the method of level curves uses only the domain of a function of two variables to study its behavior. Therefore the concept of level sets can be useful to study the qualitative behavior of functions of three variables. In general, the equation $f(x, y, z) = k$ defines a surface in space, but not necessarily so as in the case of functions of two variables. The level sets of the function $f(x, y, z) = x^2 + y^2 + z^2$ are concentric spheres $x^2 + y^2 + z^2 = k$ for $k > 0$, the level set for $k = 0$ contains just one point (the origin), and the level sets are empty for $k < 0$.

Intuitively, a surface in space can be obtained by a continuous deformation (without breaking) of a part of a plane, just like a curve is obtained by a continuous deformation of a line segment. Let S be a nonempty point set in space. A *neighborhood* of a point P of S is a collection of all points of S whose distance from P is less than a number $\delta > 0$. In particular, a neighborhood of a point in a plane is a disk centered at that point and the boundary circle does not belong to the neighborhood. If every point of a subset D of a plane has a neighborhood that is contained in D, then the set D is called *open*. In other words, for every point P of an open region D in a plane there is a disk of a sufficiently small radius that is centered at P and contained in D. *A point set S is a surface in space if every point of S has a neighborhood that can be obtained by a continuous deformation (or a deformation without breaking) of an open set in a plane and this deformation has a continuous inverse.* This is analogous to the definition of a curve as a point set in space given in Section **10.3**.

When the level sets of a function of three variables are surfaces (or collections of surfaces), they are called *level surfaces*. The shape of the level surfaces may be studied, for example, by the method of cross sections with coordinate planes. A collection of level surfaces S_i, $f(x, y, z) = k_i$, $k_{i+1} - k_i = \Delta k$, $i = 1, 2, ...$, can be depicted in the domain of f. If P_0 is a point on S_i and P is the point on S_{i+1} that is the closest to P_0, then the ratio $\Delta k/|P_0P|$ determines the maximal rate of change of f at P. So the closer the level surfaces S_i are to one another, the faster the function changes (see the left panel of Fig. **16.4**).

Example **16.6**. *Sketch and/or describe the level surfaces of the function* $f(x, y, z) = z/(1 + x^2 + y^2)$.

Solution: The domain is the entire space, and the range contains all real numbers. The equation $f(x, y, z) = k$ can be written in the form

$$z - k = k(x^2 + y^2)\,.$$

If $k \neq 0$, this equation defines a circular paraboloid whose symmetry axis is the z axis and whose vertex is at $(0, 0, k)$. For $k = 0$, the level surface is the

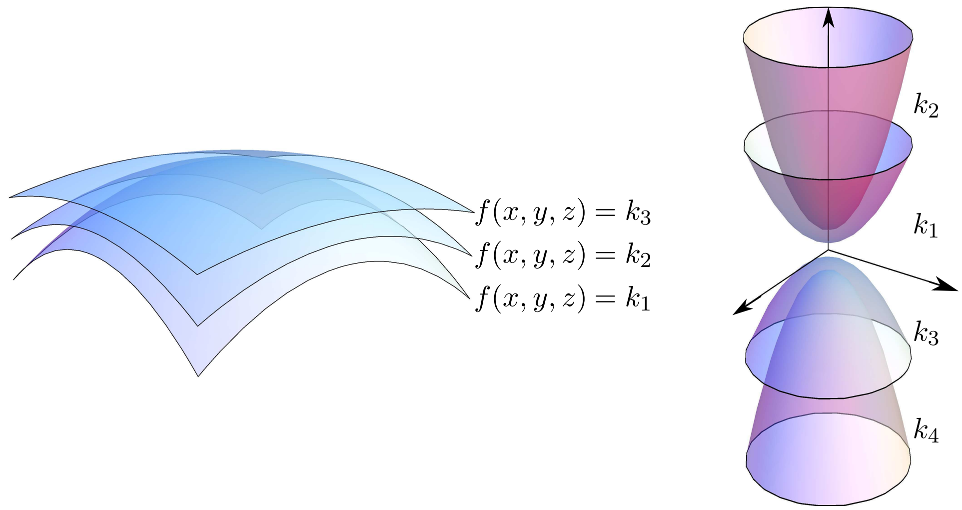


FIGURE **16.4**. **Left**: A level surface of a function f of three variables is a surface in the domain of f on which the function attains a constant value k, i.e., it is defined by the equation $f(x,y,z)=k$. Here three level surfaces are depicted.
Right: Level surfaces of the function studied in Example **16.6**. Here $k_2>k_1>0$ and $k_4<k_3<0$. The level surface $f(x,y,z)=0$ is the xy plane, $z=0$.

xy plane ($z=0$). For $k>0$, the level surfaces are paraboloids above the xy plane, i.e., they are concave upward (see the right panel of Fig. **16.4**). The paraboloid rises faster with increasing k. For $k<0$, the paraboloids are below the xy plane (i.e., they are concave downward). □

16.5. Applications to mechanics. Consider a motion of a particle of mass m along a line under a force $F=-V'(x)$ where $x=x(t)$ is the position of the particle on the line as a function of time t and $V(x)$ is a continuously differentiable function, called the potential energy. By Newton's law,

$$ma = F \quad \Rightarrow \quad mx''(t) = -V'(x(t)) \quad \Rightarrow \quad mx''(t)+V'(x(t)) = 0\,.$$

Let us multiply this equation of motion by the velocity $x'(t)$. Then by the chain rule

$$mx''(t)x'(t)+V'(x(t))x'(t)=0 \quad \Rightarrow \quad \frac{d}{dt}\Big(\frac{1}{2}m(x'(t))^2+V(x(t))\Big)=0\,.$$

Put $p=mx'$ which is called a *momentum* of the particle. Then it follows from the above equation that the quantity

$$E(x,p)=\frac{p^2}{2m}+V(x)$$

remains constant during the motion:

$$\frac{d}{dt}\Big(\frac{(p(t))^2}{2m} + V(x(t))\Big) = 0 \quad \Rightarrow \quad E(x(t), p(t)) = \text{const}\,.$$

The function $E(x, p)$ is called the *total energy* of the particle (the sum of the kinetic and potential energies). The level sets of the energy function $E(x, p) = k$ are called a *phase space portrait* of a dynamical system. Given the initial conditions $x(0) = x_0$, $p(0) = p_0$, the system evolves so that it remains in the set $E(x, p) = E(x_0, p_0)$. In particular, if a level set is a curve that contain the initial state point (x_0, p_0), then the solution of the equation of motion $(x(t), p(t))$ traverses the curve $E(x, p) = E(x_0, p_0)$. So, *the phase space portrait describes all possible motions (for all possible initial conditions) of a given dynamical system.*

For example, according to Hooke's law, small vibrations of a mass attached to a spring are described by the equation

$$mx''(t) = -\lambda x(t)\,,$$

where $\lambda > 0$ and $x(t)$ is the coordinate of the position of the mass relative to the equilibrium position that is set at $x = 0$. Since $V'(x) = \lambda x$, the potential elastic energy is $V(x) = \lambda x^2/2$ (adopting the convention that the minimal value of the energy is set to zero, $V(0) = 0$). The phase space portrait of this dynamical system is a contour map of the energy function

$$E(x, p) = \frac{p^2}{2m} + \frac{\lambda x^2}{2}$$

It consists of ellipses, larger initial energies correspond to wider ellipses. It follows from the phase space portrait that *the motion remains bounded for any initial conditions* because for any initial energy E_0:

$$-(2E_0/\lambda)^{1/2} \le x(t) \le (2E_0/\lambda)^{1/2}$$

since $(x(t), p(t))$ traverses the ellipse $E(x, p) = E_0$.

16.6. Exercises.

1–13. Find and sketch the domain of each of the following functions:

1. $f(x, y) = x/y\,.$
2. $f(x, y) = x/(x^2 + y^2)\,.$
3. $f(x, y) = x/(y^2 - 4x^2)\,.$
4. $f(x, y) = \ln(9 - x^2 - (y/2)^2)\,.$
5. $f(x, y) = \sqrt{1 - (x/2)^2 - (y/3)^2}\,.$
6. $f(x, y) = \sqrt{4 - x^2 - y^2} + 2x \ln y\,.$
7. $f(x, y) = \sqrt{4 - x^2 - y^2} + x \ln y^2\,.$
8. $f(x, y) = \sqrt{4 - x^2 - y^2} + \ln(x^2 + y^2 - 1)\,.$
9. $f(x, y, z) = x/(yz)$
10. $f(x, y, z) = x/(x - y^2 - z^2)$
11. $f(x, y, z) = \ln(1 - z + x^2 + y^2)\,.$

12. $f(x, y, z) = \sqrt{x^2 - y^2 - z^2} + \ln(1 - x^2 - y^2 - z^2)$.
13. $f(t, \mathbf{r}) = (t^2 - \|\mathbf{r}\|^2)^{-1}$, $\mathbf{r} = \langle x_1, x_2, ..., x_m \rangle$.

14–19. For each of the following functions, sketch a contour map and use it to sketch the graph:

14. $f(x, y) = x^2 + 4y^2$.
15. $f(x, y) = xy$.
16. $f(x, y) = x^2 - y^2$.
17. $f(x, y) = \sqrt{x^2 + 9y^2}$.
18. $f(x, y) = \sin x$.
19. $f(x, y) = y^2 + (1 - \cos x)$.

20–25. Describe and sketch the level sets of each of the following functions:

20. $f(x, y, z) = x + 2y + 3z$.
21. $f(x, y, z) = x^2 + 4y^2 + 9z^2$.
22. $f(x, y, z) = z + x^2 + y^2$.
23. $f(x, y, z) = x^2 + y^2 - z^2$.
24. $f(x, y, z) = \ln(x^2 + y^2 - z^2)$.
25. $f(x, y, z) = \ln(z^2 - x^2 - y^2)$.

26–32. Sketch the level sets of each of the following functions. Here $\min(a, b)$ and $\max(a, b)$ denote the smallest number and the largest number of a and b, respectively, and $\min(a, a) = \max(a, a) = a$.

26. $f(x, y) = |x| + y$.
27. $f(x, y) = |x| + |y| - |x + y|$.
28. $f(x, y) = \min(x, y)$.
29. $f(x, y) = \max(|x|, |y|)$.
30. $f(x, y) = \text{sign}(\sin(x)\sin(y))$; here $\text{sign}(a)$ is the sign function, it has the values 1 and -1 for positive and negative a, respectively.
31. $f(x, y, z) = (x + y)^2 + z^2$.
32. $f(x, y) = \tan^{-1}\left(\frac{2ay}{x^2+y^2-a^2}\right)$, $\quad a > 0$.

33–36. Explain how the graph $z = g(x, y)$ can be obtained from the graph of $f(x, y)$ if

33. $g(x, y) = k + f(x, y)$, where k is a constant;
34. $g(x, y) = mf(x, y)$, where m is a nonzero constant;
35. $g(x, y) = f(x - a, y - b)$, where a and b are constants;
36. $g(x, y) = f(px, qy)$, where p and q are nonzero constants.

37–39. Given a function $f(x, y)$, sketch the graphs of the function $g(x, y)$ defined in Exercises **33–36**. Analyze carefully various cases for values of the constants (for example, $m > 0$, $m < 0$, $p > 1$, $0 < p < 1$, and $p = -1$, etc.)

37. $f(x, y) = x^2 + y^2$.
38. $f(x, y) = xy$.
39. $f(x, y) = (a^2 - x^2 - y^2)^2$.

40. Find $f(u)$ if $f(x/y) = \sqrt{x^2 + y^2}/x$, $x > 0$.
41. Find $f(x, y)$ if $f(x + y, y/x) = x^2 - y^2$.
42. Let $z = \sqrt{y} + f(\sqrt{x} - 1)$. Find the functions z and f if $z = x$ when $y = 1$.
43. Graph the function $F(t) = f(\cos t, \sin t)$ where $f(x, y) = 1$ if $y \geq x$ and $f(x, y) = 0$ if $y < x$. Give a geometrical interpretation of the graph of F as an intersection of two surfaces.
44–47. Let $f(u)$ be a continuous function for all real u. Investigate the relation between the shape of the graph of f and the shape of the following surfaces:

44. $z = f(y - ax)$.
45. $z = f(\sqrt{x^2 + y^2})$.
46. $z = f(-\sqrt{x^2 + y^2})$.
47. $z = f(x/y)$.

48. Suppose that a potential energy of a particle of mass m is $V(x) = -\lambda(x^2 - a^2)^2$. Sketch the phase space portrait of this dynamical system in two cases $\lambda > 0$ and $\lambda < 0$. Determine the set of all initial conditions at which the motion remains bounded in each case.
49. A *pendulum* is a weight suspended on a rigid rod from a pivot so that it can swing freely. Consider a planar motion of the pendulum. Then its position can be described by the angle $\theta(t)$ counted counterclockwise from its equilibrium position as a function of time t. If L is the length of the pendulum and m is its mass, then its total energy is known to be $E = \frac{1}{2}mL^2\dot{\theta}^2 + mgL(1 - \cos\theta)$ where $\dot{\theta} = d\theta/dt$. Put $p = mL\dot{\theta}$ (the momentum of the pendulum). Sketch the phase space portrait of the pendulum. Determine the set of all initial conditions at which the pendulum cannot make a full 2π turn about the pivot point.

17. Limits and Continuity

The function $f(x) = \sin(x)/x$ is defined for all reals except $x = 0$. So the domain D of the function contains points arbitrarily close to the point $x = 0$, and therefore the limit of $f(x)$ can be studied as $x \to 0$. It is known from Calculus I that $\sin(x)/x \to 1$ as $x \to 0$. A similar question can be asked for functions of several variables. For example, the domain of the function

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

is the entire plane except the point $(x,y) = (0,0)$. The domain contains points arbitrary close to the origin and one can compute the values of the function at points that are successively closer to the origin to determine whether these values approach a particular number. For example, one can take a straight line $x = t$, $y = t$ and investigate the values of the function on it as $t \to 0$:

$$\lim_{t\to 0} f(t,t) = \lim_{t\to 0} \frac{\sin(2t^2)}{2t^2} = \lim_{u\to 0} \frac{\sin u}{u} = 1\,.$$

In contrast to the one-dimensional case, there are infinitely many curves passing through the origin along which the limit values of the function can be studied. It is then natural to think of the limit, if it exists, as the number to which the values of the function $f(P)$ as P approaches the origin along *any* curve. It is not difficult to show that the values of the above function indeed approach 1 along any curve through $(0,0)$. Let $x = x(t)$, $y = y(t)$ be a parametric curve through the origin such that $x(0) = 0$ and $y(0) = 0$ (recall that $x(t)$ and $y(t)$ are continuous on an interval containing $t = 0$). Then the distance from the origin $R(t) = \sqrt{x^2(t) + y^2(t)}$ tends to 0 as $t \to 0$ by continuity of $x(t)$ and $y(t)$. Therefore

$$\lim_{t\to 0} f(x(t), y(t)) = \lim_{R\to 0} \frac{\sin(R^2)}{R^2} = 1$$

for any parametric curve through the point $(0,0)$.

Consider another example: $f(x,y) = \sqrt{xy}$. The point $(1,-1)$ is not in the domain of the function. Furthermore, consider a disk of a radius less than 1 that is centered at this point. Any such disk contains no point of the domain of f. Evidently, the limit of f does not make any sense at $(1,-1)$ as one cannot investigate the values of the function at points arbitrary close to $(1,-1)$. The domain of the function $f(x,y) = \sqrt{-x^2-y^2}$ consists of a single point $(0,0)$. The limit of this function at the origin also does not make any sense as the function has no value at any point different from the origin. So, first of all one has to describe points for which the very question about the limit of a function makes sense. As noted before, the domain of a function f of several variables is a point set in an n-dimensional Euclidean space. The distance between two points $\mathbf{x} = (x_1, x_2, ..., x_m)$ and

$\mathbf{y} = (y_1, y_2, ..., y_m)$ is the Euclidean distance

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_m - y_m)^2}\,.$$

DEFINITION **17.1**. (Neighborhood of a Point)
A neighborhood *of a point* $\mathbf{r}_0$ *in a Euclidean space is an open ball of radius* $\delta > 0$ *centered at* $\mathbf{r}_0$*:*

$$N_\delta(\mathbf{r}_0) = \{\mathbf{r} \mid \|\mathbf{r} - \mathbf{r}_0\| < \delta\}\,.$$

The number δ *is called the* radius *of* $N_\delta(\mathbf{r}_0)$*.*

In other words, a neighborhood of a point P is the collection of all points whose distance from P is less than some positive number. A point P may or may not be in the domain of a function f. The limit of f at a point P would not make any sense if there is a neighborhood of P which contains no points of the domain of the function other than possibly the point P itself because it would be impossible to investigate the values of the function at points arbitrary close to P.

DEFINITION **17.2**. (Limit Point of a Set).
A point $\mathbf{r}_0$ *is said to be a* limit point *of a set* D *if any neighborhood* $N_\delta(\mathbf{r}_0)$ *contains a point* $\mathbf{r}$ *of* D *and* $\mathbf{r} \neq \mathbf{r}_0$*.*

Let the domain of a function f be a set D. A limit point $\mathbf{r}_0$ of D may or may not be in D (or f may or may not have a value at $\mathbf{r}_0$), but the limit of the function f always makes sense at a limit point of the domain D because a neighborhood of $\mathbf{r}_0$ of an arbitrary small radius contains a point of D.

17.1. Limits of Functions of Several Variables. Intuitively, if the values a function $f(\mathbf{r})$ near a point $\mathbf{r}_0$ *get arbitrary close* to a number c and *stay arbitrary close* to c for *all* $\mathbf{r} \neq \mathbf{r}_0$ in a suitably small neighborhood of $\mathbf{r}_0$, then f is said to have the limit at $\mathbf{r}_0$ that is equal to c.

For example, the values of $f(x, y) = x^3 y$ get arbitrary close to $c = 0$ if x and y are suitably small. If $R = (x^2 + y^2)^{1/2}$ is the distance from the origin to a point (x, y), then $|x| \leq R$ and $|y| \leq R$. Therefore

$$|f(x, y)| = |x|^3 |y| \leq R^4$$

which shows that the values of $f(x, y)$ stay arbitrary close to zero throughout a neighborhood of the origin of a suitably small radius, and it is concluded that the function $f(x, y) = x^3 y$ has the limit $c = 0$ at the point $(0, 0)$.

Consider the function

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

The origin is not in the domain of the function, but it is a limit point of the domain. So, one can consider the limit of this function when (x, y) approach $(0, 0)$. The values of the function along the coordinate axes vanish, $f(0, y) = f(x, 0)$. One can say that the values of f get arbitrary close to

0 in any neighborhood of the origin. However, *they do not stay arbitrary close to* 0 *throughout any neighborhood* of the origin. Indeed, the values of the function on the line $y = x$ are equal to $\frac{1}{2} = f(x, x)$. Evidently, $\frac{1}{2}$ is not arbitrary close to 0, but the line $y = x$ passes through any neighborhood of the origin. Thus, it is concluded that the function f has no limit at the origin.

The precise definition of limits. Although in the above examples it was possible to give a precise meaning to the words "arbitrary close" and "stay arbitrary close" in a "suitably small neighborhood", the general situation is not that simple. The precise definition of the limit is analogous to that given in Calculus I for functions of a single variable.

Definition **17.3**. (Limit of a Function of Several Variables). *Let f be a function of several variables whose domain is a set D in a Euclidean space. Let $\mathbf{r}_0$ be a limit point of D. Then the limit of $f(\mathbf{r})$ at $\mathbf{r}_0$ is said to be a number c if, for every number $\varepsilon > 0$, there exists a number $\delta > 0$ such that if $\mathbf{r}$ is in D and $0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta$, then $|f(\mathbf{r}) - c| < \varepsilon$. In this case, one writes*

$$\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = c \quad \text{or} \quad f(\mathbf{r}) \to c \quad \text{as} \quad \mathbf{r} \to \mathbf{r}_0$$

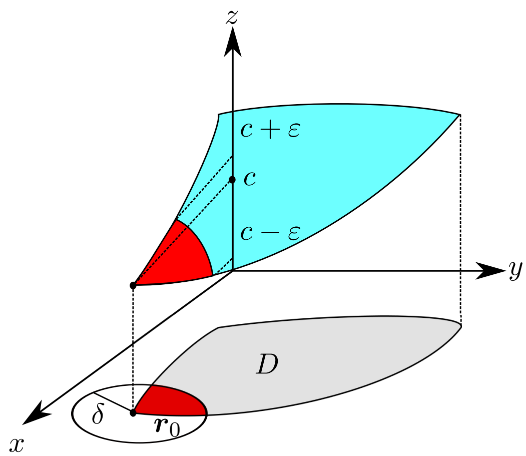
pronounced "the limit of $f(\mathbf{r})$ as $\mathbf{r}$ approaches (or goes to) $\mathbf{r}_0$ is c".

The number $|f(\mathbf{r}) - c|$ determines a deviation of the value of f at the point $\mathbf{r}$ from the number c. So, an *arbitrary* positive number ε sets a numerical measure for how "close" the values of f to c are. The existence of the limit means that no matter how small the number ε is set to be, one can find a ball of a sufficiently small radius δ and centered at the limit point $\mathbf{r}_0$ such that the values of the function at *all points* in this ball (except its center $\mathbf{r}_0$) deviate from the limit value c no more than ε, that is,

$$c - \varepsilon < f(\mathbf{r}) < c + \varepsilon \quad \text{whenever} \quad 0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta\,.$$

In other words, the existence of such δ guarantees that the values of f "stay arbitrary close" to c throughout a "suitably small neighborhood of $\mathbf{r}_0$". The condition $0 < \|\mathbf{r} - \mathbf{r}_0\|$ ensures that $\mathbf{r}$ does not coincide with $\mathbf{r}_0$. Note that f is not even defined at $\mathbf{r}_0$ if $\mathbf{r}_0$ is not in D. In the case of a function of two variables, this definition is illustrated in the left panel of Fig. **17.1**.

Suppose that the limit of f exists at $\mathbf{r}_0$ and equals c. Then the conditions stated in Definition **17.3** also imply that the limit of values of f along any curve through $\mathbf{r}_0$ is also c. Indeed, take a curve that ends at the limit point $\mathbf{r}_0$ and fix $\varepsilon > 0$ (see the right panel of Fig. **17.1**). Then, by the existence of the limit c, there is a ball of radius $\delta = \delta(\varepsilon, \mathbf{r}_0) > 0$ centered at $\mathbf{r}_0$ such that the values of f lie in the interval $c - \varepsilon < f(\mathbf{r}) < c + \varepsilon$ for all points $\mathbf{r} \neq \mathbf{r}_0$ in the ball and hence for all points of the portion of the curve in the ball. For any $\delta > 0$, the ball $N_\delta(\mathbf{r}_0)$ contains points of the curve other than $\mathbf{r}_0$. Since $\varepsilon > 0$ can be chosen arbitrary small, the limit along any curve through $\mathbf{r}_0$



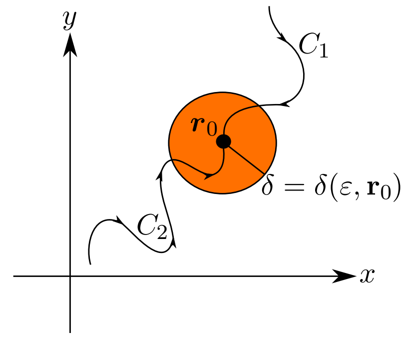


FIGURE **17.1**. **Left**: An illustration of Definition **17.3** in the case of a function f of two variables. Given a positive number ε, consider two horizontal planes $z = c - \varepsilon$ and $z = c + \varepsilon$. Then one can always find a number $\delta > 0$ and the disk N_δ centered at $\mathbf{r}_0$ such that the portion of the graph $z = f(\mathbf{r})$ above the intersection N_δ with D lies between the planes: $c - \varepsilon < f(\mathbf{r}) < c + \varepsilon$. The radius $\delta > 0$ of N_δ depends, generally, on ε and the limit point $\mathbf{r}_0$. **Right**: The independence of the limit of a path along which the limit point $\mathbf{r}_0$ is approached. For every path leading to $\mathbf{r}_0$, there is a part of it that lies in N_δ. The values of f along this part of the path deviates from c no more than any preassigned number $\varepsilon > 0$.

must be c. This is to be compared with the one-dimensional analog: if the limit of $f(x)$ exists as $x \to x_0$, then the right $x \to x_0^+$ and left $x \to x_0^-$ limits exist and are equal (and vice versa). The limit of a function along a curve is rigorously studied in the next section.

Suppose that $f(\mathbf{r}) = c$ (a constant) for all $\mathbf{r} \neq \mathbf{r}_0$, while f is not defined at $\mathbf{r}_0$. Then

$$\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = c\,.$$

Indeed, let us fix $\varepsilon > 0$. Then in *any* neighborhood $N_\delta(\mathbf{r}_0)$ of radius $\delta > 0$, the deviation of values of f from c vanishes:

$$|f(\mathbf{r}) - c| = |c - c| = 0 < \varepsilon\,, \quad 0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta\,.$$

In this simple case, the radius δ of a neighborhood in which f deviates from the limit c no more than ε appears to be independent of the value of ε. *This is not so in general.*

EXAMPLE **17.1**. *Show that*

$$\lim_{(x,y,z)\to(0,0,0)}(x^2y + yz^2 - 6z^3) = 0.$$

SOLUTION: The distance between $\mathbf{r} = (x, y, z)$ and the limit point $\mathbf{r}_0 = (0, 0, 0)$ is

$$R = \|\mathbf{r} - \mathbf{r}_0\| = \sqrt{x^2 + y^2 + z^2}\,.$$

Then

$$|x| \le R\,, \quad |y| \le R\,, \quad |z| \le R\,.$$

Let us find an upper bound on the deviation of values of the function from the limit value $c = 0$ in terms of R:

$$|f(\mathbf{r}) - c| = |x^2y + yz^2 - 6z^3| \le |x^2y| + |yz^2| + 6|z^3| \le 8R^3,$$

where the inequality $|a \pm b| \le |a| + |b|$ and $|ab| = |a||b|$ have been used. Next let us show the existence of a neighborhood of $\mathbf{r}_0$ with the properties stated in Definition **17.3**. Fix $\varepsilon > 0$. To establish the existence of $\delta > 0$, note that the inequality $8R^3 < \varepsilon$ or $R < \sqrt[3]{\varepsilon}/2$ guarantees that $|f(\mathbf{r}) - c| < \varepsilon$. So,

$$\text{if} \quad 8R^3 < \varepsilon\,, \quad \text{then} \quad |f(\mathbf{r}) - c| < \varepsilon$$

Given $\varepsilon > 0$, one has to find a ball $R < \delta$ in which the latter inequality holds. Therefore one can take

$$8\delta^3 = \varepsilon \quad \text{or} \quad \delta = \sqrt[3]{\varepsilon}/2$$

so that the condition $R < \delta$ is equivalent to $8R^3 < \varepsilon$ and, hence,

$$|f(\mathbf{r}) - c| < \varepsilon \quad \text{whenever} \quad 0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta = \sqrt[3]{\varepsilon}/2\,.$$

For example, put $\varepsilon = 10^{-6}$. Then, in the interior of a ball of radius $\delta = 0.005$, the values of the function can deviate from $c = 0$ no more than 10^{-6}. □

Note that *the choice of a particular value of $\delta > 0$ is not unique.* In the previous example, one could take any number $0 < \delta \le \sqrt[3]{\varepsilon}/2$ to fulfill the conditions for the existence of the limit. *The radius δ of a neighborhood in which a function f deviates no more than ε from the value of the limit depends on ε and, in general, on the limit point $\mathbf{r}_0$, that is, $\delta = \delta(\varepsilon, \mathbf{r}_0)$.*

EXAMPLE **17.2**. *Let $f(x, y) = xy$. Show that*

$$\lim_{(x,y)\to(x_0,y_0)} f(x, y) = x_0y_0$$

for any point (x_0, y_0).

SOLUTION: The distance between $\mathbf{r} = (x, y)$ and $\mathbf{r}_0 = (x_0, y_0)$ is

$$R = \sqrt{(x - x_0)^2 + (y - y_0)^2}\,.$$

Therefore

$$|x - x_0| \le R \quad \text{and} \quad |y - y_0| \le R\,.$$

Consider the identity

$$xy - x_0y_0 = (x - x_0)(y - y_0) + x_0(y - y_0) + (x - x_0)y_0\,.$$

Then the deviation of f from the limit value $c = x_0y_0$ is bounded as

$$\begin{aligned} |f(x,y)-c| &\le |x-x_0||y-y_0| + |x_0||y-y_0| + |x-x_0||y_0| \\ &\le R^2 + (|x_0|+|y_0|)R = R^2 + 2aR \\ &= (R+a)^2 - a^2\,, \\ a &= \frac{1}{2}(|x_0|+|y_0|)\,. \end{aligned}$$

Now fix $\varepsilon > 0$ and demand that R is such that

$$0 < (R+a)^2 - a^2 < \varepsilon \quad \Rightarrow \quad 0 < R < \sqrt{\varepsilon + a^2} - a\,.$$

Therefore the function f deviates from $c = x_0y_0$ no more than ε in a neighborhood of $\mathbf{r}_0$ of radius $\delta = \sqrt{\varepsilon + a^2} - a$:

$$|xy - x_0y_0| < \varepsilon \quad \text{whenever} \quad \sqrt{(x-x_0)^2 + (y-y_0)^2} < \delta = \sqrt{\varepsilon + a^2} - a$$

The radius of the neighborhood depends on ε and the limit point $\mathbf{r}_0$. □

17.2. An alternative definition of the limit.

Definition 17.4. *A sequence of points $\mathbf{r}_n$, $n = 1, 2, ...$, in a Euclidean space is said to converge to a point $\mathbf{r}_0$ if*

$$\lim_{n\to\infty} \|\mathbf{r}_0 - \mathbf{r}_n\| = 0$$

or, for any $\varepsilon > 0$ there exists an integer N such that

$$\|\mathbf{r}_0 - \mathbf{r}_n\| < \varepsilon \quad \text{for all} \quad n > N$$

In this case, $\mathbf{r}_n$ is also said to approach $\mathbf{r}_0$ as $n \to \infty$ and one writes $\lim_{n\to\infty} \mathbf{r}_n = \mathbf{r}_0$ or $\mathbf{r}_n \to \mathbf{r}_0$ as $n \to \infty$.

In other words, a sequence of points approaches a particular point if the distance between the particular point and the points of the sequence tends to zero. One can also say that only *finitely* many points ($n \le N$) of a convergent sequence are outside of *any* neighborhood $N_\varepsilon(\mathbf{r}_0)$ of the limit point.

As noted in the beginning of this section, for functions of several variables, there are infinitely many ways of how the limit point can be approached from within the domain of the function. More precisely, there are infinitely many sequences of points approaching the same limit point. One can take values of the function $f(\mathbf{r}_n)$ on each such sequence and investigate the limit of the numerical sequence $f(\mathbf{r}_n)$. The question of interest is: What is the relation between the limit of the function at a limit point and the limits of the values of the function on various sequences of points converging to the limit point? The following theorem answers this question.

Theorem 17.1. *Let $\mathbf{r}_0$ be a limit point of the domain D of a function f. Then*

$$\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = c \quad \text{if and only if} \quad \lim_{n\to\infty} f(\mathbf{r}_n) = c$$

for every sequence of points $\mathbf{r}_n$ in D that converges to $\mathbf{r}_0$ and $\mathbf{r}_n \ne \mathbf{r}_0$.

PROOF. Suppose that $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = c$. Choose a sequence of points $\mathbf{r}_n$ that converges to $\mathbf{r}_0$. Fix $\varepsilon > 0$. By Definition **17.3** there is $\delta > 0$ such that $|f(\mathbf{r}) - c| < \varepsilon$ whenever $0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta$. By Definition **17.4**, for any such number $\delta > 0$ there also exists an integer N such that $0 < \|\mathbf{r}_n - \mathbf{r}_0\| < \delta$ for all $n > N$. Therefore

$$|f(\mathbf{r}_n) - c| < \varepsilon \quad \text{for all} \quad n > N$$

which means that the numerical sequence of values of the function $f(\mathbf{r}_n)$ converges to the number c.

Conversely, suppose that the numerical sequence of values of the function $f(\mathbf{r}_n)$ converges to c for *every* sequence of points $\mathbf{r}_n \neq \mathbf{r}_0$ that converges to the point $\mathbf{r}_0$. One has to show that $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = c$. Suppose that this conclusion is *false*. Then by negating Definition **17.3** there should exist *some* $\varepsilon > 0$ such that for *every* $\delta > 0$ one can find a point $\mathbf{r}$ in D (depending on δ) for which

$$|f(\mathbf{r}) - c| \geq \varepsilon \quad \text{but} \quad 0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta\,.$$

Since this property should be true for any δ, it is true if $\delta = \delta_n = 1/n$, where $n = 1, 2, 3, ...$ For each δ_n there is a point $\mathbf{r} = \mathbf{r}_n$ with the above property. This means that there is a sequence of points $\mathbf{r}_n$ approaching $\mathbf{r}_0$ because $\|\mathbf{r}_n - \mathbf{r}_0\| < 1/n \to 0$ as $n \to \infty$, but the corresponding sequence of values of the function $f(\mathbf{r}_n)$ does not converge to c. This is a *contradiction*. So, $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = c$. □

The intuitive idea that the limit of a function, if it exists, should not depend on the way the limit point is approached is now rigorously established in Theorem **17.1**. The conditions under which the limit exists stated in Definition **17.3** and in Theorem **17.1** have been proved to be *equivalent*. So Theorem **17.1** could have been used as a *definition* of the limit of a function of several variables: *the limit of a function f exists at a limit point $\mathbf{r}_0$ of the domain of f and equals c if the limit of the numerical sequence of values of f on every sequence of points not containing $\mathbf{r}_0$ and converging to $\mathbf{r}_0$ exists and equals c.* Then Definition **17.3** becomes a theorem to be proved in this approach.

17.3. Properties of the Limit. The basic properties of limits of functions of one variable discussed in Calculus I are extended to the case of functions of several variables.

THEOREM **17.2**. (Properties of the Limit).
Let f and g be functions of several variables that have a common domain. Let c be a number. Suppose that $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = p$ *and* $\lim_{\mathbf{r}\to\mathbf{r}_0} g(\mathbf{r}) = q$.

Then the following properties hold:

$$\lim_{\mathbf{r}\to\mathbf{r}_0}(cf(\mathbf{r})) = c\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = cp,$$
$$\lim_{\mathbf{r}\to\mathbf{r}_0}(g(\mathbf{r})+f(\mathbf{r})) = \lim_{\mathbf{r}\to\mathbf{r}_0} g(\mathbf{r}) + \lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = q+p,$$
$$\lim_{\mathbf{r}\to\mathbf{r}_0}(g(\mathbf{r})f(\mathbf{r})) = \lim_{\mathbf{r}\to\mathbf{r}_0} g(\mathbf{r}) \lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = qp,$$
$$\lim_{\mathbf{r}\to\mathbf{r}_0}\frac{g(\mathbf{r})}{f(\mathbf{r})} = \frac{\lim_{\mathbf{r}\to\mathbf{r}_0} g(\mathbf{r})}{\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})} = \frac{q}{p}, \quad \text{if} \quad p \neq 0.$$

A proof of Theorem **17.2** is left to the reader as an exercise. As a hint, note that, by Theorem **17.1**, the above properties of the limit can be equivalently restated in terms of the basic limits laws for numerical sequences which have been established in Calculus II.

Squeeze Principle. The solution to Example **17.1** employs a rather general strategy to verify whether a particular number c is the limit of $f(\mathbf{r})$ as $\mathbf{r}\to\mathbf{r}_0$.

THEOREM **17.3**. (Squeeze Principle).
Let the functions of several variables g, f, and h have a common domain D and $g(\mathbf{r}) \le f(\mathbf{r}) \le h(\mathbf{r})$ for any $\mathbf{r}\in D$. If the limits of $g(\mathbf{r})$ and $h(\mathbf{r})$ as $\mathbf{r}\to\mathbf{r}_0$ exist and equal a number c, then the limit of $f(\mathbf{r})$ as $\mathbf{r}\to\mathbf{r}_0$ exists and equals c, that is,

$$g(\mathbf{r}) \le f(\mathbf{r}) \le h(\mathbf{r}) \text{ and } \lim_{\mathbf{r}\to\mathbf{r}_0} g(\mathbf{r}) = \lim_{\mathbf{r}\to\mathbf{r}_0} h(\mathbf{r}) = c \;\Rightarrow\; \lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = c.$$

PROOF. From the hypothesis of the theorem, it follows that

$$0 \le f(\mathbf{r}) - g(\mathbf{r}) \le h(\mathbf{r}) - g(\mathbf{r})\,.$$

Put

$$F(\mathbf{r}) = f(\mathbf{r}) - g(\mathbf{r})\,, \quad H(\mathbf{r}) = h(\mathbf{r}) - g(\mathbf{r})\,.$$

Then

$$0 \le F(\mathbf{r}) \le H(\mathbf{r})$$

implies $|F(\mathbf{r})| \le |H(\mathbf{r})|$ (the positivity of F is essential for this conclusion). By the hypothesis of the theorem and the properties of the limit,

$$H(\mathbf{r}) = h(\mathbf{r}) - g(\mathbf{r}) \to c - c = 0 \quad \text{as} \quad \mathbf{r}\to\mathbf{r}_0\,.$$

Hence, for any $\varepsilon > 0$, there is a number δ such that

$$0 \le |F(\mathbf{r})| \le |H(\mathbf{r})| < \varepsilon \quad \text{whenever} \quad 0 < \|\mathbf{r}-\mathbf{r}_0\| < \delta\,.$$

By Definition **17.3** this means that $\lim_{\mathbf{r}\to\mathbf{r}_0} F(\mathbf{r}) = 0$. By the basic properties of the limit, it is then concluded that

$$f(\mathbf{r}) = F(\mathbf{r}) + g(\mathbf{r}) \to 0 + c = c \quad \text{as} \quad \mathbf{r}\to\mathbf{r}_0\,.$$

$\square$

Alternatively, the squeeze principle for limits of functions can be established from the squeeze principle for numerical sequences by using Theorem

17.1. The details are left to the reader as an exercise. A particular case of the squeeze principle is also useful.

COROLLARY **17.1**. (Simplified Squeeze Principle).
If there exists a function h of one variable such that

$$|f(\mathbf{r}) - c| \le h(R) \to 0 \quad \text{as} \quad \|\mathbf{r} - \mathbf{r}_0\| = R \to 0^+,$$

then $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = c$.

The condition $|f(\mathbf{r}) - c| \le h(R)$ is equivalent to

$$c - h(R) \le f(\mathbf{r}) \le c + h(R)$$

which is a particular case of the hypothesis in the squeeze principle. In Example **17.1**, $h(R) = 8R^3$. In general, the condition $h(R) \to 0$ as $R \to 0^+$ implies that, for any $\varepsilon > 0$, there is an interval $0 < R < \delta(\varepsilon)$ in which $h(R) < \varepsilon$, where the number δ can be found by solving the equation $h(\delta) = \varepsilon$. Hence, $|f(\mathbf{r}) - c| < \varepsilon$ whenever $\|\mathbf{r} - \mathbf{r}_0\| = R < \delta(\varepsilon)$.

EXAMPLE **17.3**. *Show that*

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0, \quad \text{where} \quad f(x,y) = \frac{x^3y - 3x^2y^2}{x^2 + y^2 + x^4}.$$

SOLUTION: Let $R = \sqrt{x^2 + y^2}$ (the distance from the limit point $(0,0)$). Then $|x| \le R$ and $|y| \le R$. Therefore,

$$\frac{|x^3y - 3x^2y^2|}{x^2 + y^2 + x^4} \le \frac{|x|^3|y| + 3x^2y^2}{x^2 + y^2 + x^4} \le \frac{4R^4}{R^2 + x^4} = \frac{4R^2}{1 + (x^4/R^2)} \le 4R^2 .$$

It follows from this inequality that

$$-4(x^2 + y^2) \le f(x,y) \le 4(x^2 + y^2) ,$$

and, by the squeeze principle, $f(x,y)$ must tend to 0 because $\pm 4(x^2 + y^2) = \pm 4R^2 \to 0$ as $R \to 0$. In Definition **17.3**, given $\varepsilon > 0$, a number δ may be chosen as $\sqrt{\varepsilon}/2$ (or smaller). □

17.4. Continuity of Functions of Several Variables. Continuous functions of several variables are defined by analogy with continuous functions of a single variable.

DEFINITION **17.5**. (Continuity).
A function f of several variables with domain D is said to be continuous *at a point $\mathbf{r}_0$ in D if*

$$\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = f(\mathbf{r}_0) .$$

The function f is said to be continuous on D *if it is continuous at every point of D.*

EXAMPLE **17.4**. *Let $f(x,y) = 1$ if $y \ge x$ and let $f(x,y) = 0$ if $y < x$. Determine the region on which f is continuous.*

SOLUTION: The function is continuous at every point (x_0, y_0) if $y_0 \neq x_0$. Indeed, if $y_0 > x_0$, then $f(x_0, y_0) = 1$. On the other hand, for every such point one can find a disk of a sufficiently small radius δ and centered at $(x_0, y_0))$ that lies in the region $y > x$. Therefore, for any $\varepsilon > 0$,

$$|f(\mathbf{r}) - f(\mathbf{r}_0)| = 1 - 1 = 0 < \varepsilon \quad \text{whenever} \quad (x - x_0)^2 + (y - y_0)^2 < \delta^2 ,$$

which means that $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = f(\mathbf{r}_0) = 1$. The same line of reasoning applies to establish the continuity of f at any point (x_0, y_0), where $y_0 < x_0$.

If $\mathbf{r}_0 = (x_0, x_0)$, that is, the point lies on the line $y = x$, then $f(\mathbf{r}_0) = 1$. Any disk centered at such $\mathbf{r}_0$ is split into two parts by the line $y = x$. In one part ($y \geq x$), $f(\mathbf{r}) = 1$, whereas in the other part ($y < x$), $f(\mathbf{r}) = 0$. So, for $0 < \varepsilon < 1$, there is no disk of radius $\delta > 0$ in which $|f(\mathbf{r}) - f(\mathbf{r}_0)| = |f(\mathbf{r}) - 1| < \varepsilon$ because $|f(\mathbf{r}) - 1| = 1$ for $y < x$ in any such disk. The function is not continuous along the line $y = x$ in its domain. □

Discontinuity at a limit point. Suppose that $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = c$. If the limit point $\mathbf{r}_0$ lies in the domain of the function f, then the function has a value $f(\mathbf{r}_0)$ and this value may or may not coincide with the limit value c. In fact, the limit value c does not generally give any information about the possible value of the function at the limit point. For example, if $f(\mathbf{r}) = 1$ everywhere except one point $\mathbf{r}_0$ at which $f(\mathbf{r}_0) = f_0$. Then in every neighborhood $0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta$, $f(\mathbf{r}) = 1$ and, hence, the limit of f as $\mathbf{r} \to \mathbf{r}_0$ exists and equals $c = 1$. When $f_0 \neq 1$, the limit value does not coincide with the value of the function at the limit point. The values of f suffer a jump *discontinuity* when $\mathbf{r}$ reaches $\mathbf{r}_0$, and one says that f is *discontinuous* at $\mathbf{r}_0$.

A discontinuity also occurs when the limit of f as $\mathbf{r} \to \mathbf{r}_0$ does not exist while f has a value at the limit point. Furthermore, if $\mathbf{r}_0$ is a limit point of the domain D of f, but is not in D, then f is *continuously extendable* to a larger domain $D \cup \{r_0\}$ (the union of D and the point $\mathbf{r}_0$) if the limit $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})$ exists (the function is *defined* at $\mathbf{r}_0$ by its limit value at $\mathbf{r}_0$). A function f is said to be *discontinuous at a limit point* $\mathbf{r}_0$ *of its domain* D if

(i) $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})$ does not exist, or
(ii) $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})$ exists, and $\mathbf{r}_0$ is in D, but $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) \neq f(\mathbf{r}_0)$.

Note that the notion of "discontinuity at a limit point" of a function does not always mean that the function is not continuous at a limit point because the limit point may not be in the domain of the function (the function has no value at that point). It means that the function is *either* not continuous at a limit point (if the limit point is in the domain) *or* not continuously extendable to the limit point (if the limit point is not in the domain). For example, one says that the function $f(x) = 1/x$ is discontinuous at $x = 0$. Strictly speaking, this function is *continuous* on its domain (all $x \neq 0$). Therefore one can also say that f is a continuous function of x that is discontinuous at $x = 0$! This terminological paradox is eliminated if one keeps in mind that, in this case, the term "discontinuous at $x = 0$" refers to the fact that $f(x)$ is *not continuously extendable* to the limit point $x = 0$ of its domain. The function $g(x) = x^2/x$ is also not defined at $x = 0$ (the rule makes no sense at $x = 0$). But it can be continuously extended to $x = 0$ by $g(0) = 0$ and therefore has no discontinuity at $x = 0$.

17.5. Properties of continuous functions. The following theorem is a simple consequence of the basic properties of the limit.

THEOREM **17.4**. (Properties of Continuous Functions).
If f and g are continuous on D and q is a number, then $qf(\mathbf{r})$, $f(\mathbf{r}) + g(\mathbf{r})$, and $f(\mathbf{r})g(\mathbf{r})$ are continuous on D, and $f(\mathbf{r})/g(\mathbf{r})$ is continuous at any point on D for which $g(\mathbf{r}) \neq 0$.

The use of the definition to establish the continuity of a function defined by an algebraic rule is not always convenient. The following two theorems are helpful when studying the continuity of a given function. For an ordered m-tuple $\mathbf{r} = \langle x_1, x_2, ..., x_m \rangle$, the function

$$f(\mathbf{r}) = x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m} ,$$

where $k_1, k_2, ..., k_m$ are nonnegative integers, is called a *monomial of degree* $N = k_1 + k_2 + \cdots + k_m$. For example, for two variables, monomials of degree $N = 3$ are

$$x^3 , \quad x^2 y , \quad xy^2 , \quad y^3 .$$

A function f that is a linear combination of monomials is called a *polynomial function*. For example, the function

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

is a polynomial of three variables. The ratio of two polynomial functions is called a *rational function*.

THEOREM **17.5**. (Continuity of Polynomial and Rational Functions).
Let f and g be polynomial functions of several variables. Then they are continuous everywhere, and the rational function $f(\mathbf{r})/g(\mathbf{r})$ is continuous at any point $\mathbf{r}_0$ if $g(\mathbf{r}_0) \neq 0$.

PROOF. Let $f(\mathbf{r}) = c$ be a constant function. Take a sequence of points $\mathbf{r}_n$ converging to $\mathbf{r}_0$ and $\mathbf{r}_n \neq \mathbf{r}_0$. Then the numerical sequence $f(\mathbf{r}_n) = c$ converges to c. By Theorem **17.1**, $f(\mathbf{r}) \to c = f(\mathbf{r}_0)$ as $\mathbf{r} \to \mathbf{r}_0$. So, a constant function is continuous everywhere. Let $g_j(\mathbf{r}) = x_j$ (the j^{th} coordinate of a point $\mathbf{r}$). Let $\mathbf{r}_0 = \langle a_1, a_2, ..., a_m \rangle$. Since $|x_j - a_j| \leq \|\mathbf{r} - \mathbf{r}_0\| = R$, the function g_j is continuous everywhere by the squeeze principle:

$$|g_j(\mathbf{r}) - g_j(\mathbf{r}_0)| = |x_j - a_j| \leq R \to 0 \quad \text{as} \quad R \to 0^+ .$$

A monomial of any degree is a product of the functions g_j, $j = 1, 2, ..., m$, or a constant function. By the properties of continuous functions (Theorem **17.4**), any monomial is a function continuous everywhere. A polynomial is a linear combination of monomials and therefore is also a function continuous everywhere. A rational function is continuous as the ratio of two continuous functions (provided the denominator does not vanish). □

THEOREM **17.6**. (Continuity of a Composition).
Let $g(u)$ be continuous on the interval $u \in [a, b]$ and let h be a function of several variables that is continuous on D and has the range $[a, b]$. The composition $f(\mathbf{r}) = g(h(\mathbf{r}))$ is continuous on D.

This theorem follows from Theorem **17.1** and Definition **17.5**. A proof is left to the reader as an exercise.

Basic functions studied in Calculus I, $\sin u$, $\cos u$, e^u, $\ln u$, and so on, are continuous functions on their domains. If $f(\mathbf{r})$ is a continuous function of several variables, the elementary functions whose argument is replaced by $f(\mathbf{r})$ are continuous functions. In combination with the properties of continuous functions, the composition rule defines a large class of continuous functions of several variables, which is sufficient for many practical applications.

EXAMPLE **17.5**. *Find the limit*

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{e^{xz}\cos(xy+z^2)}{x+yz+3xz^4+(xyz-2)^2}$$

SOLUTION: The function is a ratio. The denominator is a polynomial and hence continuous. Its limit value is $(-2)^2 = 4 \neq 0$. The function e^{xz} is a composition of the exponential e^u and the polynomial $u = xy$. So it is continuous. Its value is 1 at the limit point. Similarly, $\cos(xy + z^2)$ is continuous as a composition of $\cos u$ and the polynomial $u = xy + z^2$. Its value is 1 at the limit point. The ratio of continuous functions is continuous and the limit is 1/4. □

17.6. Study Problems.

Problem **17.1**. *Consider two rational functions $f(x, y) = x^2/(x^2+y^2)$ and $g(x, y) = x^4/(x^2 + y^2)$. Find all points of discontinuity of these functions, if any.*

SOLUTION: The denominator $x^2 + y^2$ vanishes only at the origin $(0, 0)$. Therefore f and g are continuous everywhere but the origin (Theorem **17.5**). The origin is not in the domain of these function but it is a limit point of the domain. Let us verify whether these function are continuously extendable to the whole plane. Since $|x| \leq \sqrt{x^2 + y^2} = R$, by the squeeze principle

$$|g(x, y)| = \frac{x^4}{R^2} \leq \frac{R^4}{R^2} = R^2 \to 0 \quad \text{as} \quad R \to 0^+$$

it is concluded that $g(x, y) \to 0$ as $(x, y) \to (0, 0)$. So, g has no discontinuity at the origin (it is continuously extendable to the whole plane by $g(0, 0) = 0$).

Take a sequence of points $(x_n, 0)$ where $x_n \to 0$ and as $n \to \infty$ and $x_n \neq 0$ (it converges to the origin from within the domain of f). Then

$$\lim_{n\to 0} f(x_n, 0) = \lim_{n\to\infty} \frac{x_n^2}{x_n^2} = 1\,.$$

On the other hand, for a sequence of points $(0, y_n)$ where $y_n \to 0$ and as $n \to \infty$ and $y_n \neq 0$ (it also converges to the origin from within the domain of f), a different limit value of f is obtained:

$$\lim_{n\to 0} f(0, y_n) = \lim_{n\to\infty} \frac{0}{y_n^2} = 0\,.$$

By Theorem **17.1**, the limit of $f(x, y)$ at $(0, 0)$ does not exists and f is discontinuous at the limit point $(0, 0)$ of its domain (it is not continuously extendable to the whole plane). □

17.7. Exercises.
1–5. Use Definition **17.3** of the limit to verify each of the following limits (i.e., given $\varepsilon > 0$, find a neighborhood of the limit point with the properies specified in the definition):

1. $$\lim_{\mathbf{r}\to\mathbf{0}} \frac{x^3 - 4y^2x + 5y^3}{x^2 + y^2} = 0$$

2. $$\lim_{\mathbf{r}\to\mathbf{0}} \frac{x^3 - 4y^2x + 5y^3}{3x^2 + 4y^2} = 0$$

3. $$\lim_{\mathbf{r}\to\mathbf{0}} \frac{x^3 - 4y^4 + 5y^3x^2}{3x^2 + 4y^2} = 0$$

4. $$\lim_{\mathbf{r}\to\mathbf{0}} \frac{x^3 - 4y^2x + 5y^3}{3x^2 + 4y^2 + y^4} = 0$$

5. $$\lim_{\mathbf{r}\to\mathbf{0}} \frac{3x^3 + 4y^4 - 5z^5}{x^2 + y^2 + z^2} = 0$$

6–8. Use the squeeze principle to prove the following limits and find a neighborhood of the limit point in which the deviation of the function from the limit value does not exceed a small given number ε (Hint: $|\sin u| \leq |u|$):

6. $$\lim_{\mathbf{r}\to\mathbf{0}} y\sin(x/\sqrt{y}) = 0$$

7. $$\lim_{\mathbf{r}\to\mathbf{0}} [1 - \cos(y/x)]x^2 = 0$$

8. $$\lim_{\mathbf{r}\to\mathbf{0}} \frac{\cos(xy)\sin(4x\sqrt{y})}{\sqrt{xy}} = 0$$

9. Suppose that $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = 2$ and $\mathbf{r}_0$ is in the domain of f. If nothing else is known about the function, what can be said about the value $f(\mathbf{r}_0)$? If, in addition, f is known to be continuous at $\mathbf{r}_0$, what can be said about the value $f(\mathbf{r}_0)$?
10–19. Find the points of discontinuity of each of the following functions:

10. $f(x, y) = yx/(x^2 + y^2)$;
11. $f(x, y, z) = yxz/(x^2 + y^2 + z^2)$;
12. $f(x, y) = \sin(\sqrt{xy})$;
13. $f(x, y) = \cos(\sqrt{xyz})/(x^2y^2 + 1)$;
14. $f(x, y) = (x^2 + y^2)\ln(x^2 + y^2)$;
15. $f(x, y) = 1$ if either x or y is rational and $f(x, y) = 0$ elsewhere. ;
16. $f(x, y) = (x^2 - y^2)/(x - y)$ if $x \neq y$ and $f(x, x) = 2x$;
17. $f(x, y) = (x^2 - y^2)/(x - y)$ if $x \neq y$ and $f(x, x) = x$;
18. $f(x, y, z) = 1/[\sin(x)\sin(z - y)]$;
19. $f(x, y) = \sin\left(\frac{1}{xy}\right)$.

20–22. Each of the following functions has the value at the origin $f(0, 0) = c$. Determine whether there is a particular value of c at which the function is continuous at the origin if for $(x, y) \neq (0, 0)$:

20. $f(x, y) = \sin\Big(1/(x^2 + y^2)\Big)$;
21. $f(x, y) = (x^2 + y^2)^\nu \sin\Big(1/(x^2 + y^2)\Big)$, $\nu > 0$;
22. $f(x, y) = x^n y^m \sin\Big(1/(x^2 + y^2)\Big)$, $n \geq 0$, $m \geq 0$, and $n + m > 0$.

23-27. Use the properties of continuous functions to find the following limits

23. $\lim_{\mathbf{r}\to\mathbf{0}} \dfrac{(1 + x + yz^2)^{1/3}}{2 + 3x - 4y + 5z^2}$
24. $\lim_{\mathbf{r}\to\mathbf{0}} \sin(x\sqrt{y})$
25. $\lim_{\mathbf{r}\to\mathbf{0}} \dfrac{\sin(x\sqrt{y})}{\cos(x^2y)}$
26. $\lim_{\mathbf{r}\to\mathbf{0}} [e^{xyz} - 2\cos(yz) + 3\sin(xy)]$
27. $\lim_{\mathbf{r}\to\mathbf{0}} \ln(1 + x^2 + y^2z^2)$

28. Use Theorem **17.1** and the properties of limits of numerical sequences to prove Theorem **17.2**.
29. Use Theorem **17.1** to prove Theorem **17.6**.

18. A General Strategy to Study Limits

The definition of the limit gives only the criterion for whether a number c is the limit of $f(\mathbf{r})$ as $\mathbf{r} \to \mathbf{r}_0$. In practice, however, a possible value of the limit is typically unknown. Some studies are needed to make an "educated" guess for a possible value of the limit. Here a procedure to study limits is outlined that might be helpful. In what follows, the limit point is often set to the origin $\mathbf{r}_0 = \mathbf{0}$. This is not a limitation because one can always translate the origin of the coordinate system to any particular point by shifting the values of the argument, for example,

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = \lim_{(x,y)\to(0,0)} f(x+x_0, y+y_0)\,,$$

or, in general,

$$\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = \lim_{\mathbf{r}\to\mathbf{0}} f(\mathbf{r}+\mathbf{r}_0)\,.$$

18.1. Step 1: Continuity Argument. The simplest scenario in studying the limit happens when the function f in question is continuous at the limit point:

$$\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = f(\mathbf{r}_0)\,.$$

For example,

$$\lim_{(x,y)\to(1,2)} \frac{xy}{x^3-y^2} = -\frac{2}{3}$$

because the function in question is a rational function that is continuous if $x^3 - y^2 \neq 0$. The latter is indeed the case for the limit point $(1,2)$. If the continuity argument does not apply, then it is helpful to check the following.

18.2. Step 2: Composition Rule.

THEOREM **18.1**. (Composition Rule for Limits).
Let $g(t)$ be a function continuous at t_0. Suppose that the function f is the composition $f(\mathbf{r}) = g(h(\mathbf{r}))$ so that $\mathbf{r}_0$ is a limit point of the domain of f and $h(\mathbf{r}) \to t_0$ as $\mathbf{r} \to \mathbf{r}_0$. Then

$$\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = \lim_{t\to t_0} g(t) = g(t_0).$$

The proof is omitted as it is similar to the proof of the composition rule for limits of single-variable functions given in Calculus I. The significance of this theorem is that, under the hypotheses of the theorem, a tough problem of studying a multivariable limit is reduced to the problem of the limit of a function of a single argument. The latter problem can be studied, by, for example, analyzing a local behavior of the function by a Taylor polynomial approximation or by l'Hospital's rule. Although there is a generalization of l'Hospital's rule for multivariable limits, but it is far more complicated and much less practical to use as compared with the one-variable case. Various asymptotic approximations of the behavior of functions involved near the limit point such as, e.g., Taylor polynomial approximations, are much more practical to study multivariable limits.

EXAMPLE **18.1**. *Find*

$$\lim_{(x,y)\to(0,0)} \frac{\cos(xy)-1}{x^2y^2}.$$

SOLUTION: The function in question is $g(t) = (\cos t - 1)/t^2$ for $t \neq 0$, where the argument t is replaced by the function $h(x,y) = xy$. The function h is a polynomial and hence continuous. In particular, $h(x,y) \to h(0,0) = 0$ as $(x,y) \to (0,0)$. The function $g(t)$ is continuous for all $t \neq 0$ and its value at $t = 0$ is not defined. Since $\cos t = 1 - t^2/2 + O(t^4)$ for small t,

$$\lim_{t\to 0} \frac{\cos t - 1}{t^2} = \lim_{t\to 0} \frac{-t^2/2 + O(t^4)}{t^2} = \lim_{t\to 0}\left(-\frac{1}{2} + O(t^2)\right) = -\frac{1}{2}.$$

So the function $g(t)$ is continuously extendable at $t = 0$. By setting $g(0) = -1/2$, the function $g(t)$ becomes continuous at $t = 0$ and the hypotheses of the composition rule are fulfilled. Therefore the two dimensional limit in question exists and equals $-1/2$. $\square$

18.3. Step 3: Limits Along Curves. Recall the following result about the limit of a function of one variable. The limit of $f(x)$ as $x \to x_0$ exists and equals c if and only if the corresponding right and left limits of $f(x)$ exist and equal c:

$$\lim_{x\to x_0^+} f(x) = \lim_{x\to x_0^-} f(x) = c \quad \Longleftrightarrow \quad \lim_{x\to x_0} f(x) = c\,.$$

In other words, if the limit exists, it does not depend on the direction from which the limit point is approached. Consequently, this fact allows us to state a useful criterion for non-existence of the limit. If the left and right limits exist but do not coincide, or at least one of them does not exist, then the limit does not exist. A similar criterion for non-existence of a multivariable limit can be found.

DEFINITION **18.1**. (Parametric Curve in a Euclidean Space). *A parametric curve in a Euclidean space* $\mathbb{R}^m$ *is a continuous vector function* $\mathbf{r}(t) = \langle x_1(t), x_2(t), ..., x_m(t)\rangle$*, where* $t \in [a,b]$.

This is a natural generalization of the concept of a parametric curve in a plane or space defined by parametric equations $x_i = x_i(t)$, $i = 1, 2, ..., m$, where $x_i(t)$ are continuous functions on $[a,b]$. For example, the parametric curve $\mathbf{r}(t) = \mathbf{v}t$, $\|\mathbf{v}\| \neq 0$, is the line through the origin parallel to the vector $\mathbf{v}$. (or a part of this line if the range of the parameter t is restricted to an interval $[a,b]$).

DEFINITION **18.2**. (Limit Along a Curve). *Let* $\mathbf{r}_0$ *be a limit point of the domain* D *of a function* f*. Suppose there is a parametric curve* $\mathbf{r}(t)$*,* $t_0 \leq t \leq b$*, such that* $\mathbf{r}(t)$ *is in* D *if* $t > t_0$ *and*

$\mathbf{r}(t_0) = \mathbf{r}_0$. *Let* $F(t) = f(\mathbf{r}(t))$, $t > t_0$, *be the values of* f *on the curve. The limit*

$$\lim_{t\to t_0^+} F(t) = \lim_{t\to t_0^+} f(x_1(t), x_2(t), ..., x_m(t))$$

is called the limit of f along the curve $\mathbf{r}(t)$ *if it exists.*

THEOREM **18.2**. *If the limit of* $f(\mathbf{r})$ *exists as* $\mathbf{r} \to \mathbf{r}_0$ *and is equal to* c, *then the limit of* f *along any curve through* $\mathbf{r}_0$ *exists and is equal to* c *.*

PROOF. Fix $\varepsilon > 0$. By Definition **17.3** the existence of the limit c of $f(\mathbf{r})$ at $\mathbf{r}_0$ means that one can find a number $\delta > 0$ such that

$$|f(\mathbf{r}) - c| < \varepsilon \quad \text{whenever} \quad 0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta$$

The vector function $\mathbf{r}(t)$ is continuous at t_0, that is, $\lim_{t\to t_0^+} \mathbf{r}(t) = \mathbf{r}(t_0) = \mathbf{r}_0$ by the hypothesis. It follows then from Definition **10.2** of the limit of a vector function that for the number δ found above there is a number $\delta' > 0$ such that

$$\|\mathbf{r}(t) - \mathbf{r}_0\| < \delta \quad \text{whenever} \quad t_0 < t < t + \delta' .$$

These two relations imply that for any number $\varepsilon > 0$ one can find a number δ' such that

$$|f(\mathbf{r}(t)) - c| = |F(t) - c| < \varepsilon \quad \text{whenever} \quad t_0 < t < t + \delta' .$$

By the definition of the one-variable limit (Calculus I), this means that $F(t) \to c$ as $t \to t_0^+$. So, the limit along any curve exists and is equal to c. $\square$

In regard of this theorem, recall the discussion of the right panel in Fig. **17.1** in Section **17.1**. An immediate consequence of this theorem is a useful criterion for non-existence of a multi-variable limit.

COROLLARY **18.1**. (Criterion for Nonexistence of the Limit). *Let* f *be a function of several variables on* D *and* $\mathbf{r}_0$ *be a limit point of* D. *If there is a curve along which the limit of* f *at* $\mathbf{r}_0$ *does not exist, then the multivariable limit* $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})$ *does not exist either. If there are two curves along which the limits of* f *at* $\mathbf{r}_0$ *exist but do not coincide, then the multivariable limit* $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})$ *does not exist.*

Repeated limits. Let $(x, y) \neq (0, 0)$. Consider a curve C_1 that consists of two straight line segments $(x, y) \to (x, 0) \to (0, 0)$ and a curve C_2 that consists of two straight line segments $(x, y) \to (0, y) \to (0, 0)$. Both the curves connect (x, y) with the origin. The limits along C_1 and C_2,

$$\lim_{y\to 0}\Big(\lim_{x\to 0} f(x, y)\Big) \quad \text{and} \quad \lim_{x\to 0}\Big(\lim_{y\to 0} f(x, y)\Big)$$

are called the *repeated* limits. Suppose that all points of C_1 and C_2 are *within* the domain of f except the point $(0, 0)$. Then Theorem **18.2** and Corollary **18.1** establish the relations between the repeated limits and the two-variable limit $\lim_{(x,y)\to(0,0)} f(x,y)$. In particular, suppose that the function is continuous with respect to x if $y \neq 0$ is fixed and it is also continuous with respect to y if $x \neq 0$ is fixed. Then $f(x,y) \to f(0,y)$ as $x \to 0$ for $y \neq 0$ and $f(x,y) \to f(x,0)$ as $y \to 0$ for $x \neq 0$. The repeated limits become

$$\lim_{y\to 0} f(0,y) \quad \text{and} \quad \lim_{x\to 0} f(x,0)$$

If at least one of them does not exists or they exist but are not equal, then by Corollary **18.1** the two-variable limit does not exist. If they exist and are equal, then the two-variable limit *may or may not* exist. A further investigation of the two-variable limit is needed.

EXAMPLE **18.2**. *Find the limit*

$$\lim_{(x,y)\to(0,0)} \frac{\sin(x^2 - y^2)}{x^2 + y^2}$$

or show that the limit does not exist.

SOLUTION: The domain of the function in question is the entire plane with origin removed. The function is continuous with respect to x if $y \neq 0$ is fixed and with respect to y if $x \neq 0$ is fixed. Therefore the repeated limits are

$$\lim_{y\to 0}\Big(\lim_{x\to 0} \frac{\sin(x^2 - y^2)}{x^2 + y^2}\Big) = \lim_{y\to 0} \frac{\sin(-y^2)}{y^2} = -\lim_{y\to 0} \frac{\sin(y^2)}{y^2} = -1$$

$$\lim_{x\to 0}\Big(\lim_{y\to 0} \frac{\sin(x^2 - y^2)}{x^2 + y^2}\Big) = \lim_{x\to 0} \frac{\sin(x^2)}{x^2} = 1$$

The repeated limits exist but do not coincide. Therefore the two-variable limit does not exist. □

The following should be emphasized. Suppose that the domain of the function f does not include the coordinate axes (or their parts with the origin), but the coordinates axes consist of limit points of the domain (e.g., the domain of f contains points (x,y) with $x > 0$ and $y > 0$). Then the repeated limits make sense. However, the hypotheses of Corollary **18.1** are not fulfilled (the curves along which the limits are taken are not in the domain of the function). Examples given in Exercises **1-3** illustrate the situation in this case. In particular, *the non-existence of the repeated limits does not imply the non-existence of the two-variable limit* in this case. An example is given in Exercise **3**.

Limits Along Straight Lines. Let the limit point be the origin $\mathbf{r}_0 = \mathbf{0}$. The simplest curve through $\mathbf{r}_0$ is a straight line $x_i = v_i t$, where $t \to 0^+$ for some numbers v_i, $i = 1, 2, ..., m$, that do not vanish simultaneously (or in the vector form $\mathbf{r}(t) = t\mathbf{v}$, $\mathbf{v} \neq \mathbf{0}$). In particular, in the case of two variables it is convenient to take

$$x = t\,,\ y = at \qquad \text{or} \qquad x = t\cos\theta\,,\ y = t\sin\theta$$

where a and θ are parameters. If the multivariable limit of f exists, then the limit of f along every straight line (in the domain of f) should exist and be the same:

$$\lim_{\mathbf{r}\to\mathbf{0}} f(\mathbf{r}) = c \quad \Rightarrow \quad \lim_{t\to 0^+} f(t\mathbf{v}) = c \quad \text{for any} \quad \mathbf{v} \neq \mathbf{0}\,.$$

Consequently, if the limit along a particular line does not exists, or there are two lines along which the limits exist but are not equal, then the multivariable limit does not exist:

$$\begin{array}{lcl} \lim_{t\to 0^+} f(t\mathbf{v}) \text{ does not exist} & & \\ \text{or} & \Rightarrow & \lim\limits_{\mathbf{r}\to\mathbf{0}} f(\mathbf{r}) \text{ does not exist} \\ \lim_{t\to 0^+} f(t\mathbf{v}_1) \neq \lim_{t\to 0^+} f(t\mathbf{v}_2) & & \end{array}$$

where $\mathbf{v}_1$ and $\mathbf{v}_2$ are not parallel ($\mathbf{v}_1 \neq s\mathbf{v}_2$, $s > 0$).

EXAMPLE **18.3**. *Investigate the two-variable limit*

$$\lim_{(x,y)\to(0,0)} \frac{xy^3}{x^4 + 2y^4}.$$

SOLUTION: Consider the limits along straight lines $x = t$, $y = at$ (or $y = ax$, where a is the slope) as $t \to 0^+$:

$$\lim_{t\to 0^+} f(t, at) = \lim_{t\to 0^+} \frac{a^3t^4}{t^4(1 + 2a^4)} = \frac{a^3}{1 + 2a^4}\,.$$

So the limit along a straight line depends on the slope of the line. Therefore, the two-variable limit does not exist. □

EXAMPLE **18.4**. *Investigate the limit*

$$\lim_{(x,y)\to(0,0)} \frac{\sin(\sqrt{xy})}{x + y}.$$

SOLUTION: The domain of the function consists of the first and third quadrants as $xy \geq 0$ except the origin. Lines approaching $(0, 0)$ from within the domain are $x = t$, $y = at$, $a \geq 0$ and $t \to 0$. The line $x = 0$, $y = t$ also lies in the domain (the line with an infinite slope). The limit along a straight line approaching the origin from within the first quadrant is

$$\lim_{t\to 0^+} f(t, at) = \lim_{t\to 0^+} \frac{\sin(t\sqrt{a})}{t(1 + a)} = \lim_{t\to 0^+} \frac{t\sqrt{a} + O(t^3)}{t(1 + a)} = \frac{\sqrt{a}}{1 + a},$$

where $\sin u = u + O(u^3)$, $u = t\sqrt{a}$, has been used to calculate the limit. The limit depends on the slope of the line, and hence the two-variable limit does not exist. □

Limits Along Power Curves (Optional). If the limit along straight lines exists and is independent of the choice of the line, the numerical value of this limit provides a desired "educated" guess for the actual multivariable limit. However, this has yet to be proved by means of either the definition of the multivariable limit or, for example, the squeeze principle. The latter comprises the final step of the analysis of limits (Step 4; see below).

The following should be stressed. *If the limits along all straight lines happen to be the same number, this does not mean that the multivariable limit exists and equals that number because there might exist other curves through the limit point along which the limit attains a different value or does not even exist.*

EXAMPLE **18.5**. *Investigate the limit*

$$\lim_{(x,y)\to(0,0)} \frac{y^3}{x}.$$

SOLUTION: The domain of the function is the whole plane with the y axis removed ($x \neq 0$). The limit along a straight line

$$\lim_{t\to 0^+} f(t, at) = \lim_{t\to 0^+} \frac{a^3t^3}{t} = a^3 \lim_{t\to 0^+} t^2 = 0$$

vanishes for any slope; that is, it is independent of the choice of the line. However, the two-variable limit does not exist! Consider the power curve

$$x = t\,, \quad y = at^{1/3}\,,$$

through the origin. The limit along this curve can attain any value by varying the parameter a:

$$\lim_{t\to 0^+} f(t, at^{1/3}) = \lim_{t\to 0^+} \frac{a^3t}{t} = a^3.$$

Thus, the multivariable limit does not exist. □

In general, limits along power curves are convenient for studying limits of rational functions because the values of a rational function of several variables on a power curve are given by a rational function of the curve parameter t. One can then adjust, if possible, the power parameter of the curve so that the leading terms of the top and bottom power functions match in the limit $t \to 0^+$. For instance, in the example considered, put $x = t$ and $y = at^n$. Then $f(t, at^n) = (a^3t^{3n})/t$. The powers of the top and bottom functions in this ratio match if $3n = 1$; hence, for $n = 1/3$, the limit along the power curve depends on the parameter a and can be any number.

18.4. Step 4: Using the Squeeze Principle. If Steps 1 and 2 do not apply to the multivariable limit in question, then an "educated" guess for a possible value of the limit is helpful. This is the outcome of Step 3. If limits along a family of curves (e.g., straight lines) happen to be the same number c, then this number is the sought-for "educated" guess. The definition of the multivariable limit or the squeeze principle can be used to prove or disprove that c is the multivariable limit.

EXAMPLE **18.6**. *Find the limit or prove that it does not exist:*

$$\lim_{(x,y)\to(0,0)} \frac{\sin(xy^2)}{x^2+y^2}.$$

SOLUTION:
Step 1. The function is not defined at the origin. The continuity argument does not apply.
Step 2. No substitution exists to transform the two-variable limit to a one-variable limit.
Step 3. Put $(x, y) = (t, at)$, where $t \to 0^+$. The limit along straight lines

$$\begin{aligned}\lim_{t\to 0^+} f(t,at) &= \lim_{t\to 0^+} \frac{\sin(a^2t^3)}{t^2(1+a^2)} = \lim_{t\to 0^+} \frac{a^2t^3 + O(t^9)}{t^2(1+a^2)} \\ &= \lim_{t\to 0^+} \left(\frac{a^2t}{1+a^2} + O(t^7)\right) = 0\end{aligned}$$

vanishes (here $\sin u = u + O(u^3)$, $u = a^2t^3$, has been used to calculate the limit).
Step 4. If the two-variable limit exists, then it must be equal to 0. This can be verified by means of the simplified squeeze principle; that is, one has to verify that there exists $h(R)$ such that $|f(x, y) - c| = |f(x, y)| \le h(R) \to 0$ as $R = \sqrt{x^2 + y^2} \to 0^+$. A key technical trick here is the inequality

$$|\sin u| \le |u|,$$

which holds for any real u. One has

$$|f(x,y) - 0| = \frac{|\sin(xy^2)|}{x^2+y^2} \le \frac{|xy^2|}{x^2+y^2} \le \frac{R^3}{R^2} = R \to 0,$$

where the inequalities $|x| \le R$ and $|y| \le R$ have been used. Thus, the two-variable limit exists and equals 0. $\square$

For two-variable limits, it is sometimes convenient to use polar coordinates centered at the limit point $x - x_0 = R\cos\theta$, $y - y_0 = R\sin\theta$. The idea is to find out whether the deviation of the function $f(x, y)$ from c (the "educated" guess from Step 3) can be bounded by $h(R)$ *uniformly* for all $\theta \in [0, 2\pi]$:

$$|f(x,y) - c| = |f(x_0 + R\cos\theta, y_0 + R\sin\theta) - f_0| \le h(R) \to 0$$

as $R \to 0^+$. This technical task can be accomplished with the help of the basic properties of trigonometric functions, for example, $|\sin\theta| \le 1$, $|\cos\theta| \le 1$, and so on.

In Example **18.5**, Step 3 gives an "educated" guess for the limit $c = 0$ (by studying the limits along straight lines). Then

$$|f(x,y) - c| = \frac{|y|^3}{|x|} = \frac{|R^3 \sin^3\theta|}{|R\cos\theta|} = R^2 \sin^2(\theta)|\tan\theta|\,.$$

Despite that the deviation of f from 0 is proportional to $R^2 \to 0$ as $R \to 0^+$, it cannot be made as small as desired uniformly for all θ by decreasing R because $\tan\theta$ is not a bounded function. There is a sector in the plane corresponding to angles near $\theta = \pi/2$ where $\tan\theta$ can be larger than any number whereas $\sin^2\theta$ is *strictly* positive in it (its value is close to 1) so that the deviation of f from 0 can be as large as desired no matter how small R is. So, for any $\varepsilon > 0$, the inequality $|f(\mathbf{r}) - c| < \varepsilon$ is violated in that sector of *any* disk $0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta$, and hence the limit does not exist. In other words, even though values of f are close to 0 for *some points* near $(0,0)$, they *do not stay close to* 0 *everywhere* near $(0,0)$, and hence 0 cannot be the limit of f at $(0,0)$.

Remark. For multivariable limits with $m > 2$, a similar approach exists. If, for simplicity, $\mathbf{r}_0 = \mathbf{0}$. Then put $x_i = Ru_i$, where the variables u_i satisfy the condition $u_1^2 + u_2^2 + \cdots + u_m^2 = 1$ so that $\|\mathbf{r}\| = R$ or in the vector form $\mathbf{r} = R\mathbf{u}$ where $\|\mathbf{u}\| = 1$. For $m = 2$, $u_1 = \cos\theta$ and $u_2 = \sin\theta$. For $m \geq 3$, the variables u_i can be viewed as the directional cosines, that is, the cosines of the angles between $\mathbf{u}$ and unit vectors $\hat{\mathbf{e}}_i$ parallel to the coordinate axes, $u_i = \mathbf{u} \cdot \hat{\mathbf{e}}_i$. Then one has to investigate whether there is $h(R)$ such that

$$|f(Ru_1, Ru_2, ..., Ru_m) - f_0| \leq h(R) \to 0\,, \quad R \to 0^+\,.$$

This technical, often rather difficult, task may be accomplished using the inequalities $|u_i| \leq 1$ and some specific properties of the function f. As noted, the variables u_i are the directional cosines. They can also be trigonometric functions of the angles in the spherical coordinate system in an n-dimensional Euclidean space.

18.5. Infinite limits and limits at infinity. Suppose that the limit of a multivariable function f does not exist as $\mathbf{r} \to \mathbf{r}_0$. There are two particular cases, which are of interest, when f tends to either positive or negative infinity.

Definition **18.3**. (Infinite limits)
The limit of $f(\mathbf{r})$ as $\mathbf{r} \to \mathbf{r}_0$ is said to be the positive infinity if for any number $M > 0$ there exists a number $\delta > 0$ such that $f(\mathbf{r}) > M$ whenever $0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta$. Similarly, the limit is said to be the negative infinity if for any number $M < 0$ there exists a number $\delta > 0$ such that $f(\mathbf{r}) < M$ whenever $0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta$. In these cases, one writes, respectively,

$$\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = +\infty \quad \text{and} \quad \lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = -\infty\,.$$

For example,

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{x^2+y^2} = +\infty\,.$$

Indeed, put $R = \sqrt{x^2+y^2}$. Then, for any $M > 0$, the inequality $f(\mathbf{r}) = 1/R^2 > M$ can be written in the form $R < 1/\sqrt{M}$. Therefore the values of f in the disk $0 < \|\mathbf{r}\| < \delta = 1/\sqrt{M}$ are larger than any preassigned positive number M.

The squeeze principle has a natural extension to infinite limits. Suppose that functions $g(\mathbf{r})$ and $f(\mathbf{r})$ have a common domain and satisfy the following conditions for all points in the domain, then

$$\begin{aligned} g(\mathbf{r}) \le f(\mathbf{r}) \quad &\text{and} \quad \lim_{\mathbf{r}\to\mathbf{r}_0} g(\mathbf{r}) = +\infty \quad \Rightarrow \quad \lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = +\infty \\ f(\mathbf{r}) \le g(\mathbf{r}) \quad &\text{and} \quad \lim_{\mathbf{r}\to\mathbf{r}_0} g(\mathbf{r}) = -\infty \quad \Rightarrow \quad \lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r}) = -\infty \end{aligned}$$

Furthermore, if the limit of a function f is infinite, the values of the function f must approach the infinity along any curve through the limit point. For example, the limit

$$\lim_{\mathbf{r}\to\mathbf{0}} f(x,y) = \lim_{\mathbf{r}\to\mathbf{0}} \frac{y}{x^2+y^2} \quad \text{does not exist}$$

because the limits along straight lines $(x,y) = (t, at)$ are different:

$$\lim_{t\to 0^+} f(t,at) = \lim_{t\to 0^+} \frac{a}{t(1+a^2)} = \begin{cases} +\infty & \text{if } a > 0 \\ 0 & \text{if } a = 0 \\ -\infty & \text{if } a < 0 \end{cases}$$

If, however, the domain of f is *restricted* to the half-plane $y > 0$, then the limit exists and is ∞. Indeed,

$$\text{for all } x \text{ and } y > 0, \quad f(x,y) = \frac{y}{x^2+y^2} \ge \frac{y}{y^2} = \frac{1}{y} \to +\infty \text{ as } y \to 0^+\,,$$

and the conclusion follows from the squeeze principle.

For functions of one variable x, one can define the limits at infinity, i.e., when $x \to +\infty$ or $x \to -\infty$. In both the cases, the values of the function $f(x)$ are investigated in *neighborhoods* of the infinities $\pm\infty$, defined as $N_\delta(\infty) = \{x \,|\, x > \delta\}$ and $N_\delta(-\infty) = \{x \,|\, x < -\delta\}$ for some $\delta > 0$. Similarly, in a Euclidean space, one can investigate the values of the function in *neighborhoods of infinity*:

$$N_\delta(\infty) = \{\mathbf{r} \mid \|\mathbf{r}\| > \delta > 0\}\,.$$

If the domain D of a function $f(\mathbf{r})$ is an unbounded region (D is not contained in a ball of a sufficiently large radius), then a neighborhood of infinity in D consists of all points of D whose distance from the origin exceeds a positive number δ, $\|\mathbf{r}\| > \delta$. A smaller neighborhood is obtained by increasing δ.

DEFINITION 18.4. (Limit at infinity)
Let f be a function on an unbounded region D. A number c is the limit of a function f at infinity if for any number $\varepsilon > 0$ there exists a number $\delta > 0$ such that $|f(\mathbf{r}) - c| < \varepsilon$ whenever $\|\mathbf{r}\| > \delta$ in D and, in this case, one writes

$$\lim_{\mathbf{r}\to\infty} f(\mathbf{r}) = c \qquad \text{or} \qquad f(\mathbf{r}) \to c \quad \text{as} \quad \mathbf{r} \to \infty\,.$$

Infinite limits at infinity are defined similarly.

DEFINITION 18.5. (Infinite Limits at Infinity)
Let f be a function on an unbounded region D. Then

$$\lim_{\mathbf{r}\to\infty} f(\mathbf{r}) = +\infty \quad \text{or} \quad \lim_{\mathbf{r}\to\infty} f(\mathbf{r}) = -\infty$$

if for any number $M > 0$ there exists a number $\delta > 0$ such that, respectively,

$$f(\mathbf{r}) > M \quad \text{or} \quad f(\mathbf{r}) < -M \quad \text{whenever} \quad \|\mathbf{r}\| > \delta \quad \text{in } D\,.$$

The squeeze principle has a natural extension to the infinite limits and at infinity. If $g(\mathbf{r}) \le f(\mathbf{r})$ for all $\mathbf{r}$ and $g(\mathbf{r}) \to \infty$ as $\mathbf{r} \to \infty$, then $f(\mathbf{r}) \to \infty$. If $f(\mathbf{r}) \le g(\mathbf{r})$ for all $\mathbf{r}$ and $g(\mathbf{r}) \to -\infty$ as $\mathbf{r} \to \infty$, then $f(\mathbf{r}) \to -\infty$.

18.6. Study Problems.

Problem 18.1. *Find the limit* $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})$ *or show that it does not exist, where*

$$f(\mathbf{r}) = f(x,y,z) = (x^2 + 2y^2 + 4z^2)\ln(x^2 + y^2 + z^2)\,, \quad \mathbf{r}_0 = \mathbf{0}\,.$$

SOLUTION:
Step 1. The continuity argument does not apply because f is not defined at $\mathbf{r}_0$.
Step 2. No substitution is possible to transform the limit to a one-variable limit.
Step 3. Put $\mathbf{r}(t) = t\mathbf{v}$ where $\mathbf{v} = \langle a, b, c\rangle$ is a unit vector, $a^2 + b^2 + c^2 = 1$. Then

$$f(\mathbf{r}(t)) = At^2\ln(t^2) = 2At^2\ln t$$

where $A = a^2 + 2b^2 + 4c^2 > 0$ and $t > 0$. By l'Hospital's rule

$$\lim_{t\to 0^+} t^2\ln t = \lim_{t\to 0^+}\frac{\ln t}{t^{-2}} = \lim_{t\to 0^+}\frac{t^{-1}}{-2t^{-3}} = -\frac{1}{2}\lim_{t\to 0^+} t = 0$$

and therefore $f(\mathbf{r}(t)) \to 0$ as $t \to 0^+$. So, if the limit exists, then it must be equal to 0.
Step 4. Put $R^2 = x^2 + y^2 + z^2$. Since the limit $R \to 0^+$ is of interest, one can always assume that $R < 1$ so that $\ln R^2 = 2\ln R < 0$. By making use of the inequalities $|x| \le R$, $|y| \le R$, $|z| \le R$, one has

$$R^2 \le x^2 + 2y^2 + 4z^2 \le 7R^2\,.$$

By multiplying the latter inequality by $\ln R^2 < 0$,

$$R^2\ln R^2 \ge f(\mathbf{r}) \ge 7R^2\ln(R^2)\,.$$

Since $t\ln t \to 0$ as $t = R^2 \to 0^+$, the limit exists and equals 0 by the squeeze principle. □

Problem **18.2**. *Prove that the limit* $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})$ *exists, where*

$$f(\mathbf{r}) = f(x,y) = \frac{1-\cos(x^2y)}{x^2+2y^2}\,, \quad \mathbf{r}_0 = \mathbf{0}\,,$$

and find a disk centered at $\mathbf{r}_0$ *in which values of* f *deviate from the limit no more than* $\varepsilon = 0.5\times 10^{-4}$.

SOLUTION:
Step 1. The continuity argument does not apply because f is not defined at $\mathbf{r}_0$.
Step 2. No substitution is possible to transform the limit to a one-variable limit.
Step 3. Put $\mathbf{r}(t) = \langle t, at\rangle$. Then

$$\begin{aligned}\lim_{t\to 0^+} f(\mathbf{r}(t)) &= \lim_{t\to 0^+}\frac{1-\cos(at^3)}{t^2(1+2a^2)} = \frac{1}{1+2a^2}\lim_{t\to 0^+}\frac{1-(1-\frac{1}{2}(at^3)^2+O(t^{12}))}{t^2}\\ &= \frac{1}{1+2a^2}\lim_{t\to 0^+}\left(\frac{a^2t^4}{2}+O(t^{10})\right) = 0\,,\end{aligned}$$

where $\cos u = 1-\frac{1}{2}u^2+O(u^4)$, $u = at^3$, have been used to evaluate the limit. Therefore, if the limit exists, it must be equal to 0.
Step 4. It follows from the inequality $|\sin u| \le |u|$ that $\sin^2 u \le u^2$ and, hence,

$$1-\cos u = 2\sin^2(u/2) \le 2(u/2)^2 = u^2/2\,.$$

Put $R^2 = x^2+y^2$. Then, by making use of the above inequality with $u = x^2y$ together with $|x| \le R$ and $|y| \le R$, the following chain of inequalities is obtained:

$$|f(\mathbf{r})-0| \le \frac{(x^2y)^2/2}{x^2+2y^2} = \frac{(x^2y)^2/2}{R^2+y^2} \le \frac{(x^2y)^2/2}{R^2} \le \frac{1}{2}\frac{R^6}{R^2} = \frac{R^4}{2} \to 0$$

as $R \to 0^+$. By the squeeze principle, the limit exists and equals 0. It follows from $|f(\mathbf{r})| \le R^4/2$ that $|f(\mathbf{r})| < \varepsilon$ whenever $R^4/2 < \varepsilon$ or $R = \|\mathbf{r}-\mathbf{r}_0\| < \delta(\varepsilon) = (2\varepsilon)^{1/4} = 0.1$. In the disk $0 < \|\mathbf{r}\| < 0.1$ the values of the function deviate from 0 no more than $0.5\cdot 10^{-4}$. □

Problem **18.3**. *Find the limit* $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})$ *or show that it does not exist, where*

$$f(\mathbf{r}) = f(x,y) = \frac{x^2y}{x^2-y^2}\,, \quad \mathbf{r}_0 = \mathbf{0}\,.$$

SOLUTION:
Step 1. The continuity argument does not apply because f is not defined at $\mathbf{r}_0$.
Step 2. No substitution is possible to transform the limit to a one-variable limit.

Step 3. The domain D of the function is the whole plane with the lines $y = \pm x$ excluded. So put $\mathbf{r}(t) = \langle t, at\rangle$, where $a \neq \pm 1$. Then

$$f(\mathbf{r}(t)) = \frac{at^3}{t^2(1-a^2)} = \frac{at}{1-a^2} \to 0 \quad \text{as} \quad t \to 0^+ .$$

So, if the limit exists, then it must be equal to 0.
Step 4. In polar coordinates, $x = R\cos\theta$ and $y = R\sin\theta$, where $\|\mathbf{r}-\mathbf{r}_0\| = R$,

$$f(\mathbf{r}) = \frac{R^3\cos^2\theta\sin\theta}{R^2(\cos^2\theta - \sin^2\theta)} = \frac{1}{2}\,\frac{R\cos\theta\sin(2\theta)}{\cos(2\theta)} = \frac{R\cos\theta}{2}\,\tan(2\theta).$$

Therefore, in any disk $0 < \|\mathbf{r}-\mathbf{r}_0\| < \delta$, there is a sector corresponding to the polar angle $\pi/4 < \theta < \pi/4 + \Delta\theta$ in which the deviation $|f(\mathbf{r}) - 0|$ can be made larger that any positive number by taking $\Delta\theta > 0$ small enough (no matter how small R is) because $\tan(2\theta)$ is not bounded in this interval. Hence, for any $\varepsilon > 0$, there is no $\delta > 0$ such that $|f(\mathbf{r})| < \varepsilon$ whenever $0 < \|\mathbf{r}-\mathbf{r}_0\| < \delta$ (naturally, $\mathbf{r}$ also lies in the domain D). Thus, the limit does not exist.
Step 3 (Optional). The nonexistence of the limit established in Step 4 implies that there should exist curves along which the limit differs from 0. It is instructive to demonstrate this explicitly. Any such curve should lie within one of the narrow sectors containing the lines $y = \pm x$ (where $\tan(2\theta)$ takes large values). So put, for example,

$$x = t\,, \quad y = t - at^n \quad \text{or} \quad \mathbf{r}(t) = \langle t, t - at^n\rangle\,,$$

where $t \geq 0$, $n > 1$, and $a \neq 0$ is a number. Observe that the line $(x, y) = (t, t)$ (or $y = x$) is tangent to the curve $\mathbf{r}(t)$ at the origin because $\mathbf{r}'(0) = \langle 1, 1\rangle$ for $n > 1$. The term $-at^n$ in $\mathbf{r}(t)$ models a small deviation of the curve from the line $y = x$ in the vicinity of which the function f is expected to be unbounded. Then

$$f(\mathbf{r}(t)) = \frac{t^3 - at^{n+2}}{2at^{n+1} - a^2t^{2n}}\,.$$

This function tends to a number as $t \to 0^+$ if n is chosen to match the leading (smallest) powers of the top and bottom of the ratio in this limit: $3 = n + 1$ or $n = 2$. Thus, for $n = 2$,

$$f(\mathbf{r}(t)) = \frac{t^3 - at^4}{2at^3 - a^2t^4} = \frac{1 - at}{2a - a^2t} \to \frac{1}{2a} \quad \text{as} \quad t \to 0^+\,,$$

and $f(\mathbf{r}(t))$ diverges for $n > 2$ in this limit. □

Problem 18.4. *Find*

$$\lim_{\mathbf{r}\to\infty} \frac{\ln(x^2 + y^4)}{x^2 + 2y^2}$$

or show that the limit does not exist.

Solution: Step 1. Does not apply.
Step 2. No substitution exists to reduce the limit to a one-variable limit.

Step 3. Put $(x, y) = (t, at)$ and let $t \to \infty$. Then $\|\mathbf{r}\| \to \infty$ as $t \to \infty$. Recall that
$$\ln(1 + u) = u + O(u^2)$$
for small u. Therefore for $a \neq 0$ and $u = 1/(a^4t^2)$ the numerator in the function for large t behaves as
$$\begin{aligned}\ln(t^2 + a^4t^4) &= \ln\Big[a^4t^4\Big(1 + \frac{1}{a^4t^2}\Big)\Big] = \ln(a^4t^4) + \ln\Big(1 + \frac{1}{a^4t^2}\Big)\\ &= \ln(a^4t^4) + \frac{1}{a^4t^2} + O(t^{-4}) = \ln(a^4t^4) + O(t^{-2})\\ &= 4\ln t + \ln(a^4) + O(t^{-2})\,.\end{aligned}$$
Then the function for large values of t and $a \neq 0$ behaves as
$$f(t, at) = \frac{4}{1 + 2a^2}\frac{\ln t}{t^2} + O(t^{-2}) \to 0 \quad \text{as} \quad t \to \infty$$
since $\ln t/t^2 \to 0$ as $t \to \infty$. Recall that $\ln t$ grows *slower than any power function t^n, $n > 0$*, so that $\ln t/t^n \to 0$ as $t \to \infty$. If $a = 0$, then $f(t, 0) = 2\ln t/t^2 \to 0$ as $t \to \infty$. So the limit along all straight lines is 0.

Step 4. To prove that the limit is indeed 0, put $R = \sqrt{x^2 + y^2}$ so that $|x| \le R$ and $|y| \le R$. Since the limit $R \to \infty$ is of interest, take $R > 1$ so that $R^4 > R^2$. Then owing to the monotonicity of the logarithm function
$$\ln(x^2 + y^4) \le \ln(R^2 + R^4) \le \ln(2R^4)$$
Therefore
$$|f(x, y) - 0| \le \frac{\ln(2R^4)}{x^2 + 2y^2} = \frac{\ln(2R^4)}{R^2 + y^2} \le \frac{4\ln R + \ln 2}{R^2} \to 0 \quad \text{as} \quad R \to \infty$$
Thus, by the squeeze principle the limit is indeed 0. □

18.7. Exercises.

1–3. Prove the following statements:

1. Let $f(x, y) = (x - y)/(x + y)$. Then
$$\lim_{x\to 0}(\lim_{y\to 0} f(x, y)) = 1\,, \quad \lim_{y\to 0}(\lim_{x\to 0} f(x, y)) = -1$$
but the limit of $f(x, y)$ as $(x, y) \to (0, 0)$ does not exist.

2. Let $f(x, y) = x^2y^2/(x^2y^2 + (x - y)^2)$. Then
$$\lim_{x\to 0}(\lim_{y\to 0} f(x, y)) = \lim_{y\to 0}(\lim_{x\to 0} f(x, y)) = 0$$
but the limit of $f(x, y)$ as $(x, y) \to (0, 0)$ does not exist.

3. Let $f(x, y) = (x + y)\sin(1/x)\sin(1/y)$. Then the repeated limits
$$\lim_{x\to 0}(\lim_{y\to 0} f(x, y)) \quad \text{and} \quad \lim_{y\to 0}(\lim_{x\to 0} f(x, y))$$
do not exist, but the limit of $f(x, y)$ exists and equals 0 as $(x, y) \to (0, 0)$. Does the result contradict to Theorem **18.2**? Explain.

4–17. Find each of the following limits or show that it does not exist:

4. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{\cos(xy+z)}{x^4+y^2z^2+4}\,;$

5. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{\sin(xy)-xy}{(xy)^3}\,;$

6. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{\sqrt{xy^2+1}-1}{xy^2}\,;$

7. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{\sin(xy^3)}{x^2}\,;$

8. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{x^3+y^5}{x^2+2y^2}\,;$

9. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{e^{\|\mathbf{r}\|}-1-\|\mathbf{r}\|}{\|\mathbf{r}\|^2}\,;$

10. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{x^2+\sin^2 y}{x^2+2y^2}\,;$

11. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{xy^2+x\sin(xy)}{x^2+2y^2}\,;$

12. $\displaystyle \lim_{(x,y)\to(1,0)} \frac{\ln(x+e^y)}{\sqrt{x^2+y^2}}\,;$

13. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} (x^2+y^2)^{x^2y^2}\,;$

14. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{xy}\tan\Big(\frac{xy}{1+xy}\Big)\,;$

15. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \ln\Big(\frac{\sin(x^2-y^2)}{x^2-y^2}\Big)^2\,;$

16. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{\sqrt{xy+1}-1}{y\sqrt{x}}\,;$

17. $\displaystyle \lim_{\mathbf{r}\to\mathbf{0}} \frac{x^by^a}{x^a+y^b}\,, \quad 0<b<a, \quad x,y>0\,;$

18. Let

$$f(x,y)=\frac{|x|+|y|-|x+y|}{(x^2+y^2)^k} \quad \text{if} \quad x^2+y^2\neq 0, \quad \text{and} \quad f(0,0)=c\,.$$

Find all values of constants c and $k>0$ at which the function is continuous at the origin.

19. Let $f(x,y)=x^2y/(x^4+y^2)$ if $x^2+y^2\neq 0$ and $f(0,0)=0$. Show that f is continuous along any straight line through the origin, i.e., $F(t)=f(x(t),y(t))$ is continuous for all t where $x(t)=t\cos\theta$, $y(t)=t\sin\theta$ for any fixed θ, nevertheless f is not continuous at $(0,0)$. Hint: Investigate the limits of f along power curves through the origin.

20. Let $f(x,y)$ be continuous in a rectangle $a<x<b$, $c<y<d$. Let $g(x)$ be continuous on the interval (a,b) and takes values in (c,d). Prove that the function $F(x)=f(x,g(x))$ is continuous on (a,b).

21. Investigate the limits of the function $f(x,y)=x^2e^{-(x^2-y)}$ along the rays $x(t)=\cos(\theta)t$, $y(t)=\sin(\theta)t$, as $t\to\infty$ for all $0\le\theta\le 2\pi$. Are the values of the function arbitrary small for all $\|\mathbf{r}\|>\delta$ if δ is large enough? Does the limit $\lim_{\mathbf{r}\to\infty} f(x,y)$ exist?

22–31. Find the limit or show that it does not exist:

22. $\displaystyle\lim_{\mathbf{r}\to\mathbf{0}} \frac{\sin(x^2+y^2+z^2)}{x^4+y^4+z^4}\,;$

23. $\displaystyle\lim_{\mathbf{r}\to\mathbf{0}} \frac{x^2+2y^2+3z^2}{x^4+y^2z^4}\,;$

24. $\displaystyle\lim_{\mathbf{r}\to\infty} \frac{\ln(x^2y^2z^2)}{x^2+y^2+z^2}\,;$

Hint : Consider the limits along the curves

$x=y=z=t$ and $x=e^{-t^2},\ y=z=t$

25. $\displaystyle\lim_{\mathbf{r}\to\infty} \frac{e^{3x^2+2y^2+z^2}}{(x^2+2y^2+3z^2)^{2012}}\,;$

26. $\displaystyle\lim_{\mathbf{r}\to\mathbf{0}} \frac{z}{x^2+y^2+z^2}\,;$

27. $\displaystyle\lim_{\mathbf{r}\to\mathbf{0}} \frac{z}{x^2+y^2+z^2}\quad \text{if}\quad z<0\,;$

28. $\displaystyle\lim_{\mathbf{r}\to\infty} \frac{x^2+y^2}{x^2+y^4}\,;$

29. $\displaystyle\lim_{\mathbf{r}\to\infty} \sin\left(\frac{\pi x}{2x+y}\right)\,;$

30. $\displaystyle\lim_{\mathbf{r}\to\infty} (x^2+y^2)e^{-|x+y|}\,;$

31. $\displaystyle\lim_{\mathbf{r}\to\infty} \left(\frac{xy}{x^2+y^2}\right)^{x^2}.$

32. Find the repeated limits

$$\lim_{x\to 1}\Big(\lim_{y\to 0}\log_x(x+y)\Big)\quad\text{and}\quad \lim_{y\to 0}\Big(\lim_{x\to 1}\log_x(x+y)\Big)$$

What can be said about the corresponding two-variable limit?

19. Partial Derivatives

The derivative $f'(x_0)$ of a function $f(x)$ at $x = x_0$ contains important information about the local behavior of the function near $x = x_0$. It defines the slope of the tangent line $y = L(x)$, $L(x) = f(x_0) + f'(x_0)(x - x_0)$ to the graph $y = f(x)$ at x_0. For x close enough to x_0, values of f can be well approximated by the linearization $L(x)$, that is, $f(x) \approx L(x)$. In particular, if $f'(x_0) > 0$, f increases near x_0, and, if $f'(x_0) < 0$, f decreases near x_0. Furthermore, the second derivative $f''(x_0)$ supplies more information about f near x_0, namely, its concavity.

It is therefore important to develop a similar concept for functions of several variables in order to study their local behavior. A significant difference is that, given a point in the domain, the rate of change is going to depend on the direction in which it is measured. For example, if $f(\mathbf{r})$ is the height of a hill as a function of position $\mathbf{r}$ in the base of the hill, then the slopes from west to east and from south to north may be different. This observation leads to the concept of partial derivatives. Let x and y be the coordinates from west to east and from south to north, respectively. The graph of f is the surface $z = f(x, y)$. At a fixed point $\mathbf{r}_0 = (x_0, y_0)$, the height changes as $h(x) = f(x, y_0)$ along the west–east direction, and as $g(y) = f(x_0, y)$ along the south–north direction. Their graphs are intersections of the surface $z = f(x, y)$ with the coordinate planes $x = x_0$ and $y = y_0$, that is, $z = f(x_0, y) = g(y)$ and $z = f(x, y_0) = h(x)$. The slope along the west–east direction is then $h'(x_0)$, and along the south–north direction, is $g'(y_0)$. These slopes are called *partial derivatives* of f and denoted as

$$\frac{\partial f}{\partial x}(x_0, y_0) = \frac{d}{dx} f(x, y_0)\Big|_{x=x_0},$$
$$\frac{\partial f}{\partial y}(x_0, y_0) = \frac{d}{dy} f(x_0, y)\Big|_{y=y_0}.$$

The partial derivatives are also denoted as

$$\frac{\partial f}{\partial x}(x_0, y_0) = f'_x(x_0, y_0)\,, \quad \frac{\partial f}{\partial y}(x_0, y_0) = f'_y(x_0, y_0).$$

The subscript of f' indicates the variable with respect to which the derivative is calculated. The above analysis of the geometrical significance of partial derivatives is illustrated in Fig. **19.1** The concept of partial derivatives can easily be extended to functions of more than two variables.

19.1. Partial Derivatives of a Function of Several Variables. Let D be a subset of an n-dimensional Euclidean space.

Definition **19.1**. (Interior Point of a Set).
A point $\mathbf{r}_0$ is said to be an interior point of D if there is a neighborhood $N_\delta(\mathbf{r}_0) = \{\mathbf{r} \mid \|\mathbf{r} - \mathbf{r}_0\| < \delta\}$ of radius $\delta > 0$ that lies in D.

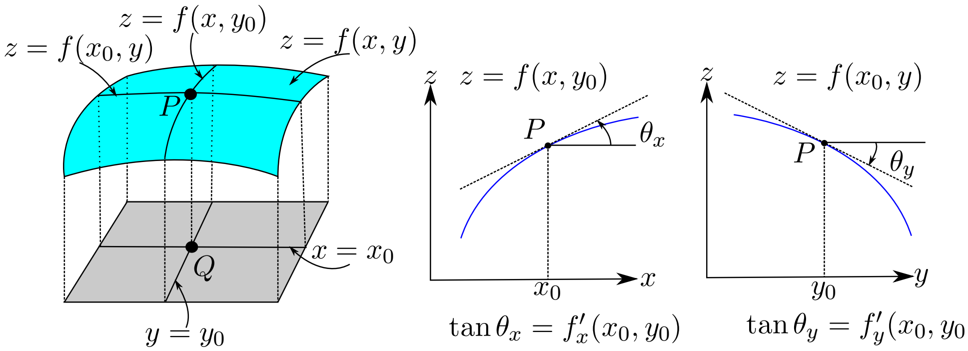


FIGURE **19.1**. Geometrical significance of partial derivatives. **Left**: The graph $z = f(x, y)$ and its cross sections by the coordinate planes planes $x = x_0$ and $y = y_0$. The point $Q = (x_0, y_0, 0)$ is in the domain of f and the point $P = (x_0, y_0, f(x_0, y_0))$ lies on the graph. **Middle**: The cross section $z = f(x, y_0)$ of the graph in the plane $y = y_0$ and the tangent line to it at the point P. The slope $\tan\theta_x$ of the tangent line is determined by the partial derivative $f'_x(x_0, y_0)$ at the point Q. **Right**: The cross section $z = f(x_0, y)$ of the graph in the plane $x = x_0$ and the tangent line to it at the point P. The slope $\tan\theta_y$ of the tangent line is determined by the partial derivative $f'_y(x_0, y_0)$ at the point Q. Here $\theta_y < 0$ as it is counted clockwise.

In other words, $\mathbf{r}_0$ is an interior point of D if there is a positive number $\delta > 0$ such that all points whose distance from $\mathbf{r}_0$ is less than δ also lie in D. For example, if D is a set of points in a plane whose coordinates are integers, then D has no interior points at all because the points of a disk of radius $0 < \delta < 1$ centered at any point $\mathbf{r}_0$ of D do not belong to D except $\mathbf{r}_0$. If $D = \{(x, y) \,|\, x^2 + y^2 \le 1\}$, then any point of D that does not lie on the circle $x^2 + y^2 = 1$ is an interior point.

DEFINITION **19.2**. (Open Sets).
A set D in a Euclidean space is said to be open if all points of D are interior points of D.

An open set is an extension of the notion of an open interval (a, b) to the multivariable case. In particular, the whole Euclidean space is open.

Recall that any vector in space may be written as a linear combination of three unit vectors, $\mathbf{r} = \langle x, y, z\rangle = x\hat{\mathbf{e}}_1 + y\hat{\mathbf{e}}_2 + z\hat{\mathbf{e}}_3$, where $\hat{\mathbf{e}}_1 = \langle 1, 0, 0\rangle$, $\hat{\mathbf{e}}_2 = \langle 0, 1, 0\rangle$, and $\hat{\mathbf{e}}_3 = \langle 0, 0, 1\rangle$. Similarly, using the rules for adding m-tuples and multiplying them by real numbers, one can write

$$\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle = x_1\hat{\mathbf{e}}_1 + x_2\hat{\mathbf{e}}_2 + \cdots + x_m\hat{\mathbf{e}}_m,$$

where $\hat{\mathbf{e}}_i$ is the m-tuple whose components are zeros except the ith one, which is equal to 1. Obviously, $\|\hat{\mathbf{e}}_i\| = 1$, $i = 1, 2, ..., m$.

DEFINITION **19.3**. (Partial Derivatives at a Point).
Let f be a function of several variables $(x_1, x_2, ..., x_m)$. Let D be the domain of f and let $\mathbf{r}_0$ be an interior point of D. If the limit

$$f'_{x_i}(\mathbf{r}_0) = \lim_{h\to 0} \frac{f(\mathbf{r}_0 + h\hat{\mathbf{e}}_i) - f(\mathbf{r}_0)}{h}$$

exists, then it is called the partial derivative of f with respect to x_i at $\mathbf{r}_0$.

The reason the point $\mathbf{r}_0$ needs to be an interior point is simple. By the definition of the one-variable limit, h can be negative or positive. So the points $\mathbf{r}_0 + h\hat{\mathbf{e}}_i$, $i = 1, 2, ..., m$, must be in the domain of the function because otherwise $f(\mathbf{r}_0 + h\hat{\mathbf{e}}_i)$ is not even defined. This is guaranteed if $\mathbf{r}_0$ is an interior point because all points $\mathbf{r}$ in a neighborhood $N_\delta(\mathbf{r}_0)$ of sufficiently small radius δ are in D; the points $\mathbf{r}_0 + h\hat{\mathbf{e}}_i$ lie in D if $|h| < \delta$.

Remark. It is also common to omit "prime" in the notations for partial derivatives. For example, the partial derivatives of $f(x, y)$ at a point (x_0, y_0) are denoted as $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$. In what follows, the notation introduced in Definition **19.3** will be used.

Let $\mathbf{r}_0 = \langle a_1, a_2, ..., a_m\rangle$, where a_i are fixed numbers. Consider the function $F(x_i)$ of one variable x_i (i is fixed), which is obtained from $f(\mathbf{r})$ by setting all the variables to a number $x_j = a_j$ except the i^{th} one (i.e., $x_j = a_j$ for all $j \neq i$). By the definition of the ordinary derivative, the partial derivative $f'_{x_i}(\mathbf{r}_0)$ exists if and only if the derivative $F'(a_i)$ exists because

$$f'_{x_i}(\mathbf{r}_0) = \lim_{h\to 0} \frac{F(a_i + h) - F(a_i)}{h} = \frac{dF(x_i)}{dx_i}\Big|_{x_i=a_i} \tag{19.1}$$

just like in the case of two variables discussed at the beginning of this section. This rule is practical for calculating partial derivatives as it reduces the problem to computing ordinary derivatives.

EXAMPLE **19.1**. *Find the partial derivatives of $f(x, y, z) = x^3 - y^2 z$ at the point $(1, 2, 3)$.*

SOLUTION: By the rule (**19.1**),

$$\begin{aligned} f'_x(1, 2, 3) &= \frac{d}{dx} f(x, 2, 3)\Big|_{x=1} = \frac{d}{dx}(x^3 - 12)\Big|_{x=1} = 3, \\ f'_y(1, 2, 3) &= \frac{d}{dy} f(1, y, 3)\Big|_{y=2} = \frac{d}{dy}(1 - 3y^2)\Big|_{y=2} = -12, \\ f'_z(1, 2, 3) &= \frac{d}{dz} f(1, 2, z)\Big|_{z=3} = \frac{d}{dz}(1 - 4z)\Big|_{z=3} = -4. \end{aligned}$$

□

Geometrical Significance of Partial Derivatives. From the rule (**19.1**), it follows that *the partial derivative $f'_{x_i}(\mathbf{r}_0)$ defines the rate of change of the function f when only the variable x_i changes while the other variables are* *kept fixed.* If, for instance, the function f in Example **19.1** defines the temperature in degrees Celsius as a function of position whose coordinates are given in meters, then, at the point $(1, 2, 3)$, the temperature increases at the rate 3 degrees Celsius per meter in the direction of the x axis, and it decreases at the rates -12 and -4 degrees Celsius per meter in the direction of the y and z axes, respectively.

19.2. Partial Derivatives as Functions. Suppose that the partial derivatives of f exist at all points of a set D. Then each partial derivative can be viewed as a function of several variables on D. These functions are denoted as $f'_{x_i}(\mathbf{r})$, where $\mathbf{r} \in D$. They can be found by the same rule (**19.1**) if, when differentiating with respect to x_i, all other variables are not set to any specific values but rather viewed as independent of x_i (i.e., $dx_j/dx_i = 0$ for all $j \neq i$). This agreement is reflected by the notation

$$f'_{x_i}(x_1, x_2, ..., x_m) = \frac{\partial}{\partial x_i} f(x_1, x_2, ..., x_m);$$

that is, the symbol $\partial/\partial x_i$ means differentiation with respect to x_i while regarding all other variables as numerical parameters independent of x_i.

Example **19.2**. *Find $f'_x(x, y)$ and $f'_y(x, y)$ if $f(x, y) = x \sin(xy)$.*

Solution: Assuming first that y is a numerical parameter independent of x, one obtains

$$\begin{aligned} f'_x(x, y) = \frac{\partial}{\partial x} f(x, y) &= \Big(\frac{\partial}{\partial x} x\Big) \sin(xy) + x \frac{\partial}{\partial x} \sin(xy) \\ &= \sin(xy) + xy \cos(xy) \end{aligned}$$

by the product rule for the derivative. If now the variable x is viewed as a numerical parameter independent of y, one obtains

$$f'_y(x, y) = \frac{\partial}{\partial y} f(x, y) = x \frac{\partial}{\partial y} \sin(xy) = x^2 \cos(xy).$$

□

19.3. Basic Rules of Differentiation. Since a partial derivative is just an ordinary derivative with one additional agreement that all other variables are viewed as numerical parameters, the basic rules of differentiation apply to partial derivatives. Let f and g be functions of several variables and let c be a number. Then

$$\frac{\partial}{\partial x_i}(cf) = c \frac{\partial f}{\partial x_i}, \qquad \frac{\partial}{\partial x_i}(f + g) = \frac{\partial f}{\partial x_i} + \frac{\partial g}{\partial x_i},$$

$$\frac{\partial}{\partial x_i}(fg) = \frac{\partial f}{\partial x_i} g + f \frac{\partial g}{\partial x_i}, \qquad \frac{\partial}{\partial x_i}\Big(\frac{f}{g}\Big) = \frac{\frac{\partial f}{\partial x_i} g - f \frac{\partial g}{\partial x_i}}{g^2}.$$

Let $h(u)$ be a differentiable function of one variable and let $g(\mathbf{r})$ be a function of several variables whose range lies in the domain of h. Then one can define the composition $f(\mathbf{r}) = h(g(\mathbf{r}))$. Assuming that the partial derivatives of g exist, the chain rule holds

$$\frac{\partial f}{\partial x_i} = h'(g)\,\frac{\partial g}{\partial x_i}. \tag{19.2}$$

EXAMPLE **19.3**. *Find the partial derivatives of the function* $f(\mathbf{r}) = \|\mathbf{r}\|^{-1}$, *where* $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$.

SOLUTION: Put $h(g) = g^{-1/2}$ and $g = g(\mathbf{r}) = x_1^2 + x_2^2 + \cdots + x_m^2 = \|\mathbf{r}\|^2$. Then $f(\mathbf{r}) = h(g(\mathbf{r}))$. Since $h'(g) = (-1/2)g^{-3/2}$ and $\partial g/\partial x_i = 2x_i$, the chain rule gives

$$\frac{\partial}{\partial x_i}\frac{1}{\|\mathbf{r}\|} = \Big(-\frac{1}{2g^{3/2}}\Big)(2x_i) = -\frac{x_i}{\|\mathbf{r}\|^3}\,, \quad i = 1, 2, ..., m\,.$$

□

19.4. Exercises.

1-7. Find the specified partial derivatives of each of the following functions:

1. $f(x, y) = (x - y)/(x + y)$, $f'_x(1, 2)$, $f'_y(1, 2)$;
2. $f(x, y, z) = (xy + z)/(z + y)$, $f'_x(1, 2, 3)$, $f'_y(1, 2, 3)$, $f'_z(1, 2, 3)$;
3. $f(\mathbf{r}) = (x_1 + 2x_2 + \cdots + mx_m)/(1 + \|\mathbf{r}\|^2)$, $f'_{x_i}(\mathbf{0})$, $i = 1, 2, ..., m$;
4. $f(x, y, z) = x\sin(yz)$, $f'_x(1, 2, \pi/2)$, $f'_y(1, 2, \pi/2)$, $f'_z(1, 2, \pi/2)$;
5. $f(x, y) = x + (y - 1)\sin^{-1}(\sqrt{x/y})$, $f'_x(1, 1)$, $f'_y(1, 1)$;
6. $f(x, y) = (x^3 + y^3)^{1/3}$, $f'_x(0, 0)$, $f'_y(0, 0)$;
7. $f(x, y) = \sqrt{|xy|}$, $f'_x(0, 0)$, $f'_y(0, 0)$.

8-23. Find the partial derivatives of each of the following functions:

8. $f(x, y) = (x + y^2)^n$;
9. $f(x, y) = x^y$;
10. $f(x, y) = xe^{(x+2y)^2}$;
11. $f(x, y) = \sin(xy)\cos(x^2 + y^2)$;
12. $f(x, y, z) = \ln(x + y^2 + z^3)$;
13. $f(x, y, z) = xy^2\cos(z^2x)$;
14. $f(\mathbf{r}) = (a_1x_1 + a_2x_2 + \cdots + a_mx_m)^n = (\mathbf{a}\cdot\mathbf{r})^n$;
15. $f(x, y) = \tan^{-1}(y/x)$;
16. $f(x, y) = \sin^{-1}(x/\sqrt{x^2 + y^2})$;
17. $f(x, y, z) = x^{y^z}$;
18. $f(x, y, z) = x^{y/z}$;
19. $f(x, y) = \tan(x^2/y)$;
20. $f(x, y, z) = \sin(x\sin(y\sin z))$;
21. $f(x, y) = (x + y^2)/(x^2 + y)$;
22. $f(x, y, z) = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{r})$, where $\mathbf{a}$ and $\mathbf{b}$ are constant vectors;
23. $f(x, y, z) = \|\mathbf{a}\times\mathbf{r}\|$, where $\mathbf{a}$ is a constant vector.

24-28. For each of the following functions, determine whether the function f increases or decreases at a specified point P_0 when one variable increases, while the others are fixed:

24. $f(x,y) = xy/(x+y)$, $P_0 = (1,2)$;
25. $f(x,y) = (x^2 - 2y^2)^{1/3}$, $P_0 = (1,1)$;
26. $f(x,y) = x^2 \sin(xy)$, $P_0 = (-1,\pi)$;
27. $f(x,y,z) = zx/(x+y^2)$, $P_0 = (1,1,1)$;
28. $f(x,y,z) = (x+yz)/\sqrt{x+2y+z^2}$, $P_0 = (1,2,2)$.

20. Higher-Order Partial Derivatives

Since partial derivatives of a function are also functions of several variables, they can be differentiated with respect to any variable. For example, for a function of two variables, all possible second partial derivatives are

$$\begin{aligned} \frac{\partial f}{\partial x} \quad &\longmapsto \quad \frac{\partial}{\partial x}\frac{\partial f}{\partial x} = \frac{\partial^2 f}{\partial x^2}\,, \quad \frac{\partial}{\partial y}\frac{\partial f}{\partial x} = \frac{\partial^2 f}{\partial y\,\partial x}, \\ \frac{\partial f}{\partial y} \quad &\longmapsto \quad \frac{\partial}{\partial x}\frac{\partial f}{\partial y} = \frac{\partial^2 f}{\partial x\,\partial y}\,, \quad \frac{\partial}{\partial y}\frac{\partial f}{\partial y} = \frac{\partial^2 f}{\partial y^2}. \end{aligned}$$

Throughout the text, brief notations for higher-order partial derivatives will also be used. For example,

$$\frac{\partial^2 f}{\partial x^2} = (f'_x)'_x = f''_{xx}\,, \quad \frac{\partial^2 f}{\partial x\,\partial y} = (f'_y)'_x = f''_{yx}$$

and similarly for f''_{yy} and f''_{xy}. Partial derivatives of the third order are defined as partial derivatives of second partial derivatives, and so on.

Example **20.1**. *For the function $f(x, y) = x^4 - x^2y + y^2$, find all second and third partial derivatives.*

Solution: The first partial derivatives are

$$f'_x = 4x^3 - 2xy\,, \qquad f'_y = -x^2 + 2y\,.$$

Then the second partial derivatives are

$$\begin{aligned} f''_{xx} &= (4x^3 - 2xy)'_x = 12x^2 - 2y\,, & f''_{yy} &= (-x^2 + 2y)'_y = 2, \\ f''_{xy} &= (4x^3 - 2xy)'_y = -2x\,, & f''_{yx} &= (-x^2 + 2y)'_x = -2x\,. \end{aligned}$$

The third partial derivatives are found similarly:

$$\begin{aligned} f'''_{xxx} &= (12x^2 - 2y)'_x = 24x\,, & f'''_{yyy} &= (2)'_y = 0, \\ f'''_{xxy} &= (12x^2 - 2y)'_y = -2\,, & f'''_{xyx} &= f'''_{yxx} = (-2x)'_x = -2, \\ f'''_{yyx} &= (2)'_x = 0\,, & f'''_{yxy} &= f'''_{xyy} = (-2x)'_y = 0. \end{aligned}$$

□

In contrast to the one-variable case, there are higher-order partial derivatives of a new type that are obtained by differentiating with respect to different variables in different orders, like f''_{xy} and f''_{yx}. In the above example, it has been found that

$$\begin{aligned} f''_{xy} &= f''_{yx}, \\ f'''_{xxy} &= f'''_{xyx} = f'''_{yxx}, \\ f'''_{xyy} &= f'''_{yyx} = f'''_{yxy}; \end{aligned}$$

that is, the result is *independent of the order in which the partial derivatives have been taken.* Is this a peculiarity of the function considered or a general property of higher-order partial derivatives? The following theorem answers this question.

THEOREM **20.1**. (Clairaut's Theorem).
Let f be a function of several variables $(x_1, x_2, ..., x_m)$ that is defined on an open ball D in a Euclidean space. If the second partial derivatives $f''_{x_i x_j}$ and $f''_{x_j x_i}$, where $j \neq i$, are continuous functions on D, then $f''_{x_i x_j} = f''_{x_j x_i}$ at any point of D.

A consequence of Clairaut's theorem can be proved. It asserts that:

The result of partial differentiation does not depend on the order in which the partial derivatives have been taken if all higher-order partial derivatives in question are continuous.

It is not always necessary to calculate higher-order partial derivatives in all possible orders to verify the hypothesis of Clairaut's theorem (i.e., the continuity of the partial derivatives). Partial derivatives of polynomials are polynomials and hence continuous. By the quotient rule for partial derivatives, rational functions have continuous partial derivatives (where the denominator does not vanish). Derivatives of basic elementary functions like the sine and cosine and exponential functions are continuous. So compositions of these functions with multivariable polynomials or rational functions have continuous partial derivatives of any order. In other words, the continuity of higher-order partial derivatives can often be established by different, simpler means.

EXAMPLE **20.2**. *Find the third derivatives f'''_{xyz}, f'''_{yzx}, f'''_{zxy}, and so on, for all permutations of x, y, and z, if $f(x, y, z) = \sin(x^2 + yz)$.*

SOLUTION: The sine and cosine functions are continuously differentiable as many times as desired. The argument of the sine function is a multivariable polynomial. By the chain rule, $(\sin g)'_x = g'_x \cos g$ and similarly for the other partial derivatives. Therefore partial derivatives of any order must be products of polynomials and the sine and cosine functions whose argument is a polynomial. Therefore, they are continuous in the entire space. The hypothesis of Clairaut's theorem is satisfied, and hence all the partial derivatives in question coincide and are equal to

$$\begin{aligned} f'''_{xyz} &= (f'_x)''_{yz} = (2x\cos(x^2+yz))''_{yz} = (-2xz\sin(x^2+yz))'_z \\ &= -2x\sin(x^2+yz) - 2xyz\cos(x^2+yz)\,. \end{aligned}$$

□

20.1. Reconstruction of a Function from Its Partial Derivatives. One of the standard problems in calculus is finding a function $f(x)$ if its derivative

$f'(x) = G(x)$ is known. A sufficient condition for the existence of a solution is the continuity of $G(x)$. In this case,

$$f'(x) = G(x) \quad \Longrightarrow \quad f(x) = \int G(x)\,dx\,.$$

A similar problem can be posed for a function of several variables. Given the first partial derivatives

$$f'_{x_i}(\mathbf{r}) = G_i(\mathbf{r})\,, \quad i = 1, 2, ..., m, \tag{20.1}$$

find $f(\mathbf{r})$ if it exists. The existence of such f is a more subtle question in the case of several variables. Suppose that the partial derivatives $\partial G_i/\partial x_j$ exist and are continuous functions in an open ball. Then taking the partial derivative $\partial/\partial x_j$ of both sides of (**20.1**) and applying Clairaut's theorem, one infers that

$$f''_{x_i x_j} = f''_{x_j x_i} \quad \Longrightarrow \quad \frac{\partial G_i}{\partial x_j} = \frac{\partial G_j}{\partial x_i}. \tag{20.2}$$

Thus, the conditions (**20.2**) on the functions G_i must be fulfilled; otherwise, f satisfying (**20.1**) does not exist. The conditions (**20.2**) are called *integrability conditions* for the system of equations (**20.1**).

EXAMPLE **20.3**. *Suppose that* $f'_x(x,y) = 2x + y$ *and* $f'_y(x,y) = 2y - x$. *Does such a function* f *exist?*

SOLUTION: The first partial derivatives of f,

$$f'_x(x,y) = G_1(x,y) = 2x + y\,, \quad f'_y(x,y) = G_2(x,y) = 2y - x\,,$$

are polynomials, and hence their partial derivatives are continuous in the entire plane. In order for f to exist, the integrability condition

$$\frac{\partial G_1}{\partial y} = \frac{\partial G_2}{\partial x}$$

must hold in the entire plane. This is not so because

$$\frac{\partial G_1}{\partial y} = \frac{\partial}{\partial y}(2x + y) = 1\,, \quad \frac{\partial G_2}{\partial x} = \frac{\partial}{\partial x}(2y - x) = -1 \quad \Rightarrow \quad \frac{\partial G_1}{\partial y} \neq \frac{\partial G_2}{\partial x}\,.$$

Thus, no such f exists. □

Suppose now that the integrability conditions (**20.2**) are satisfied. How is a solution f to (**20.1**) to be found? Evidently, one has to calculate an antiderivative of the partial derivative. In the one-variable case, an antiderivative on an interval is defined up to an additive constant. This is not so in the multivariable case. For example, consider the equation

$$f'_x(x,y) = 3x^2 y\,.$$

An antiderivative of f'_x is a function whose *partial* derivative with respect to x is $3x^2y$. It is easy to verify that x^3y satisfy this requirement. It is obtained by taking an antiderivative of $3x^2y$ with respect to x while viewing y as a numerical parameter independent of x. Since f'_x as a function of x is defined for all x (that is, on a single interval) at each fixed y, one can always add a constant to a particular antiderivative, $x^3y + c$, and obtain another solution. The key point to observe is that the constant may be a function of y only. Indeed, if $c = g(y)$, then $c'_x = (g(y))'_x = 0$. Thus,

$$f'_x(x, y) = 3x^2y \quad \Rightarrow \quad f(x, y) = x^3y + g(y)$$

for some function $g(y)$. If, in addition, the other partial derivative f'_y is given, then an explicit form of $g(y)$ can be found. Put, for example,

$$\begin{aligned} f'_x(x, y) &= 3x^2y \\ f'_y(x, y) &= x^3 + 2y \end{aligned}$$

The integrability conditions are fulfilled:

$$\begin{aligned} (f'_x)'_y &= (3x^2y)'_y = 3x^2 \\ (f'_y)'_x &= (x^3 + 2y)'_x = 3x^2 \end{aligned} \quad \Rightarrow \quad f''_{xy}(x, y) = f''_{yx}(x, y)$$

So a function with the said partial derivatives does exist. The obtained general solution to the first equation is substituted into the second equation to find an equation for the unknown function $g(y)$:

$$\begin{aligned} f(x, y) &= x^3y + g(y) \\ f'_y(x, y) &= x^3 + 2y \end{aligned} \quad \Rightarrow \quad (x^3y + g(y))'_y = x^3 + 2y \quad \Rightarrow \quad g'(y) = 2y$$

Therefore $g(y) = y^2 + c$ and $f(x, y) = x^3y + y^2 + c$.

Remark. In the above example, note the cancellation of the x^3 term in the equation for $g(y)$. This is a direct consequence of the fulfilled integrability condition. Had one tried to apply the same procedure to a similar problem without checking the integrability conditions, one could have found that, in general, no such $g(y)$ exists. In Example **20.3**, the equation $f'_x = 2x + y$ has a general solution $f(x, y) = x^2 + yx + g(y)$. Its substitution into the second equation $f'_y = 2y - x$ yields

$$\begin{aligned} f(x, y) &= x^2 + yx + g(y) \\ f'_y(x, y) &= 2y - x \end{aligned} \quad \Rightarrow \; x + g'(y) = 2y - x \; \Rightarrow \; g'(y) = 2y - 2x\,.$$

The derivative of $g(y)$ cannot depend on x and hence no such $g(y)$ exists.

EXAMPLE **20.4**. *Find* $f(x, y, z)$ *if* $f'_x = yz + 2x$, $f'_y = xz + 3y^2$, *and* $f'_z = xy + 4z^3$ *or show that it does not exist.*

SOLUTION: The integrability conditions have to be verified first:

$$\begin{aligned} (G_1)'_y &= (G_2)'_x\,, & G_1 &= yz + 2x & & & (yz + 2x)'_y &= (xz + 3y^2)'_x \\ (G_1)'_z &= (G_3)'_x\,, & G_2 &= xz + 3y^2 & &\Rightarrow & (yz + 2x)'_z &= (xy + 4z^3)'_x \\ (G_2)'_z &= (G_3)'_y\,, & G_3 &= xy + 4z^3 & & & (xz + 3y^2)'_z &= (xy + 4z^3)'_y \end{aligned}$$

The calculation of the partial derivatives in the last three equations yields, respectively, $z = z$, $y = y$, and $x = x$ so that the integrability conditions are satisfied. Therefore f exists. Taking the antiderivative with respect to x in the first equation, one finds

$$f'_x = yz + 2x \quad \Rightarrow \quad f(x, y, z) = xyz + x^2 + g(y, z)\,,$$

for some $g(y, z)$. The substitution of f into the second equations yields

$$\begin{aligned} f'_y = xz + 3y^2 \quad &\Rightarrow \quad xz + g'_y(y, z) = xz + 3y^2 \\ &\Rightarrow \quad g'_y(y, z) = 3y^2 \quad \Rightarrow \quad g(y, z) = y^3 + h(z) \\ &\Rightarrow \quad f(x, y, z) = xyz + x^2 + y^3 + h(z), \end{aligned}$$

for some $h(z)$. The substitution of f into the third equation yields

$$\begin{aligned} f'_z = xy + 4z^3 \quad &\Rightarrow \quad xy + h'(z) = xy + 4z^3 \\ &\Rightarrow \quad h'(z) = 4z^3 \\ &\Rightarrow \quad h(z) = z^4 + c \\ &\Rightarrow \quad f(x, y, z) = xyz + x^2 + y^3 + z^4 + c, \end{aligned}$$

where c is a constant. $\square$

The procedure of reconstructing f from its first partial derivatives as well as the integrability conditions (**20.2**) will be important when discussing *conservative vector fields* and the *potential* of a conservative vector field.

20.2. Partial Differential Equations. The relation between a function of several variables and its partial derivatives (of any order) is called a *partial differential equation*. Partial differential equations are a key tool to study various phenomena in nature. Many fundamental laws of nature can be stated in the form of partial differential equations.

Diffusion Equation. Let $n(\mathbf{r}, t)$, where $\mathbf{r} = \langle x, y, z\rangle$ is the position vector in space and t is time, be a concentration of a substance, say, in air or water or even in a solid. Even if there is no macroscopic motion in the medium, the concentration changes with time due to thermal motion of the molecules. This process is known as *diffusion*. In some simple situations, the rate at which the concentration changes with time at a point is

$$n'_t = k(n''_{xx} + n''_{yy} + n''_{zz}),$$

where the parameter k is a diffusion constant. So the concentration as a function of the spatial position and time must satisfy the above partial differential equation.

Wave Equation. Sound in air is propagating disturbances of the air density. If $u(\mathbf{r}, t)$ is the deviation of the air density from its constant (nondisturbed) value u_0 at the spatial point $\mathbf{r} = \langle x, y, z\rangle$ and at time t, then it can be shown that disturbances of the air density small compared to u_0 satisfy the *wave equation*:

$$u''_{tt} = c^2(u''_{xx} + u''_{yy} + u''_{zz}),$$

where c is the speed of sound in the air (its values depends on u_0 and the air pressure). Light is an electromagnetic wave. Its propagation is also described by the wave equation, where c is the speed of light in vacuum (or in a medium, if light goes through a medium) and u is the amplitude of electric or magnetic fields.

Laplace and Poisson Equations. The equation

$$u''_{xx} + u''_{yy} + u''_{zz} = f,$$

where f is a given non-zero function of position $\mathbf{r} = \langle x, y, z\rangle$ in space, is called the *Poisson equation*. In the special case when $f = 0$, this equation is known as the *Laplace equation*. The Poisson and Laplace equations are used to determine static electromagnetic fields created by static electric charges and currents.

EXAMPLE **20.5**. *Let $h(q)$ be a twice-differentiable function of a variable q. Show that $u(\mathbf{r}, t) = h(ct - \hat{\mathbf{n}} \cdot \mathbf{r})$ is a solution of the wave equation for any fixed unit vector $\hat{\mathbf{n}}$.*

SOLUTION: Let $\hat{\mathbf{n}} = (n_1, n_2, n_3)$, where $n_1^2 + n_2^2 + n_3^2 = 1$ as $\hat{\mathbf{n}}$ is the unit vector. Put

$$q = ct - \hat{\mathbf{n}} \cdot \mathbf{r} = ct - n_1 x - n_2 y - n_3 z\,.$$

By the chain rule (**19.2**), $u'_t = q'_t h'(q)$ and similarly for the other partial derivatives. Therefore

$$\begin{aligned} u'_t &= ch'(q) \\ u'_x &= -n_1 h'(q) \\ u'_y &= -n_2 h'(q) \\ u'_z &= -n_3 h'(q) \end{aligned} \quad \Rightarrow \quad \begin{aligned} u''_{tt} &= c^2 h''(q) \\ u''_{xx} &= n_1^2 h''(q) \\ u''_{yy} &= n_2^2 h''(q) \\ u''_{zz} &= n_3^2 h''(q) \end{aligned}$$

Then

$$c^2(u''_{xx} + u''_{yy} + u''_{zz}) = c^2(n_1^2 + n_2^2 + n_3^2)h''(q) = c^2 h''(q) = u''_{tt}$$

which means that the wave equation is satisfied for any h. $\square$

Consider the level sets of the solution of the wave equation discussed in this example. They correspond to a fixed value of $q = q_0$. So, for each moment of time t, the disturbance of the air density $u(\mathbf{r}, t)$ has a constant value $h(q_0)$ in the plane

$$\hat{\mathbf{n}} \cdot \mathbf{r} = ct - q_0 = d(t)\,.$$

All planes with different values of the parameter d are parallel as they have the same normal vector $\hat{\mathbf{n}}$. Since here $d(t)$ is a function of time, the plane on which the air density has a fixed value moves along the vector $\hat{\mathbf{n}}$ at the rate $d'(t) = c$. Thus, a disturbance of the air density propagates with speed c. This is the reason that the constant c in the wave equation is called the *speed of sound*. Evidently, the same line of arguments applies to electromagnetic waves; that is, they move through space at the speed of light. The speed of sound in the air is about 342 meters per second, or about 768 mph. The speed of light is $3 \cdot 10^8$ meters per second, or 186 miles per second. If a lightning strike occurs a mile away during a thunderstorm, it can be seen almost instantaneously, while the thunder will be heard in about 5 seconds later. Conversely, if one sees a lightning and starts counting seconds until the thunder is heard, then one could determine the distance to the lightning. The sound travels 1 mile in about 4.7 seconds.

20.3. Study Problems.

Problem **20.1**. *Consider the function*

$$f(x,y) = \frac{x^3y - xy^3}{x^2+y^2} \quad \text{if} \quad (x,y) \neq (0,0) \quad \text{and} \quad f(0,0) = 0.$$

*Find $f'_x(x,y)$ and $f'_y(x,y)$ for $(x,y) \neq (0,0)$. Use the rule (**19.1**) to find $f'_x(0,0)$ and $f'_y(0,0)$ and, thereby, to establish that f'_x and f'_y exist everywhere. Use the rule (**19.1**) again to show that $f''_{xy}(0,0) = -1$ and $f''_{yx}(0,0) = 1$, that is, $f''_{xy}(0,0) \neq f''_{yx}(0,0)$. Does this result contradict Clairaut's theorem?*

SOLUTION: Using the quotient rule for differentiation, one finds

$$f'_x(x,y) = \frac{x^4y + 4x^2y^3 - y^5}{(x^2+y^2)^2}\,, \quad f'_y(x,y) = \frac{x^5 - 4x^3y^2 - xy^4}{(x^2+y^2)^2}$$

if $(x,y) \neq (0,0)$. Note that, owing to the symmetry $f(x,y) = -f(y,x)$, the partial derivative f'_y is obtained from f'_x by changing the sign of the latter and swapping x and y. The partial derivatives at $(0,0)$ are found by the rule (**19.1**):

$$f'_x(0,0) = \frac{d}{dx} f(x,0)\Big|_{x=0} = 0\,, \quad f'_y(0,0) = \frac{d}{dy} f(0,y)\Big|_{y=0} = 0.$$

The first-order partial derivatives are continuous functions for all $(x,y) \neq (0,0)$. Put $\sqrt{x^2+y^2} = R$. Then using $|x| \leq R$ and $|y| \leq R$,

$$|f'_x(x,y)| = \frac{|x^4y + 4x^2y^3 - y^5|}{R^4} \leq \frac{R^5 + 4R^5 + R^5}{R^4} = 6R \to 0$$

as $R \to 0^+$. Therefore by the squeeze principle, $f'_x(x,y) \to 0 = f'_x(0,0)$ as $(x,y) \to (0,0)$. So, $f'_x(x,y)$ is continuous everywhere. The continuity of $f'_y(x,y)$ at the origin is established similarly.

Next, one has

$$\begin{aligned}
f''_{xy}(0,0) &= \frac{d}{dy} f'_x(0,y)\Big|_{y=0} = \lim_{h\to 0} \frac{f'_x(0,h)-f'_x(0,0)}{h} \\
&= \lim_{h\to 0} \frac{-h-0}{h} = -1, \\
f''_{yx}(0,0) &= \frac{d}{dx} f'_y(x,0)\Big|_{x=0} = \lim_{h\to 0} \frac{f'_y(h,0)-f'_y(0,0)}{h} \\
&= \lim_{h\to 0} \frac{h-0}{h} = 1.
\end{aligned}$$

The result does not contradict Clairaut's theorem because $f''_{xy}(x,y)$ and $f''_{yx}(x,y)$ are not continuous at $(0,0)$. By using the quotient rule to differentiate $f'_x(x,y)$ with respect to y, an explicit form of $f''_{xy}(x,y)$ can be obtained for $(x,y) \neq (0,0)$. By taking the limit of $f''_{xy}(x,y)$ as $(x,y) \to (0,0)$ along the straight line $(x,y) = (t, at)$, $t \to 0^+$, one infers that the limit depends on the slope a and hence the two-dimensional limit does not exist, that is, $\lim_{(x,y)\to(0,0)} f''_{xy}(x,y) \neq f''_{xy}(0,0) = -1$ and f''_{xy} is not continuous at $(0,0)$. The technical details are left to the reader. □

Problem **20.2**. *Find the value of a constant a for which the function*

$$u(\mathbf{r},t) = t^{-3/2} e^{-ar^2/t}, \quad r = \|\mathbf{r}\|,$$

satisfies the diffusion equation for all $t > 0$.

Solution: Note that u depends on the combination $r^2 = x^2 + y^2 + z^2$. To find the partial derivatives of u, it is convenient to use the chain rule:

$$\begin{aligned}
\frac{\partial u}{\partial x} &= \frac{\partial u}{\partial r^2}\frac{\partial r^2}{\partial x} = 2x\frac{\partial u}{\partial r^2} = -\frac{2ax}{t}u, \\
u''_{xx} &= \frac{\partial}{\partial x}\Big(\frac{\partial u}{\partial x}\Big) = -\frac{2a}{t}u - \frac{2ax}{t}\frac{\partial u}{\partial x} = \Big(-\frac{2k}{t} + \frac{4a^2x^2}{t^2}\Big)u.
\end{aligned}$$

To obtain u''_{yy} and u''_{zz}, note that r^2 is symmetric with respect to permutations of x, y, and z. Therefore, u''_{yy} and u''_{zz} are obtained from u''_{xx} by replacing, in the latter, x by y and x by z, respectively. Hence, the right side of the diffusion equation reads

$$k(u''_{xx} + u''_{yy} + u''_{zz}) = \Big(-\frac{6ka}{t} + \frac{4ka^2r^2}{t^2}\Big)u.$$

Using the product rule to calculate the partial derivative with respect to time, one finds for the left side of the diffusion equation

$$u'_t = -\frac{3}{2}t^{-5/2}e^{-ar^2/t} + t^{-3/2}e^{-ar^2/t}\frac{ar^2}{t^2} = \Big(-\frac{3}{2t} + \frac{ar^2}{t^2}\Big)u.$$

Since both sides must be equal for *all* values of $t > 0$ and r^2, the comparison of the last two expressions yields *two* conditions:

$$u'_t = k(u''_{xx} + u''_{yy} + u''_{zz}) \quad \Rightarrow \quad \begin{cases} 6ka = \frac{3}{2} \\ \quad a = 4ka^2 \end{cases} \quad \Rightarrow \quad a = \frac{1}{4k},$$

where the equality means matching the coefficients at $1/t$ and the second one is needed to match the coefficients at r^2/t^2. The only common solution of these equations is $a = 1/(4k)$. □

20.4. Exercises.

1–6. Find all second partial derivatives of each of the following functions and verify Clairaut's theorem:

1. $f(x, y) = \tan^{-1}(xy)$;
2. $f(x, y, z) = x\sin(zy^2)$;
3. $f(x, y, z) = x^3 + zy + z^2$;
4. $f(x, y, z) = (x + y)/(x + 2z)$;
5. $f(x, y) = \cos^{-1}(\sqrt{x/y})$;
6. $f(x, y) = x^y$;
7. $f(x_1, x_2, ..., x_m) = (\mathbf{a} \cdot \mathbf{r})(\mathbf{b} \cdot \mathbf{r})$,where $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$ and $\mathbf{a}$ and $\mathbf{b}$ are constant vectors.

8–12. Show without explicit calculations of higher-order partial derivatives why the hypotheses of Clairaut's theorem are satisfied for the following functions:

8. $f(x, y, z) = \sin(x^2 + y - z)\cos(xy)$;
9. $f(x, y) = \sin(x + y^2)/(x^2 + y^2), \quad x^2 + y^2 \neq 0$;
10. $f(x, y, z) = e^{x^2yz}(y^2 + zx^4)$;
11. $f(x, y) = \ln(1 + x^2 + y^4)/(x^2 - y^2), \quad x^2 \neq y^2$;
12. $f(x, y, z) = (x + yz^2 - xz^5)/(1 + x^2y^2z^4)$.

13-20. Find the indicated partial derivatives of each of the following functions:

13. $f(x, y) = x^n + xy + y^m, \quad f'''_{xxy}, f'''_{xyx}, f'''_{yyx}, f'''_{xyy}$; here n and m are positive integers;
14. $f(x, y, z) = x\cos(yx) + z^3, \quad f'''_{xyz}, f'''_{xxz}, f'''_{yyz}$;
15. $f(x, y, z) = \sin(xy)e^z, \ \partial f^5/\partial z^5, \ f^{(4)}_{xyzz}, f^{(4)}_{zyxz}, f^{(4)}_{zxzy}$;
16. $f(x, y, z, t) = \sin(x + 2y + 3z - 4t), \quad f^{(4)}_{abcd}$ where $abcd$ denotes all permutations of $xyzt$;
17. $f(x, y) = e^{xy}(y^2 + x), \quad f^{(4)}_{abcd}$ where $abcd$ denotes all permutations of $xxxy$;
18. $f(x, y, z) = \tan^{-1}\left(\frac{x+y+z-xyz}{1-xy-xz-yz}\right), \quad f^{(3)}_{abc}$ where abc denotes all permutations of xyz;
19. $f(x, y, z, t) = \ln\left((x - y)^2 + (z - t)^2\right)^{-1/2}, \quad f^{(4)}_{abcd}$ where $abcd$ are all permutations of $xyzt$;

20. $f(x,y) = e^x \sin(y)$, $\frac{\partial^{n+m} f}{\partial^n x \partial^m y}(0,0)$, where n and m are positive integers.

21-25. Given partial derivatives, find the function or show that it does not exist:

21. $f'_x = 3x^2y$, $f'_y = x^3 + 3y^2$;
22. $f'_x = yz + 3x^2$, $f'_y = xz + 4y$, $f'_z = xy + 1$;
23. $f'_{x_k} = kx_k$, $k = 1, 2, ..., m$;
24. $f'_x = xy + z$, $f'_y = x^2/2$, $f'_z = x + y$;
25. $f'_x = \sin(xy) + xy\cos(xy)$, $f'_y = x^2\cos(xy) + 1$.

26-32. Verify that a given function is a solution of the indicated differential equation:

26. $f(t,x) = A\sin(ct - x) + B\cos(ct + x)$, $c^{-2}f''_{tt} - f''_{xx} = 0$ if A, B, and c are constants;
27. $f(x,y,t) = g(ct - ax - by) + h(ct + ax + by)$, $f''_{tt} = c^2(f''_{xx} + f''_{yy})$ if a, b, and c are constants such that $a^2 + b^2 = 1$, and g and h are twice differentiable functions;
28. $f(x,y) = \ln(x^2 + y^2)$, $f''_{xx} + f''_{yy} = 0$;
29. $f(x,y) = \ln(e^x + e^y)$, $f'_x + f'_y = 1$ and $f''_{xx}f''_{yy} - (f''_{xy})^2 = 0$;
30. $f(\mathbf{r}) = \exp(\mathbf{a} \cdot \mathbf{r})$, where $\mathbf{a} \cdot \mathbf{a} = 1$ and $\mathbf{r} \in \mathbb{R}^m$, $f''_{x_1x_1} + f''_{x_2x_2} + \cdots + f''_{x_mx_m} = f$;
31. $f(\mathbf{r}) = \|\mathbf{r}\|^{2-m}$, where $\mathbf{r} \in \mathbb{R}^m$, $f''_{x_1x_1} + f''_{x_2x_2} + \cdots + f''_{x_mx_m} = 0$ for $\|\mathbf{r}\| \neq 0$;
32. $f(x,y,z) = \sin(k\|\mathbf{r}\|)/\|\mathbf{r}\|$, $f''_{xx} + f''_{yy} + f''_{zz} + k^2 f = 0$ (Helmholtz equation), where k is a constant.

33. Find a relation between constants a, b and c such that the function $u(x,y,t) = \sin(ax + by + ct)$ satisfies the wave equation $u''_{tt} - u''_{xx} - u''_{yy} = 0$. Give a geometrical description of such a relation, e.g., by setting values of c on a vertical axis, and the values of a and b on two horizontal axes.

34. Let $f(x,y,z) = u(t)$ where $t = xyz$. Show that $f^{(3)}_{xyz} = F(t)$ and find $F(t)$ in terms of $u(t)$.

35-36. Find $(f'_x)^2 + (f'_y)^2 + (f'_z)^2$ and $f''_{xx} + f''_{yy} + f''_{zz}$ for each of the following functions:

35. $f = x^3 + y^3 + z^3 - 3xyz$;
36. $f = (x^2 + y^2 + z^2)^{-1/2}$.

37-38. Let the action of K on a function f be defined by $Kf = xf'_x + yf'_y$. Find Kf, $K^2f = K(Kf)$, and $K^3f = K(K^2f)$ if

37. $f = x/(x^2 + y^2)$;
38. $f = \ln\sqrt{x^2 + y^2}$.

39. Let $f(x,y) = xy^2/(x^2 + y^2)$ if $(x,y) \neq (0,0)$ and $f(0,0) = 0$. Do $f''_{xy}(0,0)$ and $f''_{yx}(0,0)$ exist?

40–43. If $f = f(x, y)$ and $g = g(x, y, z)$, find the most general solution to each of the following equations:

40. $f''_{xx} = 0\,;$
41. $f''_{xy} = 0\,;$
42. $\partial^n f/\partial y^n = 0\,;$
43. $g'''_{xyz} = 0\,.$

44–46. Find $f(x, y)$ that satisfies the given conditions:

44. $f'_y = x^2 + 2y, \quad f(x, x^2) = 1\,;$
45. $f''_{yy} = 4, \quad f(x, 0) = 2, \quad f'_y(x, 0) = x\,;$
46. $f''_{xy} = x + y, \quad f(x, 0) = x, \quad f(0, y) = y^2\,.$

21. Differentiability of Multivariable Functions

A differentiable one-variable function $f(x)$ can be approximated near $x = x_0$ by its linearization

$$L(x) = f(x_0) + f'(x_0)(x - x_0)$$

or the tangent line. Put $x = x_0 + \Delta x$. Then by the definition of the derivative $f'(x_0)$,

$$\begin{aligned}\lim_{\Delta x\to 0} \frac{f(x) - L(x)}{\Delta x} &= \lim_{\Delta x\to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} - f'(x_0)\\ &= f'(x_0) - f'(x_0) = 0\end{aligned}$$

This relation implies that the error of the linear approximation goes to 0 faster than the deviation $\Delta x = x - x_0$ of x from x_0, that is,

$$(\mathbf{21.1})\qquad f(x) = L(x) + \varepsilon(\Delta x)\,\Delta x, \quad \text{where} \quad \varepsilon(\Delta x) \to 0 \quad \text{as} \quad \Delta x \to 0\,.$$

For example, if $f(x) = x^2$, then its linearization at $x = 1$ is

$$L(x) = 1 + 2(x - 1)\,,$$

It follows that near $x = 1$

$$f(1 + \Delta x) - L(1 + \Delta x) = (\Delta x)^2 \quad \Rightarrow \quad \varepsilon(\Delta x) = \Delta x\,.$$

Conversely, consider a line through the point $(x_0, f(x_0))$ and demand that the condition (**21**.**1**) holds. If n is the slope of the line, then

$$L(x) = f(x_0) + n(x - x_0) = f(x_0) + n\Delta x$$

and (**21**.**1**) implies that

$$\lim_{\Delta x\to 0} \frac{f(x) - L(x)}{\Delta x} = \lim_{\Delta x\to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} - n = 0$$

By the definition of the derivative $f'(x_0)$, the existence of this limit implies the existence of $f'(x_0)$ and the equality $n = f'(x_0)$. Thus, among all linear approximations of f near x_0, *only* the line with the slope $n = f'(x_0)$ is a good approximation in the sense that the error of the approximation decreases faster than Δx with decreasing Δx. The above analysis shows that:

A function $f(x)$ is differentiable at $x = x_0$ if and only if it has a good linear approximation at $x = x_0$, that is, there is a linear function $L(x)$ such that

$$\lim_{x\to x_0} \frac{f(x) - L(x)}{x - x_0} = 0\,.$$

21.1. Tangent plane to a surface. Let us try to extend the concept of a good linear approximation at a point to functions of two variables. Consider the graph $z = f(x, y)$ of a continuous two-variable function. Then the graph is a surface. Let a point $P_0 = (x_0, y_0, z_0)$ be on this surface, that is, $z_0 = f(x_0, y_0)$. A linear approximation to f at (x_0, y_0), or the most general linear function $L(x, y)$ with the property $L(x_0, y_0) = f(x_0, y_0) = z_0$, has the form

$$L(x, y) = z_0 + n_1(x - x_0) + n_2(y - y_0)\,.$$

The graph

$$z = L(x, y) \quad \Rightarrow \quad n_1(x - x_0) + n_2(y - y_0) - (z - z_0) = 0$$

is a plane through the point P_0 on the graph $z = f(x, y)$ with a normal vector $\mathbf{n} = \langle n_1, n_2, -1\rangle$. How should the slopes in the x and y directions, n_1 and n_2, be chosen in order for $L(x, y)$ to be a good linear approximation?

To answer this question, one might employ the following strategy. Consider the curve of intersection of the graph with the coordinate plane $x = x_0$. Its equation is $z = f(x_0, y)$. Then the vector function

$$\mathbf{r}_1(t) = \langle x_0\,,\ t\,,\ f(x_0, t)\rangle$$

traces out the curve of intersection. The curve goes through the point P_0 and $\mathbf{r}_1(y_0) = \overrightarrow{OP_0}$. A tangent vector at the point P_0 to this curve is

$$\mathbf{v}_1 = \mathbf{r}_1'(y_0) = \langle 0\,,\ 1\,,\ f_y'(x_0, y_0)\rangle$$

(see Fig. **21.1**). Similarly, the graph $z = f(x, y)$ intersects the coordinate plane $y = y_0$ along the curve $z = f(x, y_0)$ whose parametric equations are

$$\mathbf{r}_2(t) = \langle t, y_0, f(t, y_0)\rangle\,.$$

A tangent vector to this curve at the point P_0 is

$$\mathbf{v}_2 = \mathbf{r}_2'(x_0) = \langle 1, 0, f_x'(x_0, y_0)\rangle\,.$$

In order for these tangent vectors to exist, it has to be assumed that the function f has partial derivatives at (x_0, y_0). The lines through P_0 and parallel to the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are not parallel because $\mathbf{v}_1$ and $\mathbf{v}_2$ are not parallel. By construction, these lines are good linear approximations to two graphs, $z = f(x_0, y)$ and $z = f(x, y_0)$, in the surface $z = f(x, y)$. *It is therefore natural to assume the plane $z = L(x, y)$ that provides a good linear approximation to the surface $z = f(x, y)$ at a point P_0 must contain the lines that provide good linear approximations to two particular curves through P_0 in the surface.* Since these lines are not parallel, they define a unique plane that contains them.

Let us find this plane. A normal of this plane must be perpendicular to both vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ and, by the geometrical properties of the cross product, may be taken as

$$\mathbf{n} = \mathbf{v}_1 \times \mathbf{v}_2 = \langle f_x'(x_0, y_0)\,,\ f_y'(x_0, y_0)\,,\ -1\rangle\,.$$

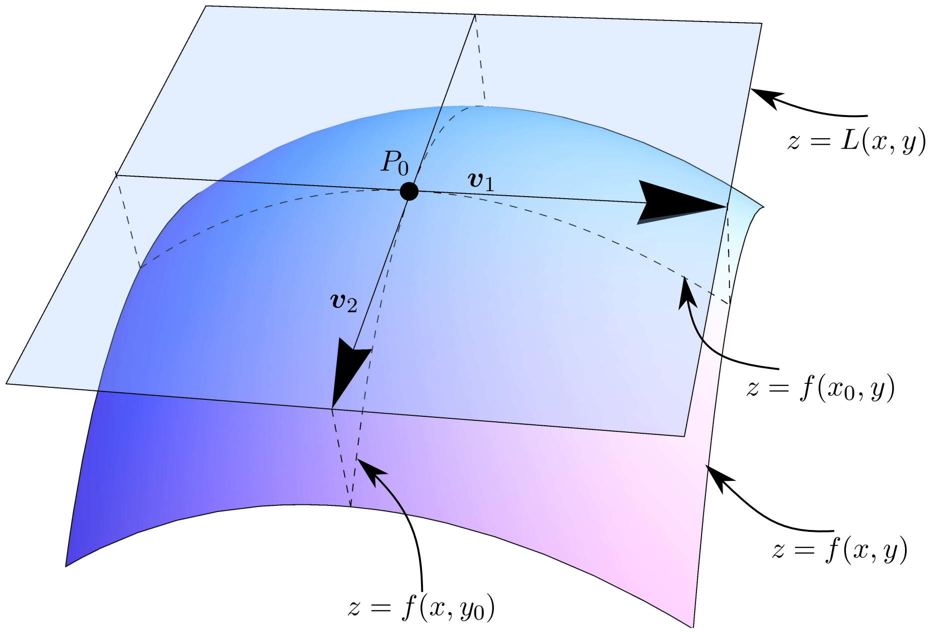


FIGURE **21.1**. The tangent plane to the graph $z = f(x, y)$ at the point $P_0 = (x_0, y_0, f(x_0, y_0))$. The curves $z = f(x_0, y)$ and $z = f(x, y_0)$ are the cross sections of the graph by the coordinate planes $x = x_0$ and $y = y_0$, respectively. Vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are tangent to the curves of the cross sections at the point P_0. The plane through P_0 and parallel to these vectors is the tangent plane to the graph. Its normal is $\mathbf{n} = \mathbf{v}_1 \times \mathbf{v}_2$.

The equation of the plane through P_0 and perpendicular to a vector $\mathbf{n}$ has the form

$$\mathbf{n} \cdot \langle x - x_0,\ y - y_0,\ z - z_0 \rangle = 0\,,$$

$$z = z_0 + f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0)\,. \tag{21.2}$$

The latter suggests that the linear approximation in which the coefficients are equal to the corresponding partial derivatives

$$L(x, y) = f(x_0, y_0) + n_1(x - x_0) + n_2(y - y_0)\,,$$
$$n_1 = f'_x(x_0, y_0)\,, \quad n_2 = f'_y(x_0, y_0)\,,$$

is the sough-after good linear approximation to the function $f(x, y)$ near the point (x_0, y_0). Since the graph of the linearization is a tangent plane to the graph, a good linear approximation is also called *a tangent plane approximation* in the case of functions of two variables.

EXAMPLE **21.1**. *Find the tangent plane to the paraboloid* $z = 3 - x^2/4 - y^2$ *through the point* $P_0 = (2, 1, 1)$.

Solution: The paraboloid is the graph of the function

$$f(x,y) = 3 - \frac{x^2}{4} - y^2\,, \qquad f(2,1) = 1$$

The normal of the tangent plane is determined by the partial derivatives

$$\begin{aligned} f'_x(2,1) &= \frac{d}{dx} f(x,1)\Big|_{x=2} = \Big(2 - \frac{x^2}{4}\Big)'\Big|_{x=2} = -\frac{x}{2}\Big|_{x=2} = -1\,, \\ f'_y(2,1) &= \frac{d}{dy} f(2,y)\Big|_{y=1} = (2-y^2)'\Big|_{y=1} = -2y\Big|_{y=1} = -2\,. \end{aligned}$$

Therefore the equation for the tangent plane (**21.2**) reads

$$z = 1 - (x-2) - 2(y-1) \quad \text{or} \quad z = 5 - x - 2y\,.$$

□

21.2. Continuity and the Existence of Partial Derivatives. Let us examine how good is the tangent plane approximation in Example **21.1**. Following the analogy with the case of functions of one variable, one has to find how fast the error of the approximation decreases with decreasing the distance between (x, y) and $(x_0, y_0) = (2, 1)$. Put $x = 2 + \Delta x$ and $y = 1 + \Delta y$. A linear approximation is given by the linear function $L(x, y) = 5 - x - 2y = 1 - \Delta x - 2\Delta y$ (the graph $z = L(x, y)$ is the tangent plane to the surface $z = f(x, y)$ at the point $(2, 1)$). Then

$$\begin{aligned} f(x,y) - L(x,y) &= f(2+\Delta x, 1+\Delta y) - 1 + \Delta x + 2\Delta y \\ &= 3 - \frac{1}{4}(2+\Delta x)^2 - (1+\Delta y)^2 - 1 + \Delta x + 2\Delta y \\ &= -\frac{1}{4}(\Delta x)^2 - (\Delta y)^2\,. \end{aligned}$$

The distance between (x, y) and (x_0, y_0) is $R = \sqrt{(\Delta x)^2 + (\Delta y)^2}$. Therefore

$$\frac{|f(x,y) - L(x,y)|}{R} = \frac{(\Delta x)^2/4 + (\Delta y)^2}{R} \le \frac{R^2/4 + R^2}{R} = \frac{5}{4}R \to 0$$

as $R \to 0$. Thus, by the squeeze principle, the error of the approximation decreases *faster* than R with decreasing R (it decreases as R^2), and the constructed linear approximation is a good one just like in the one-variable case.

It seems that the mere existence of partial derivatives at a point is sufficient for the existence of a good linear approximation at that point. Unfortunately, *this conclusion is wrong for functions of more than one variable*!

Here is a simple counterexample. It is not difficult to see that the graph of the function in Example **21.1** is the paraboloid concave downward with vertex at $(0, 0, 3)$. So the function attains its maximal value at $(0, 0)$. The partial derivatives vanish at this point $f'_x(0, 0) = f'_y(0, 0) = 0$. Hence, the tangent plane through the vertex of the paraboloid is the horizontal plane $z = 3$. Let us change the function in the following way:

$$f(x,y) \quad \to \quad g(x,y) = \begin{cases} 0 & , \quad x > 0\,,\ y > 0 \\ 3 - x^2/4 - y^2 & , \quad \text{otherwise} \end{cases}$$

In other words, the function f has been changed by setting its values to zero in the first quadrant, *excluding* the coordinate axes on which it retains its old values

$$g(0,y) = f(0,y) = 3 - y^2\,, \quad g(x,0) = f(x,0) = 3 - x^2/4$$

Therefore

$$g'_y(0,0) = f'_y(0,0) = 0\,, \quad g'_x(0,0) = f'_x(0,0) = 0\,.$$

It follows that the tangent plane to the graph $z = g(x,y)$ at $(0,0)$ is also the horizontal plane $z = 3$. But this horizontal plane is no good approximation at all! The difference $g(x,y) - 3$ remains equal to -3 for all $x > 0$, $y > 0$, no matter how close the point (x,y) to the origin $(0,0)$. It appears that the function $g(x,y)$ is not even continuous at $(0,0)$, despite the existence of its partial derivatives at $(0,0)$. The graph of f looks like a hill of the shape of a paraboloid, while the graph of g looks like the same hill with its south-east part (corresponding to the first quadrant) flatten, creating a cliff, so that one can still walk on the edge of the cliff (corresponding to the positive coordinate axes). At the top of the hill with such a cliff the slopes are still determined by (zero) slopes along the edges of the cliff, while the height drops abruptly in the south-east direction. It is also interesting to point out if the coordinate system relative to which the height of the hill is determined was rotated, say, by 45°, one of the partial derivatives relative to the new variables would not exist at the origin.

EXAMPLE **21.2**. *Consider the function*

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

Show that this function is not continuous at $(0,0)$*, but that the partial derivatives* $f'_x(0,0)$ *and* $f'_y(0,0)$ *exist.*

SOLUTION: In order to check the continuity, one has to calculate the limit $\lim_{(x,y)\to(0,0)} f(x,y)$. If it exists and equals $f(0,0) = 0$, then the function is continuous at $(0,0)$. This limit does not exist. Along lines $(x,y) = (t,at)$, the function has constant value

$$f(t,at) = \frac{at^2}{t^2 + a^2t^2} = \frac{a}{1+a^2}\,, \quad t \neq 0\,.$$

Therefore the limit of $f(x,y)$ as $(x,y) \to (0,0)$ does not exist and the function is not continuous at $(0,0)$. Note also that the non-existence of the limit means that the function is not even continuously extendable at $(0,0)$. In other words, there is no value of $f(0,0)$ at which f would become continuous.

The partial derivatives in question can be found by the rule **19.1**:

$$f(x,0) = 0 \quad \Rightarrow \quad f'_x(0,0) = \frac{d}{dx} f(x,0)\Big|_{x=0} = 0\,,$$
$$f(0,y) = 0 \quad \Rightarrow \quad f'_y(0,0) = \frac{d}{dy} f(0,y)\Big|_{y=0} = 0\,.$$

Thus, the partial derivatives exist at the origin, $f'_x(0,0) = f_y(0,0) = 0$ but f is not continuous at the origin. □

In the example considered, the plane defined by Eq. **21.2** at $(x_0, y_0) = (0,0)$ is $z = 0$. The deviation of the graph $z = f(x,y)$ from the xy plane is *finite* along the lines $(x,y) = (t, at)$, $a \neq 0$ no matter how close is t to 0. So, it does not even decrease with decreasing the distance from the origin! Thus, the tangent plane $z = 0$ defined by Eq. (**21.2**) *no approximation* at all to the graph of the function near the origin.

The above examples illustrate the following important relation between the continuity and the existence of partial derivatives:

- *The existence of partial derivatives at a point does not imply that the function is continuous at that point*;

In full contrast to the one-variable case, this implies that the deviation of the plane defined by Eq. (**21.2**) from the graph $z = f(x,y)$ may not be small no matter how close is the point (x,y) to (x_0, y_0) even though f has partial derivatives at (x_0, y_0).

This is quite a departure from the one-variable case, where the existence of the derivative at a point implied continuity at that point and was necessary and sufficient for the existence of a good linear approximation. This is the reason that functions of several variables require a generalization of the concept of differentiability. *Differentiability of a function of several variables at a point will be understood in the sense that the function has a good linear approximation at that point.*

21.3. Differentiability of multivariable functions. Consider a function of two variables $f(x,y)$ and a point (x_0, y_0) in its domain. As already noted, the most general linear function $L(x,y)$ with the property $L(x_0,y_0) = f(x_0,y_0)$ has the form

$$L(x,y) = f(x_0,y_0) + n_1(x - x_0) + n_2(y - y_0)$$

where n_1 and n_2 are arbitrary numbers. It defines a linear approximation to $f(x,y)$ near (x_0,y_0) in the sense that $L(x_0,y_0) = f(x_0,y_0)$. This notion is extended to functions of any number of variables. Given a multivariable function $f(\mathbf{r})$, a linear function

$$L(\mathbf{r}) = f(\mathbf{r}_0) + \mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0)$$

is said to be a *linear approximation* to f near $\mathbf{r}_0$ in the sense that $L(\mathbf{r}_0) = f(\mathbf{r}_0)$. The dot product is defined in an $m-$dimensional Euclidean space if f is a function of m variables. The vector $\mathbf{n}$ is an arbitrary vector so that $L(\mathbf{r})$

is the most general linear function satisfying the condition $L(\mathbf{r}_0) = f(\mathbf{r}_0)$. Note that in the case of two variables $x_1 = x$ and $x_2 = y$, $\mathbf{n} = \langle n_1, n_2\rangle$ and $\mathbf{r} - \mathbf{r}_0 = \langle x - x_0, y - y_0\rangle$ so that $\mathbf{n}\cdot(\mathbf{r} - \mathbf{r}_0) = n_1(x - x_0) + n_2(y - y_0)$.

DEFINITION **21.1**. (Differentiable Functions).
The function f of several variables $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$ on an open set D is said to be differentiable at a point $\mathbf{r}_0$ of D if there exists a linear approximation $L(\mathbf{r})$ such that

$$\lim_{\mathbf{r}\to\mathbf{r}_0} \frac{f(\mathbf{r}) - L(\mathbf{r})}{\|\mathbf{r} - \mathbf{r}_0\|} = 0. \tag{21.3}$$

If f is differentiable at all points of D, then f is said to be differentiable on D.

In what follows, functions *differentiable on an open set* will be called *differentiable*. By this definition, *the differentiability of a function is independent of a coordinate system* chosen to label points of D (a linear function remains linear under general rotations and translations of the coordinate system and the distance $\|\mathbf{r} - \mathbf{r}_0\|$ is also invariant under these transformations). For functions of single variable $f(x)$ the existence of a linear approximation at x_0 with the property (**21.3**) is equivalent to the existence of the derivative $f'(x_0)$. Indeed, put $x - x_0 = \Delta x$. Then the condition (**21.3**) is equivalent to (**21.1**):

$$\lim_{\Delta x\to 0} \frac{f(x) - L(x)}{|\Delta x|} = 0 \quad\Leftrightarrow\quad \lim_{\Delta x\to 0} \frac{f(x) - L(x)}{\Delta x} = 0 \tag{21.4}$$

and, hence, to the existence of $f'(x_0)$ as shown above. Note that the existence of the first limit implies that the corresponding left ($\Delta x \to 0^-$) and right ($\Delta x \to 0^+$) limits are equal to 0. Since $|\Delta x|$ and Δx differ only by sign, the right and left limits corresponding to the second limit also vanish and, hence, so does the second limit. The converse is established similarly.

By Definition **21.1**, *a function f is differentiable at a point $\mathbf{r}_0$ if it has a good linear approximation* in the sense that the error of the approximation decreases faster than the distance $\|\mathbf{r} - \mathbf{r}_0\|$ as $\mathbf{r} \to \mathbf{r}_0$:

$$f(\mathbf{r}) = L(\mathbf{r}) + \varepsilon(\mathbf{r})\|\mathbf{r} - \mathbf{r}_0\|, \quad \text{where} \quad \varepsilon(\mathbf{r}) \to 0 \quad \text{as} \quad \mathbf{r} \to \mathbf{r}_0\,. \tag{21.5}$$

To make Definition **21.1** of differentiability consistent, one has to show that a good linear approximation is unique if it exists. This is true for functions of one variable as has been already shown.

THEOREM **21.1**. (Uniqueness of a good linear approximation)
*A linear approximation L to a multivariable function f near a point $\mathbf{r}_0$ that satisfies the property (**21.3**) is unique if it exists.*

PROOF. Suppose that, contrary to the conclusion of the theorem, there are two linear approximations $L_1(\mathbf{r}) = f(\mathbf{r}_0) + \mathbf{n}_1\cdot(\mathbf{r} - \mathbf{r}_0)$ and $L_2(\mathbf{r}) =$

$f(\mathbf{r}_0) + \mathbf{n}_2 \cdot (\mathbf{r} - \mathbf{r}_0)$ that satisfy the condition (**21.3**) for which $\mathbf{n}_1 \neq \mathbf{n}_2$. Making use of the identity

$$L_2(\mathbf{r}) - L_1(\mathbf{r}) = [f(\mathbf{r}) - L_1(\mathbf{r})] - [f(\mathbf{r}) - L_2(\mathbf{r})]\,,$$

it is concluded that

$$\lim_{\mathbf{r}\to\mathbf{r}_0} \frac{L_2(\mathbf{r}) - L_1(\mathbf{r})}{\|\mathbf{r} - \mathbf{r}_0\|} = \lim_{\mathbf{r}\to\mathbf{r}_0} \frac{f(\mathbf{r}) - L_1(\mathbf{r})}{\|\mathbf{r} - \mathbf{r}_0\|} - \lim_{\mathbf{r}\to\mathbf{r}_0} \frac{f(\mathbf{r}) - L_2(\mathbf{r})}{\|\mathbf{r} - \mathbf{r}_0\|} = 0$$

Note that owing to the existence of the limit (**21.3**) for both linear functions L_1 and L_2, the limit of the difference equals the difference of the limits (the basic law of limits). On the other hand,

$$L_2(\mathbf{r}) - L_1(\mathbf{r}) = (\mathbf{n}_2 - \mathbf{n}_1) \cdot (\mathbf{r} - \mathbf{r}_0)\,.$$

Put $\mathbf{n} = \mathbf{n}_2 - \mathbf{n}_1$; note that $\mathbf{n} \neq \mathbf{0}$ since $\mathbf{n}_1 \neq \mathbf{n}_2$ by the assumption. Then

$$0 = \lim_{\mathbf{r}\to\mathbf{r}_0} \frac{L_2(\mathbf{r}) - L_1(\mathbf{r})}{\|\mathbf{r} - \mathbf{r}_0\|} = \lim_{\mathbf{r}\to\mathbf{r}_0} \frac{\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0)}{\|\mathbf{r} - \mathbf{r}_0\|}\,.$$

If a multivariable limit exists, then its value does not depend on a curve along which the limit point is approached. In particular, take the straight line parallel to $\mathbf{n}$, $\mathbf{r} = \mathbf{r}_0 + \mathbf{n}t$, $t \to 0^+$, in the above relation. Then along this line,

$$\frac{\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0)}{\|\mathbf{r} - \mathbf{r}_0\|} = \frac{t\mathbf{n} \cdot \mathbf{n}}{|t|\|\mathbf{n}\|} = \frac{\mathbf{n} \cdot \mathbf{n}}{\|\mathbf{n}\|}$$

because $t > 0$, and, hence,

$$0 = \lim_{t\to 0^+} \frac{\mathbf{n} \cdot \mathbf{n}}{\|\mathbf{n}\|} = \frac{\mathbf{n} \cdot \mathbf{n}}{\|\mathbf{n}\|} = \|\mathbf{n}\| \quad \Rightarrow \quad \mathbf{n} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{n}_1 = \mathbf{n}_2$$

which is a contradiction. Thus, a good linear approximation in the sense (**21.3**) is unique if it exists, $L_1(\mathbf{r}) = L_2(\mathbf{r})$. □

21.4. Properties of Differentiable Functions. In the one-variable case, a function $f(x)$ is differentiable at x_0 if and only if it has the derivative $f'(x_0)$. Also, the existence of the derivative at x_0 implies continuity at x_0 (recall Calculus I). In the multivariable case the relations between differentiability, continuity, and the existence of partial derivatives are more subtle.

THEOREM **21.2**. (Properties of Differentiable Functions).
If f is differentiable at a point $\mathbf{r}_0$, then it is continuous at $\mathbf{r}_0$ and its partial derivatives exist at $\mathbf{r}_0$.

PROOF. A linear function is continuous (it is a polynomial of degree one). Therefore $L(\mathbf{r}) \to L(\mathbf{r}_0) = f(\mathbf{r}_0)$ as $\mathbf{r} \to \mathbf{r}_0$. By taking the limit $\mathbf{r} \to \mathbf{r}_0$ in (**21.5**), it is concluded that $f(\mathbf{r}) \to f(\mathbf{r}_0)$. Hence, f is continuous at $\mathbf{r}_0$. If the multivariable limit (**21.3**) exists, then the limit along any curve through the limit point has the same value. In particular, take a straight line parallel to the jth coordinate axis. If $\hat{\mathbf{e}}_j$ is the unit vector parallel to this axis, then the vector equation of the line is $\mathbf{r} = \mathbf{r}(t) = \mathbf{r}_0 + t\hat{\mathbf{e}}_j$. Then $\|\mathbf{r} - \mathbf{r}_0\| = |t| \to 0$ as $t \to 0$ along the line, and

$$\begin{aligned} f(\mathbf{r}(t)) - L(\mathbf{r}(t)) &= f(\mathbf{r}_0 + t\hat{\mathbf{e}}_j) - f(\mathbf{r}_0) - \mathbf{n} \cdot \hat{\mathbf{e}}_j t \\ &= f(\mathbf{r}_0 + t\hat{\mathbf{e}}_j) - f(\mathbf{r}_0) - n_j t \end{aligned}$$

where n_j is the jth component of the vector $\mathbf{n}$ parallel to the line. Making use of Eq. **21.4** with $\Delta x = t$, it is concluded that the condition (**21.3**) implies

$$\lim_{t \to 0} \frac{f(\mathbf{r}_0 + t\hat{\mathbf{e}}_j) - f(\mathbf{r}_0)}{t} - n_j = 0 \quad \Rightarrow \quad n_j = f'_{x_j}(\mathbf{r}_0)$$

according to Definition **19.3** of partial derivatives at a point. The existence of partial derivatives is guaranteed by the existence of the limit (**21.3**). □

The proof also shows that the components of the vector $\mathbf{n}$ that defines a good linear approximation at a point are partial derivatives at that point:

$$f \text{ is differentiable at } \mathbf{r}_0 \quad \Rightarrow \quad L(\mathbf{r}) = f(\mathbf{r}_0) + \mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0), \quad n_j = f'_{x_j}(\mathbf{r}_0)$$

In particular, if $f(x, y)$ is differentiable at (x_0, y_0), then there is a unique good linear approximation defined by the tangent plane (**21.2**).

How to verify differentiability. It follows from Theorem **21.2** that a function is not differentiable at a point if its partial derivatives do not exist at the point. Therefore the existence of partial derivatives is a necessary (but not sufficient) condition for differentiability. This observation allows us to formulate a practical procedure to verify differentiability of a given function $f(\mathbf{r})$ at a point $\mathbf{r}_0$:

- *Find partial derivatives of f at $\mathbf{r}_0$. If at least one of them does not exist, f is not differentiable at $\mathbf{r}_0$.*
- *Construct the linear function $L(\mathbf{r}) = \mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0)$ where the components of the vector $\mathbf{n}$ are the corresponding partial derivatives of f at $\mathbf{r}_0$.*
- *Investigate the limit (**21.3**). If it exists and is equal to 0, then f is differentiable at $\mathbf{r}_0$. Otherwise, f is not differentiable at $\mathbf{r}_0$.*

Continuity and the Existence of Partial Derivatives vs Differentiability. Theorem **21.2** states that both the *continuity* and the *existence of partial derivatives* are *necessary* for differentiability. Is the converse of Theorem **21.2** true? in other words, are the continuity and the existence of partial derivatives sufficient for differentiability? Unfortunately, this in not so:

- *The existence of partial derivatives of a continuous function at a point does not imply differentiability at that point.*
- *A good linear approximation in the sense (**21.3**) (or (**21.5**)) may not exist even if a function is continuous and has partial derivatives at a point.*

The following example proves this assertion.

EXAMPLE **21.3**. *Let*

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

Show that f is continuous at $(0,0)$ and has the partial derivatives $f'_x(0,0)$ and $f'_y(0,0)$, but that it is not differentiable at $(0,0)$.

SOLUTION: The continuity is verified by the squeeze principle. Put $R = \sqrt{x^2+y^2}$. Then $|xy| = |x||y| \leq R^2$. Therefore

$$|f(x,y)| \leq \frac{R^2}{R} = R \to 0 \quad \text{as} \quad R \to 0^+ \,,$$

By the squeeze principle,

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0 = f(0,0)$$

and, hence, f is continuous at $(0,0)$. The partial derivatives are found in the same fashion as in Example **21.2**:

$$\begin{array}{lllll} f(x,0) = 0 & \Rightarrow & f'_x(x,0) = 0 & \Rightarrow & f'_x(0,0) = 0 \\ f(0,y) = 0 & \Rightarrow & f'_y(0,y) = 0 & \Rightarrow & f'_y(0,0) = 0 \end{array}$$

The continuity of f and the existence of its partial derivatives at $(0,0)$ suggest that if a linear approximation with the property (**21.3**) exists, then $L(x,y) = 0$ by Theorem **21.2**. However, $L(x,y) = 0$ does not satisfy (**21.3**). Indeed, in this case

$$\frac{f(x,y) - L(x,y)}{\|\mathbf{r}\|} = \frac{f(x,y)}{\sqrt{x^2+y^2}} = \frac{xy}{x^2+y^2}\,, \quad (x,y) \neq (0,0)$$

which coincides with the function from Example **21.2** for $(x,y) \neq (0,0)$. In Example **21.2** it has been shown that this function takes constant nonzero values along straight lines through the origin (that do not coincide with the coordinate axes). It does not approach 0 as $(x,y) \to (0,0)$. Therefore the characteristic property (**21.3**) cannot be fulfilled and the function in question is not differentiable at $(0,0)$. The plane $z = 0$ is not a good linear approximation to f at $(0,0)$. □

The following theorem establishes a *sufficient* condition for differentiability (its proof is omitted).

THEOREM **21.3**. (Differentiability and Partial Derivatives).
Let f be a function on an open set D of a Euclidean space. Then f is differentiable on D if its partial derivatives exist and are continuous functions on D.

Thus, in order for a function to be differentiable at a point, its partial derivatives should exist in a neighborhood of the point and be *continuous* at that point. The continuity of partial derivatives ensures the existence a good linear approximation **21.5**. This is a useful criterion for differentiability which is easy to verify in most cases.

EXAMPLE **21.4**. *Find the region in which the function* $e^{xz}\cos(yz)$ *is differentiable.*

SOLUTION: The function e^{xz} is the composition of the exponential e^u and the polynomial $u = xz$. By the chain rule **19.2**, its partial derivatives are continuous everywhere (e.g., $(e^{xz})'_x = ze^{xz}$). Similarly, the partial derivatives of $\cos(yz)$ are also continuous everywhere. By the product rule for partial derivatives, the partial derivatives of $e^{xz}\cos(yz)$ are continuous everywhere. By Theorem **21.3**, the function is differentiable everywhere and a good linear approximation exists everywhere. □

Remark. Theorem **21.3** provides only a *sufficient* condition for differentiability. It is important to emphasize that:

There are functions differentiable at a point whose partial derivatives exist in a neighborhood of the point but are not continuous at that point.

An example is discussed in Study Problem **21.1**.

21.5. Linearization of a function. The concept of differentiability is important for approximations. Only differentiable functions have a good linear approximation. Owing to the uniqueness of the good linear approximation, it is convenient to give it a name.

DEFINITION **21.2**. (Linearization of a Multivariable Function). *Let* f *be a function of* m *variables* $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$ *on* D *that is differentiable at an interior point* $\mathbf{r}_0 = \langle a_1, a_2, ..., a_m\rangle$ *of* D. *Put* $n_i = f'_{x_i}(\mathbf{r}_0)$, $i = 1, 2, ..., m$. *The linear function*

$$L(\mathbf{r}) = f(\mathbf{r}_0) + n_1(x_1 - a_1) + n_2(x_2 - a_2) + \cdots + n_m(x_m - a_m)$$

is called the linearization *of* f *at* $\mathbf{r}_0$.

If Δx_i denotes the deviation of x_i from a_i, then

$$(\mathbf{21.6})\quad L(\mathbf{r}) = f(\mathbf{r}_0) + n_1\Delta x_1 + n_2\Delta x_2 + \cdots + n_m\Delta x_m\,, \quad n_i = f'_{x_i}(\mathbf{r}_0)\,.$$

EXAMPLE **21.5**. *Find all points at which the function* $f(x, y) = x^2 + 3y^2$ *has a good linear approximation. Find the tangent plane to the graph* $z = f(x, y)$ *at the point* $(2, 1, 7)$ *or show that no such plane exists.*

SOLUTION: The tangent plane at a point (x_0, y_0, z_0) of the graph $z = f(x, y)$ exits if and only if $f(x, y)$ is differentiable at (x_0, y_0). The polynomial function $f(x, y) = x^2 + 3y^2$ has continuous partial derivatives at any point and, by Theorem **21.3**, f is differentiable everywhere. The components of a normal of the tangent plane are

$$n_1 = f'_x(2, 1) = 2x\Big|_{(2,1)} = 4\,, \quad n_2 = f'_y(2, 1) = 6y\Big|_{(2,1)} = 6\,, \quad n_3 = -1\,.$$

An equation of the tangent plane is $4(x - 2) + 6(y - 1) - (z - 7) = 0$ or $4x + 6y - z = 7$. □

Example **21.6**. *Use the linearization to estimate the number* $[(2.03)^2 + (1.97)^2 + (0.94)^2]^{1/2}$.

Solution: Let $f(x, y, z) = [x^2 + y^2 + z^2]^{1/2}$. This function has continuous partial derivatives everywhere except the origin because it is a composition of the polynomial $g = x^2 + y^2 + z^2$ and the power function: $f = (g)^{1/2}$. By Theorem **21.3**, f is differentiable everywhere except the origin. The number in question is the value of this function at $(x, y, z) = (2.03, 1.97, 0.94)$. This point is close to $(x_0, y_0, z_0) = (2, 2, 1)$ at which $f(2, 2, 1) = 3$. Since f is differentiable at (x_0, y_0, z_0), its linearization can be used to approximate values of f in a neighborhood of (x_0, y_0, z_0) The deviations are

$$\begin{aligned} \Delta x &= x - x_0 = 2.03 - 2 = 0.03\,, \\ \Delta y &= y - y_0 = 1.97 - 2 = -0.03\,, \\ \Delta z &= z - z_0 = 0.94 - 1 = -0.06\,. \end{aligned}$$

The partial derivatives are

$$\begin{aligned} f'_x(x_0, y_0, x_0) &= \frac{x}{(x^2 + y^2 + z^2)^{1/2}}\Big|_{(2,2,1)} = \frac{2}{3}\,, \\ f'_x(x_0, y_0, x_0) &= \frac{y}{(x^2 + y^2 + z^2)^{1/2}}\Big|_{(2,2,1)} = \frac{2}{3}\,, \\ f'_x(x_0, y_0, x_0) &= \frac{z}{(x^2 + y^2 + z^2)^{1/2}}\Big|_{(2,2,1)} = \frac{1}{3}\,. \end{aligned}$$

The linear approximation (**21.6**) gives

$$f(x, y, z) \approx L(x, y, z) = 3 + (2/3)\,\Delta x + (2/3)\,\Delta y + (1/3)\,\Delta z = 2.98.$$

The calculator value is 2.98084 (rounded to 5 decimal places). □

Example **21.7**. *Is there a point on the saddle surface* $z = xy$ *at which the tangent plane is parallel to the plane* $3x + 2y + z = 4$*? Find this tangent plane if it exists.*

Solution: The saddle surface is the graph of the function $f(x, y) = xy$ which is differentiable everywhere in the xy plane because it is a polynomial. So, the surface has a tangent plane at each point. Let a point (x_0, y_0, z_0) be on the surface, that is, $z_0 = x_0 y_0$. Then a normal of the tangent plane at this point is

$$\langle f'_x(x_0, y_0)\,,\ f'_y(x_0, y_0)\,,\ -1\rangle = \langle y_0\,,\ x_0\,,\ -1\rangle\,.$$

Two plane are parallel if their normals are parallel. Therefore x_0 and y_0 must such that the above vector is parallel t the vector $\langle 3, 2, 1\rangle$ (the normal of the given plane). Two vectors are parallel if and only if they are proportional. So there should exists a number s such that

$$\langle y_0\,,\ x_0\,,\ -1\rangle = s\langle 3\,,\ 2\,,\ 1\rangle \quad \Rightarrow \quad \begin{cases} y_0 = 3s \\ x_0 = 2s \\ -1 = s \end{cases} \quad \Rightarrow \quad \begin{aligned} x_0 &= -2 \\ y_0 &= -3 \end{aligned}$$

The equation of the tangent plane through the point $(-2, -3, 6)$ is

$$z = 6 - 3(x + 2) - 2(y + 3) \quad \text{or} \quad 3x + 2y + z = 6\,.$$

$\square$

21.6. Applications to a system of nonlinear equations. Consider two differentiable functions $f(x, y)$ and $g(x, y)$. Suppose that level sets $f(x, y) = a$ and $g(x, y) = b$ of these functions have a common point (x_0, y_0), that is, the system of equations has a solution:

$$\text{(21.7)} \qquad \begin{cases} f(x, y) = a \\ g(x, y) = b \end{cases} \quad \Rightarrow \quad (x, y) = (x_0, y_0)\,.$$

One might think of levels sets as curves so that a solution is a point of intersection of the curves. Clearly the values of x_0 and y_0 depend on a and b. Suppose the value of a is changed by adding a small number Δa and b is changed by adding a small number Δb. What are the new values of x_0 and y_0? In other words, what is a point of intersection of two level curves that are "close" to the given level curves if the point of intersection of the latter is known? It is generally a difficult task to find an exact solution of a system of nonlinear equations. So an approximation should be used. Since the functions are differentiable, their values in a neighborhood of (x_0, y_0) are well approximated by their linearization. So the system of nonlinear equation may be *linearized* to obtain an *approximate* solution. Let the new solution be $x = x_0 + \Delta x$ and $y = y_0 + \Delta y$ where Δx and Δy are to be determined. Then by linearizing the function f and g at (x_0, y_0),

$$\begin{cases} f(x_0, y_0) + f'_x(x_0, y_0)\Delta x + f'_y(x_0, y_0)\Delta y = a + \Delta a \\ g(x_0, y_0) + g'_x(x_0, y_0)\Delta x + g'_y(x_0, y_0)\Delta y = b + \Delta b \end{cases}$$

Since $f(x_0, y_0) = a$ and $g(x_0, y_0) = b$, the deviations Δx and Δy satisfies the system of linear equations:

$$\begin{cases} f'_x(x_0, y_0)\Delta x + f'_y(x_0, y_0)\Delta y = \Delta a \\ g'_x(x_0, y_0)\Delta x + g'_y(x_0, y_0)\Delta y = \Delta b \end{cases}$$

which is easy to solve. Recall that such system of linear equations has a unique solution if

$$\det \begin{pmatrix} f'_x(x_0, y_0) & f'_y(x_0, y_0) \\ g'_x(x_0, y_0) & g'_y(x_0, y_0) \end{pmatrix} \neq 0$$

The concept of using linearization to approximate a solution of a system of nonlinear equations is readily extended to the case of more than two variables.

Multivariable Newton's method. In Calculus I, Newton's method was introduced to find an approximate solution to a nonlinear equation. There is its multi-variable generalization. It employs the same logic as in the problem just discussed. Suppose that the system (**21.7**) with $a = b = 0$ has a solution (x_0, y_0) (there is no loss of generality here because the constants a and b may always be included into the definitions of f and g). Let a point (x_1, y_1) be such that the values $f(x_1, y_1)$ and $g(x_1, y_1)$ are small ((x_1, y_1) is an initial guess for the solution). Put $x_2 = x_1 + \Delta x_1$ and $y_2 = y_1 + \Delta y_1$, linearize the functions f and g at (x_1, y_1), and demand that Δx_1 and Δy_1 satisfy the linearized system (**21.7**):

$$\begin{cases} f'_x(x_1, y_1)\Delta x_1 + f'_y(x_1, y_1)\Delta y_1 = -f(x_1, y_1) \\ g'_x(x_1, y_1)\Delta x_1 + g'_y(x_1, y_1)\Delta y_1 = -g(x_1, y_1) \end{cases} .$$

The point (x_2, y_2) is expected to be closer to the true solution (x_0, y_0). To obtain an even better approximation, the procedure may be repeated for a point (x_2, y_2) (the system is linearized at (x_2, y_2) and solved to obtained a better approximation (x_3, y_3)). This recursive procedure can be used to construct a *sequence* of points (x_n, y_n), $n = 1, 2, \ldots$, where $x_{n+1} = x_n + \Delta x_n$ and $y_{n+1} = y_n + \Delta y_n$ with Δx_n and Δy_n being the solution of the linearized system

$$\begin{cases} f'_x(x_n, y_n)\Delta x_n + f'_y(x_n, y_n)\Delta y_n = -f(x_n, y_n) \\ g'_x(x_n, y_n)\Delta x_n + g'_y(x_n, y_n)\Delta y_n = -g(x_n, y_n) \end{cases} , \quad n = 1, 2, \ldots$$

Just like in the case of one-variable Newton's method, it can be proved that if (x_1, y_1) is sufficiently close to the true solution (x_0, y_0), the sequence of points $\mathbf{r}_n = \langle x_n, y_n \rangle$ converges to $\mathbf{r}_0 = \langle x_0, y_0 \rangle$, that is, $\|\mathbf{r}_n - \mathbf{r}_0\| \to 0$ as $n \to \infty$, and, hence, $f(x_n, y_n) \to 0$ and $g(x_n, y_n) \to 0$ as $n \to \infty$. The exact sufficient conditions under which this sequence converges to the true solution are studied in more advanced calculus courses. In practice, it is often sufficient to verify that the absolute values $|f(x_n, y_n)|$ and $|g(x_n, y_n)|$ become smaller than a preassigned small positive number (say, 10^{-6}) and use the point at which this occurs as an approximate solution. A generalization to the case of more than two variables follows the same line of reasoning.

Remark. Note that Newton's method is based on the assumption that *the system has a solution* near the initial point. So before initiating Newton's algorithm, the existence of a solution must be established by some means (e.g., by analyzing level curves of the functions f and g and showing that they intersect).

EXAMPLE **21.8**. *Use one iteration of Newton's method initiated at the point* $(x_1, y_1) = (0, 0)$ *to find an approximate solution to the system*

$$\begin{cases} e^{xy} + 2x + y = 1.1 \\ \sin(x + y) + x - 2y = 0.3 \end{cases}$$

Assume that a solution exists near the initial point. Verify that the obtained approximate solution is better than the initial point.

SOLUTION: Put $f(x,y) = e^{xy} + 2x + y$ and $g(x,y) = \sin(x+y) + x - 2y$ so that at the initial point $f(0,0) = 1$ and $g(0,0) = 0$. The functions have continuous partial derivatives everywhere and, hence, differentiable everywhere. In particular,

$$f'_x(0,0) = ye^{xy} + 2\Big|_{(0,0)} = 2\,, \qquad f'_y(0,0) = xe^{xy} + 1\Big|_{(0,0)} = 1\,,$$

$$g'_x(0,0) = \cos(x+y) + 1\Big|_{(0,0)} = 2\,, \quad g'_y(0,0) = \cos(x+y) - 2\Big|_{(0,0)} = -1\,.$$

Let $x = x_1 + \Delta x = \Delta x$ and $y = y_1 + \Delta y = \Delta y$. By linearizing the equations at $(0,0)$, the system of linear equations is obtained for Δx and Δy and solved:

$$\begin{cases} 2\Delta x + \Delta y = 1.1 - f(0,0) = 0.1 \\ 2\Delta x - \Delta y = 0.3 - g(0,0) = 0.3 \end{cases} \quad \Rightarrow \quad \Delta x = 0.1 \quad \Delta y = -0.1\,.$$

So an approximation to the root is the point $(0.1, -0.1)$. If (x_0, y_0) is a true solution, then $f(x_0, y_0) = 1.1$ and $g(x_0, y_0) = 0.3$. It follows then

$$f(0.1, -0.1) \approx 1.09\,, \quad g(0.1, -0.1) = 0.3\,.$$

Thus, the values of the functions appear to be closer to their values at the true solution. So, the point $(0.1, -0.1)$ is a better approximation to the true solution than the initial guess $(0,0)$. □

21.7. Study Problems.

Problem **21.1**. *Let*

$$f(x,y) = \begin{cases} (x^2+y^2)\sin\left(\frac{1}{x^2+y^2}\right) & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}.$$

Show that f is differentiable at $(0,0)$ (and hence that $f'_x(0,0)$ and $f'_y(0,0)$ exist), but that f'_x and f'_y are not continuous at $(0,0)$.

SOLUTION: By the definition of partial derivatives

$$f'_x(0,0) = \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} = \lim_{h\to 0} h\sin(1/h^2) = 0$$

which follows from the squeeze principle: $0 \leq |h\sin(1/h^2)| \leq |h| \to 0$ where the inequality $|\sin u| \leq 1$ has been used. Similarly, $f'_y(0,0) = 0$. Since the partial derivatives exist, to show differentiability of f at $(0,0)$, one has to verify whether the linear function $L(x,y) = f(0,0) + f'_x(0,0)x + f'_y(0,0)y = 0$ satisfies the condition (**21.3**). Put $R = \sqrt{x^2+y^2}$. Then

$$\lim_{(x,y)\to(0,0)} \frac{f(x,y) - L(x,y)}{\sqrt{x^2+y^2}} = \lim_{R\to 0^+} R\sin(1/R^2) = 0$$

by the squeeze principle: $0 \le |R\sin(1/R^2)| \le R \to 0$ as $R \to 0^+$ where the inequality $|\sin u| \le 1$ has been used. Thus, $L(x,y) = 0$ is indeed a good linear approximation and the function f is differentiable at the origin.

To investigate the continuity of partial derivatives, one has to study the limits $f'_x(x,y)$ and $f'_y(x,y)$ at the origin. For $(x,y) \ne (0,0)$,

$$f'_x(x,y) = 2x\sin\left(\frac{1}{x^2+y^2}\right) - \frac{2x}{x^2+y^2}\cos\left(\frac{1}{x^2+y^2}\right)$$

The first term in this expression converges to 0 by the squeeze principle: $0 \le |2x\sin(1/r^2)| \le 2|x| \to 0$ as $(x,y) \to (0,0)$, whereas the second term can take arbitrary large values in any neighborhood of the origin. To see this, consider a sequence of points that converges to the origin:

$$(x_n, y_n) = \left(\frac{1}{\sqrt{\pi n}}\,,\ 0\right), \quad n = 1, 2, ..., \quad R_n^2 = x_n^2 + y_n^2 = \frac{1}{\pi n}$$

so that $R_n \to 0$ as $n \to \infty$. Then $\cos(1/R_n^2) = (-1)^n$, $\sin(1/R_n^2) = 0$, and

$$f'_x(x_n, y_n) = 2R_n \sin\left(\frac{1}{R_n^2}\right) - \frac{2}{R_n}\cos\left(\frac{1}{R_n^2}\right) = 2(-1)^{n+1}\sqrt{\pi n}\,.$$

So $f'_x(x_n, y_n)$ can take arbitrary large positive and negative values as $n \to \infty$ and the limit $\lim_{(x,y)\to(0,0)} f'_x(x,y)$ does not exist, which means that the partial derivative $f'_x(x,y)$ is not continuous at the origin. Owing to the symmetry $f(x,y) = f(y,x)$, the same conclusion holds for $f'_y(x,y)$. $\square$

21.8. Exercises.

1. Let $g(x,y) = 4 - x^2 - y^2$. Sketch the graph $z = g(x,y)$ for the disk $x^2 + y^2 \le 4$. Find $g'_x(1,1)$ and $g'_y(1,1)$. Consider the plane $z = g(1,1) + g'_x(1,1)(x-1) + g'_y(1,1)(y-1) = L(x,y)$. Is this plane a good linear approximation to the function near $(1,1)$? Is the function differentiable at $(1,1)$?

2. Let $f(x,y) = 4-x^2-y^2$ if (x,y) is not in $D = \{(x,y)\,|\, x < y < 2x-1\,, x > 1\}$ and $f(x,y) = 0$ if (x,y) is in D. Sketch D and the graph $z = f(x,y)$ for the disk $x^2+y^2 \le 4$. Use the definition of partial derivatives to find $f'_x(1,1)$ and $f'_y(1,1)$. Consider the plane $z = f(1,1)+f'_x(1,1)(x-1)+f'_y(1,1)(y-1) = L(x,y)$. Is this plane a good linear approximation to the function near $(1,1)$? Is the function differentiable at $(1,1)$? Hint: Investigate the difference $f(x,y) - L(x,y)$ along straight lines through the point $(1,1)$.

3. Let $f(x,y) = xy^2$ if $(x,y) \ne (0,0)$ and $f(0,0) = 1$ and let $g(x,y) = \sqrt{|x|} + \sqrt{|y|}$. Are the functions f and g differentiable at $(0,0)$?

4. Let $f(x,y) = xy^2/(x^2+y^2)$ if $x^2+y^2 \ne 0$ and $f(0,0) = 0$. Show that f is continuous and has bounded partial derivatives f'_x and f'_y, but that f is not differentiable at $(0,0)$. Investigate continuity of partial derivatives near $(0,0)$.

5. Show that the function $f(x,y) = \sqrt{|xy|}$ is continuous at $(0,0)$ and has the partial derivatives $f'_x(0,0)$ and $f'_y(0,0)$, but that f is not differentiable at $(0,0)$. Investigate continuity of the partial derivatives f'_x and f'_y at the origin.

6. Let $f(x,y) = x^3/(x^2+y^2)$ if $(x,y) \neq (0,0)$ and $f(0,0) = 0$. Show that f is continuous, has partial derivatives at $(0,0)$, but that f is not differentiable at $(0,0)$.

7–13. Find the set on which each of the following functions is differentiable:

7. $f(x,y) = y\sqrt{x}$;
8. $f(x,y) = \frac{xy}{2x+y}$;
9. $f(x,y,z) = \sin(xy+z)e^{zy}$;
10. $f(x,y,z) = \sqrt{x^2+y^2-z^2}$;
11. $f(\mathbf{r}) = \ln(1 - \|\mathbf{r}\|)$, where $\mathbf{r} = \langle x_1, x_2, ..., x_m \rangle$;
12. $f(x,y) = \sqrt[3]{x^3+y^3}$;
13. $f(\mathbf{r}) = e^{-1/\|\mathbf{r}\|^2}$ if $\mathbf{r} \neq \mathbf{0}$ and $f(\mathbf{0}) = 0$, where $\mathbf{r} = \langle x, y \rangle$
Hint: Show $f'_x(\mathbf{0}) = f'_y(\mathbf{0}) = 0$ using the definition of partial derivatives. Calculate the partial derivatives at $\mathbf{r} \neq \mathbf{0}$. Investigate continuity of partial derivatives.

14–19. The line through a point P_0 of a surface perpendicular to the tangent plane at P_0 is called the *normal line*. Show that each of the following surfaces has a tangent plane at the given point P_0, and find an equation of the tangent plane and symmetric equations of the normal line to each of the following surfaces at the specified point:

14. $z = x^2 + 3y - y^3x, \quad P_0 = (1,2,-1)$;
15. $z = \sqrt{x^3y}, \quad P_0 = (1,4,2)$;
16. $z = y\ln(x^2-3y), \quad P_0 = (2,1,0)$;
17. $y = \tan^{-1}(xz^2), \quad P_0 = (1, \frac{\pi}{4}, -1)$;
18. $x = z\cos(y-z), \quad P_0 = (1,1,1)$;
19. $z - y + \ln z - \ln x = 0, \quad P_0 = (1,1,1)$.

20–22. Show that each of the following functions is differentiable at the given point $\mathbf{r}_0$, and find its linearization at $\mathbf{r}_0$:

20. $f(x,y) = \frac{2y+3}{4x+1}, \quad \mathbf{r}_0 = \mathbf{0}$;
21. $f(x,y,z) = z^{1/3}\sqrt{x+\cos^2(y)}, \quad \mathbf{r}_0 = \langle 0,0,1 \rangle$;
22. $f(\mathbf{r}) = \sin(\mathbf{n}\cdot\mathbf{r}), \quad \mathbf{r}_0 = \langle a_1, a_2, ..., a_m \rangle$, where $\mathbf{r} = \langle x_1, x_2, ..., x_m \rangle$ and $\mathbf{n} \neq \mathbf{0}$ is a constant vector orthogonal to $\mathbf{r}_0$.

23–26. Use the linearization to approximate the following numbers. Then use a calculator to find the numbers. Compare the results.

23. $\sqrt{20 - 7x^2 - y^2}$, where $(x,y) = (1.08, 1.95)$;
24. xy^2z^3 , where $(x,y,z) = (1.002, 2.003, 3.004)$;
25. $\dfrac{(1.03)^2}{\sqrt[3]{0.98\sqrt[4]{(1.05)^2}}}$;
26. $(0.97)^{1.05}$.

27-29. Use one iteration of Newton's method initiated at the given point P_0 to approximate the solution of each of the following system of nonlinear equations. Assume that the solution exists near P_0. Explain why Newton's method is applicable in each case. Check whether the approximate solution is better than the initial guess P_0.

27. $x^2 + xy + 2y^2 = 4.2\,,\quad x^2 - x^3y + y^3 = 0.9\,,\quad P_0 = (1,1)\,;$

28. $\ln(1 + y + 2x) + x^2 + y = 3.2\,,\quad x^3 + y^2 = 2.7\,,\quad P_0 = (-1,2)\,;$

29. $y\sin(y - x) + x = 1.2\,,\quad x\cos(y^2 - x) + y = 0.8\,,\quad P_0 = (1,1).$

30-31. The existence of partial derivatives at a point is not sufficient for continuity of the function at that point (Example **21.2**). Prove the following assertions:

30. Suppose that a function $f(x,y)$ is continuous with respect to x at each fixed y and has a bounded partial derivative $f'_y(x,y)$, i.e., $|f'_y(x,y)| \le M$ for some $M > 0$ and all (x,y). Then f is continuous.

31. Let $f(x,y)$ be defined on open set D and have bounded partial derivatives, $|f'_x(x,y)| \le M_1$ and $|f'_y(x,y)| \le M_2$ for all (x,y) in D and some positive M_1 and M_2. Then f is continuous on D.

22. Chain Rules and Implicit Differentiation

22.1. Chain Rules. Consider the function

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

whose domain is the entire plane. Let $x = x(t)$, $y = y(t)$ be parametric equations of a curve such that the functions $x(t)$ and $y(t)$ are differentiable for all t. Then the function

$$F(t) = f(x(t), y(t)) = x^3(t) + x(t)y^2(t)$$

defines the values of f along the curve. The function $F(t)$ is differentiable and its derivative is

$$\begin{aligned}\frac{dF(t)}{dt} &= \frac{d}{dt}\Big(x^3(t)\Big) + \frac{d}{dt}\Big(x(t)y^2(t)\Big)\\ &= 3x^2(t)\frac{dx(t)}{dt} + \frac{dx(t)}{dt}y^2(t) + x(t)\frac{d}{dt}\Big(y^2(t)\Big)\\ &= \Big(3x^2(t) + y^2(t)\Big)\frac{dx(t)}{dt} + 2x(t)y(t)\frac{dy(t)}{dt}\end{aligned}$$

By noting that $f'_x(x, y) = 3x^2 + y^2$ and $f'_y(x, y) = 2xy$, the above relation can be written in the form

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}, \quad \text{where} \quad x = x(t)\,,\ y = y(t)\,. \tag{22.1}$$

Equation (**22.1**) is an example of *chain rules* for functions of several variables. The function $F(t)$ is the composition of the function $f(x, y)$ and two functions $x = x(t)$, $y = y(t)$. If the partial derivatives f'_x and f'_y are known at a point (x_0, y_0) and the rates of change $x'(t_0)$ and $y'(t_0)$ are also known, where $x(t_0) = x_0$ and $y(t_0) = y_0$, then the rate of change $F'(t_0)$ can be computed by the *chain rule*:

$$F'(t_0) = f'_x(x_0, y_0)x'(t_0) + f'_y(x_0, y_0)y'(t_0)\,. \tag{22.2}$$

Now consider the function

$$f(x, y) = \frac{y^3}{x^2 + y^2}\,, \quad \text{if} \quad (x, y) \neq (0, 0) \quad \text{and} \quad f(0, 0) = 0\,.$$

It has partial derivatives everywhere. Indeed, for $(x, y) \neq (0, 0)$ it is a rational function and, hence, is differentiable. The partial derivatives at the origin are easy to find:

$$\begin{aligned}f'_x(x, 0) &= \frac{d}{dx}f(x, 0) = \frac{d}{dx}\,0 = 0 \quad \Rightarrow \quad f'_x(0, 0) = 0\,,\\ f'_y(0, y) &= \frac{d}{dy}f(0, y) = \frac{d}{dy}\,y = 1 \quad \Rightarrow \quad f'_y(0, 0) = 1\,.\end{aligned}$$

Consider a straight line through the origin $x = x(t) = t\cos\theta$, $y = y(t) = t\sin\theta$, where θ is a numerical parameter that defines a vector $\mathbf{u} = \langle\cos\theta, \sin\theta\rangle$

parallel to the line. The composition of f and the linear functions $x(t)$ and $y(t)$ reads

$$F(t) = f(x(t), y(t)) = \frac{t^3 \sin^3\theta}{t^2(\cos^2\theta + \sin^2\theta)} = t\sin^3\theta \quad \text{for} \quad t \neq 0\,.$$

For $t = 0$, put $F(0) = f(0,0) = 0$ so that the above equation defines $F(t)$ for all t. Therefore

$$F'(t) = \sin^3\theta \quad \Rightarrow \quad F'(0) = \sin^3\theta\,.$$

On the other hand, $x'(0) = \cos\theta$ and $y'(0) = \sin\theta$ so that the chain rule (**22.2**) gives a different result:

$$f'_x(0,0)x'(0) + f'_y(0,0)y'(0) = \sin\theta \neq \sin^3\theta = F'(0)\,.$$

So the chain rule (**22.2**) fails to give the correct rate of change of the composition at the point $(0,0)$ despite that the function f has partial derivatives at $(0,0)$. It is not difficult to verify that the chain rule (**22.1**) holds for all $(x,y) \neq (0,0)$ and any smooth parametric curve $x = x(t)$, $y = y(t)$ (calculations are pretty much the same as in the first example of a polynomial f).

Why does the chain rule fail at the origin? The considered function is *not differentiable* at the origin and this is the reason for the chain rule to fail! Note that since the partial derivatives $f'_x(0,0) = 0$ and $f'_y(0,0) = 1$ exist, a good linear approximation at the origin should have the form

$$L(x,y) = f(0,0) + f'_x(0,0)x + f'_y(0,0)y = y\,.$$

However, this linear function does not satisfy the condition (**21.3**) and therefore f is not differentiable at the origin. Indeed, for $(x,y) \neq (0,0)$

$$\frac{f(x,y) - L(x,y)}{\sqrt{x^2+y^2}} = \frac{y^3 - y(x^2+y^2)}{(x^2+y^2)^{3/2}} = -\frac{yx^2}{(x^2+y^2)^{3/2}} \not\to 0$$

as $(x,y) \to (0,0)$. The limit of this function at the origin cannot be zero because its limit along the line $x = y = t$, $t \to 0^+$, does not vanish and is equal to $-2^{-3/2} \neq 0$.

Thus, the chain rule (**22.1**) holds if the function $f(x,y)$ and the functions $x = x(t)$ and $y = y(t)$ are differentiable. In the one-variable case, the differentiability is equivalent to the existence of the derivative so that the chain rule $df/dt = f'(x)x'(t)$, $x = x(t)$, holds if $f'(x)$ and $x'(t)$ exist. In contrast to the one-variable case, the mere existence of partial derivatives is not sufficient to validate the chain rule in the multi-variable case, and a stronger condition on f (differentiability of f) is required.

THEOREM **22.1**. (Chain Rule).
Let f be a differentiable function of n variables $\mathbf{r} = \langle x_1, x_2, ..., x_n\rangle$. Suppose that each variable x_i is, in turn, a differentiable function of m variables $\mathbf{u} = \langle u_1, u_2, ..., u_m\rangle$. The composition of $x_i = x_i(\mathbf{u})$ with $f(\mathbf{r})$ defines f as

a function of $\mathbf{u}$. *Then its rate of change with respect to* u_j, $j = 1, 2, ..., m$, *reads*

$$\frac{\partial f}{\partial u_j} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial u_j} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial u_j} + \cdots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial u_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial u_j}\,. \tag{22.3}$$

The proof of this theorem is given at the end of this section. For $n = m = 1$, this is the familiar chain rule for functions of one variable $df/du = f'(x)x'(u)$. If $n = 1$ and $m > 1$, it is the chain rule (**19.2**) established earlier. If $n = 2$ and $m = 1$, the chain rule (**22.1**) is obtained.

Theorem **21.3** established sufficient conditions for differentiability of a function of several variables and offers a practical criterion for applicability of the chain rule:

If a function $f(x_1, x_2, ..., x_n)$ *has continuous partial derivatives in an open* $n-$*ball and the functions* $x_j = x_j(u_1, u_2, ..., u_m)$, $j = 1, 2, ..., n$, *also have continuous partial derivatives in an open* $m-$*ball* D, *then the chain rule* (**22.3**) *holds in* D *(assuming that the composition of the functions exists in* D*).*

The continuity of partial derivatives can be verified by simpler means than the existence of a good linear approximation (differentiability) for f and the functions x_j.

Example **22.1**. *Let* $f(x, y) = x^3 + xy^2$. *Find the rates of change* f'_r *and* f'_θ *as functions of the polar coordinates* (r, θ) *where* $x = r\cos\theta$, $y = r\sin\theta$.

Solution: The chain rule for polar coordinates is a particular case of the chain rule (**22.3**) for $n = m = 2$, where $\mathbf{r} = (x, y)$ and $\mathbf{u} = (r, \theta)$. The function $f(x, y)$ is a polynomial and hence has continuous partial derivatives everywhere. The functions $x = r\cos\theta$ and $y = r\sin\theta$ also have continuous partial derivatives for all (r, θ). Therefore the chain rule holds and

$$\begin{aligned}
\frac{\partial f}{\partial r} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} = (3x^2 + y^2)\cos\theta + 2xy\sin\theta = 3r^2\cos\theta,\\
\frac{\partial f}{\partial \theta} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = -(3x^2 + y^2)r\sin\theta + 2xyr\cos\theta = -r^3\sin\theta,
\end{aligned}$$

where x and y have been expressed in the polar coordinates to obtain the final expressions. Note that f can also be expressed first as a function of (r, θ),

$$f(x, y) = x(x^2 + y^2) = r^3\cos\theta\,.$$

The partial derivatives f'_r and f'_θ of this function obviously coincide with those obtained by the chain rule. □

Example **22.2**. *Let a function* $f(x, y, z)$ *be differentiable at* $\mathbf{r}_0 = \langle 1, 2, 3\rangle$ *and have the following rates of change,* $f'_x(\mathbf{r}_0) = 1$, $f'_y(\mathbf{r}_0) = 2$, *and* $f'_z(\mathbf{r}_0) =$

-2. *Suppose that $x = x(t,s) = t^2 s$, $y = y(t,s) = s + t$, and $z = z(t,s) = 3s$. Find the rates of change of f with respect to t and s at the point $\mathbf{r}_0$.*

SOLUTION: In the chain rule (**22.3**), put $\mathbf{r} = \langle x, y, z\rangle$ and $\mathbf{u} = \langle t, s\rangle$. The point $\mathbf{r}_0 = \langle 1, 2, 3\rangle$ corresponds to the point $\mathbf{u}_0 = \langle 1, 1\rangle$ in the new variables:

$$\begin{array}{lclclcl} x = t^2 s & & 1 = t^2 s & & 1 = t^2 s & & 1 = 1 \\ y = s + t & \Rightarrow & 2 = s + t & \Rightarrow & 1 = t & \Rightarrow & 1 = t \\ z = 3s & & 3 = 3s & & 1 = s & & 1 = s \end{array}$$

The partial derivatives of the old variables with respect to the new ones are $x'_t = 2ts$, $y'_t = 1$, $z'_t = 0$, $x'_s = t^2$, $y'_s = 1$, and $z'_s = 3$. They are continuous everywhere and hence the functions $x(s,t)$, $y(s,t)$, and $z(s,t)$ are differentiable at the point $(1,1)$ by Theorem **21.3** and the chain rule holds. By the chain rule,

$$\begin{aligned} f'_t(\mathbf{r}_0) &= f'_x(\mathbf{r}_0)x'_t(\mathbf{u}_0) + f'_y(\mathbf{r}_0)y'_t(\mathbf{u}_0) + f'_z(\mathbf{r}_0)z'_t(\mathbf{u}_0) \\ &= 1\cdot 2 + 2\cdot 1 + (-2)\cdot 0 = 4, \\ f'_s(\mathbf{r}_0) &= f'_x(\mathbf{r}_0)x'_s(\mathbf{u}_0) + f'_y(\mathbf{r}_0)y'_s(\mathbf{u}_0) + f'_z(\mathbf{r}_0)z'_s(\mathbf{u}_0) \\ &= 1\cdot 1 + 2\cdot 1 + (-2)\cdot 3 = -3. \end{aligned}$$

□

EXAMPLE **22.3**. *Let $f(x,y,z) = z^2(1 + x^2 + y^2)^{-1}$. Find the rate of change of f along the parametric curve $\mathbf{r}(t) = (\sin t, \cos t, e^t)$ in the direction of increasing t. Determine whether the function f is decreasing or increasing along the curve.*

SOLUTION: The function f is differentiable because its partial derivatives,

$$f'_x = -\frac{2xz^2}{(1+x^2+y^2)^2}\,,\quad f'_y = -\frac{2yz^2}{(1+x^2+y^2)^2}\,,\quad f'_z = \frac{2z}{1+x^2+y^2}\,,$$

are continuous everywhere. The components of $\mathbf{r}(t)$ are also differentiable: $x'(t) = \cos t$, $y'(t) = -\sin t$, $z'(t) = e^t$. By the chain rule (**22.3**) for $n = 3$ and $m = 1$,

$$\begin{aligned} \frac{df}{dt} &= f'_x(\mathbf{r}(t))x'(t) + f'_y(\mathbf{r}(t))y'(t) + f'_z(\mathbf{r}(t))z'(t) \\ &= -\frac{2xz^2\cos t}{(1+x^2+y^2)^2} + \frac{2yz^2\sin t}{(1+x^2+y^2)^2} + \frac{2ze^t}{1+x^2+y^2} \\ &= -\frac{2e^{2t}\sin t\cos t}{4} + \frac{2e^{2t}\cos t\sin t}{4} + \frac{2e^{2t}}{2} \\ &= e^{2t} > 0\,. \end{aligned}$$

The function f has a positive rate of change along the curve and therefore it is increasing along the curve in the direction of increasing t. □

22.2. Geometrical significance of the chain rules. Let f be a differentiable function of several variables and C be a smooth curve in the domain of f. Then one can *restrict* the function f to the curve C, that is, define a function on the curve C by values of f on C. Let $\mathbf{r} = \mathbf{r}(s)$ be a natural parameterization of C in some rectangular coordinate system. Then

$$\frac{df}{ds} = f'_{x_1}(\mathbf{r}(s))\, x'_1(s) + f'_{x_2}(\mathbf{r}(s))\, x'_2(s) + \cdots + f'_{x_m}(\mathbf{r}(s))\, x'_m(s)$$

is the rate of change of f along C at a point $\mathbf{r}(s)$ with increasing s.

For example, suppose f is the temperature in a room measured in Fahrenheits as a function of position and C is a smooth curve. Suppose that $df/ds = 5$ at a point P_0 of C corresponding to $s = s_0$. This means that the temperature along C is increasing at P_0 at the rate 5 F° per unit length in the direction of increasing s along C. Alternatively, let $\mathbf{r} = \mathbf{r}(t)$ be the trajectory of a particle (t is time) such that the particle passes P_0 at time $t = t_0$. If $df/dt = 3$ at P_0, then the particle "sees" that the temperature is increasing at P_0 (at $t = t_0$) at the rate 3 F° per unit time.

Let a function $f(x, y)$ be differentiable at a point $P_0 = (x_0, y_0) \neq (0, 0)$. This point can be viewed as the intersection of two coordinate lines $x = x_0$ and $y = y_0$. Consider the polar coordinates $x = r\cos\theta$, $y = r\sin\theta$. Then for every pair $(x, y) \neq (0, 0)$ one can find a unique pair (r, θ) where $0 < r < \infty$ and $0 \le \theta < 2\pi$. In particular, the point P_0 has polar coordinates (r_0, θ_0). All points for which $r = r_0$ form a circle of radius r_0 centered at the origin, and all points in the plane for which $\theta = \theta_0$ form a ray extended from the origin. It makes the angle θ_0 with the x axis counted counterclockwise. So the point P_0 can also be viewed as the intersection of the *coordinate lines (or curves)* of the polar coordinates, the circle $r = r_0$ and the ray $\theta = \theta_0$ (the curves along which the polar coordinates have constant values). Parametric equations of the circle and the ray are, respectively,

$$\begin{aligned} C_1 :&\quad x = x(\theta) = r_0\cos\theta\,, \quad y = y(\theta) = r_0\sin\theta\,, \quad 0 \le \theta \le 2\pi\,,\\ C_2 :&\quad x = x(r) = r\cos\theta_0\,, \quad y = y(r) = r\sin\theta_0\,, \quad 0 \le r < \infty\,. \end{aligned}$$

Consequently, the rates $f'_\theta(x_0, y_0)$ and $f'_r(x_0, y_0)$ found in Example **22.1** define the rates of change of f at the point P_0 along the circle C_1 and the ray C_2, respectively, in the direction of increasing the parameters θ and r.

A similar interpretation holds for Example **22.2**. The relations between (x, y, z) and (t, s) define a surface S in space as a graph of a function of two variables:

$$x = t^2 s\,, \quad y = s + t\,, \quad z = 3s \quad \Rightarrow \quad x = g(y, z) = \tfrac{1}{27} z(3y - z)^2\,.$$

To obtain $g(y, z)$, the variables (t, s) are expressed in terms of y and z using the last two equations, $s = z/3$ and $t = y - s = y - z/3$, and then substituted into $x = t^2 s$. So the function $f(x, y, z)$ is *restricted* to this surface when (x, y, z) are replaced by the given expressions in terms of (t, s). The point $\mathbf{r}_0 = \langle 1, 2, 3\rangle$ lies in the surface. The fixed value of the parameter $t = t_0 = 1$

defines a curve C_1 in the surface S that passes through P_0, whereas the fixed value of the other parameter $s = s_0 = 1$ defines another curve C_2 in S through P_0. Parametric equations of these curves are

$$C_1: \quad \mathbf{r} = \mathbf{r}_1(s) = \langle s, s+1, 3s\rangle\,; \qquad C_2: \quad \mathbf{r} = \mathbf{r}_2(t) = \langle t^2, t+1, 3\rangle\,.$$

They are obtained by setting, respectively, $t = 1$ and $s = 1$ in the relations between (x, y, z) and (t, s). The rate $f_t'(\mathbf{r}_0) = 4 > 0$ shows that the function f restricted to the curve C_2 is *increasing* at the point $\mathbf{r}_0$ at a rate of 4 units per unit increment of the parameter t, while f restricted to the curve C_1 is *decreasing* at a rate of 3 units per unit increment of the parameter s because $f_s'(\mathbf{r}_0) = -3 < 0$. Note that the parameters t and s are not the arclength parameter of these curves. So, the unit increments of t and s do not correspond to the unit increment of the arclength along the curves.

22.3. Chain rules for higher-order partial derivatives. The chain rule (**22.3**) can be used to calculate higher order partial derivatives with respect to new variables if the functions in Theorem **22.1** have *continuous partial derivatives of the needed order* (partial derivatives of lower orders should be differentiable functions). For example, let

$$f = f(x, y)\,, \quad x = x(t)\,, \quad y = y(t)\,.$$

Then, assuming that f has continuous partial derivatives up to second order and the functions $x(t)$ and $y(t)$ are twice differentiable,

$$\begin{aligned}
\frac{d^2 f}{dt^2} &= \frac{d}{dt}\frac{df}{dt} = \frac{d}{dt}\left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}\right) \\
&= \frac{d}{dt}\left(\frac{\partial f}{\partial x}\right)\frac{dx}{dt} + \frac{\partial f}{\partial x}\frac{d^2x}{dt^2} + \frac{d}{dt}\left(\frac{\partial f}{\partial y}\right)\frac{dy}{dt} + \frac{\partial f}{\partial y}\frac{d^2y}{dt^2} \\
&= \left(\frac{\partial^2 f}{\partial x^2}\frac{dx}{dt} + \frac{\partial^2 f}{\partial y \partial x}\frac{dy}{dt}\right)\frac{dx}{dt} + \frac{\partial f}{\partial x}\frac{d^2x}{dt^2} \\
&\quad + \left(\frac{\partial^2 f}{\partial x \partial y}\frac{dx}{dt} + \frac{\partial^2 f}{\partial y^2}\frac{dy}{dt}\right)\frac{dy}{dt} + \frac{\partial f}{\partial y}\frac{d^2y}{dt^2} \\
&= \frac{\partial^2 f}{\partial x^2}\left(\frac{dx}{dt}\right)^2 + 2\frac{\partial^2 f}{\partial y \partial x}\frac{dy}{dt}\frac{dx}{dt} + \frac{\partial^2 f}{\partial y^2}\left(\frac{dy}{dt}\right)^2 + \frac{\partial f}{\partial x}\frac{d^2x}{dt^2} + \frac{\partial f}{\partial y}\frac{d^2y}{dt^2}
\end{aligned}$$

where Clairaut's theorem $f''_{xy} = f''_{yx}$ was used (owing to the assumption of continuity of the partial derivatives).

EXAMPLE **22.4**. *If $g(u, v) = f(x, y)$ where $x = (u^2 - v^2)/2$ and $y = uv$, find g''_{uv}. Assume that f has continuous second partial derivatives. If $f_y'(1, 2) = 1$, $f''_{xx}(1, 2) = f''_{yy}(1, 2) = 2$, and $f''_{xy}(1, 2) = 3$, find the value of g''_{uv} at $(x, y) = (1, 2)$.*

SOLUTION: One has $x_u' = u$, $x_v' = -v$, $y_u' = v$, and $y_v' = u$. Then

$$g_u' = f_x' x_u' + f_y' y_u' = f_x' u + f_y' v\,.$$

The derivative $g''_{uv} = (g'_u)'_v$ is calculated by applying the chain rule to the function g'_u:

$$\begin{aligned} g''_{uv} &= u(f'_x)'_v + v(f'_y)'_v + f'_y \\ &= u(f''_{xx}x'_v + f''_{xy}y'_v) + v(f''_{yx}x'_v + f''_{yy}y'_v) + f'_y \\ &= u(-vf''_{xx} + uf''_{xy}) + v(-vf''_{yx} + uf''_{yy}) + f'_y \\ &= uv(f''_{yy} - f''_{xx}) + (u^2 - v^2)f''_{xy} + f'_y = y(f''_{yy} - f''_{xx}) + 2xf''_{xy} + f'_y\,. \end{aligned}$$

where $f''_{xy} = f''_{yx}$ has been used. The value of g''_{uv} at the point in question is $2 \cdot (2-2) + 2 \cdot 3 + 1 = 7$. □

22.4. Implicit Differentiation. Let $F(x, y, z)$ be a function of three variables. Suppose that the equation $F(x, y, z) = 0$ can be solved for the variable z to obtain z as a function of two variables:

$$F(x, y, z) = 0 \quad \Longrightarrow \quad z = z(x, y)$$

The characteristic property of the function $z(x, y)$ is that

$$(\mathbf{22.4}) \qquad F(x, y, z(x, y)) = 0 \quad \text{for all } (x, y)\,.$$

In other words, a function of two variables $z(x, y)$ is not defined by an explicit rule but rather by an equation to be solved to find the values of the function. For example,

$$(\mathbf{22.5}) \qquad \begin{aligned} &F(x, y, z) = z^2 - 2xz + y = 0 \\ &\Rightarrow \quad z = z_\pm(x, y) = x \pm \sqrt{x^2 - y}\,, \quad x^2 \geq y\,. \end{aligned}$$

In this case, the equation has a solution if $x^2 \geq y$ which is the domain of the functions $z_\pm(x, y)$. The solution is not unique. There are two functions, $z_+(x, y)$ and $z_-(x, y)$, defined by the given equation. For every pair (x, y) there are two values of $z = z_\pm(x, y)$ that satisfy the equation $F(x, y, z) = 0$. From the geometrical point of view, one can say that the level set on which the function F take zero value consists of two surfaces $z = z_\pm(x, y)$, just like the sphere $x^2 + y^2 + z^2 = 1$ can be viewed as the level set of $F(x, y, z) = x^2 + y^2 + z^2$ which is the union of two graphs

$$x^2 + y^2 + z^2 = 1 \quad \Leftrightarrow \quad z = z_\pm(x, y) = \pm\sqrt{1 - x^2 - y^2}$$

representing the upper and lower hemispheres.

To resolve the uniqueness problem, one can impose an additional condition on the solution. Suppose that a point $P_0 = (x_0, y_0, z_0)$ lies in the level set $F(x, y, z) = 0$, that is, $F(x_0, y_0, z_0) = 0$. Then it is required that the solution $z(x, y)$ satisfy the condition $z(x_0, y_0) = z_0$. Geometrically, this means that the graph $z = z(x, y)$ coincides with the level surface $F(x, y, z) = 0$ in a *neighborhood* of a particular point P_0 in the level surface. In Example (**22.5**), let us take a particular point $P_0 = (x_0, y_0, z_0) = (2, 3, 1)$. This point lies in the level set $F(2, 3, 1) = 0$. Then the graph $z = z_-(x, y)$ coincides with the level set $F(x,y,z)=0$ in a neighborhood of P_0 because $z_-(2,3)=1$, while the other does not because $z_+(2,3)=3\neq 1$:

$$\begin{aligned} z^2-2xy+y&=0\\ z(2,3)&=1\end{aligned} \quad\Rightarrow\quad z=z_-(x,y)=x-\sqrt{x^2-y}\,.$$

In the example considered an *explicit* form of the solution was not so difficult to find by solving the quadratic equation. But in most cases, an explicit solution is not possible to obtain. For example, finding an explicit form of the function $z=z(x,y)$ satisfying an addition condition and the equation

$$(\mathbf{22.6})\qquad F(x,y,z)=z^5-2xz+y=0\,,\quad z(2,3)=1\,,\quad F(2,3,1)=0\,,$$

requires solving an equation of the fifth order. Only equations up to the fourth order can be solved analytically. Does a solution exist? Is it unique? It will be shown later that Eq. **22.6** does have a unique solution near the point $(2,3,1)$.

Let us put aside these questions for a moment and suppose that, given a function $F(x,y,z)$, the equation $F(x,y,z)=0$ is proved to have a unique solution when (x,y) lie in some open region D in the xy plane. In this case, the function $z(x,y)$ with the property (**22.4**) for all (x,y) in D is said to be defined *implicitly* on D. Although an analytic form of an implicitly defined function is unknown, its rates of change can be found and provide important information about its local behavior.

Suppose that F is differentiable. Furthermore, the root $z(x,y)$ is also assumed to be differentiable on D. Since relation (**22.4**) holds for all (x,y) in D, the partial derivatives of its left side must also vanish in D. They can be computed by the chain rule (**22.3**), where $n=3$, $m=2$, $\mathbf{r}=\langle x,y,z\rangle$, $\mathbf{u}=\langle u,v\rangle$, and the relations between old and new variables are $x=u$, $y=v$, and $z=z(u,v)$. One has $x'_u=1$, $x'_v=0$, $y'_u=0$, $y'_v=1$, and $z'_u(u,v)=z'_x(x,y)$ and $z'_v(u,v)=z'_y(x,y)$ because $x=u$ and $y=v$. Therefore Eq. (**22.4**) implies that

$$\begin{aligned}
\frac{\partial}{\partial u}F(x,y,z(x,y)) &= \frac{\partial F}{\partial x}+\frac{\partial F}{\partial z}\frac{\partial z}{\partial x}=0 \quad\Longrightarrow\quad z'_x=-\frac{F'_x}{F'_z},\\
\frac{\partial}{\partial v}F(x,y,z(x,y)) &= \frac{\partial F}{\partial y}+\frac{\partial F}{\partial z}\frac{\partial z}{\partial y}=0 \quad\Longrightarrow\quad z'_y=-\frac{F'_y}{F'_z},
\end{aligned}$$

These equations are called the *implicit differentiation* equations. They determine the rates of change of an implicitly defined function of two variables. Note that in order for these equations to make sense, the condition $F'_z\neq 0$ must be imposed.

In Example (**22.5**)

$$
\begin{aligned}
&F'_x = -2z\,, \quad F'_y = 1\,, \quad F'_z = 2z - 2x \\
\Rightarrow \quad & z'_x(x,y) = -\frac{F'_x}{F'_z}\Big|_{z=z(x,y)} = \frac{z(x,y)}{z(x,y)-x}\,, \\
\Rightarrow \quad & z'_y(x,y) = -\frac{F'_y}{F'_z}\Big|_{z=z(x,y)} = -\frac{1}{2(z(x,y)-x)}
\end{aligned}
$$

In order to obtain the partial derivatives as explicit functions of (x, y), the root $z = z(x, y)$ has to be substituted into the implicit differentiation equations. For example, if one takes the root $z = z_-(x, y)$ then

$$
\begin{aligned}
& z(x,y) = x - \sqrt{x^2 - y} \\
\Rightarrow \quad & z'_x(x,y) = \frac{\sqrt{x^2-y} - x}{\sqrt{x^2-y}} = 1 - \frac{x}{\sqrt{x^2-y}}\,, \quad z'_y(x,y) = \frac{1}{2\sqrt{x^2-y}}
\end{aligned}
$$

Note that the implicit differentiation equations give the same answer for partial derivatives that can be obtained by *explicit differentiation*:

$$
\begin{aligned}
z'_x(x,y) &= \frac{\partial}{\partial x}\Big(x - \sqrt{x^2-y}\Big) = 1 - \frac{x}{\sqrt{x^2-y}}\,, \\
z'_y(x,y) &= \frac{\partial}{\partial y}\Big(x - \sqrt{x^2-y}\Big) = \frac{1}{2\sqrt{x^2-y}}\,.
\end{aligned}
$$

The advantage of using the implicit differentiation equations become apparent when no explicit form of the root $z(x, y)$ can be found. Let $z(x, y)$ be defined by Eq. (**22.6**). Then $F(x, y, z) = z^5 - 2xz + y$ and

$$
\begin{aligned}
(\mathbf{22.7}) \qquad & F'_x = -2z\,, \quad F'_y = 1\,, \quad F'_z = 5z^4 - 2x\,, \\
\Rightarrow \quad & z'_x(x,y) = -\frac{F'_x}{F'_z} = \frac{2z}{5z^4 - 2x}\,, \quad z'_y(x,y) = -\frac{F'_y}{F'_z} = -\frac{1}{5z^4-2x}\,.
\end{aligned}
$$

One might ask: What is the point of these equations if an explicit form $z = z(x, y)$ is not known and, hence, the values of the partial derivative cannot be calculated at any point (x, y)? It is true that an explicit form of the partial derivatives is impossible to obtain, but *their values can be computed at any point that is known to lie in the level set* $F(x, y, z) = 0$. For example, $F(2, 3, 1) = 0$. So the point $(2, 3, 1)$ lies in the level set $F(x, y, z) = 0$. If the solution $z = z(x, y)$ satisfying the condition $z(2, 3) = 1$ is proved to be unique, then

$$
\begin{aligned}
z'_x(2,3) &= -\frac{F'_x(2,3,1)}{F'_z(2,3,1)} = -\frac{-2}{5-4} = 2\,, \\
z'_y(2,3) &= -\frac{F'_y(2,3,1)}{F'_z(2,3,1)} = -\frac{1}{5-4} = -1\,.
\end{aligned}
$$

The values of the partial derivatives allows us to conclude that the function $z(x, y)$ *increases with increasing* x *and decreases with increasing* y at the point $(2, 3)$. This conclusion about the behavior of the function $z(x, y)$ near the point $(2, 3)$ has been reached *without* an explicit form of $z(x, y)$.

The implicit function theorem. The questions about the very existence and uniqueness of $z(x, y)$ for a given $F(x, y, z)$ and the differentiability of $z(x, y)$ have been left unanswered in the above analysis of the implicit differentiation. The following theorem addresses them all.

THEOREM **22.2**. (Implicit Function Theorem).
Let F be a function of $m + 1$ variables, $F(\mathbf{r}, z)$, where $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$ and z is real such that F and F'_z are continuous in an open ball B in $\mathbb{R}^{m+1}$. Suppose that there exists a point $(\mathbf{r}_0, z_0)$ in B such that $F(\mathbf{r}_0, z_0) = 0$ and $F'_z(\mathbf{r}_0, z_0) \neq 0$. Then there exists an open neighborhood D of $\mathbf{r}_0$ in $\mathbb{R}^m$, an open interval I, and a unique function $z(\mathbf{r})$ on D with the range I such that for $\mathbf{r}$ in D and u in I, $F(\mathbf{r}, u) = 0$ if and only if $u = z(\mathbf{r})$. Moreover, the function $z(\mathbf{r})$ is continuous. If, in addition, F is differentiable in B, then the function $z(\mathbf{r})$ is differentiable in D and

$$z'_{x_i}(\mathbf{r}) = -\frac{F'_{x_i}(\mathbf{r}, z(\mathbf{r}))}{F'_z(\mathbf{r}, z(\mathbf{r}))}.$$

for all $\mathbf{r}$ in D.

The proof of this theorem goes beyond the scope of this course. It includes proofs of the existence and uniqueness of $z(\mathbf{r})$ and its differentiability. Once these facts are established, a derivation of the implicit differentiation equations follows the same way as in the $m = 2$ case:

$$\frac{\partial F}{\partial x_i} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x_i} = 0 \quad \Longrightarrow \quad z'_{x_i}(\mathbf{r}) = -\frac{F'_{x_i}(\mathbf{r}, z(\mathbf{r}))}{F'_z(\mathbf{r}, z(\mathbf{r}))}.$$

EXAMPLE **22.5**. *Show that the equation (**22.6**) has a unique continuous solution $z = z(x, y)$ near $(x, y) = (2, 3)$ and that the function $z(x, y)$ is differentiable near $(2, 3)$.*

SOLUTION: Let us verify the hypotheses of the implicit function theorem. The function $F(x, y, z) = z^5 - 2xz + y$ and its partial derivative $F'_z = 5z^4 - 2x$ are polynomials and, hence, continuous everywhere. In particular, they are continuous in a ball centered at $(2, 3, 1)$. Furthermore, $F(2, 3, 1) = 0$ and $F'_z(2, 3, 1) = 1 \neq 0$. By the implicit function theorem the equation $F(x, y, z) = 0$ has a unique solution $z = z(x, y)$ in some disk D centered at $(2, 3)$ such that $z(2, 3) = 1$ and the function $z(x, y)$ is continuous in the disk.

Next, the partial derivatives of $F(x, y, z)$ are polynomials and, hence, continuous everywhere and, in particular, in a ball centered at $(2, 3, 1)$. The function $F(x, y, z)$ is differentiable in the ball because it has continuous partial. By the implicit function theorem the solution $z = z(x, y)$ is a differentiable function in the disk D, and its partial derivatives are given by the implicit differentiation equations (see (**22.7**)). $\square$

Geometrical significance of the implicit differentiation. Let us analyze a geometrical significance of the implicit differentiation equations. Suppose that $F(x, y, z)$ has continuous partial derivatives and $F'_z \neq 0$. The equation $F(x, y, z) = 0$ defines a level set of the function F. This level set is a surface. Indeed, in a neighborhood of any particular point, the level set coincides with the graph $z = z(x, y)$ of a differentiable function $z(x, y)$ (which is a surface) as follows from the implicit function theorem. Let $\mathbf{r}_0 = \langle x_0, y_0, z_0 \rangle$ be a point in this level surface (that is, $z_0 = z(x_0, y_0)$). Since the graph coincides with the level surface $F(x, y, z) = 0$ (at least in some neighborhood of (x_0, y_0)), the tangent plane to the graph

$$z = z_0 + z'_x(x_0, y_0)(x - x_0) + z'_y(x_0, y_0)(y - y_0)\,,$$
$$z'_x(x_0, y_0) = -\frac{F'_x(\mathbf{r}_0)}{F'_z(\mathbf{r}_0)}\,, \quad z'_y(x_0, y_0) = -\frac{F'_y(\mathbf{r}_0)}{F'_z(\mathbf{r}_0)}$$

is also a *tangent plane to the level surface* $F(x, y, z) = 0$ *at the point* (x_0, y_0, z_0). By multiplying this equation by $F'_z(\mathbf{r}_0)$, it can be written in the form

$$F'_x(\mathbf{r}_0)(x - x_0) + F'_y(\mathbf{r}_0)(y - y_0) + F'_z(\mathbf{r}_0)(z - z_0) = 0\,. \tag{22.8}$$

Therefore the vector

$$\mathbf{n} = \langle F'_x(\mathbf{r}_0)\,,\ F'_y(\mathbf{r}_0)\,,\ F'_z(\mathbf{r}_0) \rangle$$

is a *normal vector to the tangent plane to a level surface* $F(x, y, z) = 0$ *at a point* (x_0, y_0, z_0). In fact, it is a normal vector to a level surface $F(x, y, z) = k$ for any k (from the range of F) because the partial derivatives of $F(x, y, z)$ and $F(x, y, z) - k$ coincide (naturally, the point (x_0, y_0, z_0) must satisfy the condition $F(x_0, y_0, z_0) = k$).

If the solution $z(x, y)$ to $F(x, y, z) = 0$ is differentiable at (x_0, y_0), then its values in a neighborhood of this point can be approximated by its linearization:

$$z(x_0 + \Delta x, y_0 + \Delta y) \approx z_0 + z'_x(x_0, y_0)\Delta x + z'_y(x_0, y_0)\Delta y\,.$$

In other words, having found a root z_0 of the equation $F(x, y, z) = 0$ at a particular point (x_0, y_0), the linearization of the implicit function $z(x, y)$ allows us to approximate the root at any point near (x_0, y_0).

For example, the linearization of the function $z(x, y)$ implicitly defined by (**22.6**) at the point $(x_0, y_0) = (2, 3)$ has the form

$$z(x, y) \approx 1 + 2(x - 2) + (-1)(y - 3) = 1 + 2\Delta x - \Delta y\,, \tag{22.9}$$

where $x = 2 + \Delta x$ and $y = 3 + \Delta y$. Let us substitute this approximation of $z(x, y)$ into (**22.6**). Using the binomial expansion

$$(1 + a)^5 = 1 + 5a + 10a^2 + 10a^3 + 5a^4 + a^5\,.$$

it is not difficult to find that

$$\begin{aligned} z^5 = (1 + 2\Delta x - \Delta y)^5 &= 1 + 10\Delta x - 5\Delta y + \text{t.o.h.o.} \\ -2zx = -2(1 + 2\Delta x - \Delta y)(2 + \Delta x) &= -4 - 10\Delta x + 4\Delta y + \text{t.o.h.o.} \\ y &= 3 + \Delta y \end{aligned}$$

where t.o.h.o. stands for *terms of higher orders* and denotes terms proportional $(\Delta x)^n(\Delta y)^m$ with $n + m \geq 2$ (terms that are non-linear functions of Δx and Δy). By adding these three relations, it is concluded that the characteristic condition (**22.4**) holds for the linearization of the solution to Eq. (**22.6**) *up to terms of higher orders*:

$$F(x, y, z) = F(2 + \Delta x, 3 + \Delta y, 1 + 2\Delta x - \Delta y) = 0 + \text{t.o.h.o.}$$

In other words, when values of terms of higher orders can be neglected as compared to the values of Δx and Δy, then the linearization of the implicit function $z(x, y)$ approximates the root of the equation $F(x, y, z) = 0$ and this approximation becomes better with decreasing Δx and Δy.

EXAMPLE **22.6**. *Use the linearization of an implicit function to approximate the root of Eq. (*22.6*) at $x = 1.99$ and $y = 3.02$. Find the plane tangent to the surface defined by (*22.6*) at the point $(2, 3, 1)$.*

SOLUTION: It follows from Example **22.5** that Eq. (**22.6**) defines a differentiable function $z = z(x, y)$ near $(2, 3)$ Therefore its values near $(2, 3)$ may be approximated by the linearization of $z(x, y)$ at $(2, 3)$. Set $\Delta x = -0.01$ and $\Delta y = 0.02$. Then using (**22.9**)

$$z(1.99, 3.02) = z(2 + \Delta x, 3 + \Delta y) \approx 1 + 2\Delta x - \Delta y = 1 - 0.04 = 0.96$$

The tangent plane in questions coincides with the tangent plane to the graph of the implicit function $z = z(x, y)$ at $(2, 3, 1)$. Therefore setting $\Delta x = x - 2$ and $\Delta y = y - 3$, the equation of the tangent plane is obtained:

$$z = 1 + 2(x - 2) - (y - 3) \quad \text{or} \quad 2x - y - z = 0\,.$$

Alternatively, the tangent plane can be obtained from (**22.8**). Using (**22.7**) $F'_x(2, 3, 1) = -2$, $F'_y(2, 3, 1) = 1$, $F'_z(2, 3, 1) = 1$. Therefore the tangent plane to the surface (**22.6**) is

$$-2(x - 2) + (y - 3) + (z - 1) = 0 \quad \text{or} \quad 2x - y - z = 0\,.$$

□

EXAMPLE **22.7**. *Show that the equation $z(3x - y) = \pi \sin(xyz)$ has a unique differentiable solution $z = z(x, y)$ in a neighborhood of $(1, 1)$ such that $z(1, 1) = \pi/2$ and find the rates of change $z'_x(1, 1)$ and $z'_y(1, 1)$. Use the linearization of $z(x, y)$ at $(1, 1)$ to approximate $z(1.1, 0.9)$*

SOLUTION: Put $F(x, y, z) = \pi \sin(xyz) - z(3x - y)$. Then the existence and uniqueness of the solution can be established by verifying the hypotheses of the implicit function theorem in which $\mathbf{r} = \langle x, y\rangle$, $\mathbf{r}_0 = \langle 1, 1\rangle$, and $z_0 =$

$\pi/2$. First, note that the function F is the sum of a polynomial and the sine function of a polynomial. So F is continuous everywhere. Its partial derivative

$$F'_z(x,y,z) = \pi xy\cos(xyz) - 3x + y$$

is also continuous everywhere (by the same reasoning) and satisfies the condition

$$F'_z(1,1,\pi/2) = -2 \neq 0\,.$$

By the first part of the implicit function theorem there is an open disk in the xy plane containing the point $(1,1)$ in which the equation $F(x,y,z)=0$ has a unique continuous solution $z = z(x,y)$.

Next, the function F is differentiable everywhere because its partial derivatives are continuous everywhere

$$F'_x = \pi yz\cos(xyz) - 3z\,, \quad F'_y = \pi xz\cos(xyz) + z\,.$$

Therefore, by the second part of the implicit function theorem, the solution $z = z(x,y)$ is a differentiable function in a neighborhood of $(1,1)$ and by the implicit differentiation equations,

$$z'_x(1,1) = -\frac{F'_x(1,1,\pi/2)}{F'_z(1,1,\pi/2)} = -\frac{3\pi}{4}\,, \qquad z'_y(1,1) = -\frac{F'_y(1,1,\pi/2)}{F'_z(1,1,\pi/2)} = \frac{\pi}{4}.$$

Since $z(x,y)$ is differentiable at $(1,1)$, there is a good linear approximation at $(1,1)$ (see Eq. (**21.3**)). Put $x = 1+\Delta x$ and $y = 1+\Delta y$ where $\Delta x = 0.1$ and $\Delta y = -0.1$. Then

$$\begin{aligned} z(1+\Delta x, 1+\Delta y) &\approx z(1,1) + z'_x(1,1)\Delta x + z'_y(1,1)\Delta y \\ &= \frac{\pi}{2} - \frac{3\pi}{40} - \frac{\pi}{40} = \frac{2\pi}{5}\,. \end{aligned}$$

□

Remark. If the function $F(\mathbf{r}, z)$ has sufficiently many continuous higher order partial derivatives, then higher order partial derivatives of the solution $z(\mathbf{r})$ to the equation $F(\mathbf{r}, z) = 0$ can be obtained by differentiating of the implicit differentiation equations. An example is given in Study Problem **22.1**.

22.5. Proof of Theorem 22.1. Since the functions $x_i(\mathbf{u})$ are differentiable, the partial derivatives $\partial x_i/\partial u_j$ exist and moreover, by Eq. (**21.5**),

$$\Delta x_i = x_i(\mathbf{u} + \hat{\mathbf{e}}_j h) - x_i(\mathbf{u}) = \frac{\partial x_i}{\partial u_j} h + \varepsilon_i(h)|h|, \quad \varepsilon_i(h) \to 0, \quad h \to 0\,,$$

for every i. Define the vector $\Delta\mathbf{r}_h = (\Delta x_1, \Delta x_2, ..., \Delta x_n)$. It has the property that $\Delta\mathbf{r}_h \to \mathbf{0}$ as $h \to 0$. If $F(\mathbf{u}) = f(x_1(\mathbf{u}), x_2(\mathbf{u}), ..., x_n(\mathbf{u}))$, then by the definition of the partial derivatives

$$\frac{\partial f}{\partial u_j} = \lim_{h\to 0} \frac{F(\mathbf{u} + \hat{\mathbf{e}}_j h) - F(\mathbf{u})}{h} = \lim_{h\to 0} \frac{f(\mathbf{r} + \Delta\mathbf{r}_h) - f(\mathbf{r})}{h}$$

if the limit exists. By the hypothesis, the function f is differentiable and, hence, has partial derivatives $\partial f/\partial x_i$ which determine a linear approximation (**21.6**) with the property (**21.5**):

$$f(\mathbf{r}+\Delta\mathbf{r}_h)-f(\mathbf{r})=\frac{\partial f}{\partial x_1}\Delta x_1+\frac{\partial f}{\partial x_2}\Delta x_2+\cdots+\frac{\partial f}{\partial x_n}\Delta x_n+\varepsilon(\Delta\mathbf{r}_h)\|\Delta\mathbf{r}_h\|,$$

where $\varepsilon(\Delta\mathbf{r}_h)\to 0$ as $\Delta\mathbf{r}_h\to\mathbf{0}$ or as $h\to 0$. The substitution of this relation into the limit shows that the limit exists and the conclusion of the theorem follows. Indeed, the first n terms contain the limits

$$\lim_{h\to 0}\frac{\Delta x_i}{h}=\frac{\partial x_i}{\partial u_j}+\lim_{h\to 0}\varepsilon_i(h)\frac{|h|}{h}=\frac{\partial x_i}{\partial u_j}$$

because $|h|/h=\pm 1$ for all $h\neq 0$ and $\varepsilon_i(h)\to 0$ as $h\to 0$. The ratio $\|\Delta\mathbf{r}_h\|/|h|=[(\Delta x_1/h)^2+(\Delta x_2/h)^2+\cdots+(\Delta x_n/h)^2]^{1/2}\to M<\infty$ as $h\to 0$ where M is determined by the partial derivatives $\partial x_i/\partial u_j$. Therefore the limit of the last term vanishes:

$$\begin{aligned}\lim_{h\to 0}\varepsilon(\Delta\mathbf{r}_h)\frac{\|\Delta\mathbf{r}_h\|}{h}&=\lim_{h\to 0}\varepsilon(\Delta\mathbf{r}_h)\frac{|h|}{h}\frac{\|\Delta\mathbf{r}_h\|}{|h|}\\&=\lim_{h\to 0}\varepsilon(\Delta\mathbf{r}_h)\frac{|h|}{h}\cdot\lim_{h\to 0}\frac{\|\Delta\mathbf{r}_h\|}{|h|}=0\cdot M=0\end{aligned}$$

because $\varepsilon(\Delta\mathbf{r}_h)|h|/h=\pm\varepsilon(\Delta\mathbf{r}_h)$ if $h\neq 0$ and $\varepsilon(\Delta\mathbf{r}_h)\to 0$ as $h\to 0$. This competes the proof.

22.6. Study Problems.

Problem **22.1**. *Let the function $z(x,y)$ be defined implicitly by $z^5+zx-y=0$ in a neighborhood of $(1,2,1)$. Find all its first and second partial derivatives. In particular, give the values of these partial derivatives at $(x,y)=(1,2)$.*

Solution: Let $F(x,y,z)=z^5+zx-y$. It is a polynomial and, hence, differentiable everywhere. Then $F'_z=5z^4+x$. The function $z(x,y)$ exists and is differentiable in a neighborhood of $(1,2)$ by the implicit function theorem because $F(1,2,1)=0$ and $F'_z(1,2,1)=6\neq 0$. The first and second partial derivatives of F are continuous everywhere:

$$\begin{aligned}&F'_x=z,\;\; F'_y=-1,\; F'_z=5z^4+x\,,\\&F''_{xx}=0,\; F''_{xy}=0,\;\; F''_{xz}=1\,,\\&F''_{yy}=0,\; F''_{yz}=0,\;\; F''_{zz}=20z^3\,.\end{aligned}$$

By implicit differentiation,

$$z'_x=-\frac{F'_x}{F'_z}=-\frac{z}{5z^4+x}\,,\quad z'_y=-\frac{F'_y}{F'_z}=\frac{1}{5z^4+x}\,.$$

Since $z(x,y)$ has continuous partial derivatives, the obtained functions $z'_x(x,y)$ and $z'_y(x,y)$ also have continuous partial derivatives (the partial derivatives are differentiable functions). Taking the partial derivatives of z'_x and z'_y with respect to x and y (with the assumption that $z = z(x,y)$) and using the quotient rule for differentiation, the second partial derivatives are obtained:

$$z''_{xx} = -\frac{z'_x(5z^4+x) - z(20z^3 z'_x + 1)}{(5z^4+x)^2} = \frac{(15z^4 - x)z'_x + z}{(5z^4+x)^2},$$

$$z''_{xy} = z''_{yx} = (z'_y)'_x = -\frac{20z^3 z'_x + 1}{(5z^4+x)^2}, \quad z''_{yy} = -\frac{20z^3 z'_y}{(5z^4+x)^2}.$$

The explicit form of z'_x and z'_y may be substituted into these relations to express the second partial derivatives via x, y, and z. At the point $(1,2)$, the values of the first partial derivatives are

$$z'_x(1,2) = -\frac{1}{6}, \qquad z'_y(1,2) = \frac{1}{6}.$$

Using these values and $z(1,2) = 1$, the values of the second partial derivatives are evaluated:

$$z''_{xx}(1,2) = -\frac{1}{27}, \quad z''_{xy}(1,2) = \frac{7}{108}, \quad z''_{yy}(1,2) = -\frac{5}{54}.$$

$\square$

22.7. Exercises.

1. Let $f(x,y) = 4 - x^2 - y^2$ if (x,y) is not in $D = \{(x,y) | x < y < 2x - 1\,, x > 1\}$ and $f(x,y) = 0$ if (x,y) is in D. Sketch D and the graph $z = f(x,y)$ for the disk $x^2 + y^2 \le 4$. Use the definition of partial derivatives to find $f'_x(1,1)$ and $f'_y(1,1)$. Consider a smooth simple curve through the point $(1,1)$ with parametric equations $x = x(t)$, $y = y(t)$ so that $x(0) = y(0) = 1$ and $x'(0)$ and $y'(0)$ do not vanish. Put $F(t) = f(x(t), y(t))$. Is it true that $F'(0) = f'_x(1,1)x'(0) + f'_y(1,1)y'(0)$ for any smooth simple curve through $(1,1)$? If not, give an example of a curve for which the chain rule holds and an example of a curve for which it does not hold.

2. Use the chain rule to find dz/dt if $z = \sqrt{1 + x^2 + 2y^2}$ and $x = 2t^3$, $y = \ln t$.

3. Use the chain rule to find $\partial z/\partial s$ and $\partial z/\partial t$ if $z = e^{-x}\sin(xy)$ and $x = ts$, $y = \sqrt{s^2 + t^2}$.

4–5. Use the chain rule to write the partial derivatives of each of the following functions F with respect to the new variables:

4. $F = f(x,y)$, $x = x(u,v,w)$, $y = y(u,v,w)$;

5. $F = f(x,y,z,t)$, $x = x(u,v)$, $y = y(u,v)$, $z = z(w,s)$, $t = t(w,s)$.

6. Find the rates of change $\partial z/\partial u$, $\partial z/\partial v$, $\partial z/\partial w$ when $(u,v,w) = (2,1,1)$ if $z = x^2 + yx + y^3$ and $x = uv^2 + w^3$, $y = u + v\ln w$.

7. For a differentiable function $f(x,y,z)$ find the rates of change f'_u, f'_v, f'_w when $(x,y,z) = (0,2,-1)$ if $x = 2/u - v + w$, $y = vuw$, $z = w^3$, and $f'_x(0,2,-1) = 2$, $f'_y(0,2,-1) = 1$, $f'_z(0,2,-1) = -1$.

8. If $z(u,v) = f(x,y)$ where $x = e^u \cos v$ and $y = e^u \sin v$, show that $z''_{xx} + z''_{yy} = e^{-2u}(z''_{uu} + z''_{vv})$.
9. If $z(u,v) = f(x,y)$ where $x = u^2 + v^2$ and $y = 2uv$, find all the second order partial derivatives of $z(u,v)$.
10. If $z(u,v) = f(x,y)$ where $x = u+v$ and $y = u-v$, show that $(z'_x)^2 + (z'_y)^2 = z'_u z'_v$.
11–16. Find the first and second partial derivatives of the function g in terms of the first and second partial derivatives of the function f:

11. $g(x,y,z) = f(x^2+y^2+z^2)$;
12. $g(x,y) = f(x, x/y)$;
13. $g(x,y,z) = f(x, xy, xyz)$;
14. $g(x,y) = f(x/y, y/x)$;
15. $g(x,y,z) = f(x+y+z, x^2+y^2+z^2)$;
16. $g(x,y) = f(x+y, xy)$.

17. Find $g''_{xx} + g''_{yy} + g''_{zz}$ if $g(x,y,z) = f(x+y+z, x^2+y^2+z^2)$
18. Let $x = r\cos\theta$ and $y = r\sin\theta$. Show that

$$f''_{xx} + f''_{yy} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial^2\theta}$$

19. Let $x = \rho\sin\phi\cos\theta$, $y = \rho\sin\phi\sin\theta$, $z = \rho\cos\phi$. The variables (ρ,ϕ,θ) are called *spherical coordinates*. Show that

$$f''_{xx} + f''_{yy} + f''_{zz} = \frac{1}{\rho^2}\frac{\partial}{\partial\rho}\left(\rho^2\frac{\partial f}{\partial\rho}\right) + \frac{1}{\rho^2\sin\phi}\frac{\partial}{\partial\phi}\left(\sin\phi\frac{\partial f}{\partial\phi}\right) + \frac{1}{\rho^2\sin^2\phi}\frac{\partial^2 f}{\partial\theta^2}$$

20. Prove that if a function $f(x,y)$ satisfies the Laplace equation $f''_{xx}+f''_{yy} = 0$, then the function $g(x,y) = f(x/(x^2+y^2), y/(x^2+y^2))$, $x^2+y^2 > 0$, also satisfies the Laplace equation.
21. Prove that if a function $f(x,t)$ satisfies the diffusion equation $f'_t = a^2 f''_{xx}$, $a > 0$, then the function

$$g(x,t) = \frac{1}{a\sqrt{t}}e^{-x^2/(4a^2t)}f\Big(\frac{x}{a^2t}, -\frac{x}{a^4t}\Big), \quad t > 0$$

also satisfies the diffusion equation.
22. Prove that if $f(x,y,z)$ satisfies the Laplace equation $f''_{xx}+f''_{yy}+f''_{zz} = 0$, then the function

$$g(x,y,z) = \frac{1}{r}f\Big(\frac{a^2x}{r^2}, \frac{a^2y}{r^2}, \frac{a^2z}{r^2}\Big), \quad r = \sqrt{x^2+y^2+z^2} \neq 0,$$

also satisfies the Laplace equation for any real a.
23. Show that the function $g(x,y) = x^n f(y/x^2)$, where f is a differentiable function, satisfies the equation $xg'_x + 2yg'_y = ng$.
24. Show that the function $g(x,y,z) = x^n f(y/x^a, z/x^b)$, where f is a differentiable function, satisfies the equation $xg'_x + ayg'_y + bzg'_z = ng$.
25–27. Assume in each case below that the given equation determines z

implicitly as a function of x and y. Find the first partial derivatives of this function $z(x, y)$:

25. $x + 2y + 3z = e^z$;
26. $x - z = \tan^{-1}(yz)$;
27. $x/z = \ln(z/y) + 1$.

28–30. Assume in each case below that the given equation determines z implicitly as a function of x and y. Find the second partial derivatives of this function $z(x, y)$:

28. $x + 2y + 3z = \ln z$;
29. $x + z = \tan^{-1}(yz)$;
30. $zx = e^{yz}$.

31–33. Verify the hypotheses of the implicit function theorem to show that in each case below the given equation determines z implicitly as a differentiable function of x and y in a neighborhood of the given point $P_0 = (x_0, y_0, z_0)$. Use the linearization to approximate the specified value of this function $z(x, y)$.

31. $z^3 - xz + y = 0\,, \quad P_0 = (3, -2, 2)\,, \quad z(2.8, -2.3)$;
32. $\ln z + xz - 2y = 0\,, \quad P_0 = (2, 1, 1)\,, \quad z(0.9, 1.2)$;
33. $yz \ln(1 + xz) - x \ln(1 + zy) = 0\,, \quad P_0 = (1, 1, 1)\,, \quad z(0.8, 1.1)$.

34. Find f'_x and f'_y where $f = (x + z)/(y + z)$ and z is defined implicitly by the equation $ze^z = xe^x + ye^y$.

35. Show that the function $z(x, y)$ defined by the equation $F(x-az, y-bz) = 0$, where F is a differentiable function of two variables and a and b are constants, satisfies the equation $az'_x + bz'_y = 1$.

36. Let the temperature of the air at a point (x, y, z) be $T(x, y, z)$ degrees Celsius. Suppose that T is a differentiable function. An insect flies through the air so that its position as a function of time t, in seconds, is given by $x = \sqrt{1 + t}$, $y = 2t$, $z = t^2 - 1$. If $T'_x(2, 6, 8) = 2$, $T'_y(2, 6, 8) = -1$, and $T'_z(2, 6, 8) = 1$, how fast is the temperature rising (or decreasing) on the insect's path as it flies through the point $(2, 6, 8)$?

37. Let the concentration of a chemical in a medium at a point (x, y, z) be $f(x, y, z)$ gram per cubic centimeter. Suppose that f is a differentiable function. Let the curve C be the intersection of the surfaces $z = xy/a$, $y = x^2/b$, where (x, y, z) are given in centimeters and $a = b = 1$ cm. Find the rate of change of the concentration of the chemical along the curve C at the point $(2, 4, 8)$ in the direction away from the origin if the rates of change along the coordinate axes are $f'_x(2, 4, 8) = 1\ g/cm^4$, $f'_y(2, 4, 8) = 2\ g/cm^4$, and $f'_z(2, 4, 8) = -1\ g/cm^4$.

38. Assume $f = f(x, y, z)$ to be a differentiable function. Put $x = 2uv$, $y = u^2 - v^2 + w$, $z = u^3vw$. Find the partial derivatives f'_u, f'_v, and f'_w at the point $u = v = w = 1$, if $f'_x = a$, $f'_y = b$, and $f'_z = c$ at $(x, y, z) = (2, 1, 1)$.

39. Let a rectangular box have the dimensions x, y, and z that change with time. Suppose that at a certain instant the dimensions are $x = 1$ m,

$y = z = 2$ m, and x and y are increasing at the rate 2 m/s and z is decreasing at the rate 3 m/s. At that instance, find the rates at which the volume, the surface area, and the largest diagonal are changing.

40. A function is said to be homogeneous of degree n if, for any number t, it has the property $f(tx, ty) = t^n f(x, y)$. Give an example of a polynomial function that is homogeneous of degree n. Show that a homogeneous differentiable function satisfies the equation $xf'_x + yf'_y = nf$ and that $f'_x(tx, ty) = t^{n-1} f'_x(x, y)$.

41. Suppose that the equation $F(x, y, z) = 0$ defines implicitly the functions $z = z(x, y)$, or $y = y(x, z)$, or $x = x(y, z)$. Assuming that F is differentiable and the partial derivatives F'_x, F'_y, and F'_z do not vanish, prove that

$$\frac{\partial z}{\partial x}\frac{\partial x}{\partial y}\frac{\partial y}{\partial z} = -1\,.$$

42. Let $x^2 = vw$, $y^2 = uw$, $z^2 = uv$, and $f(x, y, z) = F(u, v, w)$. If f and F are differentiable functions, show that $xf'_x + yf'_y + zf'_z = uF'_u + vF'_v + wF'_w$.

43. Simplify $z'_x \sec x + z'_y \sec y$ if $z = \sin y + f(\sin x - \sin y)$ where f is a differentiable function.

23. The Differential and Taylor Polynomials

23.1. The differential of multivariable functions. Given m variables $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$, one can introduce another m *independent* variables $d\mathbf{r} = \langle dx_1, dx_2, ..., dx_m\rangle$ called *differentials* of variables $\mathbf{r}$. For example, the volume of a rectangular box with dimensions x, y, and z is a function of three variables $V(x, y, z) = xyz$. The dimensions of the box are measured with errors dx, dy, and dz, respectively. The errors depend on a method of measurements (e.g., rulers with different grids produce different errors). So the errors $d\mathbf{r} = \langle dx, dy, dz\rangle$ can be viewed as variables whose values are independent of values of the dimensions $\mathbf{r} = \langle x, y, z\rangle$. In a way, differentials $d\mathbf{r}$ can be regarded as variations of the variables $\mathbf{r}$ that are independent of any particular value of $\mathbf{r}$.

DEFINITION **23.1**. (Differential).
Let f be a differentiable function of m variables. The function of $2m$ variables, $\mathbf{r}$ *and* $d\mathbf{r}$*,*

$$df(\mathbf{r}) = f'_{x_1}(\mathbf{r})\,dx_1 + f'_{x_2}(\mathbf{r})\,dx_2 + \cdots + f'_{x_m}(\mathbf{r})\,dx_m$$

is called the differential *of f.*

Note that the differential is a *linear* function in $d\mathbf{r}$. For example, let $f(x, y) = x\sin(xy)$. It is differentiable everywhere and its differential reads

$$\begin{aligned} df(x, y) &= f'_x(x, y)dx + f'_y(x, y)dy \\ &= \Big(\sin(xy) + xy\cos(xy)\Big)dx + x^2\sin(xy)\,dy\,. \end{aligned}$$

Similarly, the differential of a function of three variables (x, y, z) is a function of 6 variables, (x, y, z) and (dx, dy, dz). For example, let

$$f(x, y, z) = x^3y + 2y^2z^2 + xz\,.$$

Then

$$\begin{aligned} df(x, y, z) &= f'_x(x, y, z)dx + f'_y(x, y, z)dy + f'_z(x, y, z)dz \\ &= (3x^2y + z)dx + (x^3 + 4yz^2)dy + (4y^2z + x)dz\,. \end{aligned}$$

Geometrical significance of the differential. Consider the graph $y = f(x)$ of a function f of a single variable x (see Fig. **23.1** (left panel)). The differential

$$df(x_0) = f'(x_0)dx$$

at a point $x = x_0$ determines the increment of y along the tangent line

$$y = L(x) = f(x_0) + f'(x_0)(x - x_0)$$

as x changes from x_0 to $x_0 + \Delta x$ where $\Delta x = dx$:

$$dy = L(x_0 + dx) - L(x_0) = f'(x_0)dx = df(x_0)\,.$$

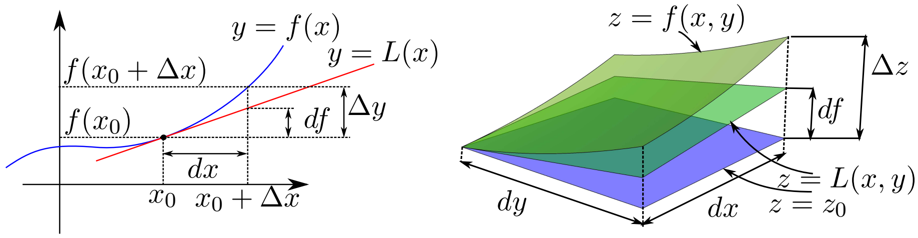


FIGURE **23.1**. Geometrical significance of the differential. **Left**: The differential of a function of one variable. It defines the increment of y along the tangent line $y = L(x)$ to the graph $y = f(x)$ at (x_0, y_0), $y_0 = f(x_0)$, when x changes from x_0 to $x_0 + \Delta x$ where $dx = \Delta x$. As $\Delta x \to 0$, the difference $\Delta y - df$ tends to 0 faster than Δx. **Right**: The differential of a function of two variables. It defines the increment of z along the tangent plane $z = L(x, y)$ to the graph $z = f(x, y)$ at (x_0, y_0, z_0), $z_0 = f(x_0, y_0)$, when (x, y) changes from (x_0, y_0) to $(x_0 + \Delta x, y_0 + \Delta y)$ where $dx = \Delta x$ and $dy = \Delta y$. The difference $\Delta z - df$ tends to 0 faster than $\|\Delta \mathbf{r}\| = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ as $\|\Delta \mathbf{r}\| \to 0$.

If $\Delta y = f(x_0 + \Delta x) - f(x_0)$ is the increment of y along the graph $y = f(x)$, then with $dx = \Delta x$ one has

$$\begin{aligned} \Delta y - df(x_0) &= \Delta y - f'(x_0)\Delta x = \left(\frac{\Delta y}{\Delta x} - f'(x_0)\right)\Delta x \\ &= \varepsilon(\Delta x)\Delta x\,, \qquad \varepsilon(\Delta x) \to 0\,, \quad \Delta x \to 0\,, \end{aligned}$$

because $\Delta y/\Delta x \to f'(x_0)$ as $\Delta x \to 0$. Thus, the difference $\Delta y - df$ tends to zero faster than Δx. Therefore, for any differentiable function f:

$$\begin{aligned} f(x_0 + dx) &= f(x_0) + df(x_0) + \varepsilon(dx)dx \\ &= f(x_0) + \Big(f'(x_0) + \varepsilon(dx)\Big)\, dx\,, \end{aligned}$$

where $\varepsilon(dx) \to 0$ as $dx \to 0$. This shows that for dx small enough, $\varepsilon(dx)$ can be neglected as compared to the number $f'(x_0) \neq 0$ and the values of f near x_0 are well approximated by $f(x_0) + df(x_0)$. The error of the approximation $\varepsilon(dx)dx$ tends to zero faster than dx.

Similarly, the differential $df(x_0, y_0)$ of a function of two variables at a point $P_0 = (x_0, y_0)$ determines the increment of z along the tangent plane

$$z = L(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0)$$

to the graph $z = f(x, y)$ at the point $(x_0, y_0, f(x_0, y_0))$, when (x, y) changes from (x_0, y_0) to $(x_0 + \Delta x, y_0 + \Delta y)$ where $dx = \Delta x$ and $dy = \Delta y$:

$$L(x_0 + dx, y_0 + dy) - L(x_0, y_0) = f_x(x_0, y_0)dx + f'_y(x_0, y_0)dy = df(x_0, y_0)$$

as depicted in the right panel of Fig. **23.1**. The increment of the function along its graph is

$$\Delta z = f(x_0 + dx, y_0 + dy) - f(x_0, y_0)\,.$$

Since f is differentiable and, hence, has a good linear approximation (see (**21.3**)), the difference $\Delta z - df(x_0, y_0)$ tends to zero faster than the distance $dr = \sqrt{(dx)^2 + (dy)^2}$ as $dr \to 0$. Therefore, similarly to the one-variable case, the differential defines the linearization of the function and approximates variations of values of the function in a neighborhood of a given point,

$$f(x_0 + dx, y_0 + dy) = f(x_0, y_0) + df(x_0, y_0) + \varepsilon(dx, dy)dr\,,$$

where $\varepsilon(dx, dy) \to 0$ as $dr = \sqrt{(dx)^2 + (dy)^2} \to 0$.

It is now straightforward to generalize the relation between the differential and the linearization to functions of any number of variables. Let $L(\mathbf{r})$ be the linearization of a function f at a point $\mathbf{r}_0$. Then

$$L(\mathbf{r}_0 + d\mathbf{r}) = f(\mathbf{r}_0) + df(\mathbf{r}_0)\,.$$

By differentiability of f (see (**21.3**)), the differential approximates variations of values of f in a neighborhood of a given point

$$f(\mathbf{r}_0 + d\mathbf{r}) = f(\mathbf{r}_0) + df(\mathbf{r}_0) + \|d\mathbf{r}\|\varepsilon(d\mathbf{r})$$

where the error of the approximation $\|d\mathbf{r}\|\varepsilon(d\mathbf{r})$ tends to 0 faster than $\|d\mathbf{r}\|$ because $\varepsilon(d\mathbf{r}) \to 0$ as $d\mathbf{r} \to \mathbf{0}$. In other words, the differential determines variations of a differentiable function f under variations $d\mathbf{r}$ of its argument, when higher powers $\|d\mathbf{r}\|^p$, $p > 1$, can be neglected as compared to $\|d\mathbf{r}\|$.

EXAMPLE **23.1**. *Find $df(x, y)$ if $f(x, y) = \sqrt{1 + x^2 y}$. In particular, evaluate $df(1, 3)$ for $(dx, dy) = (0.1, -0.2)$. What is the significance of this number?*

SOLUTION: The function has continuous partial derivatives in a neighborhood of $(1, 3)$ and hence is differentiable at $(1, 3)$. One has

$$df(x, y) = f'_x(x, y)dx + f'_y(x, y)dy = \frac{xydx}{\sqrt{1 + x^2 y}} + \frac{x^2 dy}{2\sqrt{1 + x^2 y}}\,.$$

Then

$$df(1, 3) = \frac{3}{2}dx + \frac{1}{4}dy = 0.15 - 0.05 = 0.1$$

The number $f(1, 3) + df(1, 3)$ defines the value of the linearization $L(x, y)$ of f at $(1, 3)$ for $(x, y) = (1 + dx, 3 + dy)$. It can be used to approximate $f(1 + dx, 3 + dy)$ when $\|d\mathbf{r}\| = \sqrt{(dx)^2 + (dy)^2}$ are small enough (small in comparison with, e.g., $\|d\mathbf{r}\|^2$):

$$f(1 + dx, 3 + dy) \approx L(1 + dx, 3 + dy) = f(1, 3) + df(1, 3)$$

In particular, a calculator gives $f(1+0.1, 3-0.2) - f(1,3) = 0.09476$ which is to be compared with $df(1,3) = 0.1$. Note that in this case $(dx)^2 = 0.01$ and $(dy)^2 = 0.04$ are nearly tenfold smaller than dx and dy. □

23.2. Applications to the Error Analysis. If a quantity f depends on several other quantities, say, x, y, and z, for definitiveness, i.e., f is a function $f(x,y,z)$. Suppose measurements show that $x = x_0$, $y = y_0$, and $z = z_0$. Since in practice all measurements contain errors, the value $f(x_0, y_0, z_0)$ does not have much of practical significance until its error is determined.

For example, the volume of a rectangular box with dimensions x, y, and z is the function of three variables

$$V(x,y,z) = xyz\,.$$

In practice, repetitive measurements give the values of x, y, and z from intervals

$$x_0 - \delta x \le x \le x_0 + \delta x, \quad y_0 - \delta y \le y \le y_0 + \delta y, \quad z_0 - \delta z \le z \le z_0 + \delta z,$$

where $\mathbf{r}_0 = \langle x_0, y_0, z_0 \rangle$ are the *mean values* of the dimensions for all measurements, while $\delta\mathbf{r} = \langle \delta x, \delta y, \delta z \rangle$ are *upper bounds of the absolute errors* or the *maximal uncertainties* of the measurements. To indicate the maximal uncertainty in the measured quantities, one writes

$$x = x_0 \pm \delta x\,, \quad y = y_0 \pm \delta x\,, \quad z = z_0 \pm \delta z\,.$$

Different methods of the length measurement would have different absolute error bounds (or maximal uncertainties). In other words, the dimensions x, y, and z and the bounds δx, δy, and δz are all independent variables. The values of the dimensions obtained in each measurement are

$$x = x_0 + dx\,, \quad y = y_0 + dy\,, \quad z = z_0 + dz\,,$$

where the differentials (dx, dy, dz) (or small variations) take their values in the intervals

$$-\delta x \le dx \le \delta x, \quad -\delta y \le dy \le \delta y, \quad -\delta z \le dz \le \delta z\,.$$

Let R_δ be the rectangular box spanned by all such values of $d\mathbf{r} = \langle dx, dy, dz \rangle$. The question arises:

Given the mean values $\mathbf{r}_0 = \langle x_0, y_0, z_0 \rangle$ and the absolute error bounds $\delta\mathbf{r}$, what is the absolute error bound of the volume value calculated at $\mathbf{r}_0$?

For each particular measurement, the error is

$$V(\mathbf{r}_0 + d\mathbf{r}) - V(\mathbf{r}_0) = dV(\mathbf{r}_0)$$

assuming that terms tending to 0 faster than $\|d\mathbf{r}\|$ can be neglected (errors of measurements are assumed to be small, at least, one wishes so). The components of $d\mathbf{r}$ are independent variables taking their values in the rectangular box R_δ. Then the maximal uncertainty of the calculated value of the volume is

$$\delta V = \max_{R_\delta} |dV(\mathbf{r}_0)|\,,$$

where the maximum is taken over all $d\mathbf{r}$ in R_δ. For example, if $\mathbf{r}_0 = \langle 1, 2, 3\rangle$ is in centimeters and $\delta\mathbf{r} = \langle 1, 1, 1\rangle$ is in millimeters, then the absolute error of the volume is

$$\begin{aligned}
\delta V &= \max_{R_\delta} |dV(\mathbf{r}_0)| = \max_{R_\delta} |y_0 z_0\, dx + x_0 z_0\, dy + x_0 y_0\, dz| \\
&= \max_{R_\delta} |6dx + 3dy + 2dz| = 0.6 + 0.3 + 0.2 = 1.1\ \text{ cm}^3 \\
V &= V(\mathbf{r}_0) \pm \delta V = 6 \pm 1.1\ \text{ cm}^3\,.
\end{aligned}$$

Here the maximum is reached at $dx = dy = dz = \pm 0.1\,\text{cm}$. The above analysis of error bounds can be generalized.

Definition **23.2**. (Absolute and Relative Error Bounds).
Let f be a quantity that depends on other quantities $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$ so that $f = f(\mathbf{r})$ is a differentiable function. Suppose that the values $x_i = a_i$ are known with the absolute error bounds δx_i. Put $\mathbf{r}_0 = \langle a_1, a_2, ..., a_m\rangle$, $d\mathbf{r} = \langle dx_1, dx_2, ..., dx_m\rangle$, and

$$\delta f = \max_{R_\delta} |df(\mathbf{r}_0)|\,, \quad R_\delta = \{d\mathbf{r}\ |\ -\delta x_i \le dx_i \le \delta x_i\,, i = 1, 2, ..., m\}\,.$$

where the maximum is taken over all the rectangular box R_δ. The number δf is called the absolute error bound *of the value of f at $\mathbf{r} = \mathbf{r}_0$. If, in addition, $f(\mathbf{r}_0) \ne 0$, then the number*

$$\epsilon = \frac{\delta f}{|f(\mathbf{r}_0)|} \times 100\%$$

is called the relative error bound *of the value of f at $\mathbf{r} = \mathbf{r}_0$.*

In the above example, the relative error bound of the volume measurements is $\epsilon = (1.1/6) \times 100\% \approx 18\%$. The values of each term in the linear function

$$df(\mathbf{r}_0) = \sum_{i=1}^{m} f'_{x_i}(\mathbf{r}_0)\, dx_i$$

lie in the interval

$$-|f'_{x_i}(\mathbf{r}_0)|\delta x_i \le f'_{x_i}(\mathbf{r}_0)\, dx_i \le |f'_{x_i}(\mathbf{r}_0)|\delta x_i\,, \quad i = 1, 2, ..., m$$

where the boundary values are reached when $dx_i = \pm\delta x_i$. Taking the sum over i in this inequality it is concluded that $|df(\mathbf{r}_0)|$ attains its maximal value on R_δ when $dx_i = \delta x_i$ for all i for which the coefficient $f'_{x_i}(\mathbf{r}_0)$ is positive and $dx_i = -\delta x_i$ for all i for which the coefficient $f'_{x_i}(\mathbf{r}_0)$ is negative. So the absolute error bound can be written in the form

$$\delta f = \sum_{i=1}^{m} |f'_{x_i}(\mathbf{r}_0)|\, \delta x_i .$$

23.3. Accuracy of a Linear Approximation. Values of a differentiable function $f(x)$ near x_0 can be well approximated by its linearization:

$$f(x_0 + dx) \approx f(x_0) + df(x_0) = f(x_0) + f'(x_0)dx \,,$$

where dx denotes a variation of the variable x near x_0. Suppose f is differentiable many times. The following questions are of interest.

- Can one estimate the error bound of the linear approximation?
- Can one systematically improve an approximation of f near x_0?

These questions are answered by the Taylor theorem studied in Calculus I and II. It asserts that *if $f(x)$ has continuous derivatives up to order n on an interval I containing x_0 and $f^{(n+1)}$ exists and is bounded on I, $|f^{(n+1)}(x)| \le M_{n+1}$ for some constant M_{n+1}, then*

$$f(x) = T_n(x) + \varepsilon_n(x) \,,$$
$$T_n(x) = f(x_0) + \frac{f'(x_0)}{1!}\Delta x + \frac{f''(x_0)}{2!}(\Delta x)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(\Delta x)^n \,,$$
$$|\varepsilon_n(x)| \le \frac{M_{n+1}}{(n+1)!}|\Delta x|^{n+1} \,, \quad \Delta x = x - x_0 \,. \tag{23.1}$$

The polynomial $T_n(x)$ is called the *Taylor polynomial of degree n* about x_0. The remainder $\varepsilon_n(x)$ determines the absolute error of the nth-order Taylor polynomial approximation $f(x) \approx T_n(x)$.

Suppose one wants to approximate the values of a function f in an interval

$$I: \quad x_0 - \delta x \le x \le x_0 + \delta x$$

for some $\delta x > 0$. The function f is assumed to have sufficiently many continuous derivatives in this interval. Put

$$x = x_0 + dx \,, \qquad |dx| \le \delta x \,.$$

The differential dx determines variations of the argument of f near x_0 and δx is the upper bound of these variations. Let us write an explicit form of a few first Taylor polynomials about the point x_0 (with $x - x_0 = dx$) and the upper bound of the error of the corresponding approximation of the function f in the interval I:

$$T_0(x_0 + dx) = f(x_0) \,,$$
$$|f(x_0 + dx) - T_0(x_0 + dx)| \le M_1 |dx| \,, \quad M_1 = \max_I |f'(x)|$$

The zero-degree Taylor polynomial is a constant $T_0(x) = f(x_0)$. It approximates the values of f in the interval I by the value $f(x_0)$. Naturally, the maximal error of such an approximation is determined by the maximal slope M_1 of f in the interval I.

A better approximation is obtained if the rate of change of f at x_0 is taken into account. This is achieved by the tangent line approximation which is also the first-order Taylor polynomial approximation:

$$T_1(x_0+dx) = f(x_0) + f'(x_0)dx = f(x_0) + df(x_0)\,,$$

$$|f(x_0+dx) - T_1(x_0+dx)| \leq \frac{1}{2}M_2|dx|^2\,, \quad M_2 = \max_I |f''(x)| \tag{23.2}$$

Equation (**23.2**) answers the question about the accuracy of the tangent line approximation. If the second derivative of f is bounded, then the error of the tangent line approximation decreases *quadratically* with decreasing the distance from the point in a neighborhood of which the function is approximated.

Naturally, an even better approximation can be obtained if, in addition to the slope, the concavity of the approximated function is taken into account. This is done by the second-order Taylor polynomial approximation:

$$T_2(x_0+dx) = f(x_0) + f'(x_0)dx + \frac{f''(x_0)}{2!}\,(dx)^2\,,$$
$$|f(x_0+dx) - T_2(x_0+dx)| \leq \frac{M_3}{6}\,|dx|^3\,, \quad M_3 = \max_I |f'''(x)|\,.$$

Now the error of the approximation decreases with decreasing the distance $|dx|$ even faster.

For a constant function, the zero-order Taylor polynomial approximation about any point is exact (it has no error). For a linear function, the first-order Taylor polynomial approximation about any point is exact because a linear function is uniquely determined by its value and the slope at a point. The second-order Taylor polynomial approximation of a quadratic polynomial about any point is also exact (a quadratic polynomial has a constant (fixed) concavity and, hence, is uniquely determined by the value of the concavity and by its value and the slope at a point):

$$\left\{\begin{aligned} f''(x) &= a \\ f'(x_0) &= b \\ f(x_0) &= c \end{aligned}\right. \quad \Leftrightarrow \quad f(x) = c + b(x-x_0) + \frac{1}{2}a^2(x-x_0)^2\,.$$

It is not difficult to see that the Taylor polynomial of degree n for a polynomial of the same degree about any point coincides with that polynomial. A polynomial of degree n has a constant nth derivative and therefore can be uniquely reconstructed from the value of $f^{(n)}(x) = a$ (by taking n antiderivatives), provided the values of the polynomial and the values of its derivatives up to order $n-1$ are given at a particular point (to determine the integration constants). The reconstructed polynomial coincides with the Taylor polynomial.

Let us introduce shorter (more convenient) notations for Taylor polynomials. The differential df can be viewed as the result of the action of the *operator* d on a differentiable function f:

$$d = dx\frac{d}{dx}: \quad f(x) \;\to\; df(x) = dx\frac{df(x)}{dx} = f'(x)dx$$

that is, the action of d on f means taking the derivative of f at a point x and multiplying the result by an *independent* variable dx. A repetitive action of the operator d on f produces *higher-order differentials* of f. The nth differential of f is defined by the rule

$$d^n f(x) = f^{(n)}(x)(dx)^n = \left(dx\frac{d}{dx}\right)^n f(x)$$

where the action of the powers d^n on f is understood as successive actions of the operator d, $d^n f = d^{n-1}(df)$, in which the variables dx and x are independent. For example,

$$d^2 f(x) = d(df(x)) = dx\frac{\partial}{\partial x}\Big(f'(x)dx\Big) = (dx)^2\frac{\partial}{\partial x}\, f'(x) = (dx)^2 f''(x)\,.$$

Then the Taylor polynomials about x_0 can be written as

$$T_n(x_0+dx) = f(x_0) + \frac{1}{1!}df(x_0) + \frac{1}{2!}d^2 f(x_0) + \cdots + \frac{1}{n!}d^n f(x_0)\,,$$

where $x = x_0 + dx$ or $dx = x - x_0$, and $n = 1, 2, \ldots$. An upper bound of errors of the nth order Taylor polynomial approximation in the interval I is

$$\textbf{(23.3)} \qquad |f(x_0+dx) - T_n(x_0+dx)| \le \frac{M_{n+1}}{(n+1)!}\,|dx|^{n+1}\,,$$
$$M_{n+1} = \max_I |f^{(n+1)}(x)|\,.$$

It is desirable to find a convenient way to estimate the Taylor approximation error, avoiding the calculation of constants M_n.

An estimate of the Taylor approximation error. Let us estimate first the absolute error of the linear approximation. Put $|dx| \le \delta x$ and $I = [x_0 - \delta x, x_0 + \delta x]$. Suppose in addition that $f'''(x)$ is continuous on I. By the mean value theorem (Calculus 1), for any continuously differentiable function $g(x)$ on I there exists a point x^* between x_0 and $x_0 + dx$ such that

$$g(x_0+dx) - g(x_0) = g'(x^*)dx \quad \Rightarrow \quad g(x_0+dx) = g(x_0) + g'(x^*)dx\,.$$

In particular, set $g(x) = f''(x)$ and use $|A+B| \le |A| + |B|$ to obtain

$$\begin{aligned} |f''(x_0+dx)| &\le |f''(x_0)| + |f'''(x^*)||dx| \\ &\le |f''(x_0)| + M_3|\delta x|\,, \quad M_3 = \max_I |f'''(x)|\,. \end{aligned}$$

The latter inequality holds for any dx in the left side. Therefore

$$M_2 = \max_I |f''(x)| \le |f''(x_0)| + M_3\delta x$$

because $|dx| \le \delta x$. For small enough variations of dx (small enough δx), the second term in the right side is much smaller than the first one so that M_2 can be approximated by $|f''(x_0)|$: $M_2 \approx |f''(x_0)|$. The absolute error of the linear approximation can therefore be estimated as

$$\begin{aligned}
\varepsilon_1 &= |f(x_0+dx) - T_1(x_0+dx)| = |f(x_0+dx) - f(x_0) - df(x_0)| \\
&\le \frac{M_2}{2}|dx|^2 \approx \frac{|f''(x_0)|}{2}\,|dx|^2 = \frac{1}{2}|d^2 f(x_0)|\,, \\
\varepsilon_1 &\approx \varepsilon_1^{es} = \frac{1}{2}|d^2 f(x_0)|\,.
\end{aligned}$$

Here ε_1^{es} is an estimate of the absolute error of the linear approximation of f near x_0.

Similarly by setting $g(x) = f^{(n+1)}(x)$ and assuming that f has sufficiently many continuous derivatives, the constant M_{n+1} in Eq. (**23.3**) is shown to satisfy the condition

$$M_{n+1} \le |f^{(n+1)}(x_0)| + M_{n+2}\delta x\,.$$

For small enough δx, the constant M_{n+1} can be approximated by the derivative $|f^{(n+1)}(x_0)|$. It follows from (**23.3**) that the absolute error of the Taylor polynomial approximation can be estimated as

$$\begin{aligned}
\varepsilon_n &= |f(x_0+dx) - T_n(x_0+dx| \approx \varepsilon_n^{es} \\
\varepsilon_n^{es} &= \frac{|f^{(n+1)}(x_0)|}{(n+1)!}\,|dx|^{n+1} = \frac{|d^{n+1} f(x_0)|}{(n+1)!}
\end{aligned} \tag{23.4}$$

The right side of Eq. (**23.4**) is the term that defines the next order of the Taylor polynomial approximation. If the current approximation is accurate, then a "correction" to it is expected to be small.

This observation can be quantified by estimating the relative error of the Taylor approximation

$$\epsilon_n = \frac{|f(x) - T_n(x)|}{|f(x)|} \times 100\%\,, \quad x = x_0 + dx$$

The top of this ratio is approximated by ε_n^{es}, while in the denominator one can use

$$f(x_0+dx) \approx T_n(x_0+dx) \tag{23.5}$$

so that

$$\epsilon_n^{es} = \frac{|d^{n+1} f(x_0)|}{(n+1)!|T_n(x_0+dx)|} \times 100\%\,. \tag{23.6}$$

The number ϵ_n^{es} is easier to calculate than ϵ_n. It is expected to be close to ϵ_n whenever the next term in the Taylor expansion is much smaller than the previous one. For example, the linear approximation is expected to be accurate if the term quadratic in dx in the Taylor polynomial approximation is small as compared to the linear one:

$$(\mathbf{23}.\mathbf{7}) \qquad \frac{1}{2}|d^2f(x_0)| \ll |df(x_0)| \qquad \text{or} \qquad \frac{|d^2f(x_0)|}{2|df(x_0)|} \ll 1\,,$$

where $\ll$ stands for "much less", and the absolute value is needed because the values of the first and second differential may be positive or negative. In this case, the relative error and its estimate

$$(\mathbf{23}.\mathbf{8}) \qquad \epsilon_1 \approx \epsilon_1^{es} = \frac{|d^2f(x_0)|}{2|f(x_0)+df(x_0)|} \times 100\%\,.$$

are small.

It should be understood that a rigorous bound on the error is given in the Taylor theorem (**23**.**3**). The proposed criteria can be used to *estimate* the error by order of magnitude. If, for example, the ratio (**23**.**7**) is 0.1 or less (one can say, $0.1 \ll 1$), then the next (second) order approximation would only change the approximate value of the function in the *next decimal place* as compared to the first order approximation.

To illustrate the above error analysis, let us estimate the number $e^{0.6}$ using several Taylor polynomial approximations about 0. So, put

$$f(x) = e^x\,, \quad x_0 = 0\,, \quad dx = 0.6\,.$$

Let us write first six terms of the Taylor expansion and calculate their values (up to $d^5f(x_0)$). Since $f'(x) = e^x = f(x)$ so that $f^{(n)}(0) = 1$,

$$\begin{aligned} e^{0.6} &\approx 1 + dx + \frac{(dx)^2}{2} + \frac{(dx)^3}{6} + \frac{(dx)^4}{24} + \frac{(dx)^5}{120} \\ &= 1 + 0.6 + 0.18 + 0.036 + 0.0054 + 0.000648 \\ &= 1.822058 \end{aligned}$$

The exact value (rounded to the same decimal place) is 1.822119. Let us compare the exact and approximate values of the relative errors for a few Taylor approximations, keeping two first significant digits:

$$\begin{array}{ccccc} \epsilon_0^{es} = 60\%\,, & \epsilon_1^{es} = 11\%\,, & \epsilon_2^{es} = 2.0\%\,, & \epsilon_3^{es} = 0.30\%\,, & \epsilon_4^{es} = 0.036\% \\ \epsilon_0 = 45\%\,, & \epsilon_1 = 12\%\,, & \epsilon_2 = 2.3\%\,, & \epsilon_3 = 0.34\%\,, & \epsilon_4 = 0.039\% \end{array}$$

So, the estimate of the relative error is not far off, $\epsilon_n^{es} \approx \epsilon_n$.

So, the Taylor polynomial approximations provide a systematic way to approximate a function in a neighborhood of a point if the function has sufficiently many continuous derivatives in that neighborhood. It is therefore desirable to develop an analogous method for functions of several variables.

Remark. There are differentiable functions that cannot be approximated by Taylor polynomials at some points. For example, the function

$$f(x) = \exp\Big(-\frac{1}{x^2}\Big)\,, \quad x \neq 0; \qquad f(0) = 0\,,$$

is differentiable at $x = 0$ any number of times (as it is at any other point). Using the definition of the derivative as the limit, it is not difficult to show recursively for $n = 1, 2, ...$ that

$$f^{(n)}(0) = 0 \quad \Rightarrow \quad d^n f(0) = 0\,, \quad n = 1, 2, \ldots\,,$$

(this is left to the reader as an exercise). Therefore all Taylor polynomials for f about $x = 0$ identically vanish in any neighborhood of $x = 0$. On other hand, the function is not zero for a non-zero value of its argument. Hence, $f(x)$ cannot be approximated by Taylor polynomials about $x = 0$. Of course, Taylor approximations exist about any $x_0 \neq 0$ because $d^n f(x_0)) \neq 0$ for all n if $x_0 \neq 0$.

23.4. Taylor Polynomials of Two Variables. Let f be a function of two variables x and y and the differentials dx and dy be another two independent variables. By the analogy with the one-variable case, the differential df is viewed as the result of the action of the operator d on f:

$$df(x, y) = \left(dx\frac{\partial}{\partial x} + dy\frac{\partial}{\partial y}\right) f(x, y) = dx f'_x(x, y) + dy f'_y(x, y)\,.$$

The operator d has the same properties as the differential (e.g., the product and quotient rules):

$$d(f + g) = df + dg\,, \quad d(fg) = dfg + fdg\,, \quad d\left(\frac{f}{g}\right) = \frac{dfg - fdg}{g^2}$$

for any two differentiable functions f and g ($g \neq 0$ in the quotient rule).

Definition **23.3**. *Suppose that f has continuous partial derivatives up to order n. The quantity*

$$d^n f(x, y) = \left(dx\frac{\partial}{\partial x} + dy\frac{\partial}{\partial y}\right)^n f(x, y)$$

is called the n^{th} order differential *of f, where the action of powers d^n on f is defined successively $d^n f = d^{n-1}(df)$ and the variables dx, dy, x, and y are viewed as independent when taking partial derivatives.*

The differential $d^n f$ is a function of 4 variables dx, dy, x, and y. For example,

$$\begin{aligned} d^2 f &= \left(dx\frac{\partial}{\partial x} + dy\frac{\partial}{\partial y}\right)^2 f = \left(dx\frac{\partial}{\partial x} + dy\frac{\partial}{\partial y}\right)(dx f'_x + dy f'_y) \\ &= \left((dx)^2\frac{\partial}{\partial x} + dxdy\frac{\partial}{\partial y}\right) f'_x + \left(dxdy\frac{\partial}{\partial x} + (dy)^2\frac{\partial}{\partial y}\right) f'_y \\ &= f''_{xx}(dx)^2 + 2f''_{xy}dxdy + f''_{yy}(dy)^2 \end{aligned}$$

By continuity of the partial derivatives, the order of differentiation is irrelevant (Clairaut's theorem), $f''_{xy} = f''_{yx}$. The numerical coefficients at each of the terms are binomial coefficients: $(a+b)^2 = a^2+2ab+b^2$. Similarly, using the binomial coefficients for the cube of the sum

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

the third differential is obtained:

$$d^3f = f'''_{xxx}(dx)^3 + 3f'''_{xxy}(dx)^2dy + 3f'''_{xyy}dx(dy)^2 + f'''_{yyy}(dy)^3\,,$$

where Clairaut's theorem has been used again. The n^{th} differential can be written in the form

$$d^nf = \sum_{k=0}^{n} B_k^n \,\frac{\partial^n f}{\partial x^{n-k}\partial y^k}\,(dx)^{n-k}(dy)^k\,, \qquad B_k^n = \frac{n!}{k!(n-k)!}$$

where B_k^n are the binomial coefficients:

$$(a+b)^n = \sum_{k=0}^{n} B_k^n a^{n-k}b^k\,.$$

Example **23.2**. *Find d^2f if $f(x,y) = x^2y + y^3$.*

Solution: The function is a polynomial and hence has continuous partial derivatives of any order. In particular,

$$\begin{array}{l} f''_{xx} = 2y \\ f''_{xy} = 2x \\ f''_{yy} = 6y \end{array} \quad \Rightarrow \quad \begin{aligned} d^2f &= f''_{xx}(x,y)(dx)^2 + 2f''_{xy}(x,y)dxdy + f''_{yy}(x,y)(dy)^2 \\ &= 2y(dx)^2 + 4xdxdy + 6y(dy)^2\,. \end{aligned}$$

□

Example **23.3**. *Find d^nf if $f(x,y) = e^{ax+by}$ where a and b are constants.*

Solution: Note first that the function is the composition $f(x,y) = e^u$ where $u = ax+by$. If $f(x,y) = g(u)$ where u is a function of (x,y), then by the product rule for the operator d

$$df = g'(u)du\,, \quad d^2f = d(g'(u)du) = g''(u)(du)^2 + g'(u)d^2u\,.$$

So by successive actions of the operator d, the differentials of f can be expressed in terms of the differentials of u. In the case considered, u is a linear function and therefore all second partial derivatives are identically zero (and so are all higher-order partial derivatives). Therefore $d^nu = 0$ for all $n \geq 2$. Or, notice that $du = adx + bdy$ does not depend on (x,y) and hence $d^2u = d(du) = 0$. It is then concluded that

$$\begin{aligned} df = e^u du \quad &\Rightarrow \quad d^2f = e^u(du)^2 \\ &\Rightarrow \quad d^nf = e^u(du)^n = e^{ax+by}(adx+bdy)^n\,, \end{aligned}$$

for all n. □

Definition **23.4**. (Taylor polynomials of two variables)
Let f have continuous partial derivatives up to order n. The Taylor polynomial of degree n about a point (x_0, y_0) is

$$T_n(x,y) = f(x_0,y_0) + \frac{1}{1!}\,df(x_0,y_0) + \frac{1}{2!}\,d^2f(x_0,y_0) + \cdots + \frac{1}{n!}d^nf(x_0,y_0)$$

where dx is set to $x - x_0$ and dy to $y - y_0$ after taking the differentials.

The first four Taylor polynomials are

$$\begin{aligned}
T_0(\mathbf{r}) &= f(\mathbf{r}_0),\\
T_1(\mathbf{r}) &= f(\mathbf{r}_0) + f'_x(\mathbf{r}_0)\,dx + f'_y(\mathbf{r}_0)\,dy\,,\\
T_2(\mathbf{r}) &= T_1(\mathbf{r}) + \frac{f''_{xx}(\mathbf{r}_0)}{2}\,(dx)^2 + f''_{xy}(\mathbf{r}_0)\,dxdy + \frac{f''_{yy}(\mathbf{r}_0)}{2}\,(dy)^2,\\
T_3(\mathbf{r}) &= T_2(\mathbf{r}) + \frac{f'''_{xxx}(\mathbf{r}_0)}{6}\,(dx)^3 + \frac{f'''_{xxy}(\mathbf{r}_0)}{2}\,(dx)^2dy\\
&\quad + \frac{f'''_{xyy}(\mathbf{r}_0)}{2}\,dx(dy)^2 + \frac{f'''_{yyy}(\mathbf{r}_0)}{6}\,(dy)^3\,,
\end{aligned}$$

where $\mathbf{r} = \langle x, y\rangle$, $\mathbf{r}_0 = \langle x_0, y_0\rangle$, $dx = x - x_0$, and $dy = y - y_0$. The linearization of f at $\mathbf{r}_0$ coincides with the first-degree Taylor polynomial $T_1(\mathbf{r})$.

23.5. Multivariable Taylor Polynomials. For more than two variables, Taylor polynomials are defined similarly. Let

$$\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle\,, \qquad d\mathbf{r} = \langle dx_1, dx_2, ..., dx_m\rangle\,.$$

Suppose that a function f has continuous partial derivatives up to order n. The n^{th} order differential of $f(\mathbf{r})$ is defined by

$$d^nf(\mathbf{r}) = \left(dx_1\frac{\partial}{\partial x_1} + dx_2\frac{\partial}{\partial x_2} + \cdots + dx_m\frac{\partial}{\partial x_m}\right)^n f(\mathbf{r})$$

where the variables $\mathbf{r}$ and $d\mathbf{r}$ are viewed as independent when differentiating. For example,

$$\begin{aligned}
d^2f(\mathbf{r}) &= \sum_{j=1}^{m}\sum_{i=1}^{m} f''_{x_jx_i}(\mathbf{r})\,dx_jdx_i\,,\\
d^3f(\mathbf{r}) &= \sum_{k=1}^{m}\sum_{j=1}^{m}\sum_{i=1}^{m} f'''_{x_kx_jx_i}(\mathbf{r})\,dx_kdx_jdx_i\,.
\end{aligned}$$

Owing to continuity of the partial derivatives and Clairaut's theorem, some of the terms in these multiple sums are identical. Numerical coefficients at the equal terms are the multivariable analog of the binomial coefficients that appear in the expansion of $(a_1 + a_2 + \cdots + a_m)^n$, $n = 2, 3, ...$

The Taylor polynomial of degree n about a point $\mathbf{r}_0$ is

$$T_n(\mathbf{r}) = f(\mathbf{r}_0) + \frac{1}{1!}df(\mathbf{r}_0) + \frac{1}{2!}d^2f(\mathbf{r}_0) + \cdots + \frac{1}{n!}d^nf(\mathbf{r}_0)$$

where $d\mathbf{r}$ is set to $\mathbf{r}-\mathbf{r}_0$ after calculating the differentials. Taylor polynomials are partial sum of the power series:

$$f(\mathbf{r}_0) + \sum_{n=1}^{\infty} \frac{d^n f(\mathbf{r}_0)}{n!}, \qquad d\mathbf{r} = \mathbf{r} - \mathbf{r}_0,$$

which is called the *Taylor series for f about a point* $\mathbf{r}_0$. If the series converges to $f(\mathbf{r})$ in some neighborhood of $\mathbf{r}_0$:

$$f(\mathbf{r}) = f(\mathbf{r}_0) + \sum_{n=1}^{\infty} \frac{d^n f(\mathbf{r}_0)}{n!}, \qquad d\mathbf{r} = \mathbf{r} - \mathbf{r}_0,$$

then the series is called a *power (Taylor) series representation of f near* $\mathbf{r}_0$.

23.6. Multivariable Taylor polynomial approximations.

THEOREM **23.1**. (Taylor Theorem)
Let D be an open ball centered at $\mathbf{r}_0$ and let the partial derivatives of a function f be continuous up to order n on D. Then

$$f(\mathbf{r}) = T_n(\mathbf{r}) + \varepsilon_n(\mathbf{r})$$

and the reminder ε_n satisfies the condition

$$|\varepsilon_n(\mathbf{r})| \le h_n(\mathbf{r})\|\mathbf{r} - \mathbf{r}_0\|^n \quad \text{where} \quad h_n(\mathbf{r}) \to 0 \text{ as } \mathbf{r} \to \mathbf{r}_0$$

If $n = 1$, then $T_1(\mathbf{r}) = L(\mathbf{r})$ is the linearization of f at $\mathbf{r}_0$. The Taylor theorem merely states that a function with continuous partial derivatives has a good linear approximation. This is a simple consequence of Theorem **21.3** (the continuity of partial derivatives ensures differentiability of f at $\mathbf{r}_0$, and, hence, the existence of a good linear approximation). For $n \ge 2$, it states that the approximation of f by the Taylor polynomial T_n is a good approximation in the sense that the error decreases faster than $\|\mathbf{r} - \mathbf{r}_0\|^n$. Consequently, the relative error of the approximation

$$\epsilon_n = \frac{|\varepsilon_n(\mathbf{r})|}{|f(\mathbf{r})|} \times 100\% = \frac{|f(\mathbf{r}) - T_n(\mathbf{r})|}{|f(\mathbf{r})|} \times 100\% \tag{23.9}$$

is expected to get smaller with increasing n for $\mathbf{r}$ close to $\mathbf{r}_0$ (assuming that $f(\mathbf{r}) \ne 0$).

So, *a practical significance of the Taylor theorem is that higher-order differentials of a function can be used to obtain successively better approximations of values of a function near a point if the function has continuous partial derivatives of higher orders in a neighborhood of that point.*

EXAMPLE **23.4**. *Let $f(x, y) = \sqrt{1 + x^2 y}$. Find $df(1, 3)$ and $d^2 f(1, 3)$ and use them to approximate $f(1 + 0.1, 3 - 0.2)$ by the linear and quadratic Taylor polynomials about $(1, 3)$. Compare the results with the exact value (e.g., use a calculator to find it) by finding the relative errors ϵ_1 and ϵ_2.*

SOLUTION: Put $(dx, dy) = (0.1, -0.2)$. It was found in Example **23.1** that $df(1,3) = 0.1$. The second partial derivatives are obtained by the quotient rule (see f'_x and f'_y in Example **23.1**):

$$f''_{xx}(1,3) = \frac{y(1+x^2y)^{1/2} - x^2y^2(1+x^2y)^{-1/2}}{1+x^2y}\Big|_{(1,3)} = \frac{3}{8}$$
$$f''_{xy}(1,3) = \frac{2x(1+x^2y)^{1/2} - x^3y(1+x^2y)^{-1/2}}{2(1+x^2y)}\Big|_{(1,3)} = \frac{5}{16}$$
$$f''_{yy}(1,3) = -\frac{x^4}{4}(1+x^2y)^{-3/2}\Big|_{(1,3)} = -\frac{1}{32}$$

Therefore

$$\begin{aligned} d^2f(1,3) &= f''_{xx}(1,3)(dx)^2 + 2f''_{xy}(1,3)dxdy + f''_{yy}(1,3)(dy)^2 \\ &= \frac{3}{8}\,(dx)^2 + \frac{5}{8}\,dxdy - \frac{1}{32}\,(dy)^2 = -0.01 \end{aligned}$$

The linear approximation gives

$$f(1+dx, 3+dy) \approx f(1,3) + df(1,3) = 2 + 0.1 = 2.1\,.$$

The approximation by the quadratic Taylor polynomial about $(1,3)$ gives

$$\begin{aligned} f(1+dx, 3+dy) &\approx f(1,3) + df(1,3) + \frac{1}{2}d^2f(1,3) \\ &= 2.1 - 0.01/2 = 2.095\,. \end{aligned}$$

The exact (calculator) value is $f(1+dx, 3+dy) = 2.0948$ rounded to 4 decimal places. The relative errors (rounded to two significant digits) of the obtained Taylor approximations are

$$\begin{aligned} \epsilon_1 &= \frac{|2.0948 - 2.1|}{2.0948} \times 100\% = 0.25\%\,, \\ \epsilon_2 &= \frac{|2.0948 - 2.095|}{2.0948} \times 100\% = 0.0095\%\,. \end{aligned}$$

Evidently, the quadratic approximation is better than the linear approximation. □

Just like in the one-variable case, it is desirable to have a simple method to estimate the relative error of Taylor approximations. The Taylor theorem does not allow to do so because the function h_n remains unknown. However, if the function has sufficiently many continuous partial derivatives, an upper bound on h_n can be obtained.

Suppose that one wishes to approximate a function f in a ball of radius δ centered at $\mathbf{r}_0$:

$$B_\delta: \qquad \mathbf{r} = \mathbf{r}_0 + d\mathbf{r}\,, \qquad \|d\mathbf{r}\| \le \delta\,.$$

Let us first consider the two-variable case, $\mathbf{r} = \langle x, y\rangle$. The following consequence of the Taylor theorem can be proved.

COROLLARY **23.1**. (Error bound of the linear approximation) *Suppose that a function f has continuous partial derivatives in an open disk $B_\delta = \{\mathbf{r}\,|\,\|\mathbf{r}-\mathbf{r}_0\| \le \delta$ and the second partial derivatives exist and are bounded on B_δ:*

$$|f''_{xx}| \le M_{20}\,, \quad |f''_{xy}| \le M_{11}\,, \quad |f''_{yx}| \le M_{11}\,, \quad |f''_{yy}| \le M_{02}\,,$$

all points in B_δ. Let $T_1(\mathbf{r}) = L(\mathbf{r})$ be the linearization of f at $\mathbf{r}_0$. Put $\mathbf{r} = \mathbf{r}_0 + d\mathbf{r}$. Then

$$\begin{aligned} |f(\mathbf{r}) - L(\mathbf{r})| &\le \frac{1}{2}\Big(M_{20}(dx)^2 + 2M_{11}|dxdy| + M_{02}(dy)^2\Big) \\ &\le \frac{M_2}{2}\|d\mathbf{r}\|^2 \le \frac{1}{2}M_2\delta^2 \end{aligned} \tag{23.10}$$

for all $\mathbf{r}$ in D, where $M_2 = M_{20} + 2M_{11} + M_{02}$

Inequality (**23.10**) follows from $|dx| \le \|d\mathbf{r}\|$ and similarly for the variable y. It is the two-variable analog of (**23.2**) in the one-variable case. If f has more continuous partial derivatives, then a two-variable analog of (**23.3**) can also be established from the Taylor theorem.

COROLLARY **23.2**. (Error Bound of Taylor Polynomial Approximations) *If, in addition to the hypotheses of Theorem **23.1**, the function f has partial derivatives of order $n+1$ that are bounded on B_δ; that is, there exist numbers M_{kn+1}, $k = 0, 1, 2, ..., n+1$, such that*

$$\left|\frac{\partial^{n+1} f(x,y)}{\partial x^{n+1-k}\partial y^k}\right| \le M_{kn+1}\,, \qquad \mathbf{r} = \mathbf{r}_0 + d\mathbf{r}\,, \quad \|d\mathbf{r}\| \le \delta\,.$$

Then the remainder satisfies

$$|\varepsilon_n(\mathbf{r})| \le \sum_{k=0}^{n+1} \frac{B_k^{n+1} M_{kn+1}}{(n+1)!}\,|dx|^{n+1-k}|dy|^k\,,$$

where $B_k^{n+1} = (n+1)!/(k!(n+1-k)!)$ are binomial coefficients.

The upper bound for the error can be simplified a little bit. Note $|dx| \le \|d\mathbf{r}\|$ and $|dy| \le \|d\mathbf{r}\|$. Hence, $|dx|^{n+1-k}|dy|^k \le \|d\mathbf{r}\|^{n+1}$. Making use of this inequality, one infers that

$$|\varepsilon_n(\mathbf{r}_0 + d\mathbf{r})| \le \frac{M_{n+1}}{(n+1)!}\|d\mathbf{r}\|^{n+1}, \tag{23.11}$$

where the constant $M_{n+1} = \sum_{k=0}^{n+1} B_k^{n+1} M_{kn+1}$. Inequality (**23.11**) is a two-variable analog of (**23.3**).

There are natural extensions of Corollaries **23.1** and **23.2** to multivariable functions. In fact, (**23.11**) remains valid for any number of variables, where the calculation of the constant M_{n+1} is changed. Finding an upper bound for the error of Taylor polynomial approximations requires bounds on partial derivatives of f in a neighborhood of a point. This is not generally an easy task. So, just like in the one-variable case, one has to establish a simple practical way to *estimate* the error of Taylor approximations.

Error estimates for multivariable Taylor approximations. Suppose that f has sufficiently many *continuous* partial derivatives in the ball B_δ (introduced above). In contrast to the hypotheses of Corollary **23.2**, the partial derivatives of order $n+1$ are assumed to be not only existing and bounded, but also continuous. The latter implies that all partial derivatives up order $n+1$ are *differentiable* functions. This, in turns, means that any partial derivative of order $n+1$ has a good linear approximation and, hence, its maximal value in the ball B_δ differs from its value at $\mathbf{r}_0$ by a term proportional to the radius $\delta \approx \|d\mathbf{r}\|$ of the ball (just like in the one-variable case). Therefore, by replacing the upper bounds M_{kn+1} by the absolute values of the corresponding partial derivatives at $\mathbf{r}_0$ in Corollary **23.2**, the error bound is changed by terms of order $\|d\mathbf{r}\|^{n+2}$ which are small as compared to the leading terms of order $\|d\mathbf{r}\|^{n+1}$ in (**23.11**), provided, of course, that $\|d\mathbf{r}\|$ is small enough.

Hence, just like in the one-variable case, the absolute error of a Taylor polynomial approximation

$$f(\mathbf{r}_0 + d\mathbf{r}) = T_n(\mathbf{r}_0 + d\mathbf{r}) + \varepsilon_n(d\mathbf{r})$$

can be estimated by the $(n+1)$th differential:

$$\varepsilon_n(d\mathbf{r}) \approx \varepsilon_n^{es}(d\mathbf{r}) = \frac{|d^{(n+1)}f(\mathbf{r}_0)|}{(n+1)!}$$

A Taylor polynomial approximation is sufficiently accurate if the next term of the Taylor expansion is much smaller than the former one:

$$\frac{|d^{(n+1)}f(\mathbf{r}_0)|}{(n+1)!} \ll \frac{|d^n f(\mathbf{r}_0)|}{n!} \qquad \Rightarrow \qquad \frac{|d^{(n+1)}f(\mathbf{r}_0)|}{(n+1)|d^n f(\mathbf{r}_0)|} \ll 1\,.$$

The relative error of the linear approximation for a function with continuous second partial derivatives near a point $\mathbf{r}_0$ *can be estimated as*

$$\epsilon_1^{es} = \frac{|d^2 f(\mathbf{r}_0)|}{2|f(\mathbf{r}_0) + df(\mathbf{r}_0)|} \times 100\% \qquad (\mathbf{23.12})$$

If the second differential happens to be identically zero, then the next nonzero (third) differential can be used to access the accuracy of the linear approximation. Similarly, the relative error (**23.9**) of the nth order Taylor approximation (where $\mathbf{r} = \mathbf{r}_0 + d\mathbf{r}$) is estimated as

$$\epsilon_n^{es} = \frac{|d^{(n+1)}f(\mathbf{r}_0)|}{(n+1)!|T_n(\mathbf{r}_0 + d\mathbf{r})|} \times 100\%$$

provided the function f has continuous partial derivatives up to order $n+1$. The above analysis is not specific for functions of two variables. So, the given estimates of the relative error of Taylor polynomial approximations can be used for functions of any number of variables.

Example **23.5**. *Find the linear and quadratic Taylor polynomials for the function* $f(x, y) = \sqrt{2 + xy}$ *about the point* $(x_0, y_0) = (2, 1)$*. Use the polynomials to estimate* $f(3, 1.5)$*. Estimate the relative relative error of the constant, linear, and quadratic Taylor polynomial approximations. Compare the estimates with the exact relative error of these approximations.*

Solution: Put $x = 2 + dx$ and $y = 1 + dy$ where $dx = 1$ and $dy = 0.5$. To estimate the relative accuracy of the quadratic Taylor polynomial approximation, one has to calculate $d^3 f(2, 1)$. To avoid computing partial derivatives of $f(x, y)$ up to the third order, the following technical trick is used (compare with technicalities in Example **23.4**):

$$f(2 + dx, 1 + dy) = 2\sqrt{1 + du}\,, \quad du = \frac{1}{4}(dx + 2dy + dxdy)\,.$$

Since powers of du produce powers of dx and dy, the expansion of $(1+du)^{1/2}$ in powers du gives the needed expansion in powers of dx and dy. So, put $g(u) = \sqrt{u}$. Then $(1 + du)^{1/2} = g(1 + du)$ and

$$\begin{aligned} f(2 + dx, 1 + dy) &= f(2, 1) + df(2, 1) + \frac{1}{2}d^2 f(2, 1) + \frac{1}{6}d^3 f(2, 1) + \cdots \\ &= 2g(1 + du) \\ &= 2\Big(g(1) + dg(1) + \frac{1}{2}d^2 g(1) + \frac{1}{6}d^3 g(1) + \cdots\Big) \\ &= 2 + du - \frac{1}{4}(du)^2 + \frac{1}{8}(du)^3 + \cdots\,, \end{aligned}$$

where the dots denotes terms $(du)^4$ and higher. These terms can only give contributions to $d^4 f(2, 1)$ and higher as du is proportional to dx and dy. The differentials up to order 3 are obtained by comparing the first line of the above equalities with the last one written in powers of dx and dy:

$$\begin{aligned} (du)^2 &= \frac{1}{16}\Big((dx)^2 + 4dxdy + 4(dy)^2 + 2(dx)^2 dy + 4dx(dy)^2 + \cdots\Big) \\ (du)^3 &= \frac{1}{64}(dx + 2dy)^3 + \cdots \\ &= \frac{1}{64}\Big((dx)^3 + 6(dx)^2 dy + 12dx(dy)^2 + 8(dy)^3\Big) + \cdots\,, \end{aligned}$$

where again only terms $(dx)^k (dy)^m$ with $n = k + m \le 3$ (relevant for $d^n f$, $n \le 3$) were kept; the terms of higher orders are denoted by dots. Collecting terms with $n = 1, 2, 3$, the needed differentials of f are found:

$$\begin{aligned} df(2, 1) &= \frac{1}{4}dx + \frac{1}{2}dy\,, \\ \frac{1}{2}d^2 f(2, 1) &= -\frac{1}{64}(dx)^2 + \frac{3}{16}dxdy - \frac{1}{16}(dy)^2 \\ \frac{1}{6}d^3 f(2, 1) &= \frac{(dx)^3}{512} - \frac{5(dx)^2 dy}{256} - \frac{5dx(dy)^2}{128} + \frac{(dy)^3}{64} \end{aligned}$$

Therefore the linear and quadratic Taylor polynomials are

$$\begin{aligned} T_1(x,y) &= 2 + \frac{1}{4}(x-2) + \frac{1}{2}(y-1)\,, \\ T_2(x,y) &= T_1(x,y) - \frac{1}{64}(x-2)^2 + \frac{3}{16}(x-2)(y-1) - \frac{1}{16}(y-1)^2 \end{aligned}$$

The values of the differentials and the successive Taylor approximations (up to the cubic order) are

$$\begin{aligned} f(2+dx, 1+dy) &\approx f(2,1) + df(2,1) + \frac{1}{2}d^2 f(2,1) + \frac{1}{6}d^3 f(2,1) \\ &= 2 + 0.5 + 0.0625 - 0.015625 \\ &= 2.546875 \end{aligned}$$

The exact value of the function is $f(2+1, 1+0.5) = 2.549510$ (rounded to the same decimal place as the approximate value). The constant, linear, and quadratic Taylor approximations give, respectively, 2, 2.5, and 2.5625 for the value of $f(3, 1.5)$. The estimated and exact relative errors of these approximations rounded up to two significant digits are

$$\begin{aligned} &\epsilon_0^{es} = 25\%\,, \quad \epsilon_1^{es} = 1.3\%\,, \quad \epsilon_2^{es} = 0.61\% \\ &\epsilon_0 = 27\%\,, \quad \epsilon_1 = 1.9\%\,, \quad \epsilon_2 = 0.51\%\,, \quad \epsilon_3 = 0.10\% \end{aligned}$$

Here ϵ_3 is the relative error of the cubic Taylor polynomial approximation (the value of $d^4 f(2,1)$ would be needed for its estimation). One can see that the successive Taylor approximations improve the accuracy of the approximation and the estimates of the errors are close to the exact values. □

Example **23.6**. *Use the linear approximation or the differential to estimate the amount of aluminum in a closed aluminum can with diameter 10 cm and height 10 cm if the aluminum is 0.05 cm thick. Estimate the error of the approximation.*

Solution: The volume of a cylinder of radius r and height h is $f(h,r) = \pi h r^2$. The volume of a closed cylindrical shell (or the can) of thickness δ is therefore

$$V = f(h + 2\delta, r + \delta) - f(h,r)\,,$$

where h and r are the internal height and radius of the shell. Put $dh = 2\delta = 0.1$ and $dr = \delta = 0.05$. Then $V \approx df(10,5)$. One has $f'_h = \pi r^2$ and $f'_r = 2\pi h r$; hence,

$$\begin{aligned} V &\approx df(10,5) = f'_h(10,5)\,dh + f'_r(10,5)\,dr = 25\pi\,dh + 100\pi\,dr \\ &= 7.5\pi \text{ cm}^3\,. \end{aligned}$$

To estimate the error of the linear approximation, the value of the second differential is needed (f is a polynomial and therefore all its partial derivatives of any order are continuous). The second partial derivatives are

$f''_{hh} = 0$, $f''_{hr} = 2\pi r$, and $f''_{rr} = 2\pi h$. Therefore the absolute error of the approximation can be estimated by

$$\begin{aligned}\varepsilon_1^{es} &= \frac{1}{2}|d^2 f(10,5)| = \frac{1}{2}\Big(20\pi dhdr + 20\pi (dr)^2\Big)\\ &= 0.075\pi \text{ cm}^3\end{aligned}$$

The estimated relative error is

$$\epsilon_1^{es} = \frac{\varepsilon_1^{es}}{|df(10,5)|} \times 100\% = \frac{|d^2 f(10,5)|}{2|df(10,5)|} \times 100\% = 1\%\,.$$

□

With increasing the number of variables, calculation of higher-order differentials of a function to find Taylor polynomials might be a technically tedious problem. In some special cases, however, it can be avoided. The idea has already been used in Example **23.5** for a function of two variables. The concept is further elucidated by the following example of a function of three variables (see also Study Problem **23.2**).

EXAMPLE **23.7**. *Find T_3 for the function $f(x,y,z) = \sin(xy+z)$ about the origin.*

SOLUTION: The Taylor polynomial T_3 in question is a polynomial of degree 3 in x, y, and z, which is uniquely determined by the coefficients at monomials of degree less or equal 3. Put $u = xy + z$. The variable u is small near the origin. So the Taylor polynomial approximation for f near the origin is determined by the Taylor polynomials for $\sin u$ about $u = 0$. The latter is obtained from the Maclaurin series

$$\sin u = u - \frac{1}{6}u^3 + O(u^5)\,,$$

where $O(u^5)$ contains only monomials of degree 5 and higher. Since the polynomial $u = xy + z$ vanishes at the origin, its powers u^n may contain only monomials of degree n and higher. Therefore T_3 is obtained from

$$\begin{aligned}u - \tfrac{1}{6}u^3 &= (xy+z) - \tfrac{1}{6}(xy+z)^3\\ &= z + xy - \tfrac{1}{6}\Big(z^3 + 3(xy)z^2 + 3(xy)^2 z + (xy)^3\Big)\end{aligned}$$

by retaining in the latter all monomials up to degree 3, which yields

$$T_3(x,y,z) = z + xy - \frac{1}{6}z^3\,.$$

Evidently, the procedure is far simpler than calculating 19 partial derivatives (up to the third order)! □

23.7. Study Problems.

Problem **23.1**. *Find all Taylor polynomials for*

$$P_3(x,y) = 1 + 2x - xy + y^2 + 4x^3 - y^2x$$

about $(0,0)$. *Use this example to find all Taylor polynomials for a generic polynomial* $P_k(x,y)$ *of degree* k *about* $(0,0)$.

SOLUTION: All partial derivatives of P_3 of orders higher than 3 vanish identically so $d^nP_3 = 0$ for all $n > 3$. Therefore $T_n = T_3$ for all $n > 3$. By direct calculation of the differentials $d^nP_3(0,0)$ for $n = 1, 2, 3$:

$$\begin{aligned} T_0(x,y) &= 1\,, \\ T_1(x,y) &= 1 + 2x\,, \\ T_2(x,y) &= 1 + 2x - xy + y^2\,, \\ T_3(x,y) &= 1 + 2x - xy + y^2 + 4x^3 - y^2x = P_3(x,y)\,. \end{aligned}$$

Similarly, for a generic polynomial P_k, its Taylor polynomials about $(0,0)$ of degree higher than k coincide with the Taylor polynomial of degree k, $T_n = T_k$ for all $n > k$, because $d^nP_k = 0$ for $n > k$. Any polynomial is the sum

$$P_k = Q_0 + Q_1 + \cdots + Q_k\,,$$

where Q_j is a *homogeneous* polynomial of degree $j = 0, 1, ..., k$; it contains only monomials of degree j. For example, the given polynomial P_3 is

$$\begin{aligned} &P_3 = Q_0 + Q_1 + Q_2 + Q_3\,, \\ &Q_0 = 1\,,\ Q_1 = 2x\,,\ Q_2 = -xy + y^2\,,\ Q_3 = 4x^3 - y^2x\,. \end{aligned}$$

For $n \le k$, the considered example suggests that the following general solution

$$T_n = Q_0 + Q_1 + Q_2 + \cdots + Q_n\,, \quad n = 0, 1, 2, ..., k\,.$$

Let us prove that this is indeed so.

Partial derivatives of a homogeneous polynomial of degree j are homogeneous polynomials of degree $j - 1$, where $j \ge 1$. For example, $(Q_3)'_x = 12x^2 - y^2$ and $(Q_3)'_y = -2xy$. Consequently, second partial derivatives of a homogeneous polynomial of degree j are homogeneous polynomials of degree $j - 2$ if $j \ge 2$. Clearly, the similar conclusion holds for higher order derivatives up to order j and all partial derivatives of any order higher than j vanish identically. Since only a constant homogeneous polynomial does not vanish at the origin,

$$d^nQ_j(0,0) = 0 \quad \text{for all} \quad n \ne j\,.$$

Therefore for $n \le k$, the Taylor polynomial of P_k about the origin are

$$T_n = Q_0 + dQ_1(0,0) + \frac{1}{2!}d^2Q_2(0,0) + \cdots + \frac{1}{n!}d^nQ_n(0,0)\,,$$

where dx is set to x and dy to y after calculation of the differentials (because $x_0 = y_0 = 0$). So it remains to find $d^nQ_n(0,0)$ for $n \le k$. A homogeneous polynomial of degree n is the sum of monomials of degree n. Since $d^n(f+g) = d^n f + d^n g$, it is sufficient to find $d^n Q_n(0,0)$ for $Q_n = x^{n-p}y^p$ where $0 \le p \le n$ is an integer. Next note that

$$\frac{\partial^n}{\partial x^{n-q}\partial y^q} x^{n-p}y^p\Big|_{(x,y)=(0,0)} = (n-p)!p! \quad \text{if} \quad q = p$$

and vanishes otherwise because all the derivatives of y^p vanish at $y = 0$ except the p^{th} derivative which is constant and equal to $p!$ (similarly for x^{n-p}). Using the binomial expansion of d^n, it is concluded that

$$\begin{aligned} &Q_n(x,y) = x^{n-p}y^p \qquad \Rightarrow \\ &\frac{1}{n!}d^n Q_n(0,0) = B_p^n \frac{(n-p)!p!}{n!}(dx)^{n-p}(dy)^p = (dx)^{n-p}(dy)^p \,. \end{aligned}$$

Replacing dx and dy by $x - x_0 = x$ and $y - y_0 = y$, respectively, it follows that Taylor polynomials of a polynomial $P_k = Q_0 + Q_1 + \cdots + Q_k$, where Q_j, $j = 1, 2, ..., k$, are homogeneous polynomials,

$$T_n = Q_0 + Q_1 + \cdots + Q_n \,, \quad n = 0, 1, ..., k \,.$$

In particular, for the given polynomial P_3, its Taylor polynomials about the origin are $T_0 = 1$, $T_1 = T_0 - 2x$, $T_2 = T_1 - xy + y^2$, and $T_k = P_3$ for $k \ge 3$. □

Problem **23.2**. *Find T_1, T_2 and T_3 for $f(x,y,z) = (1+xy)/(1+x+y^2+z^3)$ about the origin.*

Solution: The function f is a rational function. It is therefore sufficient to find a suitable Taylor polynomial for the function $(1+x+y^2+z^3)^{-1}$ and then multiply it by the polynomial $1 + xy$, retaining only monomials up to the degree 3. Put $u = x + y^2 + z^3$. Then

$$(1+u)^{-1} = 1 - u + u^2 - u^3 + O(u^4)$$

(as a geometric series), where $O(u^4)$ contains monomials of degree 4 and higher. Note that, for $n \ge 4$, the terms u^n contain only monomials of degree 4 and higher in variables x, y, and z and, hence, can be omitted. Up to the degree 3, one has $u^2 = x^2 + 2xy^2 + \cdots$ and $u^3 = x^3 + \cdots$. Therefore,

$$(1+xy)(1-u+u^2-u^3) = (1+xy)(1-x-y^2-z^3+x^2+2xy^2-x^3+\cdots)$$

and carrying out the multiplication and arranging the monomials in the order of increasing degrees one infers:

$$\begin{aligned} T_1(x,y,z) &= 1 - x \,, \\ T_2(x,y,z) &= T_1(x,y,z) + x^2 + xy - y^2 \,, \\ T_3(x,y,z) &= T_2(x,y,z) - x^3 - x^2y + 2xy^2 - z^3 \,. \end{aligned}$$

□

Problem **23.3**. (Multivariable Taylor and Maclaurin series)
Suppose that a function f has continuous partial derivatives of any order and the reminder in the Taylor polynomial approximation $f = T_{n-1} + \varepsilon_n$ near $\mathbf{r}_0$ converges to zero as $n \to \infty$, i.e., $\varepsilon_n \to 0$. Then the function can be represented by the Taylor series about a point $\mathbf{r}_0$:

$$f(\mathbf{r}) = f(\mathbf{r}_0) + \sum_{n=1}^{\infty} \frac{1}{n!}\, d^n f(\mathbf{r}_0)$$

where $d\mathbf{r}$ is set to $\mathbf{r} - \mathbf{r}_0$ after calculating the differentials. The Taylor series about $\mathbf{r}_0 = \mathbf{0}$ is called the Maclaurin series. Find the Maclaurin series of $f(x, y) = \sin(xy^2)$ and the set in which it converges.

SOLUTION: Since the argument of the sine is the polynomial xy^2, the Maclaurin series of f can be obtained from the Maclaurin series of $\sin u$ by setting in it $u = xy^2$. From Calculus II,

$$f(\mathbf{r}) = \sin u = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(2n-1)!}\, u^{2n-1} = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(2n-1)!}\, x^{2n-1} y^{4n-2}\,.$$

Since the Maclaurin series for $\sin u$ converges for all u, the above series converges for all (x, y). □

23.8. Exercises.

1–7. Find the differential df of each of the following functions:

1. $f(x, y) = x^3 + y^3 - 3xy(x - y)$;
2. $f(x, y) = y\cos(x^2 y)$;
3. $f(x, y) = \sin(x^2 + y^2)$;
4. $f(x, y, z) = x + yz + ye^{xyz}$;
5. $f(x, y, z) = \ln(x^x y^y z^z)$;
6. $f(x, y, z) = y/(1 + xyz)$;
7. $f(\mathbf{r}) = \sqrt{a^2 - \|\mathbf{r}\|^2}$ where a is a constant and $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$.

8. Four positive numbers, each less than 100, are rounded and then multiplied together. Use the differential to estimate the maximum possible error in the computed product that might result from the rounding.

9. A boundary stripe 10 cm wide is painted around a rectangle whose dimensions are 50 m by 100 m. Use differentials to approximate the number of square meters of paint in the stripe. Use the second differential to estimate the error bound of the approximation.

10. A rectangle has sides of $x = 6$ and $y = 8$ meters. Use differentials to estimate the change of the length of the diagonal and the area of the rectangle if x is increased by 2 cm and y is decreased by 5 cm. Use the second differential to estimate the error of the approximation.

11. Consider a sector of a disk with radius $R = 20$ cm and the angle $\theta = \pi/3$. Use the differential to determine how much should the radius be decreased in order for the area of the sector to remain the same when the angle is increased by $1°$. Use the second differential to estimate the error of the approximation.

12. Let the quantities f and g be measured with relative errors R_f and R_g. Show that the relative error the product fg is the sum $R_f + R_g$.

13. Measurements of the radius r and the height h of a cylinder are $r = 2.2 \pm 0.1$ and $h = 3.1 \pm 0.2$ in meters. Find the absolute and relative errors of the volume of the cylinder calculated from these data.

14. The adjacent sides of a triangle have the lengths $a = 100 \pm 2$ and $b = 200 \pm 5$ in meters, and the angle between them is $\theta = 60° \pm 1°$. Find the relative and absolute errors in calculation of the length of the third side of the triangle.

15. If R is the total resistance of n resistors, connected in parallel, with resistances R_j, $j = 1, 2, ..., n$, then $R^{-1} = R_1^{-1} + R_2^{-1} + \cdots R_n^{-1}$. If each resistance R_j is known with a relative error of 0.5%, what is the relative error of R?

16-18. Use the Taylor theorem and its consequences to find the upper bound of the absolute error of the linear approximation of each of the following functions about the origin in the ball of radius R (i.e. for $\|\mathbf{r}\| \leq R$):

16. $f(x, y) = \sqrt{1 + \sin(x + y)}, \quad R = \frac{1}{2}\,;$

17. $f(x, y) = \frac{1+3x}{2+y}, \quad R = 1\,;$

18. $f(x, y, z) = \ln(1 + x + 2y - 3z), \quad R = 0.1\,.$

19-24. Find the indicated differentials of a given function:

19. $f(x, y) = x - y + x^2 y\,, \quad d^n f\,, \quad n = 1, 2, ...;$

20. $f(x, y) = \ln(x + y)\,, \quad d^n f\,, \quad n = 1, 2, ...;$

21. $f(x, y) = \sin(x)\cosh(y)\,, \quad d^3 f\,;$

22. $f(x, y, z) = xyz\,, \quad d^n f\,, \quad n = 1, 2, ...;$

23. $f(x, y, z) = 1/(1 + xyz)\,, \quad d^2 f\,;$

24. $f(\mathbf{r}) = \|\mathbf{r}\|\,,$ df and $d^2 f$, where $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle\,.$

25. Let $f(\mathbf{r}) = g(u)$ where u is a linear function of $\mathbf{r}$, $u = c + \mathbf{n}\cdot\mathbf{r}$ where c is a constant and $\mathbf{n}$ is a constant vector. Show that $d^n f = g^{(n)}(u)(\mathbf{n}\cdot d\mathbf{r})^n$.

26. Let $Q_n(x, y, z)$ be a homogeneous polynomial of degree n (it contains only monomials of degree n). Show that $d^n Q_n(x, y, z) = n! Q_n(dx, dy, dz)$.

27-29. Find the Taylor polynomial T_2 about a specified point P_0 and a given function:

27. $f(x, y) = y + x^3 + 2xy^2 - x^2 y^2, \quad P_0 = (1, 1)\,;$

28. $f(x, y) = \sin(xy), \quad P_0 = (\pi/2, 1)\,;$

29. $f(x, y) = x^y, \quad P_0 = (1, 1)\,.$

30. Let $f(x, y) = x^y$. Use Taylor polynomials T_1, T_2, and T_3 about the point $(1, 1)$ to approximate $f(1.2, 0.7)$. Compare the results of the three approximations with the exact value of $f(1.2, 0.7)$ by computing the relative errors of the approximations (use a calculator value rounded to an appropriate number of decimal places).

31-35. Use the method of Example **23.7** and Study Problem **23.2** to find the indicated Taylor polynomials about the origin for each of the following functions:

31. $f(x, y) = \sqrt{1 + x + 2y}$, $\quad T_n(x, y)$, $n \le 2$;
32. $f(x, y) = \frac{xy}{1-x^2-y^2}$, $\quad T_n(x, y)$, $n \le 4$;
33. $f(x, y, z) = \sin(x + 2y + z^2)$, $\quad T_n(x, y, z)$, $n \le 3$;
34. $f(x, y, z) = e^{xy}\cos(zy)$, $\quad T_n(x, y, z)$, $n \le 4$;
35. $f(x, y) = \ln(1 + x + 2y)/(1 + x^2 + y^2)$, $\quad T_n(x, y)$, $n \le 2$.

36-37. Find polynomials of degree 2 to calculate approximate values of the following functions in a region in which $x^2 + y^2$ is small as compared with 1:

36. $f(x, y) = \cos y/\cos x$;
37. $\tan^{-1}\left(\frac{1+x+y}{1-x+y}\right)$.

38. Find a non-zero polynomial of the smallest degree to approximate a local behavior of the function $\cos(x + y + z) - \cos(x)\cos(y)\cos(z)$ near the origin.

39. Let

$$g(r) = \frac{1}{2\pi}\int_0^{2\pi} f(x_0 + r\cos\theta, y_0 + r\sin\theta)d\theta$$

where f has continuous partial derivatives up to order 4 and x_0 and y_0 are constants. Find T_4 about $r = 0$ for $g(r)$.

40. Consider the roots $z = z(x, y)$ of the equation $F(x, y, z) = z^5 + xz - y = 0$ near $(1, 2, 1)$. Use Taylor polynomials $T_1(x, y)$ and $T_2(x, y)$ about $(1, 2)$ to approximate $z(x, y)$. In particular, calculate the approximations $z_1 = T_1(0.7, 2.5)$ and $z_2 = T_2(0.7, 2.5)$ of $z(0.7, 2.5)$. Use a calculator to find the values $F(0.7, 2.5, z_1)$ and $F(0.7, 2.5, z_2)$. Their deviation from 0 determines an error of the approximations z_1 and z_2. Which of the approximations is more accurate? Hint: Use the result of Study Problem **22.1**.

24. Directional Derivative and the Gradient

24.1. Directional Derivative. Let f be a function of several variables $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$. The partial derivative $f'_{x_i}(\mathbf{r}_0)$ is the rate of change of f at a point $\mathbf{r}_0$ in the direction of the i^{th} coordinate axis. This direction is defined by the unit vector $\hat{\mathbf{e}}_i$ parallel to the corresponding coordinate axis. Let $\hat{\mathbf{u}}$ be a unit vector that does not coincide with any of the vectors $\hat{\mathbf{e}}_i$. What is the rate of change of f at $\mathbf{r}_0$ in the direction of $\hat{\mathbf{u}}$? For example, if $f(x, y)$ is the height of a mountain, where the x and y axes are oriented along the west-to-east and south-to-north directions, respectively, then it is reasonable to ask about the slopes, for example, in the southeast or northwest directions. Naturally, these slopes generally differ from the slopes f'_x and f'_y.

To answer the question about the slope in the direction of a unit vector $\hat{\mathbf{u}}$, consider a straight line through $\mathbf{r}_0$ parallel to $\hat{\mathbf{u}}$. Its vector equation is

$$\mathbf{r}(h) = \mathbf{r}_0 + h\hat{\mathbf{u}}\,,$$

where h is a parameter that labels points of the line. The values of f along the line are given by the composition

$$F(h) = f(\mathbf{r}(h))\,.$$

The numbers $F(0)$ and $F(h)$ are the values of f at a given point $\mathbf{r}_0$ and the point $\mathbf{r}(h)$, $h \neq 0$, that is at the distance $|h|$ from $\mathbf{r}_0$ along the line. So the slope is given by the derivative

$$F'(0) = \lim_{h\to 0} \frac{F(h) - F(0)}{h}\,.$$

Therefore, the following definition is natural.

Definition **24.1**. (Directional Derivative).
Let f be a function on an open set D. The directional derivative of f at point $\mathbf{r}_0$ in D in the direction of a unit vector $\hat{\mathbf{u}}$ is the limit

$$D_{\mathbf{u}}f(\mathbf{r}_0) = \lim_{h\to 0} \frac{f(\mathbf{r}_0 + h\hat{\mathbf{u}}) - f(\mathbf{r}_0)}{h}$$

if the limit exists.

The number $D_{\mathbf{u}}f(\mathbf{r}_0)$ is the rate of change of f at $\mathbf{r}_0$ in the direction of $\hat{\mathbf{u}}$. A geometrical significance of the directional derivative in the case of two-variable function is illustrated in Fig. **24.1**. Suppose that f is a *differentiable* function. By definition,

$$D_{\mathbf{u}}f(\mathbf{r}_0) = \frac{d}{dh}f(\mathbf{r}(h))\Big|_{h=0}\,, \qquad \mathbf{r}(h) = \mathbf{r}_0 + h\hat{\mathbf{u}}\,.$$

Since f is a differentiable function, the chain rule applies to compute the derivative. In the case of functions of two variables, the parametric equations of the line through (x_0, y_0) and parallel to a unit vector $\hat{\mathbf{u}} = \langle u_1, u_2\rangle$ are

$$x(h) = x_0 + hu_1\,, \quad y(h) = y_0 + hu_2\,.$$

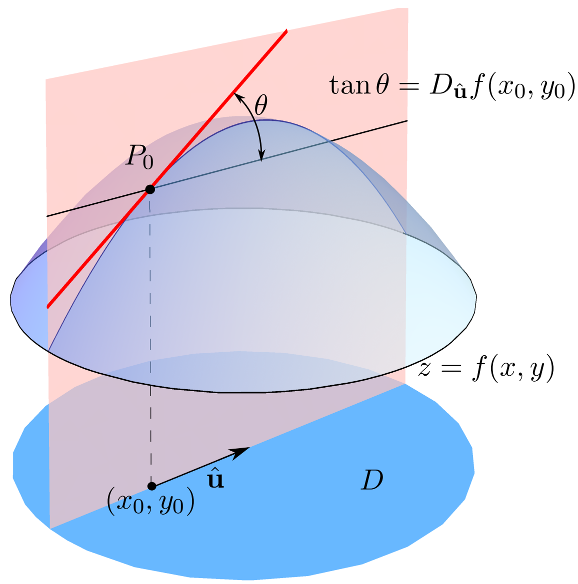


FIGURE **24.1**. Geometrical significance of the directional derivative of a function of two variables. Consider the graph $z = f(x, y)$ over a region D. The vertical plane (i.e. parallel to the z axis) through (x_0, y_0) and parallel to the unit vector $\hat{\mathbf{u}}$ intersects the graph along a curve. The slope of the tangent line to the curve of intersection at the point $P_0 = (x_0, y_0, f(x_0, y_0))$ is determined by the directional derivative $D_{\hat{\mathbf{u}}} f(x_0, y_0)$. So the directional derivative defines the rate of change of f at (x_0, y_0) in the direction $\hat{\mathbf{u}}$.

By the chain rule

$$\begin{aligned}\frac{d}{dh} f(x(h), y(h)) &= f'_x(x(h), y(h))x'(h) + f'_y(x(h), y(h))y'(h) \\ &= f'(x(h), y(h))u_1 + f'_y(x(h), y(h))u_2 \,.\end{aligned}$$

Setting $h = 0$ in this relation

$$D_{\mathbf{u}} f(x_0, y_0) = f'_x(x_0, y_0)u_1 + f'_y(x_0, y_0)u_2 \,.$$

Similarly, for any number of variables, one has

$$\frac{df(\mathbf{r}(h))}{dh} = f'_{x_1}(\mathbf{r}(h))x'_1(h) + f'_{x_2}(\mathbf{r}(h))x'_2(h) + \cdots + f'_{x_m}(\mathbf{r}(h))x'_m(h) \,.$$

Setting $h = 0$ in this relation and taking into account that $\mathbf{r}'(h) = \hat{\mathbf{u}}$ or $x'_i(h) = u_i$, where $\hat{\mathbf{u}} = \langle u_1, u_2, ..., u_m \rangle$, the following result is proved.

THEOREM **24.1**. (Directional derivative)
Suppose f is a differentiable function of several variable $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$ on an open set D. Then the directional derivative of f at a point $\mathbf{r}_0$ in D in the direction of unit vector $\hat{\mathbf{u}} = \langle u_1, u_2, ..., u_m\rangle$ is

$$D_{\mathbf{u}} f(\mathbf{r}_0) = f'_{x_1}(\mathbf{r}_0)u_1 + f'_{x_2}(\mathbf{r}_0)u_2 + \cdots + f'_{x_m}(\mathbf{r}_0)u_m\,. \tag{24.1}$$

It is important to emphasize that:

If f has partial derivatives at $\mathbf{r}_0$, but is not differentiable at $\mathbf{r}_0$, then the relation (**24.1**) *is false.*

An example is given in Study Problem **24.1**. Note that Eq. (**24.1**) follows from the chain rule, but the mere existence of partial derivatives is *not* sufficient for the chain rule to hold. Furthermore, Study Problem **24.1** also illustrates the following fact:

Even if a function has directional derivatives at a point in every direction, it may not be differentiable at that point (no good linear approximation exists at that point).

Equation (**24.1**) provides a convenient way to compute the directional derivative if f is differentiable. In turn, a practical way to check differentiability is provided by Theorem **21.3**:

Equation (**24.1**) *holds if the function f has partial derivatives in a neighborhood of $\mathbf{r}_0$ that are continuous at $\mathbf{r}_0$.*

The existence and continuity of partial derivatives are often easy to establish by analyzing an explicit form of the function in question (in this regard, recall the discussion of Clairaut's Theorem **20.1** in Section **20**). If the direction in which the rate of change of a function to be determined is specified by a nonunit vector $\mathbf{u}$, then the corresponding unit vector can be obtained by dividing it by its length $\|\mathbf{u}\|$, that is, $\hat{\mathbf{u}} = \mathbf{u}/\|\mathbf{u}\|$.

EXAMPLE **24.1**. *The height of a hill is $f(x, y) = (9 - 3x^2 - y^2)^{1/2}$, where the x and y axes in the base of the hill are directed from west to east and from south to north, respectively. A hiker is on the hill at a point corresponding to the point $\mathbf{r}_0 = \langle 1, 2\rangle$ in the base. Suppose the hiker is facing in the northwest direction. What is the slope the hiker sees?*

SOLUTION: A unit vector in the plane can always be written in the form $\hat{\mathbf{u}} = \langle\cos\varphi, \sin\varphi\rangle$, where the angle φ is counted counterclockwise from the positive x axis; that is, $\varphi = 0$ corresponds to the east direction, $\varphi = \pi/2$ to the north direction, $\varphi = \pi$ to the west direction, and so on. So for the northwest direction $\varphi = 3\pi/4$ and $\hat{\mathbf{u}} = \langle -1/\sqrt{2}, 1/\sqrt{2}\rangle = \langle u_1, u_2\rangle$. The partial derivatives,

$$f'_x(x,y) = -\frac{3x}{(9-3x^2-y^2)^{1/2}}\,,\quad f'_y(x,y) = -\frac{y}{(9-3x^2-y^2)^{1/2}}\,,$$

are continuous functions near the point $(1,2)$ (as ratios of continuous functions) and hence f is differentiable at $(1,2)$. Since $f'_x(1,2) = -3/\sqrt{2}$ and $f'_y(1,2) = -2/\sqrt{2}$, by (**24.1**), the slope is

$$D_{\mathbf{u}}f(\mathbf{r}_0) = f'_x(1,2)u_1 + f'_y(1,2)u_2 = 3/2 - 1 = 1/2\,.$$

If the hiker goes northwest, he has to climb up one unit of length per two units of length forward. □

Example **24.2**. *Find the directional derivative of* $f(x,y,z) = x^2+3xz+z^2y$ *at the point* $(1,1,-1)$ *in the direction toward the point* $(3,-1,0)$. *Does the function increase or decrease in this direction?*

Solution: Put $\mathbf{r}_0 = \langle 1,1,-1\rangle$ and $\mathbf{r}_1 = \langle 3,-1,0\rangle$. Then the vector $\mathbf{u} = \mathbf{r}_1 - \mathbf{r}_0 = \langle 2,-2,1\rangle$ points from the point $\mathbf{r}_0$ toward the point $\mathbf{r}_1$ according to the rules of vector algebra. But it is not a unit vector because its length is $\|\mathbf{u}\| = 3$. So the unit vector in the same direction is

$$\hat{\mathbf{u}} = \tfrac{1}{3}\mathbf{u} = \langle \tfrac{2}{3}, -\tfrac{2}{3}, \tfrac{1}{3}\rangle = \langle u_1, u_2, u_3\rangle\,.$$

The partial derivatives,

$$f'_x = 2x+3z\,,\quad f'_y = z^2\,,\quad f'_z = 3x+2zy\,,$$

are polynomials and hence continuous everywhere so that f is differentiable everywhere. Since $f'_x(\mathbf{r}_0) = -1$, $f'_y(\mathbf{r}_0) = 1$, and $f'_z(\mathbf{r}_0) = 1$, by Eq. (**24.1**), the directional derivative is

$$D_{\mathbf{u}}f(\mathbf{r}_0) = f'_x(\mathbf{r}_0)u_1 + f'_y(\mathbf{r}_0)u_2 + f'_z(\mathbf{r}_0)u_3 = -2/3 - 2/3 + 1/3 = -1\,.$$

The directional derivative is negative and therefore the function decreases at $\mathbf{r}_0$ in the direction toward $\mathbf{r}_1$ (the rate of change is negative in that direction). □

24.2. Level sets of a differentiable function. Level sets of a function f of two variables (x,y) are defined by the equation $f(x,y) = k$. The following question is of interest. *Under what conditions on the function f is the level set a curve?* The implicit function theorem (Theorem **22.2**) allows us to answer this question.

Let (x_0,y_0) be a particular point in the considered level set, that is, $f(x_0,y_0) = k$. Suppose that f and f'_y are continuous in an open disk D centered at (x_0,y_0) and $f'_y(x_0,y_0) \neq 0$. Then by Theorem **22.2** (where $F(x,z) = f(x,z) - k$, $z = y$, and $\mathbf{r}$ is a one-dimensional vector whose single component is x) there exists an open interval $D = (a,b)$ that contains x_0 (a neighborhood of x_0) in which the equation $F(x,y) = f(x,y) - k = 0$ has a unique solution $y = y(x)$ where $y(x)$ is a continuous function. Therefore the level set is the graph $y = y(x)$ of a continuous function in a neighborhood of (x_0, y_0) or a simple curve. Furthermore, if f is differentiable, then $y(x)$ is differentiable and the level curve is smooth. The level curve is traversed by the vector function $\mathbf{r}(t) = \langle t, y(t)\rangle$ and $\mathbf{r}'(t) = \langle 1, y'(t)\rangle \neq \mathbf{0}$ (the unit tangent vector to the graph exists and is continuous). If $f'_y(x_0, y_0) = 0$ but $f'_x(x_0, y_0) \neq 0$, then the equation $f(x, y) - k = 0$ can be solved for $x = x(y)$ where $x(y)$ is continuous in an open interval containing y_0. The level set near (x_0, y_0) is again a graph of a continuous function, that is, a curve.

Suppose now that f is a differentiable function and has partial derivatives that do not vanish simultaneously. Consider a level curve through (x_0, y_0). The curve is traversed by the vector function $\mathbf{r}_1(t) = \langle t, y(t)\rangle$ if $f'_y(x_0, y_0) \neq 0$ or by $\mathbf{r}_2(t) = \langle x(t), t\rangle$ if $f'_x(x_0, y_0) \neq 0$. The vectors $\mathbf{T}_1 = \mathbf{r}'_1(x_0) = \langle 1, y'(x_0)\rangle$ and $\mathbf{T}_2 = \mathbf{r}'_2(y_0) = \langle x'(y_0), 1\rangle$ are tangent to the curve at the point (x_0, y_0). By the implicit differentiation formula,

$$\mathbf{T}_1 = \langle 1,\ y'(x_0)\rangle\,, \quad y'(x_0) = -\frac{f'_x(x_0, y_0)}{f'_y(x_0, y_0)} \quad \text{if} \quad f'_y(x_0, y_0) \neq 0$$

$$\mathbf{T}_2 = \langle x'(y_0),\ 1\rangle\,, \quad x'(y_0) = -\frac{f'_y(x_0, y_0)}{f'_x(x_0, y_0)} \quad \text{if} \quad f'_x(x_0, y_0) \neq 0\,.$$

In either case, the tangent vectors are orthogonal to the nonzero vector

$$\mathbf{n} = \langle f'_x(x_0, y_0)\,,\ f'_y(x_0, y_0)\rangle$$

whose components are the corresponding partial derivatives of f at (x_0, y_0) because

$$\begin{aligned} \mathbf{n}\cdot\mathbf{T}_1 &= f'_x(x_0, y_0) - f'_x(x_0, y_0) = 0\,, \\ \mathbf{n}\cdot\mathbf{T}_2 &= -f'_y(x_0, y_0) + f'_y(x_0, y_0) = 0\,. \end{aligned}$$

Note also that the *nonzero* vector $\mathbf{T} = \langle -f'_y(x_0, y_0), f'_x(x_0, y_0)\rangle$ is tangent to the level curve bacause it is orthogonal to $\mathbf{n}$:

$$\mathbf{n}\cdot\mathbf{T} = -f_x(x_0, y_0)f'_y(x_0, y_0) + f_y(x_0, y_0)f'_x(x_0, y_0) = 0\,.$$

If the partial derivatives of f are continuous, then the components of the tangent vector $\mathbf{T} \neq \mathbf{0}$ are continuous functions and, hence, the level curve is smooth. Therefore the following consequence of the implicit function theorem holds.

Corollary **24.1**. *Suppose that a function f of two variables (x, y) has continuous partial derivatives in an open disk D that do not vanish simultaneously anywhere in D. Then nonempty level sets $f(x, y) = k$ are smooth curves in D.*

Now consider a level set of a function of three variables $f(x, y, z) = k$ and a point (x_0, y_0, z_0) in it. If f and f'_z are continuous in an open ball B centered at (x_0, y_0, z_0), then by Theorem **22.2** (where $F = f - k$ and $\mathbf{r} = \langle x, y\rangle$) the equation $F(x, y, z) = f(x, y, z) - k = 0$ has a unique solution $z = z(x, y)$ which is a continuous function in an open disk D centered at (x_0, y_0). Thus, the level set is the graph $z = z(x, y)$ of a continuous function of two variable, that is, it is a surface (it is obtained by a continuous deformation of an open disk in a plane). Furthermore, if f is differentiable in B, then the function $z(x, y)$ is also differentiable in D. In particular, it has the tangent plane at each point. By the implicit differentiation formula the partial derivatives of $z(x, y)$ at a point (x_0, y_0, z_0) of the level surface are

$$z'_x(x_0, y_0) = -\frac{f'_x(x_0, y_0, z_0)}{f'_z(x_0, y_0, z_0)}\,, \quad z'_y(x_0, y_0) = -\frac{f'_y(x_0, y_0, z_0)}{f'_z(x_0, y_0, z_0)}\,.$$

Therefore a normal of the tangent plane to the graph $z = z(x, y)$ at a point (x_0, y_0, z_0) is

$$\begin{aligned} &\langle -z'_x(x_0, y_0),\ -z'_y(x_0, y_0),\ 1\rangle = \frac{1}{f'_z(x_0, y_0, z_0)}\,\mathbf{n} \ \text{ if } \ f'_z(x_0, y_0, z_0) \neq 0\,, \\ &\mathbf{n} = \langle f'_x(x_0, y_0, z_0),\ f'_y(x_0, y_0, z_0),\ f'_z(x_0, y_0, z_0)\rangle\,; \end{aligned}$$

it is parallel to the nonzero vector $\mathbf{n}$ whose components are the corresponding partial derivatives at (x_0, y_0, z_0).

Just like in the case of functions of two variables, the implicit function theorem can be applied to show that a level set of a differentiable function of three variables is a surface in a neighborhood of each point of the level set if its partial derivatives do not vanish simultaneously anywhere. If $f'_z = 0$ at a point P_0, then the level set near P_0 is the graph of a differentiable function, $x = x(y, z)$ if $f'_x(P_0) \neq 0$ or $y = y(x, z)$ if $f'_y(P_0) \neq 0$. The implicit differentiation formula yields that in each case the vector $\mathbf{n}$ is a normal to the tangent plane to the level surface. The details similar to the case $f'_z(P_0) \neq 0$ are left to the reader as an exercise. In all cases, a normal vector to a level surface at a point P_0 is parallel to the vector whose components are the corresponding partial derivatives at P_0 of the function. Therefore, if the partial derivatives are continuous functions that do not vanish simultaneously, then the normal $\mathbf{n} \neq \mathbf{0}$ has continuous components along the surface. Such (level) surfaces are called *smooth*. So, the following consequence of the implicit function theorem holds.

Corollary **24.2**. (Level surfaces)
Suppose that a function f of three variables (x, y, z) has continuous partial derivatives in an open ball B that do not vanish simultaneously anywhere in B. Then nonempty level sets $f(x, y, z) = k$ are smooth surfaces in B.

24.3. The Gradient and Its Geometrical Significance. The vector whose components are the corresponding partial derivatives of a differentiable function has been shown to provide an important information about the properties of the function. It is therefore convenient to give it a name.

Definition **24.2**. (The Gradient).
Let f be a differentiable function of several variables $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$ on an open set D and let $\mathbf{r}_0$ be a point in D. The vector whose components are

partial derivatives of f at $\mathbf{r}_0$,

$$\boldsymbol{\nabla} f(\mathbf{r}_0) = \langle f'_{x_1}(\mathbf{r}_0),\ f'_{x_2}(\mathbf{r}_0),\ \dots\ ,\ f'_{x_m}(\mathbf{r}_0)\rangle\,,$$

is called the gradient *of f at the point $\mathbf{r}_0$.*

A boldface font is used for the gradient symbol $\boldsymbol{\nabla}$ to emphasize that the gradient is a *vector*. So, for two-variable functions, the gradient is a two-dimensional vector:

$$f(x, y): \qquad \boldsymbol{\nabla} f = \langle f'_x,\ f'_y\rangle\,;$$

for three-variable functions, the gradient is a three-dimensional vector;

$$f(x, y, z): \qquad \boldsymbol{\nabla} f = \langle f'_x, f'_y, f'_z\rangle\,;$$

and so on. Comparing (**24.1**) with the definition of the gradient and recalling the definition of the dot product, the directional derivative can now be written in the compact form

$$D_{\mathbf{u}} f(\mathbf{r}_0) = \boldsymbol{\nabla} f(\mathbf{r}_0)\cdot \hat{\mathbf{u}}\,. \tag{24.2}$$

This equation is the most suitable for analyzing the significance of the gradient.

Consider first the cases of two- and three-variable functions. The gradient is either a vector in a plane or space, respectively. In Example **24.1**, the gradient at the point $(1, 2)$ is

$$f(x, y) = (9 - 3x^2 - y^2)^{1/2}: \qquad \boldsymbol{\nabla} f(1, 2) = \langle -3/\sqrt{2}, -2/\sqrt{2}\rangle\,.$$

In Example **24.2**, the gradient at the point $(1, 1, -1)$ is

$$f(x, y, z) = x^2 + 3xz + z^2 y: \quad \boldsymbol{\nabla} f(1, 1, -1) = \langle -1, 1, 1\rangle\,.$$

Recall the geometrical property of the dot product $\mathbf{a}\cdot\mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\|\cos\theta$, where $\theta \in [0, \pi]$ is the angle between the nonzero vectors $\mathbf{a}$ and $\mathbf{b}$. The value $\theta = 0$ corresponds to parallel vectors $\mathbf{a}$ and $\mathbf{b}$. When $\theta = \pi/2$, the vectors are orthogonal. The vectors point in the opposite directions if $\theta = \pi$ (antiparallel vectors). Assume that $\boldsymbol{\nabla} f(\mathbf{r}_0) \neq \mathbf{0}$. Let θ be the angle between the gradient $\boldsymbol{\nabla} f(\mathbf{r}_0)$ and the unit vector $\hat{\mathbf{u}}$. Then

$$D_{\mathbf{u}} f(\mathbf{r}_0) = \boldsymbol{\nabla} f(\mathbf{r}_0)\cdot\hat{\mathbf{u}} = \|\boldsymbol{\nabla} f(\mathbf{r}_0)\|\|\hat{\mathbf{u}}\|\cos\theta = \|\boldsymbol{\nabla} f(\mathbf{r}_0)\|\cos\theta \tag{24.3}$$

because $\|\hat{\mathbf{u}}\| = 1$ (the unit vector). As the components of the gradient are fixed numbers (the values of the partial derivatives at a particular point $\mathbf{r}_0$), the directional derivative at $\mathbf{r}_0$ varies only if the vector $\hat{\mathbf{u}}$ changes. Thus,

the rates of change of a differentiable function f at a point in all directions that have the same angle θ with the gradient at that point are the same.

In the two-variable case, only two such directions are possible if $\hat{\mathbf{u}}$ is not parallel to the gradient, while in the three-variable case the rays from $\mathbf{r}_0$ in all such directions form a cone whose axis is along the gradient as depicted in the left and right panels of Fig. **24.3**, respectively. It is then concluded

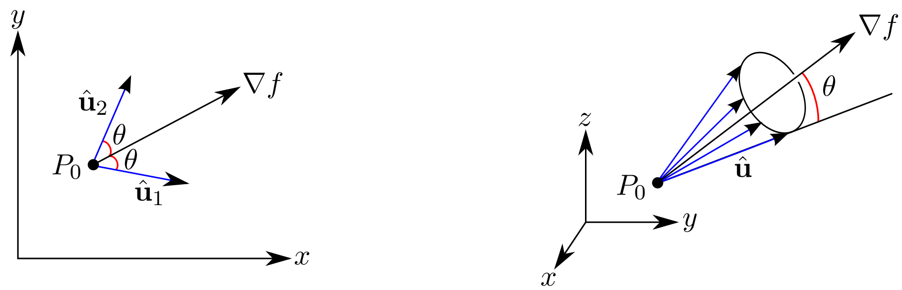


FIGURE **24.2**. **Left**: The same rate of change of a differentiable function of two variables at a point P_0 occurs in two directions that have the same angle with the gradient $\boldsymbol{\nabla} f(P_0)$.
Right: The same rate of change of a differentiable function of three variables at a point P_0 occurs in infinitely many directions that have the same angle with the gradient $\boldsymbol{\nabla} f(P_0)$ (they form a circular cone about the gradient).

that the rate of change of f attains its absolute maximum or minimum when $\cos\theta$ does. Therefore, *the maximal rate is attained in the direction of the gradient* ($\theta = 0$) *and is equal to the magnitude of the gradient* $\|\boldsymbol{\nabla} f(\mathbf{r}_0)\|$, *whereas the minimal rate of change* $-\|\boldsymbol{\nabla} f(\mathbf{r}_0)\|$ *occurs in the direction of* $-\boldsymbol{\nabla} f(\mathbf{r}_0)$, *that is, opposite to the gradient* ($\theta = \pi$).

The graph of a function of two variables $z = f(x, y)$ may be viewed as the shape of a hill. Then the gradient at a particular point shows the direction of the *steepest ascent*, while its opposite points in the direction of the *steepest descent*. In Example **24.1**, the maximal slope at the point $(1, 2)$ is

$$\|\boldsymbol{\nabla} f(\mathbf{r}_0)\| = (1/\sqrt{2})\|\langle -3, -2\rangle\| = \sqrt{13/2}\,.$$

It occurs in the direction of $\langle -3/\sqrt{2}, -2/\sqrt{2}\rangle$ or $\langle -3, -2\rangle$ (the multiplication of a vector by a positive constant does not change its direction). The unit vector in the direction of the gradient corresponds to $\varphi \approx 214°$ (somewhat in between west and southwest):

$$\langle \cos\varphi, \sin\varphi\rangle = \frac{1}{\|\boldsymbol{\nabla} f(\mathbf{r}_0)\|}\boldsymbol{\nabla} f(\mathbf{r}_0) = \frac{1}{\sqrt{13}}\langle -3, -2\rangle \quad \Rightarrow \quad \varphi \approx 214°\,.$$

If the hiker goes in this direction (in the direction of the steepest ascent), he has to climb up at an angle of

$$\tan^{-1}(\|\boldsymbol{\nabla} f(\mathbf{r}_0)\|) = \tan^{-1}(\sqrt{13/2}) \approx 69°$$

with the horizon. The hiker's original direction was $\varphi = 135°$, which makes the angle $79°$ with the direction of the steepest ascent. In this direction, he is climbing up at an angle of

$$\tan^{-1}(D_{\mathbf{u}} f(\mathbf{r}_0)) = \tan^{-1}(1/2) \approx 27°$$

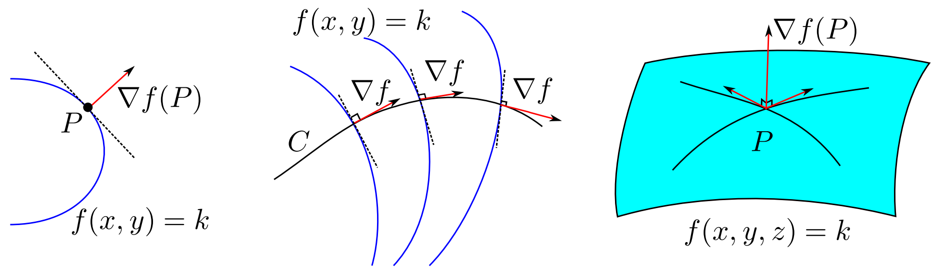


FIGURE **24.3**. **Left**: The gradient at a point P is normal to a level curve $f(x,y)=k$ through P of a function f of two variables. **Middle**: A curve C of steepest descent or ascent for a function f has the characteristic property that the gradient $\boldsymbol{\nabla} f$ is tangent to it. The level curves of f are normal C. The function f increases most rapidly along C in the direction of $\boldsymbol{\nabla} f$ and f decreases most rapidly along C in the opposite direction $-\boldsymbol{\nabla} f$. **Right**: The gradient of a function of three variables is normal to any curve through P in the level surface $f(x,y,z)=k$. So, $\boldsymbol{\nabla} f(P)$ is a normal to the tangent plane through P to the level surface.

with the horizon. The same slope also occurs in the direction $\varphi = 214° + 79° = 293°$ at hiker's current position (see Fig. **24.3** (left panel)).

Next, consider a level set $f(x,y)=k$ of a differentiable function of two variables. In Section **24.2**, it has been shown that level sets are smooth curves in an open disk D if the partial derivatives f'_x and f'_y are continuous and do not vanish simultaneously anywhere in D. Furthermore, it has been shown that *the gradient* $\mathbf{n}= \boldsymbol{\nabla} f(x_0,y_0) \neq \mathbf{0}$ *at a point* (x_0,y_0) *of a level curve of* f *is orthogonal to a tangent vector to the level curve of* f *at that point. This is often expressed by saying that the gradient of* f *is always normal to the level curves of* f.

This geometrical property of the gradient is illustrated in the left panel of Fig. **24.3**. Recall that a function $f(x,y)$ can be described by a contour map, which is a collection of level curves. Then one can define a curve through a particular point that is normal to all level curves in some neighborhood of that point. This curve is called the *curve of steepest descent or ascent* through that point. A tangent vector of this curve at any point is parallel to the gradient at that point. The values of the function increase (or decrease) most rapidly along this curve. If a hiker follows the direction of the gradient of the height, he would go along the path of steepest ascent (or the steepest descent if he follows the direction opposite to the gradient) as depicted in the middle panel of Fig. **24.3**.

Similarly, the level sets $f(\mathbf{r})=k$ of a function of three variables $\mathbf{r} = \langle x,y,z\rangle$ are smooth surfaces inside an open ball B if the partial derivatives are continuous and do not vanish simultaneously anywhere in B. As has been shown in Section **24.2**, the gradient $\mathbf{n} = \boldsymbol{\nabla} f(\mathbf{r}_0)$ at a particular point $\mathbf{r}_0$ of the level surface is a normal of the plane tangent to the surface at $\mathbf{r}_0$. In particular, if $\mathbf{r}(t)$ is a smooth curve in the level surface such that $\mathbf{r}_0 = \mathbf{r}(t_0)$, then the value of f along the curve is a constant k and f has zero rate of change along the curve. By the chain rule,

$$0 = \frac{d}{dt} f(\mathbf{r}(t)) = f'_x(\mathbf{r}(t))x'_t + f'_y(\mathbf{r}(t))y'_t + f'_z(\mathbf{r}(t))z'_t = \boldsymbol{\nabla} f(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$$

Thus, for any smooth curve in the level surface, the vector $\mathbf{r}'(t_0)$ tangent to the curve at $\mathbf{r}_0$ is orthogonal to the gradient $\boldsymbol{\nabla} f(\mathbf{r}_0)$. So the gradient is normal to every smooth curve in a level surface. This geometrical fact is expressed by saying that *the gradient is normal to level surfaces* (see the right panel of Fig. **24.3**). One can define a curve through a particular point whose tangent vector at any point is parallel to the gradient, just like the curve of steepest descent or ascent in the case of functions of two variables. This curve is normal to level surfaces of the function. The values of the function increase (or decrease) most rapidly along this curve.

All these findings are summarized in the following theorem.

THEOREM **24.2**. (Geometrical Properties of the Gradient).
Suppose that a function f is differentiable at $\mathbf{r}_0$ and $\boldsymbol{\nabla} f(\mathbf{r}_0) \neq \mathbf{0}$. Let S be the level surface (or curve) through the point $\mathbf{r}_0$. Then

(1) *The maximal rate of change of f at $\mathbf{r}_0$ occurs in the direction of the gradient $\boldsymbol{\nabla} f(\mathbf{r}_0)$ and is equal to its magnitude $\|\boldsymbol{\nabla} f(\mathbf{r}_0)\|$.*
(2) *The minimal rate of change of f at $\mathbf{r}_0$ occurs in the direction opposite to the gradient $-\boldsymbol{\nabla} f(\mathbf{r}_0)$ and equals $-\|\boldsymbol{\nabla} f(\mathbf{r}_0)\|$.*
(3) *If f has continuous partial derivatives on an open ball D containing $\mathbf{r}_0$, then the portion of S inside D is a smooth surface (or curve), and $\boldsymbol{\nabla} f$ is normal to S at $\mathbf{r}_0$.*

EXAMPLE **24.3**. *Find an equation of the tangent plane to the ellipsoid $x^2 + 2y^2 + 3z^2 = 11$ at the point $(2, 1, 1)$.*

SOLUTION: The equation of the ellipsoid can be viewed as the level surface

$$f(x, y, z) = 11$$

of the function

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

through the point $\mathbf{r}_0 = (2, 1, 1)$ because $f(2, 1, 1) = 11$. By the geometrical property of the gradient, the vector $\mathbf{n} = \boldsymbol{\nabla} f(\mathbf{r}_0)$ is normal to the plane in question because the components of $\boldsymbol{\nabla} f = \langle 2x, 4y, 6z\rangle$ are continuous. One has

$$\mathbf{n} = \boldsymbol{\nabla} f(2, 1, 1) = \langle 4, 4, 6\rangle .$$

An equation of the plane through the point $(2, 1, 1)$ and normal to $\mathbf{n}$ is

$$4(x - 2) + 4(y - 1) + 6(z - 1) = 0 \quad \text{or} \quad 2x + 2y + 3z = 9 .$$

$\square$

The gradient of a differentiable function of two variables can be found if the rates of change of the function in any two non-parallel directions are known. Similarly, in order to determine the gradient of a function of three variables, the rates of change in three non-coplanar direction must be known. To understand the claim, fist note that the components of the gradient are partial derivatives which can be viewed as the directional derivatives along the standard basis vectors. The standard basis vectors are perpendicular and, hence, they are non-parallel in a plane (two variables) and non-coplanar in space (three variables). The following example illustrates the concept in the case of two variables (the multi-variable case is similar).

EXAMPLE **24.4**. *Find the maximal rate of change of a differentiable function f of two variables at a point P and the direction in which it occurs if the function is decreasing at P in the direction of the vector $\mathbf{a} = \langle 1, 1\rangle$ at a rate of $\sqrt{2}$ and increasing in the direction of the vector $\mathbf{b} = \langle -1, 2\rangle$ at a rate of $\sqrt{5}$.*

SOLUTION: The maximal rate of change occurs in the direction of the gradient at P and is equal to the magnitude of the gradient. Thus, the problem is to find the gradient if

$$D_{\mathbf{a}}f(P) = -\sqrt{2}\,, \qquad D_{\mathbf{b}}f(P) = \sqrt{5}\,.$$

The negative sign at $\sqrt{2}$ is required as the function is decreasing in the direction of $\mathbf{a}$. Put $\boldsymbol{\nabla} f(P) = \langle X, Y\rangle$, where the components X and Y of the gradient are to be found. Since neither $\mathbf{a}$ nor $\mathbf{b}$ are unit vectors,

$$\begin{aligned} D_{\mathbf{a}}f(P) &= \frac{1}{\|\mathbf{a}\|}\,\mathbf{a}\cdot\boldsymbol{\nabla} f = \frac{1}{\sqrt{2}}(X+Y) = -\sqrt{2}\,,\\ D_{\mathbf{b}}f(P) &= \frac{1}{\|\mathbf{b}\|}\,\mathbf{b}\cdot\boldsymbol{\nabla} f = \frac{1}{\sqrt{5}}\,(-X+2Y) = \sqrt{5}\,. \end{aligned}$$

Multiplying the first equation by $\sqrt{2}$ and the second equation by $\sqrt{5}$, and then adding the equations, one finds

$$3Y = 3 \quad\Rightarrow\quad Y = 1\,; \qquad X + Y = -2 \quad\Rightarrow\quad X = -3\,.$$

The the maximal rate of change of f at P and the direction in which it occurs are, respectively,

$$\|\boldsymbol{\nabla} f\| = \|\langle -3, 1\rangle\| = \sqrt{10}\,, \qquad \boldsymbol{\nabla} f = \langle -3, 1\rangle\,.$$

□

It is worth noting that if in this problem $\mathbf{a}$ and $\mathbf{b}$ are parallel, then the problem has no solution. Indeed, two parallel vectors are proportional, $\mathbf{a} = s\mathbf{b}$ for some real s. Then $D_{\mathbf{a}}f = \pm D_{\mathbf{b}}f$ (the minus sign occurs if the vectors have opposite directions (anti-parallel vectors)), and the system of equations for the components of the gradient has no solution.

Theorem **24.2** holds for functions of more than three variables as well. Equation (**24.2**) was obtained for any number of variables, and the representation of the dot product (**24.3**) holds in any Euclidean space. Thus, the first two properties of the gradient are valid for functions of any number of variables. The implicit function theorem holds for any number of variables so the equation of a level set $f(\mathbf{r}) = k$ can be solved with respect one of the variables if the gradient does not vanish, e.g., $x_m = g(x_1, ..., x_{m-1})$ near a point where the conditions of the theorem are fulfilled (here $f'_{x_m} \neq 0$). It can be interpreted as an $(m-1)$-dimensional surface embedded in an m-dimensional Euclidean space, which is hard to visualize. To this end, it is only noted that if f has continuous partial derivatives (the components of the gradient are continuous functions) and $\mathbf{r}(t)$ is a smooth parametric curve in a level set $f(\mathbf{r}) = k$, then f has a constant value along any such curve, and, by the chain rule, it follows that

$$\frac{d}{dt} f(\mathbf{r}(t)) = \boldsymbol{\nabla} f(\mathbf{r}(t)) \cdot \mathbf{r}'(t) = 0 \quad \text{for any } t\,.$$

At any particular point $\mathbf{r}_0 = \mathbf{r}(t_0)$, tangent vectors $\mathbf{r}'(t_0)$ to all such curves through $\mathbf{r}_0$ are orthogonal to a *single* vector $\boldsymbol{\nabla} f(\mathbf{r}_0) \neq \mathbf{0}$. So these vectors form an $(m-1)$-dimensional Euclidean space (called a *tangent space* to the level surface at $\mathbf{r}_0$), just like all vectors in a plane in three dimensional Euclidean space are orthogonal to a normal of the plane. In this sense, the third property of the gradient holds for functions of more than three variables.

Remark. The gradient can be viewed as the result of the action of the operator $\boldsymbol{\nabla} = \langle \partial/\partial x_1, \partial/\partial x_2, \cdots, \partial/\partial x_m \rangle$ if $\boldsymbol{\nabla} f$ is understood in the sense of multiplication of the "vector" $\boldsymbol{\nabla}$ by a scalar f:

$$\boldsymbol{\nabla} f = \Big\langle \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, ..., \frac{\partial}{\partial x_m} \Big\rangle f = \Big\langle \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_1}, \ ..., \ \frac{\partial f}{\partial x_1} \Big\rangle\,.$$

With this notation, the operator d introduced to define the differentials of a function f has a compact form $d = d\mathbf{r} \cdot \boldsymbol{\nabla}$. The linearization $L(\mathbf{r})$ of $f(\mathbf{r})$ at $\mathbf{r}_0$ and the differentials of f also have a simple form for any number of variables:

$$L(\mathbf{r}) = f(\mathbf{r}_0) + \boldsymbol{\nabla} f(\mathbf{r}_0) \cdot (\mathbf{r} - \mathbf{r}_0)\,, \qquad d^n f(\mathbf{r}) = (d\mathbf{r} \cdot \boldsymbol{\nabla})^n f(\mathbf{r})\,.$$

24.4. Study Problems.

Problem **24.1**. (Differentiability and Directional Derivative)
Let $f(x, y) = y^3/(x^2 + y^2)$ *if* $(x, y) \neq (0, 0)$ *and* $f(0, 0) = 0$. *Show that* $D_{\mathbf{u}} f(0, 0)$ *exists for any* $\hat{\mathbf{u}}$, *but it is not given by equation* (**24.1**). *Show that this function is not differentiable at* $(0, 0)$. *In other words, the existence of the directional derivative at a point in every direction does not imply differentiability at that point.*

SOLUTION: Put $\hat{\mathbf{u}} = \langle \cos\theta, \sin\theta\rangle$ for $0 \le \theta < 2\pi$. By the definition of the directional derivative

$$D_{\mathbf{u}}f(0,0) = \lim_{h\to 0} \frac{f(h\cos\theta, h\sin\theta) - f(0,0)}{h} = \lim_{h\to 0} \frac{h^3\sin^3\theta}{h^3} = \sin^3\theta\,.$$

In particular, for $\theta = 0$, $\hat{\mathbf{u}} = (1,0)$, and $D_{\mathbf{u}}f(0,0) = f'_x(0,0) = 0$; similarly for $\theta = \pi/2$, $\hat{\mathbf{u}} = (0,1)$, and $D_{\mathbf{u}}f(0,0) = f'_y(0,0) = 1$. Therefore

$$f'_x(0,0)u_1 + f'_y(0,0)u_2 = \sin\theta \ne \sin^3\theta = D_{\mathbf{u}}f(0,0)$$

that is, the relation (**24.1**) does not hold.

Since the partial derivatives exist, the function is differentiable if the linear $L(x,y) = f'_x(0,0)x + f'_y(0,0)y = y$ is a good linear approximation to f at the origin. But the limit

$$\lim_{(x,y)\to(0,0)} \frac{f(x,y) - L(x,y)}{(x^2+y^2)^{1/2}} = -\lim_{(x,y)\to(0,0)} \frac{yx^2}{(x^2+y^2)^{3/2}}$$

does not exist because the limits along straight lines $(x,y) = (t, at)$, $t \to 0^+$, have different values $a/(1+a^2)^{3/2}$, where a is real. So, f is not differentiable at the origin, despite that it has all directional derivatives at the origin. □

Problem **24.2**. *Suppose that three level surfaces $f(x,y,z) = 1$, $g(x,y,z) = 2$, and $h(x,y,z) = 3$ are intersecting along a smooth curve C. Let P be a point on C in whose neighborhood f, g, and h have continuous partial derivatives and their gradients do not vanish at P. Find $\boldsymbol{\nabla} f \cdot (\boldsymbol{\nabla} g \times \boldsymbol{\nabla} h)$ at P.*

SOLUTION: Let $\mathbf{v}$ be a tangent vector to C at the point P (it exists because the curve is smooth). By Part (3) of Theorem **24.2**, the equations $f(x,y,z) = 1$, $g(x,y,z) = 2$, and $h(x,y,z) = 3$ define three smooth surfaces in a neighborhood of P. Since C lies in the surface $f(x,y,z) = 1$, the gradient $\boldsymbol{\nabla} f(P)$ is orthogonal to $\mathbf{v}$. Similarly, the gradients $\boldsymbol{\nabla} g(P)$ and $\boldsymbol{\nabla} h(P)$ must be orthogonal to $\mathbf{v}$. Therefore, all the gradients must be in a plane perpendicular to the vector $\mathbf{v}$. The triple product for any three coplanar vectors vanishes, and hence $\boldsymbol{\nabla} f \cdot (\boldsymbol{\nabla} g \times \boldsymbol{\nabla} h) = 0$ at P. □

Problem **24.3**. (Energy Conservation in Mechanics)
Consider Newton's second law $m\mathbf{a} = \mathbf{F}$. Suppose that the force is the gradient $\mathbf{F} = -\boldsymbol{\nabla} U$, where $U = U(\mathbf{r})$. Let $\mathbf{r} = \mathbf{r}(t)$ be the trajectory satisfying Newton's second law. Prove that the quantity $E = mv^2/2 + U(\mathbf{r})$, where $v = \|\mathbf{r}'(t)\|$ is the speed, is a constant of motion, that is, $dE/dt = 0$. This constant is called the total energy of a particle. A force that can be represented by the gradient of a function U is called a conservative force, and the function U is called a potential energy.

SOLUTION: First, note that $v^2 = \mathbf{v}\cdot\mathbf{v}$. Hence, $(v^2)' = 2\mathbf{v}\cdot\mathbf{v}' = 2\mathbf{v}\cdot\mathbf{a}$. Using the chain rule,

$$\frac{dU}{dt} = U'_x x'(t) + U'_y y'(t) + U'_z z'(t) = \mathbf{r}'\cdot\boldsymbol{\nabla} U = \mathbf{v}\cdot\boldsymbol{\nabla} U\,.$$

It follows from these two relations that

$$\frac{dE}{dt} = \frac{m}{2}(v^2)' + \frac{dU}{dt} = m\mathbf{v}\cdot\mathbf{a} + \mathbf{v}\cdot\boldsymbol{\nabla}U = \mathbf{v}\cdot(m\mathbf{a} - \mathbf{F}) = 0$$

So the total energy is conserved for the trajectory of the motion. □

24.5. Exercises.

1. Let f be a differentiable function of two variables. Specify the directions at a particular point in which the function has no rate of change? How many such directions exist if the first partial derivatives do not vanish at that point? Answer the same questions for a differentiable function of three variables.

2. Let $f(x,y) = y$ if $y \neq x^2$ and $f(x,y) = 0$ if $y = x^2$. Find $D_{\mathbf{n}}f(0,0)$ for all unit vectors $\hat{\mathbf{n}}$. Show that $f(x,y)$ is not differentiable at $(0,0)$. Is the function continuous at $(0,0)$?

3–8. For each of the following functions, find the gradient and the directional derivative at a specified point P in the direction parallel to a given vector $\mathbf{v}$. Indicate whether the function increases or decreases in that direction at P.

3. $f(x,y) = x^2y$, $P = (1,2)$, $\mathbf{v} = \langle 4, 3\rangle$;
4. $f(x,y) = x/(1+xy)$, $P = (1,1)$, $\mathbf{v} = \langle 2, 1\rangle$;
5. $f(x,y,z) = x^2y - zy^2 + xz^2$, $P = (1,2,-1)$, $\mathbf{v} = \langle 1,-2,2\rangle$;
6. $f(x,y,z) = \tan^{-1}(1 + x + y^2 + z^3)$, $P = (1,-1,1)$, $\mathbf{v} = \langle 1,1,1\rangle$;
7. $f(x,y,z) = \sqrt{x+yz}$, $P = (1,1,3)$, $\mathbf{v} = \langle 2,6,3\rangle$;
8. $f(x,y,z) = (x+y)/z$, $P = (2,1,1)$, $\mathbf{v} = \langle 2,-1,-2\rangle$.

9–13. Find the maximal and minimal rates of change of each of the following functions at a specified point P and the directions in which they occur. Find the directions in which the function has zero rate of change at P.

9. $f(x,y) = x/y^2$, $P = (2,1)$;
10. $f(x,y) = x^y$, $P = (2,1)$;
11. $f(x,y,z) = xz/(1+yz)$, $P = (1,2,3)$;
12. $f(x,y,z) = x\sin(yz)$, $P = (1,2,\pi/3)$;
13. $f(x,y,z) = x^{y^z}$, $P = (2,2,1)$.

14. Let $f(x,y) = y/(1+x^2+y)$. For a number $-1 \le p \le 1$, find all unit vectors $\hat{\mathbf{u}}$ along which the rate of change of f at $(2,-3)$ is p times the maximal rate of change of f at $(2,-3)$.

15. For the function $f(x,y,z) = \frac{1}{2}x^2 - \frac{1}{2}y^2x + z^3y$ at the point $P_0 = (1,2,-1)$ find:

(i) The maximal rate of change of f and the direction in which it occurs;
(ii) A direction in which the rate of change is half of the maximal rate of change. How many such directions exist?
(iii) The rate of change in the direction from P_0 toward the point $P = (3,1,1)$.

16. If f is a differentiable real-valued function of one real variable, and u is a differentiable real-valued function of m real variables, show that $\boldsymbol{\nabla}(f(u(\mathbf{r}))) = f'(u(\mathbf{r}))\boldsymbol{\nabla}u(\mathbf{r})$, where $\mathbf{r} = \langle x_1, x_2, \ldots, x_m\rangle$.
17. Find $\boldsymbol{\nabla}\|\mathbf{c}\times\mathbf{r}\|^2$ where $\mathbf{c}$ is a constant vector.
18. If f is a differentiable real-valued function of two real variables, and u and v are differentiable real-valued functions of m real variables, show that $\boldsymbol{\nabla}f(u(\mathbf{r}), v(\mathbf{r})) = f'_u(u(\mathbf{r}), v(\mathbf{r}))\boldsymbol{\nabla}u(\mathbf{r}) + f'_v(u(\mathbf{r}), v(\mathbf{r}))\boldsymbol{\nabla}v(\mathbf{r})$, where $\mathbf{r} = \langle x_1, x_2, \ldots, x_m\rangle$.
19. Find the directional derivative of $f(\mathbf{r}) = (x/a)^2 + (y/b)^2 + (z/c)^2$ at a point $\mathbf{r} = \langle x, y, z\rangle$ in the direction of $\mathbf{r}$. Find the points at which this derivative is equal to $\|\boldsymbol{\nabla}f\|$.
20. Find the angle between the gradients of $f = x/(x^2+y^2+z^2)$ at the points $(1,2,2)$ and $(-3,1,0)$.
21. Let $f(x,y,z) = z/\sqrt{x^2+y^2+z^2}$. Sketch the level surfaces of the function f and the function $\|\boldsymbol{\nabla}f\|$. What is the significance of the level surfaces of $\|\boldsymbol{\nabla}f\|$? Find the maximal and minimal values of f and $\|\boldsymbol{\nabla}f\|$ in the region $1 \le z \le 2$.
22. Let a curve C be defined as the intersection of the plane $\sin\theta(x-x_0) - \cos\theta(y-y_0) = 0$, where θ is a parameter, and the graph $z = f(x,y)$ of a differentiable function f is differentiable. Find $\tan\alpha$ where α is the angle between the tangent line to C at $(x_0, y_0, f(x_0,y_0))$ and the xy plane.
23. Consider the function $f(x,y,z) = 2\sqrt{z+xy}$ and three points $P_0 = (1,2,2)$, $P_1 = (-1,4,1)$, and $P_2 = (-2,-2,2)$. In which direction is the absolute value of the rate of change of f at P_0 the largest, toward P_1 or toward P_2? What is the direction in which f has the largest rate of change at P_0?
24. For the function $f(x,y,z) = xy + zy + zx$ at the point $P_0 = (1,-1,0)$ find:

(i) The maximal rate of change;
(ii) The rate of change in the direction $\mathbf{v} = \langle -1, 2, -2)\rangle$;
(iii) The angle θ between $\mathbf{v}$ and the direction in which the maximal rate of f occurs.

25. Let $f(x,y,z) = x/(x^2+y^2+z^2)^{1/2}$. Find the rate of change of f in the direction of the tangent vector to the curve $\mathbf{r}(t) = (\langle t, 2t^2, -2t^2)\rangle$ at the point $(1,2,-2)$.
26. Find all points at which the gradient of $f = x^3+y^3+z^3-3xyz$ has the given property:

(i) $\boldsymbol{\nabla}f$ is orthogonal to the z axis;
(ii) $\boldsymbol{\nabla}f$ is parallel to the z axis;
(iii) $\boldsymbol{\nabla}f = \mathbf{0}$.

27. Let $f(\mathbf{r}) = \ln\|\mathbf{r}-\mathbf{r}_0\|$, where $\mathbf{r}_0$ is a fixed vector. Find points $\mathbf{r} = \langle x, y, z\rangle$ in space where $\|\boldsymbol{\nabla}f\| = 1$.

28–33. Show that each of the following equations define a smooth surface through the specified point P in a neighborhood of P, and find the tangent plane and the normal line to the surface through P:

28. $x^2 + y^2 + z^2 = 169,\ \ P = (3, 4, 12)$;
29. $x^2 - 2y^2 + z^2 + yz = 2,\ P = (2, 1, -1)$;
30. $x = \tan^{-1}(y/z),\ P = (\pi/4, 1, 1)$;
31. $z = y + \ln(x/z),\ \ P = (1, 1, 1)$;
32. $2^{x/z} + 2^{y/z} = 8,\ \ P = (2, 2, 1)$;
33. $x^2 + 4y^2 + 3z^2 = 5,\ \ (1, -1/2, -1)$.

34. Find the points of the surface $x^2 + 2y^2 + 3z^2 + 2xy + 2zx + 4yz = 8$ at which the tangent planes are parallel to the coordinate planes.
35. Find the tangent planes to the surface $x^2 + 2y^2 + 3z^2 = 21$ that are parallel to the plane $x + 4y + 6z = 0$.
36. Find the points on the ellipsoid $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$ at which the normal line makes equal angles with the coordinate axes.
37. Consider the paraboloid $z = x^2 + y^2$.

(i) Give the parametric equations of the normal line through a point $P_0 = (x_0, y_0, z_0)$ on the paraboloid;
(ii) Consider all normal lines through points with a fixed value of z_0 (say, $z_0 = 2$). Show that all such lines intersects at one single point which lies on the $z-$axis and find the coordinates of this point.

38. Find the points on the hyperboloid $x^2 - y^2 + 2z^2 = 5$ where the normal line is parallel to the line that joints the points $(3, -1, 0)$ and $(5, 3, 8)$.
39. Find an equation of the plane tangent to the surface $x^2 + y^2 - 4z^2 = 1$ at a generic point (x_0, y_0, z_0) of the surface.
40. Find the rate of change of the function $h(x, y) = \sqrt{10 - x^2y^2}$ at the point $P_0 = (1, 1)$ in the direction toward the point $P = (-2, 5)$. Let $h(x, y)$ be the height of a mountain in a neighborhood of P_0. Would you be climbing up or getting down when you go from P_0 toward P?
41. Your Mars rover is caught by a dust storm in mountains. The visibility is zero. Your current position is $P_0 = (1, 2)$. You can escape in the direction of a cave located at $P_1 = (4, -2)$ or in the direction of the base located at $P_2 = (17, 14)$. Which way would you drive to avoid steep climbing or descending if the height in a neighborhood of P_0 can be approximated by the function $h(x, y) = xy + x^2$?
42. You are flying a small aircraft on the planet Weirdo. You have disturbed a nest of nasty everything-eating bugs. The onboard radar indicates that the concentration of the bugs is $C(x, y, z) = 100 - x^2 - 2y^2 - 3z^2$ and $C(x, y, z) = 0$ if $x^2 + 2y^2 + 3z^2 > 100$. If your current position is $(2, 3, 1)$, in which direction would you fire a mass-destruction microwave laser to kill as many poor bugs as possible near you? Find the optimal escape trajectory.
43. Suppose that the functions f and g have continuous partial derivatives, their gradients do not vanish, and their level curves $f(x, y) = 0$ and

$g(x, y) = 0$ intersect at some point P_0. The rate of change of the function f at P_0 along the curve $g(x, y) = 0$ is a half of its maximal rate of change at P_0. What is the angle at which the curves intersect (the angle between the tangent lines)?

44. Suppose that the directional derivatives $D_{\mathbf{u}}f = a$ and $D_{\mathbf{v}}f = b$ of a differentiable function f of two variables are known at a particular point P_0 for two unit non-parallel vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ that make the angles θ and ϕ with the x axis counted counterclockwise from the latter, respectively. Find the gradient of f at P_0.

45. Three tests of drilling into rock along the directions $\mathbf{u} = \langle 1, 2, 2\rangle$, $\mathbf{v} = \langle 0, 4, 3\rangle$, and $\mathbf{w} = \langle 0, 0, 1\rangle$ yielded that the gold concentration increases at the rates 3 units per meter, 3 units per meter, and 1 unit per meter, respectively. Assume that the concentration is a differentiable function. In what direction would you drill to maximize the gold yield and at what rate does the gold concentration increase in that direction?

46. A level surface of a differentiable function $f(x, y, z)$ contains the curves $\mathbf{r}_1(t) = \langle 2+3t, 1-t^2, 3-4t+t^2\rangle$ and $\mathbf{r}_2(t) = \langle 1+t^2, 2t^3-1, 2t+1\rangle$. Can this information be used to find the tangent plane to the surface at the point $(2, 1, 3)$? If so, find an equation of the plane.

47. Prove that tangent planes to the surface $xyz = a^3 > 0$ and the coordinate planes form tetrahedrons of equal volumes.

48. Prove the total length of intervals from the origin to the points of intersection of tangent planes to the surface $\sqrt{x} + \sqrt{y} + \sqrt{z} = \sqrt{a}$, $a > 0$, with the coordinate axes is constant.

49. Two surfaces are called *orthogonal* at a point of intersection if the normal lines to the surfaces at that point are orthogonal. Show that the surfaces $x^2 + y^2 + z^2 = r^2$, $x^2 + y^2 = z^2 \tan^2\phi$, and $y\cos\theta = x\sin\theta$ are pairwise orthogonal at their points of intersection for any values of the constants $r > 0$, $0 < \phi < \pi$, and $0 \le \theta < 2\pi$.

50. Find the directional derivative of $f(x, y, z)$ in the direction of the gradient of $g(x, y, z)$. Assume that the gradients $\boldsymbol{\nabla} f$ and $\boldsymbol{\nabla} g$ do not vanish and have continuous components. What is the geometrical significance of this derivative?

51. Find the angle at which the cylinder $x^2 + y^2 = a^2$ intersects the surface $bz = xy$ at a generic point of intersection (x_0, y_0, z_0).

52. A ray of light reflects from a mirrored surface at a point P just as it would reflect from the mirrored plane tangent to the surface at P (if the light travels along a vector $\mathbf{u}$, then the reflected light travels along a vector obtained from $\mathbf{u}$ by reversing the direction of the component parallel to the normal to the surface). Show that the light coming parallel to the z axis from above the xy plane will be focused by the parabolic mirror $az = x^2+y^2$, $a > 0$, to a single point. Find its coordinates. This property of parabolic mirrors is used to design telescopes. Conversely, if a point light source is put into the focus of a parabolic mirror, then the light that is coming out if the source and reflected by the mirror forms a beam parallel to the symmetry axis of the mirror. This property is used to design flashlights.

25. Maximum and Minimum Values

25.1. Critical Points of Multivariable Functions. Positions of local maxima and minima of a one-variable function play an important role when analyzing an overall behavior of the function. In Calculus I, it was shown how the derivatives can be used to find local maxima and minima. Here this analysis is extended to multivariable functions.

The following notation will be used. An open ball of radius δ centered at a point $\mathbf{r}_0$ is denoted $B_\delta = \{\mathbf{r} \mid \|\mathbf{r} - \mathbf{r}_0\| < \delta\}$; that is, it is a set of points whose distance from $\mathbf{r}_0$ is less than $\delta > 0$. A neighborhood $N_\delta(\mathbf{r}_0)$ of a point $\mathbf{r}_0$ in a set D is a set of common points of D and B_δ; that is, $N_\delta(\mathbf{r}_0)$ contains all points in D whose distance from $\mathbf{r}_0$ is less than δ.

Definition **25.1**. (Absolute and Local Maxima or Minima).
A function f on a set D is said to have a local maximum at $\mathbf{r}_0$ if there is a neighborhood $N_\delta(\mathbf{r}_0)$ such that $f(\mathbf{r}_0) \geq f(\mathbf{r})$ for all $\mathbf{r}$ in $N_\delta(\mathbf{r}_0)$. The number $f(\mathbf{r}_0)$ is called a local maximum value*. If there is a neighborhood $N_\delta(\mathbf{r}_0)$ such that $f(\mathbf{r}_0) \leq f(\mathbf{r})$ for all $\mathbf{r}$ in $N_\delta(\mathbf{r}_0)$, then f is said to have a local minimum at $\mathbf{r}_0$ and the number $f(\mathbf{r}_0)$ is called a* local minimum value*. If the inequality $f(\mathbf{r}_0) \geq f(\mathbf{r})$ or $f(\mathbf{r}_0) \leq f(\mathbf{r})$ holds for all points $\mathbf{r}$ in the domain of f, then f has an absolute maximum or absolute minimum at $\mathbf{r}_0$, respectively.*

Minimal and maximal values are also called *extremum values*. In the one-variable case, Fermat's theorem asserts that if a differentiable function has a local extremum at x_0, then its derivative vanishes at x_0. The tangent line to the graph of f at x_0 is horizontal: $y = f(x_0) + df(x_0) = f(x_0) + f'(x_0)dx = f(x_0)$. There is an extension of Fermat's theorem to the multivariable case.

Theorem **25.1**. (Necessary Condition for a Local Extremum)
If a differentiable function f has a local extremum at an interior point $\mathbf{r}_0$ of its domain D, then $df(\mathbf{r}_0) = 0$ or $\boldsymbol{\nabla} f(\mathbf{r}_0) = \mathbf{0}$ (all partial derivatives of f vanish at $\mathbf{r}_0$).

Proof. Consider a smooth parametric curve $\mathbf{r}(t)$ through the point $\mathbf{r}_0$ such that $\mathbf{r}(t_0) = \mathbf{r}_0$. Then $d\mathbf{r}(t_0) = \mathbf{r}'(t_0)dt \neq \mathbf{0}$ (the curves is smooth and, hence, has a non-zero tangent vector). The function $F(t) = f(\mathbf{r}(t))$ defines the values of f along the curve. Therefore, $F(t)$ must have a local extremum at $t = t_0$. Since f is differentiable, the differential $dF(t_0) = F'(t_0)dt$ exists by the chain rule:

$$dF(t_0) = d\mathbf{r}(t_0) \cdot \boldsymbol{\nabla} f(\mathbf{r}_0)\,.$$

By Fermat's theorem $dF(t_0) = 0$ and, hence, by the geometrical properties of the dot product

$$dF(t_0) = 0 \quad \Rightarrow \quad d\mathbf{r}(t_0) \cdot \boldsymbol{\nabla} f(\mathbf{r}_0) = 0 \quad \Rightarrow \quad d\mathbf{r}(t_0) \perp \boldsymbol{\nabla} f(\mathbf{r}_0)$$

the tangent vector $d\mathbf{r}(t_0)$ and the gradient $\boldsymbol{\nabla} f(\mathbf{r}_0)$ are orthogonal for any smooth curve through $\mathbf{r}_0$. But the tangent vector $d\mathbf{r}(t_0)$ is an arbitrary non-zero vector because the curve is arbitrarily chosen. The only vector that is orthogonal to any vector is the zero vector and the conclusion of the theorem follows: $\nabla f(\mathbf{r}_0) = \mathbf{0}$. □

In particular, for a differentiable function f of two variables, this theorem states that the tangent plane to the graph of f at a local extremum is horizontal.

A local extremum may occur at a point where the function is not differentiable. For example, $f(x, y) = \sqrt{x^2 + y^2}$ is continuous everywhere and has an absolute minimum at $(0, 0)$. Note that the graph $z = \sqrt{x^2 + y^2}$ is a cone with the vertex at the origin. However, the partial derivatives $f'_x(0, 0)$ and $f'_y(0, 0)$ do not exist because the derivative of $f(x, 0) = |x|$ does not exists at $x = 0$, and similarly, the derivative of $f(0, y) = |y|$ does not exists at $y = 0$.

Definition **25.2**. (Critical Points).
An interior point $\mathbf{r}_0$ of the domain of a function f is said to be a critical point of f if either $\nabla f(\mathbf{r}_0) = \mathbf{0}$ or the gradient does not exist at $\mathbf{r}_0$.

It should be emphasized that a function f does not always achieve a local extreme value at a critical point. In particular, *the converse of Theorem* **25.1** *is not true*. If the gradient of a function f vanishes at a point, then f may not have a local extremum at that point. For example, let

$$f(x, y) = xy\,.$$

It is differentiable everywhere and its partial derivatives are $f'_x = y$ and $f'_y = x$. They vanish at the origin,

$$\nabla f(0, 0) = \mathbf{0}\,.$$

However, the function achieves neither a local maximum value nor a local minimum value at this critical point. Indeed, consider a straight line through the origin, $x = t$, $y = at$. Then the values of f along the line are

$$F(t) = f(t, at) = at^2\,.$$

So $F(t)$ has a minimum at $t = 0$ if $a > 0$ or a maximum if $a < 0$. Thus, f cannot have a local extremum at $(0, 0)$. The graph $z = xy$ is a hyperbolic paraboloid rotated through an angle $\pi/4$ about the z axis (see Example **9.4**). It looks like a saddle.

Furthermore, consider

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

It follows that

$$f'_x(x, y) = 3x^3 - 2xy\,, \quad f'_y(x, y) = -x^2 \quad \Rightarrow \quad \nabla f(0, 0) = \mathbf{0}\,.$$

However the values of the function along *any* straight line through the origin $x = t\cos\theta$, $y = t\sin\theta$ do not achieve a local extremum value:

$$F(t) = f(t\cos\theta, t\sin\theta) = \cos^2\theta(\cos\theta - \sin\theta)t^3 = at^3$$

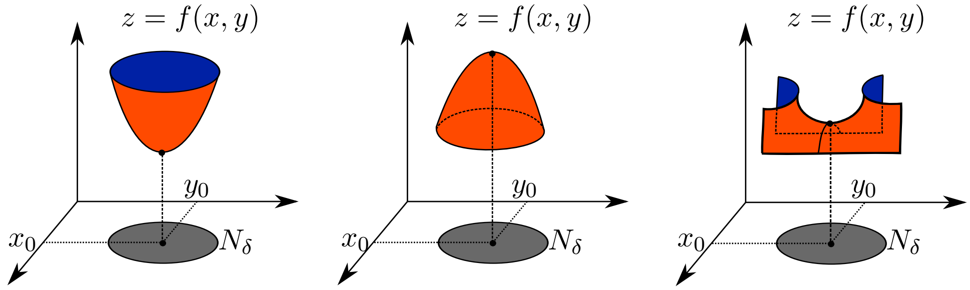


FIGURE **25.1**. **Left**: The graph $z = f(x, y)$ near a local minimum of f. The values of f are no less than $f(x_0, y_0)$ for all (x, y) in a sufficiently small neighborhood N_δ of (x_0, y_0). **Middle**: The graph $z = f(x, y)$ near a local maximum of f. The values of f do not exceed $f(x_0, y_0)$ for all (x, y) in a sufficiently small neighborhood N_δ of (x_0, y_0). **Right**: An example of the graph $z = f(x, y)$ near a saddle point (x_0, y_0) of f. In a neighborhood N_δ of (x_0, y_0), the values of f do not have a local maximum or minimum.

where the constant a can be positive, or negative, or zero. So, the curve of intersection of the surface

$$z = x^3 - x^2 y$$

with any plane containing the z axis (except the planes $x = 0$ and $x = y$) has an inflection point at the origin. Such a surface is known as the "monkey saddle".

In the two examples considered, the critical point $(0, 0)$ is an example of a *saddle point* of a function (see Fig. **25.1** (right panel)).

DEFINITION **25.3**. (Saddle Point)
A saddle point *of a function f of several variables is a critical point at which f does not achieve a local maximum or minimum value.*

Remark. The analysis of the above two examples of a saddle point might make an impression that if $\nabla f(\mathbf{r}_0) = \mathbf{0}$ and the values of f along *any* straight line through $\mathbf{r}_0$ have a local maximum (minimum) (i.e., $F(t) = f(\mathbf{r}_0 + \mathbf{v}t)$ has a local maximum (minimum) at $t = 0$ for any non-zero vector $\mathbf{v}$), then f has a local maximum (minimum) at $\mathbf{r}_0$. This conjecture is *false*! An example is given in Study Problem **25.2**.

25.2. Concavity. Recall from Calculus I that if the graph of a function f of a single variable x lies above all its tangent lines in an interval I, then f is concave upward on I. If the graph lies below all its tangent lines in I, then f is concave downward on I. Furthermore, if $f'(x_0) = 0$ (the tangent line is horizontal at x_0) and f is concave upward in an open interval I containing x_0, then $f(x) > f(x_0)$ for all $x \neq x_0$ in I and, hence, f has a local minimum.

Similarly, f has a local maximum at x_0 where $f'(x_0) = 0$ if it is concave downward in a neighborhood of x_0. If the function f is twice differentiable on I, then it is concave upward if $f''(x) > 0$ on I and it is concave downward if $f''(x) < 0$ on I. The concavity test can be restated in the form of the second differential

$$d^2f(x) = f''(x)(dx)^2$$

which is a function of two independent variables x and dx:

$$\begin{array}{lcl} d^2f(x) > 0,\ dx \neq 0 & \Rightarrow & f \text{ is concave upward} \\ d^2f(x) < 0,\ dx \neq 0 & \Rightarrow & f \text{ is concave downward} \end{array}$$

Suppose that $f'(x_0) = 0$, $f''(x_0) \neq 0$, and f'' is continuous at x_0. Recall *if a continuous function is positive (negative) at a particular point, then it is positive (negative) in a neighborhood of that point*. So the continuity of f'' ensures that $d^2f(x)$ has the same sign as $d^2f(x_0)$ for all x near x_0 and all $dx \neq 0$. Hence, the graph of f has a fixed concavity in a neighborhood of x_0. Thus, if $d^2f(x_0) < 0$ $(dx \neq 0)$, then f has a local maximum at x_0 and, if $d^2f(x_0) > 0$ $(dx \neq 0)$, then f has a local minimum at x_0. It turns out that this sufficient condition for a function to have a local extremum has a natural extension to functions of several variables.

THEOREM **25.2**. (Sufficient condition for a local extremum)
Suppose that a function f of several variables has continuous second partial derivatives in an open ball containing a point $\mathbf{r}_0$ and $\nabla f(\mathbf{r}_0) = \mathbf{0}$. Then

$$\begin{array}{lll} f \text{ has a local maximum at } \mathbf{r}_0 & \text{if} & d^2f(\mathbf{r}_0) < 0, \\ f \text{ has a local minimum at } \mathbf{r}_0 & \text{if} & d^2f(\mathbf{r}_0) > 0, \end{array}$$

for all $d\mathbf{r}$ such that $d\mathbf{r} \neq \mathbf{0}$.

The proof of this theorem is omitted. However, the conclusion of theorem can be understood as follows. Let $\mathbf{r}_0$ be an interior point of the domain of f. Suppose that $d^2f(\mathbf{r}_0) > 0$ (or $d^2f(\mathbf{r}_0) < 0$) for all $d\mathbf{r} \neq \mathbf{0}$. If f has continuous second partial derivatives, then by Taylor Theorem **23.1**, its values near $\mathbf{r}_0$ can be approximated by the second Taylor polynomial

$$\begin{aligned} f(\mathbf{r}_0 + d\mathbf{r}) &= T_2(\mathbf{r}_0 + d\mathbf{r}) + \varepsilon_2(d\mathbf{r}) \\ &= f(\mathbf{r}_0) + df(\mathbf{r}_0) + \frac{1}{2}d^2f(\mathbf{r}_0) + h_2(d\mathbf{r})\|d\mathbf{r}\|^2, \end{aligned}$$

where $h_2(d\mathbf{r}) \to 0$ as $d\mathbf{r} \to 0$. Therefore in a neighborhood $N_\delta(\mathbf{r}_0)$ of a sufficiently small radius δ, the sign of the difference between the value of the function and the value of its linearization coincides with the sign of the second differential $d^2f(\mathbf{r}_0)$:

$$f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0) - df(\mathbf{r}_0) = \frac{1}{2}d^2f(\mathbf{r}_0) + h_2(d\mathbf{r})\|d\mathbf{r}\|^2.$$

Indeed, the error $h_2\|d\mathbf{r}\|^2$ decreases to zero faster than $\|d\mathbf{r}\|^2$ with decreasing $\|d\mathbf{r}\|$ whereas $d^2f(\mathbf{r}_0)$ decreases as $\|d\mathbf{r}\|^2$ and, hence, $|d^2f(\mathbf{r}_0)| > |h_2|\|d\mathbf{r}\|^2$

for all $0 < \|d\mathbf{r}\| < \delta$ and a small enough δ. Neglecting the terms smaller than $\|d\mathbf{r}\|^2$, it is concluded that

$$f(\mathbf{r}_0 + d\mathbf{r}) - L(\mathbf{r}_0 + d\mathbf{r}) = \tfrac{1}{2}d^2 f(\mathbf{r}_0)\,, \tag{25.1}$$

where $L(\mathbf{r}_0 + d\mathbf{r}) = f(\mathbf{r}_0) + df(\mathbf{r}_0)$ is the linearization of f at $\mathbf{r}_0$. If, in addition, $\mathbf{r}_0$ is a critical point, that is, $df(\mathbf{r}_0) = 0$, then

$$f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0) = \tfrac{1}{2}d^2 f(\mathbf{r}_0)\,.$$

This equation shows that if $d^2f(\mathbf{r}_0) > 0$ (or $d^2f(\mathbf{r}_0) < 0$) for all $d\mathbf{r} \neq \mathbf{0}$, then at any point in a neighborhood $N_\delta(\mathbf{r}_0)$, the value of f is strictly greater (or strictly less) than $f(\mathbf{r}_0)$, and f attains a local maximum value (or a local minimum value) at $\mathbf{r}_0$.

Furthermore, the concept of *concavity* can be extended to the multivariable case. Let f have continuous second partial derivatives and $\mathbf{r}_0$ be an interior point of the domain of f. Equation (**25.1**) shows that if $d^2f(\mathbf{r}_0) > 0$ (or $d^2f(\mathbf{r}_0) < 0$) for all $d\mathbf{r} \neq \mathbf{0}$, then at any point in a neighborhood $N_\delta(\mathbf{r}_0)$, the value of f is strictly greater (or strictly less) than the value of its linearization at $\mathbf{r}_0$. In particular, in the two-variable case, the graph $z = f(x, y)$ lies above (or below) the tangent plane at (x_0, y_0) for all (x, y) in a small disk centered at (x_0, y_0). So, *this analysis shows that if the second differential $d^2f(\mathbf{r})$ is positive (negative) for all $\mathbf{r}$ in an open ball B and all $d\mathbf{r} \neq \mathbf{0}$, then the graph $z = f(\mathbf{r})$ over B is concave upward (downward) relative to the direction of the z axis.*

The second differential $d^2f(\mathbf{r})$ is a function of two independent variables $\mathbf{r}$ and $d\mathbf{r}$. If second partial derivatives of f are continuous, then the function d^2f is continuous. A continuous function has the property that if it is positive (negative) at a particular point, then it is positive (negative) in a neighborhood of that point. Therefore if $d^2f(\mathbf{r}_0)$ is positive (negative) for all $d\mathbf{r} \neq \mathbf{0}$ at a critical point $\mathbf{r}_0$, then $d^2f(\mathbf{r})$ is positive (negative) for all $d\mathbf{r} \neq \mathbf{0}$ and *all* $\mathbf{r}$ in a neighborhood of $\mathbf{r}_0$, that is, the graph $z = f(\mathbf{r})$ has a fixed concavity in a neighborhood of $\mathbf{r}_0$. In the two-variable case, the graph $z = f(x, y)$ near a local maximum at (x_0, y_0) looks like a paraboloid concave downward (see Fig. **25.1** (middle panel)). Similarly, the graph $z = f(x, y)$ near a local minimum at (x_0, y_0) looks like a paraboloid concave upward (see Fig. **25.1** (left panel)).

25.3. Second-Derivative Test. The differential $d^2f(\mathbf{r}_0)$ is a homogeneous quadratic polynomial in the variables $d\mathbf{r}$. Its sign is determined by its coefficients which are the second partial derivatives of f at $\mathbf{r}_0$. One can pose the question: *Under what conditions on the second partial derivatives of f at a critical point $\mathbf{r}_0$ is the second differential $d^2f(\mathbf{r}_0)$ positive or negative for all $d\mathbf{r} \neq \mathbf{0}$?* This question is answered first for functions of two variables. The general case is studied in the next section.

Suppose that a function f of two variable (x, y) has continuous second partial derivatives in an open disk centered at $\mathbf{r}_0 = \langle x_0, y_0 \rangle$. Then

$$d^2 f(\mathbf{r}_0) = a(dx)^2 + 2c dx dy + b(dy)^2 \,,$$

where $a = f''_{xx}(\mathbf{r}_0)$, $b = f''_{yy}(\mathbf{r}_0)$, and $c = f''_{xy}(\mathbf{r}_0) = f''_{yx}(\mathbf{r}_0)$ (Clairaut's theorem). The second partial derivatives can be arranged into a 2×2 symmetric matrix whose diagonal elements are a and b and whose off-diagonal elements are c. The quadratic polynomial of a variable λ,

$$P_2(\lambda) = \det \begin{pmatrix} a - \lambda & c \\ c & b - \lambda \end{pmatrix} = (a - \lambda)(b - \lambda) - c^2,$$

is called the *characteristic polynomial* of the matrix of the second partial derivatives of f at $\mathbf{r}_0$.

THEOREM **25.3**. (Second-Derivative Test).
Let $\mathbf{r}_0$ be a critical point of a function f. Suppose that the second partial derivatives of f are continuous in an open disk containing $\mathbf{r}_0$. Let $P_2(\lambda)$ be the characteristic polynomial of the matrix of the second partial derivatives of f at $\mathbf{r}_0$. Let λ_i, $i = 1, 2$, be the roots of $P_2(\lambda)$. Then

(1) *If the roots are strictly positive, $\lambda_i > 0$, then f has a local minimum at $\mathbf{r}_0$.*
(2) *If the roots are strictly negative, $\lambda_i < 0$, then f has a local maximum at $\mathbf{r}_0$.*
(3) *If the roots do not vanish but have different signs, then f has neither a local maximum nor a local minimum at $\mathbf{r}_0$ (in this case, $\mathbf{r}_0$ is a saddle point).*
(4) *If at least one of the roots vanishes, then f may have a local maximum, a local minimum, or none of the above (the second-derivative test is inconclusive).*

A proof of the second-derivative test is given in Section **25.4**. The proof shows that the characteristic polynomial $P_2(\lambda)$ *always* have two *real* roots (complex roots never occur). The roots may coincide (one real root of multiplicity 2). The roots λ_1 and λ_2 of the quadratic equation $P_2(\lambda) = 0$ or

$$\lambda^2 - (a + b)\lambda + ab - c^2 = 0$$

are known to satisfy the conditions:

$$\lambda_1 \lambda_2 = ab - c^2 \,, \qquad \lambda_1 + \lambda_2 = a + b \,.$$

Put $D = ab - c^2$. So, if $D < 0$, then $\lambda_1 \lambda_2 < 0$ and the roots have different signs. By Part (3) of the second-derivative test, $\mathbf{r}_0$ is a saddle point. If $D > 0$, then the roots have the same sign. Suppose that $D > 0$. Then either $a > 0$ or $a < 0$. In the first case,

$$D > 0 \quad \Rightarrow \quad \frac{D}{a} > 0 \quad \Rightarrow \quad b > \frac{c^2}{a} > 0 \,.$$

Therefore the roots are positive because they have the same sign and their sum is positive. By Part (1) of the second derivative test, the function has a local minimum at $\mathbf{r}_0$. If $a < 0$, then

$$D > 0 \quad \Rightarrow \quad \frac{D}{a} < 0 \quad \Rightarrow \quad b < \frac{c^2}{a} < 0\,.$$

Therefore the roots are negative because they have the same sign and their sum is negative. By Part (2) of the second derivative test, the function has a local maximum at $\mathbf{r}_0$. Thus, the following consequence of the second derivative test is established.

COROLLARY **25.1**. *Suppose that a function f satisfies the hypotheses of Theorem* **25.3**. *Put $D = ab - c^2$ where $a = f''_{xx}(\mathbf{r}_0)$, $b = f''_{yy}(\mathbf{r}_0)$, and $c = f''_{xy}(\mathbf{r}_0)$. Then*

(1) *If $D > 0$ and $a > 0$, then $f(\mathbf{r}_0)$ is a local minimum.*
(2) *If $D > 0$ and $a < 0$, then $f(\mathbf{r}_0)$ is a local maximum.*
(3) *If $D < 0$, then $\mathbf{r}_0$ is a saddle point of f.*
(4) *If $D = 0$, then the second-derivative test gives no information about the nature of the critical point $\mathbf{r}_0$.*

EXAMPLE **25.1**. *Find all critical points of the function*

$$f(x, y) = \frac{1}{3}x^3 + xy^2 - x^2 - y^2$$

and determine whether f has a local maximum, minimum, or saddle at them.

SOLUTION: Critical points. The function is a polynomial, and therefore it has continuous partial derivatives everywhere of any order. So its critical points are solutions of the system of equations

$$\nabla f(x, y) = \mathbf{0} \quad \Leftrightarrow \quad \begin{cases} f'_x = x^2 + y^2 - 2x = 0 \\ f'_y = 2xy - 2y \qquad = 0 \end{cases}$$

It is important to obtain an equivalent system of equations when transforming the system $\nabla f(\mathbf{r}) = \mathbf{0}$. It follows from the second equation that either $y = 0$ or $x = 1$. Therefore, the original system of equations is *equivalent* to two systems of equations:

$$\begin{cases} f'_x = x^2 + y^2 - 2x = 0 \\ x \ = 1 \end{cases} \quad \text{or} \quad \begin{cases} f'_x = x^2 + y^2 - 2x = 0 \\ y \ = 0 \end{cases}\,.$$

Solutions of the first system are $(1, 1)$ and $(1, -1)$. Solutions of the second system are $(0, 0)$ and $(2, 0)$. Thus, the function has four critical points.
The second-derivative test applies because the function has continuous second partial derivatives:

$$f''_{xx} = 2x - 2\,, \quad f''_{yy} = 2x - 2\,, \quad f''_{xy} = 2y.$$

Critical points $(1, \pm 1)$: The values of the second partial derivatives are:

$$a = f_{xx}(1, \pm 1) = 0\,, \quad b = f''_{yy}(1, \pm 1) = 0\,, \quad c = f''_{xy}(1, \pm 1) = \pm 2\,.$$

The characteristic equation reads

$$P_2(\lambda) = (a-\lambda)(b-\lambda) - c^2 = \lambda^2 - 4 = 0\,.$$

Its roots $\lambda_1 = 2$ and $\lambda_2 = -2$ do not vanish and have opposite signs. Therefore, the points $(1, \pm 1)$ are *saddle points* of the function.
Critical point $(0,0)$: The values of the second partial derivatives are

$$a = f''_{xx}(0,0) = -2\,, \quad b = f''_{yy}(0,0) = -2\,, \quad c = f''_{xy}(0,0) = 0\,.$$

The characteristic equation

$$P_2(\lambda) = (a-\lambda)(b-\lambda) = (-2-\lambda)^2 = 0$$

has one negative root of multiplicity 2, that is, $\lambda_1 = \lambda_2 = -2 < 0$. Therefore f has a *local maximum* at $(0,0)$.
Critical point $(2,0)$: The values of the second partial derivatives are

$$a = f''_{xx}(2,0) = 2\,, \quad b = f''_{yy}(2,0) = 2\,, \quad c = f''_{xy}(2,0) = 0\,.$$

The characteristic equation

$$P_2(\lambda) = (a-\lambda)(b-\lambda) = (2-\lambda)^2 = 0$$

has one positive root of multiplicity 2, $\lambda_1 = \lambda_2 = 2 > 0$. Therefore the function has a *local minimum* at $(2,0)$. □

EXAMPLE **25.2**. *Investigate the function* $f(x,y) = e^{x^2-y}(5-2x+y)$ *for extreme values.*

SOLUTION: The functions is defined on the whole plane and, as the product of an exponential and a polynomial, it has continuous partial derivatives of any order. So its critical points are points where its gradient vanishes, and its local extreme values, if any, can be investigated by the second-derivative test.
Critical points. Using the product rule for partial derivatives,

$$\begin{aligned} f'_x &= e^{x^2-y}\Big(2x(5-2x+y) - 2\Big) = 0 \quad \Rightarrow \quad x(5-2x+y) = 1 \\ f'_y &= e^{x^2-y}\Big((-1)(5-2x+y) + 1\Big) = 0 \quad \Rightarrow \quad 5-2x+y = 1 \end{aligned}$$

The substitution of the second equation into the first one yields $x = 1$. Then it follows from the second equation that $y = -2$. So the function has just one critical point $(1,-2)$.
Second derivative test. Using the product rule for partial derivatives,

$$\begin{aligned} f''_{xx} &= (f'_x)'_x = e^{x^2-y}\Big[2x\Big(2x(5-2x+y) - 2\Big) + 2(5-2x+y) - 4\Big] \\ f''_{yy} &= (f'_y)'_y = e^{x^2-y}\Big[(-1)\Big((-1)(5-2x+y) + 1\Big) - 1\Big] \\ f''_{xy} &= (f'_y)'_x = e^{x^2-y}\Big[2x\Big((-1)(5-2x+y) + 1\Big) + 2\Big] \end{aligned}$$

The values of the second partial derivatives at the critical point are

$$a = f''_{xx}(1,-2) = -2e^3\,, \quad b = f''_{yy}(1,-2) = -e^3\,, \quad c = f''_{xy}(1,-2) = 2e^3\,.$$

Therefore $D = ab - c^2 = -2e^6 < 0$. By Corollary **25.1**, the only critical point is a saddle point. The function has no extreme values. $\square$

25.4. Proof of Theorem 25.3. Consider a rotation of the variables (dx, dy):

$$(dx,\ dy) = (dx' \cos\phi - dy' \sin\phi,\ dy' \cos\phi + dx' \sin\phi)$$

Following the proof of Theorem **9.1** (classification of quadric cylinders), the second differential is written in the new variables (dx', dy') as

$$\begin{aligned}
d^2 f(\mathbf{r}_0) &= a(dx)^2 + 2c\,dx\,dy + b(dy)^2 = a'(dx')^2 + 2c'dx'dy' + b'(dy')^2\\
a' &= \tfrac{1}{2}\Big(a + b + (a-b)\cos(2\phi) + 2c\sin(2\phi)\Big)\\
b' &= \tfrac{1}{2}\Big(a + b - (a-b)\cos(2\phi) - 2c\sin(2\phi)\Big)\\
2c' &= 2c\cos(2\phi) - (a-b)\sin(2\phi)
\end{aligned}$$

The rotation angle is chosen so that $c' = 0$. Put $A^2 = (a-b)^2 + 4c^2$. If $\cos(2\phi) = (a-b)/A$ and $\sin(2\phi) = 2c/A$, then $c' = 0$. With this choice,

$$a' = \tfrac{1}{2}(a + b + A), \quad b' = \tfrac{1}{2}(a + b - A)$$

Next note that

$$a' + b' = a + b\,, \quad a'b' = \frac{1}{4}((a+b)^2 - A^2) = ab - c^2\,.$$

On the other hand, the roots of the quadratic equation $P_2(\lambda) = 0$ also satisfy the same conditions

$$\lambda_1 + \lambda_2 = a + b\,, \quad \lambda_1\lambda_2 = ab - c^2\,.$$

Thus, $a' = \lambda_1$, $b' = \lambda_2$, and

$$d^2 f(\mathbf{r}_0) = \lambda_1(dx')^2 + \lambda_2(dy')^2$$

If λ_1 and λ_2 are strictly positive, then $d^2f(\mathbf{r}_0) > 0$ for all $(dx, dy) \neq (0,0)$ and by Theorem **25.2** the function has a local minimum at $\mathbf{r}_0$. If λ_1 and λ_2 are strictly negative, then $d^2f(\mathbf{r}_0) < 0$ for all $(dx, dy) \neq (0,0)$ and by Theorem **25.2** the function has a local maximum at $\mathbf{r}_0$. If λ_1 and λ_2 do not vanish but have opposite signs, $\lambda_1\lambda_2 < 0$, then in a neighborhood of $\mathbf{r}_0$, the graph of f looks like

$$z = f(\mathbf{r}_0) + \lambda_1(x' - x_0')^2 + \lambda_2(y' - y_0')^2$$

where the coordinates (x', y') are obtained from (x, y) by rotation through the angle ϕ. When λ_1 and λ_2 have different signs, this surface is a hyperbolic paraboloid (a saddle), and f has neither a local minimum nor maximum. Case (4) is easily proved by examples (see Study Problem **25.3**).

25.5. Study Problems.

Problem **25.1**. *Find all critical points of the function* $f(x, y) = \sin(x)\sin(y)$ *and determine whether* f *has a local maximum, minimum, or saddle at them.*

SOLUTION: The function has continuous partial derivative of any order on the whole plane. So, its critical points are points where the gradient of f vanishes, and the second-derivative test applies to study their nature.
Critical points are solutions of the system of equations:

$$\begin{cases} f'_x = \cos(x)\sin(y) = 0 \\ f'_y = \sin(x)\cos(y) = 0 \end{cases}$$

The first equation is satisfied if either $x = \frac{\pi}{2} + \pi n$ or $y = \pi m$, where n and m are integers. So the system is equivalent two systems of equations

$$\begin{cases} x = \frac{\pi}{2} + \pi n \\ \sin(x)\cos(y) = 0 \end{cases} \quad \text{or} \quad \begin{cases} y = \pi m \\ \sin(x)\cos(y) = 0 \end{cases}$$

Since $\sin(\frac{\pi}{2} + \pi n) = (-1)^n$, the second equation of the first system is equivalent to $\cos y = 0$ or $y = \frac{\pi}{2} + \pi m$. The second equation of the second system is equivalent to $\sin x = 0$ or $x = \pi n$. Thus, for any pair of integers n and m, the points

$$\mathbf{r}_{nm} = \langle \tfrac{\pi}{2} + \pi n, \tfrac{\pi}{2} + \pi m \rangle\,, \quad \mathbf{r}'_{nm} = \langle \pi n,\ \pi m \rangle$$

are critical points of the function.
The second-derivative test has to be applied to each critical point. The second partial derivatives are

$$f''_{xx} = -\sin(x)\sin(y), \quad f''_{yy} = -\sin(x)\sin(y), \quad f''_{xy} = \cos(x)\cos(y)$$

For the critical points $\mathbf{r}_{nm}$, one has

$$\begin{aligned} a &= f''_{xx}(\mathbf{r}_{nm}) = -(-1)^{n+m}\,, \quad b = f''_{yy}(\mathbf{r}_{nm}) = -(-1)^{n+m} = a\,, \\ c &= f''_{xy}(\mathbf{r}_{nm}) = 0\,. \end{aligned}$$

The characteristic equation is $(a - \lambda)^2 = 0$ and, hence,

$$\lambda_1 = \lambda_2 = -(-1)^{n+m}\,.$$

If $n + m$ is *even*, then the roots are negative and $f(\mathbf{r}_{nm}) = 1$ is a *local maximum*. If $n + m$ is *odd*, then the roots are positive and $f(\mathbf{r}_{nm}) = -1$ is a *local minimum*. For the critical points $\mathbf{r}'_{nm}$, one has

$$a = f''_{xx}(\mathbf{r}'_{nm}) = 0\,, \quad b = f''_{yy}(\mathbf{r}'_{nm}) = 0\,, \quad c = f''_{xy}(\mathbf{r}'_{nm}) = (-1)^{n+m}\,.$$

The characteristic equation $\lambda^2 - c^2 = \lambda^2 - 1 = 0$ has two roots $\lambda = \pm 1$ of opposite signs. Thus, $\mathbf{r}'_{nm}$ are *saddle points* of f. □

Problem **25.2**. *Let*

$$f(x,y) = x^2 + y^2 - 2x^2y - \frac{4x^6y^2}{(x^4+y^2)^2} \quad \text{if} \quad (x,y) \neq (0,0)$$

and $f(0,0) = 0$. *Show that, for all* (x,y)*, the following inequality holds:* $4x^4y^2 \leq (x^4+y^2)^2$. *Use it and the squeeze principle to conclude that* f *is continuous. Next, consider a line through* $(0,0)$ *and parallel to* $\hat{\mathbf{u}} = \langle \cos\varphi, \sin\varphi \rangle$ *and the values of* f *on it:*

$$F_\varphi(t) = f(t\cos\varphi, t\sin\varphi).$$

Show that $F_\varphi(0) = 0$, $F'_\varphi(0) = 0$, *and* $F''_\varphi(0) = 2$ *for all* $0 \leq \varphi \leq 2\pi$. *Thus,* f *has a minimum at* $(0,0)$ *along any straight line through* $(0,0)$. *Show that nevertheless* f *has no minimum at* $(0,0)$ *by studying its value along the parabolic curve* $(x,y) = (t,t^2)$.

SOLUTION: For any numbers A and B,

$$0 \leq (A-B)^2 = A^2 - 2AB + B^2 \quad \Rightarrow \quad 2AB \leq A^2 + B^2\,.$$

Therefore,

$$4AB = 2AB + 2AB \leq 2AB + A^2 + B^2 = (A+B)^2\,.$$

By setting $A = x^4$ and $B = y^2$, the said inequality is established:

$$4x^4y^2 \leq (x^4+y^2)^2\,.$$

The continuity of the last term in f at $(0,0)$ has to be verified. By the above inequality,

$$0 < \frac{4x^6y^2}{(x^4+y^2)^2} \leq \frac{4x^6y^2}{4x^4y^2} = x^2 \to 0 \quad \text{as} \quad (x,y) \to (0,0).$$

Thus, by the squeeze principle $f(x,y) \to f(0,0) = 0$ as $(x,y) \to (0,0)$, and f is continuous everywhere. If $\varphi = 0$ or $\varphi = \pi$, that is, the line coincides with the x axis, $(x,y) = (t,0)$, one has $F_\varphi(t) = t^2$, from which it follows that $F_\varphi(0) = F'_\varphi(0) = 0$ and $F''_\varphi(0) = 2$. When φ is not equal to 0 or π so that $\sin\varphi \neq 0$, one has

$$F_\varphi(t) = t^2 + at^3 + \frac{bt^4}{(1+ct^2)^2} = t^2 + at^3 + bt^4(1 + O(t^2)),$$

$$a = -2\cos^2\varphi\sin\varphi, \quad b = -\frac{4\cos^6\varphi}{\sin^2\varphi}, \quad c = \frac{\cos^4\varphi}{\sin^2\varphi}.$$

A straightforward differentiation shows that $F_\varphi(0) = F'_\varphi(0) = 0$ and $F''_\varphi(0) = 2$ as stated, and $F_\varphi(t)$ has an absolute minimum at $t = 0$, or f attains an absolute minimum at $(0,0)$ along any straight line through $(0,0)$. Nevertheless, the latter *does not imply that* f *has a minimum at* $(0,0)$! Indeed, along the parabola $(x,y) = (t,t^2)$, the function f behaves as

$$f(t,t^2) = -t^4,$$

which attains an *absolute maximum* at $t = 0$. Thus, along the parabola, f has a maximum value at the origin and hence cannot have a local minimum there. The problem illustrates the remark given after Definition **25.3**. □

Problem **25.3**. *Suppose that a function f has continuous second partial derivatives and a critical point at which one of the roots of the characteristic equation or both vanish. Prove Part (4) of the second-derivative test by giving examples of f that has a local maximum, or a local minimum or a saddle at a critical point.*

SOLUTION: Consider the function

$$f(x, y) = x^2 + sy^4 ,$$

where s is a number. Since $f'_x = 2x$ and $f'_y = 4sy^3$, it has one critical point $(0, 0)$. The second partial derivatives at the critical point are

$$a = f''_{xx}(0, 0) = 2 , \quad b = f''_{yy}(0, 0) = 0 , \quad c = f''_{xy}(0, 0) = 0 .$$

Therefore, one of the roots of the characteristic equation is zero:

$$P_2(\lambda) = -(2 - \lambda)\lambda = 0 \quad \Rightarrow \quad \lambda_1 = 2 , \quad \lambda_2 = 0 .$$

If $s > 0$, then $f(x, y) \geq 0$ for all (x, y) and f has a minimum at $(0, 0)$. Let $s = -1$. Then the curves $x = \pm y^2$ divide any disk centered at the origin into four sectors with the vertex at the critical point (the origin). In the sectors containing the x axis, $f(x, y) \geq 0$, whereas in the sectors containing the y axis, $f(x, y) \leq 0$. Thus, in this case f cannot have a local extreme value at the critical point. In particular the graph of f has the shape of a saddle. By the similar reasoning, the function

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

has a maximum value at $(0, 0)$ if $s > 0$. If $s = -1$, the graph of f has the shape of a saddle. So, if one of the roots vanishes, then f may have a local maximum or a local minimum, or a saddle. The same conclusion is reached when $\lambda_1 = \lambda_2 = 0$ by studying the functions

$$f(x, y) = \pm(x^4 + sy^4)$$

along the similar lines of arguments. □

Problem **25.4**. *For functions of one variable it is impossible for a continuous function to have two local minima (or two local maxima) and no local maximum (or no local minimum). However for functions of several variables such functions exist. Show that the function*

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

has two local minima and no local maximum.

SOLUTION: Critical points. The function f is a polynomial and, hence, differentiable everywhere. Therefore its critical points are points where the gradient vanishes:

$$\begin{cases} f'_x = 2y^2(xy^2 - y - 1) = 0 \\ f'_y = 4y(y^2 - 1) + 2(2xy - 1)(xy^2 - y - 1) = 0 \end{cases}$$

It follows from the first equation that either $y = 0$ or $xy^2 - y - 1 = 0$. But the second equation has no solution if $y = 0$ because its right side is equal to $2 \neq 0$ for any x. So, $y \neq 0$ for any critical point and the above system is equivalent to

$$\begin{cases} xy^2 - y - 1 = 0 \\ \quad\; y^2 - 1 = 0 \end{cases} \quad \Rightarrow \quad (x, y) = (2, 1)\,, \quad (x, y) = (0, -1)\,.$$

Thus, the function has two critical points.
Second-derivative test. The second partial derivatives are

$$\begin{aligned} f''_{xx} &= 2y^4\,, \\ f''_{yy} &= 4(y^2 - 1) + 8y^2 + 4x(xy^2 - y - 1) + 2(2xy - 1)^2\,, \\ f''_{xy} &= 4y(xy^2 - y - 1) + 2y^2(2xy - 1)\,. \end{aligned}$$

Their values at the critical point $(2, 1)$ are

$$\begin{aligned} &a = f''_{xx}(2,1) = 2\,, \quad b = f''_{yy}(2,1) = 26\,, \quad c = f''_{xy}(2,1) = 6 \\ \Rightarrow\ &D = ab - c^2 = 52 - 36 = 16 > 0 \quad \text{and} \quad a = 2 > 0\,. \end{aligned}$$

By Corollary **25.1**, the function has a local minimum value at $(2, 1)$. The values of the second partial derivatives at the second critical point are

$$\begin{aligned} &a = f''_{xx}(0,-1) = 2\,, \quad b = f''_{yy}(0,-1) = 10\,, \quad c = f''_{xy}(0,-1) = -2 \\ \Rightarrow\ &D = ab - c^2 = 20 - 4 = 16 > 0 \quad \text{and} \quad a = 2 > 0\,. \end{aligned}$$

By Corollary **25.1**, the function has a local minimum value at $(2, 1)$. By Theorem **25.1** the function f cannot have other critical points and, hence, it has no local maximum value. □

Problem **25.5**. (The least square method)
Suppose that a scientist has a reason to believe that two quantities x and y are related linearly, $y = mx + b$ where m and b are unknown constants. The scientist performs an experiment and collects data as points on the plane (x_i, y_i), $i = 1, 2, ..., N$. Since the data contain errors, the points do not lie on a straight line. Let $d_i = y_i - (mx_i + b)$ be the vertical deviation of the point (x_i, y_i) from the line $y = mb + x$. The method of least squares determines the constants m and b by demanding that the sum of squares $\sum d_i^2$ attains its minimal value, thus providing the "best" fit to the data points. Find m and b.

SOLUTION: Consider the function $f(m,b) = \sum_{i=1}^N d_i^2$. Its critical points satisfy the equations

$$f'_b = -2\sum_{i=1}^N d_i = 0\,, \quad f'_m = -2\sum_{i=1}^N x_i d_i = 0\,,$$

because $(d_i)'_b = -1$ and $(d_i)'_m = -x_i$. The substitution of $d_i = y_i - (mx_i + b)$ into these equations yields the following system

$$m\sum_{i=1}^N x_i + bN = \sum_{i=1}^N y_i\,, \qquad m\sum_{i=1}^N x_i^2 + b\sum_{i=1}^N x_i = \sum_{i=1}^N x_i y_i$$

The solution of this system determines the slope m and the constant b of the least square linear fit to the data points:

$$m = \frac{\langle x\rangle\langle y\rangle - \langle xy\rangle}{\langle x\rangle^2 - \langle x^2\rangle}\,, \qquad b = \langle y\rangle - \langle x\rangle m\,,$$

where $\langle a\rangle$ denotes the mean value of a quantity a:

$$\langle a\rangle = \frac{1}{N}\sum_{i=1}^N a_i\,.$$

For example $\langle xy\rangle = (x_1y_1 + \cdots + x_Ny_N)/N$. Note that the second-derivative test here is not really necessary to conclude that f has a minimum value at the critical point. Explain why! □

25.6. Exercises.

1. Suppose that $(0,0)$ is a critical point of a function f with continuous second partial derivatives. In each case, what can be said about the nature of the critical point:

(i) $f''_{xx}(0,0) = -3\,, \quad f''_{xy}(0,0) = 2\,, \quad f''_{yy}(0,0) = -2\,;$
(ii) $f''_{xx}(0,0) = 3\,, \quad f''_{xy}(0,0) = 2\,, \quad f''_{yy}(0,0) = 2\,;$
(iii) $f''_{xx}(0,0) = 1\,, \quad f''_{xy}(0,0) = 2\,, \quad f''_{yy}(0,0) = 2\,;$
(iv) $f''_{xx}(0,0) = 2\,, \quad f''_{xy}(0,0) = 2\,, \quad f''_{yy}(0,0) = 2\,.$

2–27. For each of the following functions, find all critical points and determine if they are a relative maximum, a relative minimum, or a saddle point:

2. $f(x,y) = x^2 + (y-2)^2\,;$
3. $f(x,y) = x^2 - (y-2)^2\,;$
4. $f(x,y) = x^4 - 2x^2 - y^3 + 3y\,;$
5. $f(x,y) = \frac{1}{2}x^2 + \frac{1}{3}y^3 - xy\,;$
6. $f(x,y) = x^2 - xy + y^2 - 2x + y\,;$
7. $f(x,y) = \frac{1}{3}x^3 + y^2 - x^2 - 3x - y + 1\,;$
8. $f(x,y) = x^3 + y^3 - 3xy\,;$
9. $f(x,y) = xy + 50/x + 20/y$, $x > 0$, $y > 0\,;$
10. $f(x,y) = x^2 + y^2 + \frac{1}{x^2y^2}\,;$

11. $f(x,y) = \cos(x)\cos(y)$;
12. $f(x,y) = \cos x + y^2$;
13. $f(x,y) = y^3 + 6xy + 8x^3$;
14. $f(x,y) = x^3 - 2xy + y^2$;
15. $f(x,y) = xy(1-x-y)$;
16. $f(x,y) = x\cos y$;
17. $f(x,y) = xy\sqrt{1 - x^2/a^2 - y^2/b^2}$;
18. $f(x,y) = (ax+by+c)/\sqrt{1+x^2+y^2}$, $a^2+b^2+c^2 \neq 0$;
19. $f(x,y) = (5x+7y-25)e^{-x^2-xy-y^2}$;
20. $f(x,y) = \sin x + \sin y + \cos(x+y)$;
21. $f(x,y) = \frac{1}{3}x^3 + xy^2 - x^2 - y^2$;
22. $f(x,y) = \frac{1}{3}y^3 + xy + \frac{8}{3}x^3$;
23. $f(x,y) = x^2 + xy + y^2 - 4\ln x - 10\ln y$;
24. $f(x,y) = xy\ln(x^2+y^2)$;
25. $f(x,y) = x + y + \sin(x)\sin(y)$;
26. $f(x,y) = \sin(x) + \cos(y) + \cos(x-y)$;
27. $f(x,y) = x - 2y + \ln(\sqrt{x^2+y^2}) + 3\tan^{-1}(y/x)$.

28-30. Let the function $z = z(x,y)$ be defined implicitly by the given equation. Use the implicit differentiation to find extreme values of $z(x,y)$:

28. $x^2 + y^2 + z^2 - 2x + 2y - 4z - 10 = 0$;
29. $x^2 + y^2 + z^2 - xz - yz + 2x + 2y + 2z - 2 = 0$;
30. $(x^2+y^2+z^2)^2 = a^2(x^2+y^2-z^2)$.

31. Find the point on the plane $x - y + z = 4$ that is closest to the point $(1,2,3)$. Hint: minimize the distance from $(1,2,3)$ to a generic point of the plane.

32. A rectangular box without a lid is to be made from 12 m^2 of cardboard. Find the maximum volume of such a box. Hint: express one of the dimensions of the box as a function of the others using that the total area of the box faces is 12 m^2 and then maximize the volume.

26. Maximum and Minimum Values on a Set

26.1. Second-Derivative Test for Multivariable Functions. Theorem **25.2** holds for any number of variables as it is stated and there is a multi-variable analog of the second derivative test (Theorem **25.3**). As in the two-variable case, the values of the second partial derivatives at a critical point

$$f''_{x_ix_j}(\mathbf{r}_0) = D_{ij}$$

can be arranged into a $m \times m$ matrix. By Clairaut's theorem, this matrix is symmetric $D_{ij} = D_{ji}$. The polynomial of degree m,

$$P_m(\lambda) = \det \begin{pmatrix} D_{11}-\lambda & D_{12} & D_{13} & \cdots & D_{1m} \\ D_{21} & D_{22}-\lambda & D_{23} & \cdots & D_{2m} \\ D_{31} & D_{32} & D_{33}-\lambda & \cdots & D_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D_{m1} & D_{m2} & D_{m3} & \cdots & D_{mm}-\lambda \end{pmatrix},$$

is called the *characteristic* polynomial of the matrix of second partial derivatives. The following facts are established by methods of linear algebra:

(1) The characteristic polynomial of a real symmetric $m \times m$ matrix has m real roots λ_1, λ_2,..., λ_m (a root of multiplicity k is counted k times);

(2) There exists a rotation

$$d\mathbf{r} = \langle dx_1, dx_2, ..., dx_m\rangle \to d\mathbf{r}' = \langle dx'_1, dx_2, ..., dx'_m\rangle\,,$$

which is a linear homogeneous transformation that preserves the length $\|d\mathbf{r}\| = \|d\mathbf{r}'\|$, such that

$$\begin{aligned} d^2f(\mathbf{r}_0) &= \sum_{i=1}^{m}\sum_{j=1}^{m} f''_{x_ix_j}(\mathbf{r}_0)dx_idx_j = \sum_{i=1}^{m}\sum_{j=1}^{m} D_{ij}dx_idx_j \\ &= \lambda_1(dx'_1)^2 + \lambda_2(dx'_2)^2 + \cdots + \lambda_m(dx'_m)^2 \end{aligned}$$

(3) The roots of the characteristic polynomial satisfy the conditions:

$$\begin{aligned} \lambda_1 + \lambda_2 + \cdots + \lambda_m &= D_{11} + D_{22} + \cdots + D_{mm} \\ \lambda_1\lambda_2\cdots\lambda_m &= \det D \end{aligned}$$

Fact (2) implies that if all roots of the characteristic polynomial are strictly positive, then $d^2f(\mathbf{r}_0) > 0$ for all $\|d\mathbf{r}\| = \|d\mathbf{r}'\| \neq 0$ and, hence, $f(\mathbf{r}_0)$ is a local minimum by Theorem **25.2**. Similarly, if all the roots are strictly negative, then $f(\mathbf{r}_0)$ is a local maximum. Corollary **25.1** follows from Fact (3) for $m = 2$. For $m > 2$ these properties of the roots are insufficient to establish a multi-variable analog of Corollary **25.1**. Fact (3) also implies that if $\det D = 0$, then one or more roots are 0. Hence, $d^2f(\mathbf{r}_0) = 0$ for some $d\mathbf{r} \neq \mathbf{0}$, and the hypotheses of Theorem **25.2** are not fulfilled.

THEOREM **26.1**. (Second-Derivative Test for m Variables).
Let $\mathbf{r}_0$ be a critical point of f and suppose that f has continuous second partial derivatives in an open ball centered at $\mathbf{r}_0$. Let λ_i, $i = 1, 2, ..., m$, be roots of the characteristic polynomial $P_m(\lambda)$ of the matrix of second derivatives $D_{ij} = f''_{x_i x_j}(\mathbf{r}_0)$.

(1) *If all the roots are strictly positive, $\lambda_i > 0$, then f has a local minimum at $\mathbf{r}_0$.*
(2) *If all the roots are strictly negative, $\lambda_i < 0$, then f has a local maximum at $\mathbf{r}_0$.*
(3) *If two roots have different signs, then f has no local minimum or maximum at $\mathbf{r}_0$.*
(4) *If some of the roots vanish (or $\det D = 0$), then f may have a local maximum, or a local minimum, or none of the above (the test is inconclusive).*

In case (3) the difference $f(\mathbf{r}) - f(\mathbf{r}_0)$ changes its sign in a neighborhood of $\mathbf{r}_0$. It is an m dimensional analog of a *saddle point*. In general, roots of $P_m(\lambda)$ are found numerically. If some of the roots are guessed, then a synthetic division can be used to reduce the order of the equation. If $P_m(\lambda_1) = 0$, then there is a polynomial Q_{m-1} of degree $m-1$ such that $P_m(\lambda) = (\lambda - \lambda_1)Q_{m-1}(\lambda)$ so that the other roots satisfy $Q_{m-1}(\lambda) = 0$. The signs of the roots can also be established by a graphical method (an example is given in Study Problem **26.1**).

EXAMPLE **26.1**. *Investigate the function $f(x, y, z) = \frac{1}{3}x^3 + \frac{1}{2}y^2 + z^2 + xy + 2z$ for extreme values.*

SOLUTION: The functions is a polynomial so it has continuous partial derivatives of any order everywhere. So its critical points satisfy the equations:

$$\begin{cases} f'_x = x^2 + y = 0 \\ f'_y = y + x \;\; = 0 \\ f'_z = 2z + 2 = 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} x^2 = x \\ y \;\; = -x \\ z \;\; = -1 \end{cases}$$

The first equation has two solutions $x = 0$ and $x = 1$. So the function has two critical points $\mathbf{r}_1 = \langle 0, 0, -1\rangle$ and $\mathbf{r}_2 = \langle 1, -1, -1\rangle$. The second order partial derivatives are

$$f''_{xx} = 2x, \quad f''_{xy} = f''_{yy} = 1, \quad f''_{xz} = f''_{yz} = 0, \quad f''_{zz} = 2$$

For the critical point $\mathbf{r}_1$, the characteristic polynomial

$$P_3(\lambda) = \det\begin{pmatrix} -\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{pmatrix} = (2-\lambda)(\lambda^2 - \lambda - 1)$$

has the roots 2 and $(1 \pm \sqrt{5})/2$. They do not vanish but have different signs. So $\mathbf{r}_1$ is a saddle point of f (no extreme value). For the critical point $\mathbf{r}_2$, the characteristic polynomial

$$P_3(\lambda) = \det \begin{pmatrix} 2-\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{pmatrix} = (2-\lambda)(\lambda^2 - 3\lambda + 1)$$

has positive roots $2 > 0$ and $(3 \pm \sqrt{5})/2 > 0$. So $f(1, -1, -1) = -7/6$ is a local minimum. □

26.2. When the second derivative test is inconclusive. Suppose that a function f has an isolated critical point $\mathbf{r}_0$ (that is, a neighborhood of $\mathbf{r}_0$ of a small enough radius has no other critical points). Furthermore, suppose that f has continuous partial derivatives of sufficiently high orders in a neighborhood of $\mathbf{r}_0$. Since a local extreme value occurs only at critical points, by Definition **25.1** f has a local maximum value at $\mathbf{r}_0$ if

$$f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0) < 0 \quad \text{for all} \quad 0 < \|d\mathbf{r}\| < \delta$$

for some $\delta > 0$ (δ is the radius of a neighborhood of $\mathbf{r}_0$). Similarly, f has a local minimum value at $\mathbf{r}_0$ if

$$f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0) > 0 \quad \text{for all} \quad 0 < \|d\mathbf{r}\| < \delta\,.$$

By the Taylor theorem, the local behavior of a function near a critical point $\mathbf{r}_0$ is determined by higher order differentials $d^n f(\mathbf{r}_0)$ which are *polynomials* in the variable $d\mathbf{r}$:

$$f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0) = \tfrac{1}{2} d^2 f(\mathbf{r}_0) + \tfrac{1}{6} d^3 f(\mathbf{r}_0) + \cdots + \tfrac{1}{n!} d^n f(\mathbf{r}_0) + O(\|d\mathbf{r}\|^{n+1})\,.$$

If δ is small enough and $d^2 f(\mathbf{r}_0) \neq 0$ for all $d\mathbf{r} \neq \mathbf{0}$, then the higher order terms can be neglected and the sign of the difference is determined by the sign of $d^2 f(\mathbf{r}_0)$ (all roots of the characteristic polynomial are either positive (a local minimum) or negative (a local maximum)). Suppose that $d^2 f(\mathbf{r}_0) \geq 0$ for all $d\mathbf{r}$, but $d^2 f(\mathbf{r}_0) = 0$ for some $d\mathbf{r}$ (the case $d^2 f(\mathbf{r}_0) \leq 0$ can be treated similarly). It is then possible that the difference $f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0)$ could be negative for those $d\mathbf{r}$ for which $d^2 f(\mathbf{r}_0) = 0$ and the function does not have a local minimum at $\mathbf{r}_0$. To determine the sign of the difference, the sign of higher order differentials has to be investigated for such $d\mathbf{r}$, that is, the sign of $d^3 f(\mathbf{r}_0)$ for those $d\mathbf{r}$ for which $d^2 f(\mathbf{r}_0) = 0$. If $d^2 f(\mathbf{r}_0) = d^3 f(\mathbf{r}_0) = 0$ for some $d\mathbf{r}$, then the sign of $d^4 f(\mathbf{r}_0)$ for these values of $d\mathbf{r}$ has to be determined, and so on until the conclusion about the nature of the critical point is reached. It is generally easier to study the concavity of a polynomial than that of a general function.

Example **26.2**. *Show that* $(0, 0)$ *is a critical point of the function* $f(x, y) = (1 + x^4 + y^2 + y^3)^{1/2}$ *and use a suitable Taylor approximation to determine whether* f *has a local maximum, minimum, or saddle at* $(0, 0)$.

Solution: The partial derivatives

$$f'_x = \frac{2x^3}{(1+x^4+y^2+y^3)^{1/2}}\,, \quad f'_y = \frac{2y+3y^2}{2(1+x^4+y^2+y^3)^{1/2}}$$

vanish at $(x,y) = (0,0)$. Therefore it is a critical point. Put $u = x^4+y^2+y^3$. Then

$$f(x,y) = (1+u)^{1/2} = 1 + \frac{1}{2}u + \varepsilon(u)u\,, \quad \lim_{u\to 0}\varepsilon(u) = 0\,.$$

It follows that

$$f(x,y) - f(0,0) = \frac{1}{2}u(1+2\varepsilon(u))$$

For small enough u, $1+2\varepsilon(u) > 0$ and therefore the sign of this difference in a neighborhood of the critical point is determined by the sign of u. For small enough $y \neq 0$, $y^2+y^3 > 0$ and, hence, $u = x^4+y^2+y^3 > 0$ for all $(x,y) \neq (0,0)$ in a neighborhood of $(0,0)$ and the function has a local minimum at $(0,0)$. □

The reader can verify that the second derivative test is inconclusive in this problem by a direct calculation of second partial derivatives. Using the expansion $(1+u)^{1/2} = 1+\frac{1}{2}u-\frac{1}{8}u^2+O(u^3)$, the Taylor polynomials about the origin can be obtained by retaining the monomials in u and u^2 of the degrees corresponding to the degree of the Taylor polynomial as explained in Example **23.7** and Study Problem **23.2**:

$$T_2 = 1+\frac{1}{2}y^2\,, \quad T_3 = T_2 + \frac{1}{2}y^3\,, \quad T_4 = T_3 + \frac{1}{2}x^4 - \frac{1}{8}y^4\,.$$

It follows from the definition of Taylor polynomials that

$$\begin{aligned} \frac{1}{2!}d^2f(0,0) &= \frac{1}{2}(dy)^2\,, \\ \frac{1}{3!}d^3f(0,0) &= \frac{1}{2}(dy)^3\,, \\ \frac{1}{4!}d^4f(0,0) &= \frac{1}{2}(dx)^4 - \frac{1}{8}(dy)^4\,. \end{aligned}$$

Note that $d^2f(0,0) \geq 0$, but $d^2f(0,0) = 0$ for all $d\mathbf{r} = \langle dx, 0\rangle$, $dx \neq 0$. Therefore $d^2f(0,0)$ does not determine the sign of the difference $f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0)$ for the critical point $\mathbf{r}_0 = \mathbf{0}$ if $d\mathbf{r}$ is parallel to the x axis, and approximations by higher order differentials are needed. Since $d^3f(0,0)$ also vanishes for $d\mathbf{r}$ parallel to the x axis, one has to study the sign of $d^4f(0,0)$. The latter happens to be positive for $d\mathbf{r} = \langle dx, 0\rangle$, $dx \neq 0$, so that $f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0) > 0$ for all sufficiently small $\|d\mathbf{r}\| \neq 0$. Note that if the term x^4 in the function is replaced by $-x^4$, then the sign of $(dx)^4$ in $d^4f(0,0)$ becomes negative and the function does not have a local extreme value at $(0,0)$ because $f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0) < 0$ for $d\mathbf{r}$ parallel to the x axis.

In general, the directions of $d\mathbf{r}$ along which $d^2f(\mathbf{r}_0) = 0$ do not coincide with one (or more) coordinate axes. Such directions are determined by the rotation of the coordinate system that brings $d^2f(\mathbf{r}_0)$ to the standard form as noted in the beginning of the section (or in the proof of the second-derivative test for functions of two variables).

Non-isolated critical points. There are functions whose critical points are not isolated. Consider the function

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

The function is a polynomial (it has continuous partial derivatives of any order) and therefore all its critical points are solutions to the system of equations

$$\begin{cases} f'_x = 2x(y-1)^2 = 0 \\ f'_y = 2x^2(y-1) = 0 \end{cases} \quad \Rightarrow \quad (x, y) = (0, p) \quad \text{and} \quad (x, y) = (q, 1)$$

where p and q are any real numbers. Note that the critical points form the level set $f(x, y) = 0$ which consists of two perpendicular lines intersecting at the point $(0, 1)$. The second partial derivatives are

$$f''_{xx} = 2(y-1)^2 , \quad f''_{yy} = 2x^2 , \quad f''_{xy} = 4x(y-1)$$

are continuous and the second derivative test applies. For critical points $(0, p)$, their values are

$$a = f''_{xx}(0, p) = 2(p-1)^2 , \quad b = f''_{yy}(0, p) = 0 , \quad c = f''_{xy}(0, p) = 0 .$$

Therefore $D = ab - c^2 = 0$ and the second-derivative test is inconclusive. Similarly, this test is inconclusive for points $(q, 1)$:

$$a = f''_{xx}(q, 1) = 0 , \quad b = f''_{yy}(q, 1) = 2q^2 , \quad c = f''_{xy}(q, 1) = 0 \quad \Rightarrow \quad D = 0 .$$

However, $f(x, y) \geq 0$ for any point and hence in any neighborhood of any critical point $f(x, y) \geq f(0, p) = 0$ and $f(x, y) \geq f(q, 1) = 0$ and the function attains its minimal value at all critical points.

If a function f has a constant value along a curve of its critical points which is, say, a local minimum value, then the graph $z = f(\mathbf{r})$ in a neighborhood of a critical point does not look like a concave upward paraboloid, but rather it resembles a valley in mountains (local minima of the height are along the valley bottom). Consequently, if the variables $d\mathbf{r}$ are such that $\mathbf{r}_0 + d\mathbf{r}$ lies on the curve of critical points, then the difference $f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0) = 0$ vanishes for all such $d\mathbf{r}$ and so should all the differentials $d^n f(\mathbf{r}_0) = 0$. The analysis based on the sign of higher-order differentials becomes inconclusive. However, if the curve of critical points is identified, then variations of the values of the function can be studied in the directions that are not tangent to the curve. The sign of higher-order differentials can be determined for such directions of $d\mathbf{r}$ and the nature of critical points can be established.

In the above example, the critical points $\mathbf{r}_0 = \langle 0, p\rangle$ form the y axis. Therefore $d\mathbf{r} = \langle dx, 0\rangle$, $dx \neq 0$, is perpendicular to this line and

$$f(dx, p) - f(0, p) = (dx)^2(p-1)^2 > 0 , \quad p \neq 1 .$$

So, the function has a minimum at $(0, p)$, $p \neq 1$. Similarly, the critical points $(q, 1)$ form a line parallel to the x axis and the vector $d\mathbf{r} = \langle 0, dy \rangle$, $dy \neq 0$, is perpendicular to it. Then

$$f(q, 1 + dy) - f(q, 1) = q^2 (dy)^2 > 0 \,, \quad q \neq 0 \,.$$

So, the function has a minimum at $(q, 1)$, $q \neq 0$. At the point $(0, 1)$, $f(dx, 1 + dy) - f(0, 1) = (dx)^2 (dy)^2 > 0$ and, hence, the function also has a minimum.

In general, if $\hat{\mathbf{v}}$ is a unit vector tangent to a curve formed by critical points at a particular critical point $\mathbf{r}_0$, then one has to investigate $f(\mathbf{r}_0 + d\mathbf{r}) - f(\mathbf{r}_0)$ where $d\mathbf{r}$ are not independent but subject to the condition $\hat{\mathbf{v}} \cdot d\mathbf{r} = 0$.

26.3. Absolute Maximum and Minimum Values. For a function f of one variable, the *extreme value theorem* says that if f is continuous on a closed interval $[a, b]$, then f has an absolute minimum value and an absolute maximum value (Calculus I). For example, the function $f(x) = x^2$ on $[-1, 2]$ attains an absolute minimum value at $x = 0$ and an absolute maximum value at $x = 2$. The function is differentiable for all x and therefore its critical points are determined by $f'(x) = 2x = 0$. So the absolute minimum value occurs at the critical point $x = 0$ inside the interval, while the absolute maximum value occurs on the boundary of the interval that is not a critical point of f. Thus, to find the absolute maximum and minimum values of a function f in a closed interval in the domain of f, the values of f must be evaluated and compared not only at the critical points but also at the boundaries of the interval.

The situation for multivariable functions is similar. For example, the function $f(x, y) = x^2 + y^2$ whose arguments are restricted to the square $D = [0, 1] \times [0, 1]$ attains its absolute maximum and minimum values on the boundary of D as shown in the left panel of Fig. **26.1**.

Definition **26.1**. (Closed Set).
A set D in a Euclidean space is said to be closed if it contains all its limit points.

Let D be a part of the plane bounded by a simple closed curve C. Recall that any neighborhood of a limit point of D contains points of D. If a limit point of D is not an interior point of D, then it lies on the boundary curve C. So D is closed if it contains its boundary. All points of an open interval (a, b) are its limit points, but, in addition, the boundaries a and b are also its limit points, so when they are added, a closed set $[a, b]$ is obtained. Similarly, the set in the plane $D = \{(x, y) | x^2 + y^2 < 1\}$ has limit points on the circle $x^2 + y^2 = 1$ (the boundary of D), which is not in D. By adding these points, a closed set is obtained, $x^2 + y^2 \leq 1$.

Definition **26.2**. (Bounded Set).
A set D in a Euclidean space is said to be bounded if it is contained in some ball.

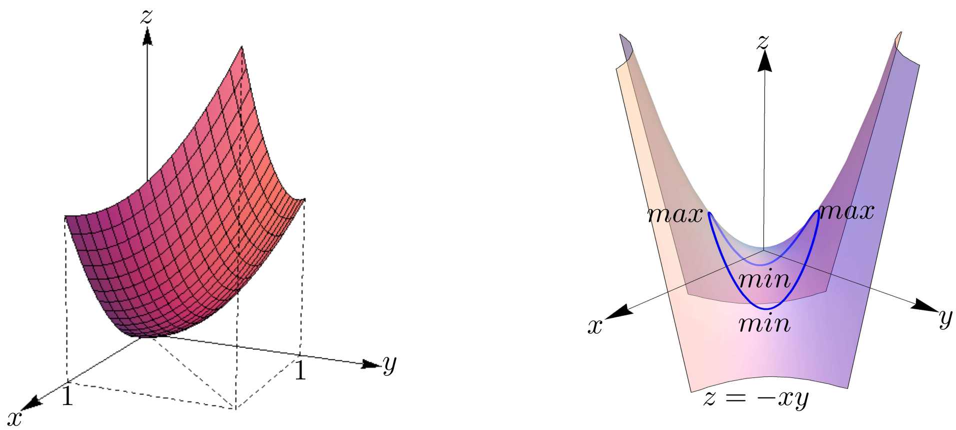


FIGURE **26.1**. **Left**: The graph $z = x^2 + y^2$ (a circular paraboloid) over the square $D = [0,1] \times [0,1]$. The function $f(x,y) = x^2+y^2$ attains its absolute maximum and minimum values on the boundary of D: $f(0,0) \le f(x,y) \le f(1,1)$ for all points in D. **Right**: The graph $z = -xy$. The values of $f(x,y) = -xy$ along the circle $x^2+y^2 = 4$ are shown by the curve on the graph. The function has two local maxima and minima the disk $x^2+y^2 \le 4$, while it has no maximum and minimum values on the entire plane.

In other words, for any two points in a bounded set, the distance between them cannot exceed some value (the diameter of the ball that contains the set).

THEOREM **26.2**. (Extreme Value Theorem).
If f is continuous on a closed, bounded set D in a Euclidean space, then f attains an absolute maximum value $f(\mathbf{r}_1)$ and an absolute minimum value $f(\mathbf{r}_2)$ at some points $\mathbf{r}_1$ and $\mathbf{r}_2$ in D.

The closedness of D is essential. For example, if the function $f(x,y) = x^2 + y^2$ is restricted to the *open* square $D = (0,1) \times (0,1)$, then it has no extreme values on D. For all (x,y) in D, $f(0,0) < f(x,y) < f(1,1)$ and there are points in D arbitrarily close to $(0,0)$ and $(1,1)$, but the points $(0,0)$ and $(1,1)$ are not in D. So, f takes values on D arbitrary close to 0 and 2, never reaching them. The boundedness of D is also crucial. For example, if the function $f(x,y) = x^2 + y^2$ is restricted to the first quadrant, $x \ge 0$ and $y \ge 0$, then f has no maximum value on D. It should be noted that the continuity of f and the closedness and boundedness of D are *sufficient* (not necessary) conditions for f to attain its absolute extreme values on D. There are non-continuous or continuous functions on an unbounded or non-closed region D (or both) that attain their extreme values on D. Such examples are given in Calculus I and functions of a single variable may always be viewed as a particular case of functions of two or more variables: $f(x, y) = g(x)$.

The minimum and maximum values of a function f on a set D will be denoted, respectively, by $\min_D f$ and $\max_D f$. One writes

$$f(\mathbf{r}_1) = \max_D f \qquad \text{and} \qquad f(\mathbf{r}_2) = \max_D f\,,$$

if the absolute maximum and minimum occur, respectively, at $\mathbf{r}_1$ and $\mathbf{r}_2$. By the extreme value theorem, it follows that the points $\mathbf{r}_1$ and $\mathbf{r}_2$ are either critical points of f (because local extrema always occur at critical points) or lie on the boundary of D. So, to find the absolute minimum and maximum values of a continuous function f on a closed, bounded set D, one has to

Step 1. Find the values of f at the critical points of f in the interior of D.
Step 2. Find the extreme values of f on the boundary of D.
Step 3. The largest of the values obtained in Steps 1 and 2 is the absolute maximum value, and the smallest of these values is the absolute minimum value.

EXAMPLE **26.3**. *Find the absolute maximum and minimum values of $f(x, y) = x^2 + y^2 + xy$ on the disk $x^2 + y^2 \le 4$ and the points at which they occur.*

SOLUTION: The function f is continuous everywhere because it is a polynomial. The disk $x^2 + y^2 \le 4$ is a closed set in the plane. So, the hypotheses of the extreme value theorem are fulfilled.

Step 1. The function has continuous partial derivatives of any order on the whole plane. Therefore critical points of f satisfy the system of equations

$$f'_x = 2x + y = 0\,, \quad f'_y = 2y + x = 0\,,$$

which has the only solution $(x, y) = (0, 0)$. The critical point happens to be in the interior of the disk. The value of f at the critical point is $f(0, 0) = 0$.
Step 2. The boundary of the disk is the circle C: $x^2 + y^2 = 4$. To find the extreme values of f on it, take the parametric equations of C:

$$x = x(t) = 2\cos t\,, \quad y = y(t) = 2\sin t\,, \quad 0 \le t \le 2\pi\,.$$

The values of the function on the boundary are

$$F(t) = f(x(t), y(t)) = 4 + 4\cos t \sin t = 4 + 2\sin(2t)\,.$$

The function $F(t)$ attains its maximal value 6 on $[0, 2\pi]$ when $\sin(2t) = 1$ or $t = \pi/4$ and $t = \pi/4 + \pi$. These values of t correspond to the points $(\sqrt{2}, \sqrt{2})$ and $(-\sqrt{2}, -\sqrt{2})$. Similarly, $F(t)$ attains its minimal value 2 on $[0, 2\pi]$ when $\sin(2t) = -1$ or $t = 3\pi/4$ and $t = 3\pi/4 + \pi$. These values of t correspond to the points $(-\sqrt{2}, \sqrt{2})$ and $(\sqrt{2}, -\sqrt{2})$. Therefore

$$f(\sqrt{2}, \sqrt{2}) = f(-\sqrt{2}, -\sqrt{2}) = \max_C f = 6\,,$$
$$f(-\sqrt{2}, \sqrt{2}) = f(\sqrt{2}, -\sqrt{2}) = \min_C f = 2\,.$$

Step 3. Then

$$\max_D f = \max\Big\{f(0,0),\ \max_C f\Big\} = \max\{0,\ 6\} = 6\,,$$
$$\min_D f = \min\Big\{f(0,0),\ \min_C f\Big\} = \min\{0,\ 2\} = 0\,.$$

The maximum of f on D occurs at the points $(\sqrt{2},\sqrt{2})$ and $(-\sqrt{2},-\sqrt{2})$. The minimum of f on D occurs at the point $(0,0)$. □

EXAMPLE **26.4**. *Find the absolute maximum and minimum values of* $f(x,y,z) = x^2+y^2-z^2+2z$ *on the closed set* $D = \{(x,y,z)\,|\,x^2+y^2 \le z \le 4\}$.

SOLUTION: The set D is the solid bounded from below by the paraboloid $z = x^2+y^2$ and from the top by the plane $z = 4$. The set D is a bounded closed set, and f is continuous everywhere because it is a polynomial.

Step 1. Since f is differentiable everywhere, its critical points satisfy the equations

$$f'_x = 2x = 0\,, \quad f'_y = 2y = 0\,, \quad f'_z = -2z+2 = 0\,.$$

which has the only solution $(x,y,z) = (0,0,1)$. The critical point happens to be in the interior of D. The value of f at it is $f(0,0,1) = 1$.
Step 2. The boundary S of D consists of two surfaces:

$$S_1 = \{(x,y,z)\,|\,z = 4,\ x^2+y^2 \le 4\}\,,$$
$$S_2 = \{(x,y,z)\,|\,z = x^2+y^2,\ x^2+y^2 \le 4\}\,.$$

The surface S_1 is the disk of radius 2 in the plane $z = 4$ and S_2 is the portion of the paraboloid above the disk $x^2+y^2 \le 4$. The values of f on S_1 are

$$F_1(x,y) = f(x,y,4) = x^2+y^2-8\,, \quad x^2+y^2 \le 4\,.$$

The problem now is to find the extreme values of the function F_1 on the disk. In principle, at this point, Steps 1, 2, and 3 have to be applied to F_1 as a function of *two variables*. These technicalities can be avoided in this particular case by noting that $F_1(x,y) = r^2-8$, where $r^2 = x^2+y^2 \le 4$. Therefore, the maximum value of F_1 is reached when $r^2 = 4$, and its minimum value is reached when $r^2 = 0$. So

$$\max_{S_1} f = -4\,, \quad \min_{S_1} f = -8\,.$$

The values of f on S_2 are

$$F_2(x,y) = f(x,y,x^2+y^2) = 3(x^2+y^2)-(x^2+y^2)^2 = 3r^2-r^4 = g(r)\,,$$

where $0 \le r \le 2$. The critical points of $g(r)$ satisfy the equation $g'(r) = 6r-4r^3 = 0$ whose solutions for $r \ge 0$ are $r = 0$, $r = \sqrt{3/2}$. The extereme values of g can occur either at the boundaries of the interval $0 \le r \le 2$ or at the critical point $r = \sqrt{3/2}$. Therefore,

$$\max_{S_2} f = \max\{g(0),\ g(\sqrt{3/2}),\ g(2)\} = \max\{0,\ 9/4,\ -4\} = 9/4\,,$$
$$\min_{S_2} f = \min\{g(0),\ g(\sqrt{3/2}),\ g(2)\} = \min\{0,\ 9/4,\ -4\} = -4\,.$$

Step 3. The maximum and minimum values of f on the boundary S are, respectively,

$$\max_S f = \max\{-4, 9/4\} = 9/4\,, \qquad \min_S = \min\{-8, -4\} = -8\,.$$

Therefore the absolute maximum and minimum values of f on D are, respectively,

$$\max_D f = \max\{1,\ 9/4\} = 9/4\,, \qquad \min_D f = \min\{1,\ -8\} = -8\,.$$

Both extreme values of f occur on the boundary of D: $f(0,0,4) = -8$ and the absolute maximum value is attained along the circle of intersection of the plane $z = 3/2$ with the paraboloid $z = x^2 + y^2$. □

26.4. Intermediate Value Theorem. Let f be a continuous function of a single variable on an interval and $f(a) < f(b)$ for some a and b in the interval. The *intermediate value theorem* states that the function f attains all values between $f(a)$ and $f(b)$ in the interval with boundary points a and b. There is an analog of this theorem for functions of several variables.

Definition **26.3**. (Connected Set)
A set D in a Euclidean space is connected if any two points of D can be connected by a curve that lies in D.

For example, the disk $x^2 + y^2 < 4$ is a connected set in two-dimensional Euclidean space (a plane). The set D in space whose points satisfy the condition $x^2 + y^2 + z^2 \neq 1$ is not connected because any curve that connects a point inside the ball of radius 1 with a point outside this ball has to intersect the sphere $x^2 + y^2 + z^2 = 1$ whose points are not in D.

Theorem **26.3**. (Intermediate Value Theorem)
Let f be a continuous function on a connected set D in a Euclidean space. For any two points $\mathbf{r}_1$ and $\mathbf{r}_2$ in D such that $f(\mathbf{r}_1) < f(\mathbf{r}_2)$ and any number c such that $f(\mathbf{r}_1) \leq c \leq f(\mathbf{r}_2)$, there is a point $\mathbf{r}_0$ in D such that $f(\mathbf{r}_0) = c$.

Proof. Since D is connected, there exists a parametric curve $\mathbf{r} = \mathbf{r}(t)$, $t_1 \leq t \leq t_2$, in D such that $\mathbf{r}(t_1) = \mathbf{r}_1$ and $\mathbf{r}(t_2) = \mathbf{r}_2$. Since f is continuous in D, its values on the curve define a continuous function on an interval,

$$F(t) = f(\mathbf{r}(t))\,, \quad t_1 \leq t \leq t_2\,,$$

such that $f(\mathbf{r}_1) = F(t_1) < F(t_2) = f(\mathbf{r}_2)$. Therefore F takes all values between $F(t_1)$ and $F(t_2)$ in the interval $[t_1, t_2]$. Then for any number $F(t_1) \leq c \leq F(t_2)$ there exists a number $t_1 \leq t_0 \leq t_2$ such that $F(t_0) = c$. The point $\mathbf{r}_0 = \mathbf{r}(t_0)$ is in D as the curve lies in D and $f(\mathbf{r}_0) = c$. □

In particular, if f is a continuous function on a closed, bounded, and connected set D in a Euclidean space, then f attains its absolute maximum and minimum values as well as all values between them in D.

26.5. Study problems.

Problem **26.1**. *Find local extreme values of the function* $f(x, y, z) = x + y^2/(4x) + z^2/y + 2/z$ *if* $x > 0$, $y > 0$, *and* $z > 0$.

SOLUTION: The function is differentiable in the specified domain. So its critical points are solutions of the system of equations

$$f'_x = 1 - \frac{y^2}{4x^2} = 0, \quad f'_y = \frac{y}{2x} - \frac{z^2}{y^2} = 0, \quad f'_z = \frac{2z}{y} - \frac{2}{z^2} = 0$$

The first equation is equivalent to $y = 2x$ (since $x > 0$ and $y > 0$). The substitution of this relation into the second equation gives $z = y$ because $y > 0$ and $z > 0$. The substitution of the latter relation into the third equation yields $z = 1$ as $z > 0$. There is only one critical point $\mathbf{r}_0 = \langle \frac{1}{2}, 1, 1\rangle$. The second partial derivatives at $\mathbf{r}_0$ are:

$$\begin{aligned}
&f''_{xx}(\mathbf{r}_0) = \frac{y^2}{2x^3}\Big|_{\mathbf{r}_0} = 4, && f''_{xy}(\mathbf{r}_0) = -\frac{y}{2x^2}\Big|_{\mathbf{r}_0} = -2,\\
&f''_{xz}(\mathbf{r}_0) = 0, && f''_{yy}(\mathbf{r}_0) = \frac{1}{2x} + \frac{2z^2}{y^3}\Big|_{\mathbf{r}_0} = 3,\\
&f''_{yz}(\mathbf{r}_0) = -\frac{2z}{y^2}\Big|_{\mathbf{r}_0} = -2, && f''_{zz}(\mathbf{r}_0) = \frac{2}{y} + \frac{4}{z^3}\Big|_{\mathbf{r}_0} = 6
\end{aligned}$$

The characteristic equation of the second derivative matrix is

$$\begin{aligned}
\det\begin{pmatrix} 4-\lambda & -2 & 0\\ -2 & 3-\lambda & -2\\ 0 & -2 & 6-\lambda\end{pmatrix} &= (4-\lambda)[(3-\lambda)(6-\lambda) - 4] - 4(6-\lambda)\\
&= -\lambda^3 + 13\lambda^2 - 46\lambda + 32 = 0
\end{aligned}$$

First of all, $\lambda = 0$ is not a root and the second derivative test is conclusive. To analyze the sings of the roots, the following method is employed. The characteristic equation is written in the form

$$\lambda(\lambda^2 - 13\lambda + 46) = 32$$

This equation determines the points of intersection of the graph $y = g(\lambda) = \lambda(\lambda^2 - 13\lambda + 46)$ with the horizontal line $y = 32$. The polynomial $g(\lambda)$ has one simple root $g(0) = 0$ because the quadratic equation $\lambda^2 - 13\lambda + 46 = 0$ has no real roots. Therefore $g(\lambda) > 0$ if $\lambda > 0$ and $g(\lambda) < 0$ if $\lambda < 0$. This implies that the intersection of the horizontal line $y = 32 > 0$ with the graph $y = g(\lambda)$ occurs only for $\lambda > 0$. Thus, all roots of the characteristic polynomial $P_3(\lambda)$ are positive and, hence, $f(1/2, 1, 1) = 4$ is a local minimum value. $\square$

Problem **26.2**. *Find the absolute extreme values of the function in the rectangle:*

$$f(x,y)=\frac{1}{2}x^2+\frac{1}{3}y^3-xy\,,\quad 0\le x\le 2\,,\quad 0\le y\le 2\,.$$

SOLUTION: The function is continuous everywhere because it is a polynomial and the given rectangle is a closed set. So the function attains its extreme values in the rectangle.

Step 1. The function has continuous partial derivative of any order. Therefore its critical points satisfy the system of equations:

$$f'_x=x-y=0\,,\quad f'_y=y^2-x=0\,.$$

Since $x=y$ by the first equation, $x^2-x=0$ by the second equation so that the function has two critical points $(0,0)$ and $(1,1)$. The point $(0,0)$ lies on the boundary of the rectangle and can be discarded because the extreme values of f on the boundary will be studied separately. The critical point $(1,1)$ lies in the interior of the rectangle and $f(1,1)=-1/6$.

Step 2. The boundary C of the rectangle consists of four straight line segments:

$$\begin{aligned}L_1&=\{(x,y)\,|\,0\le x\le 2,\ y=2\}\\ L_2&=\{(x,y)\,|\,0\le x\le 2,\ y=0\}\\ L_3&=\{(x,y)\,|\,x=2,\ 0\le y\le 2\}\\ L_4&=\{(x,y)\,|\,x=0,\ 0\le y\le 2\}\end{aligned}$$

The values of f along L_1 are

$$F_1(x)=f(x,2)=\frac{1}{2}x^2-2x+\frac{8}{3}$$

where $0\le x\le 2$. Since $F_1'(x)=x-2\le 0$, the function is decreasing and its maximum and minimum values are

$$\max_{L_1}f=F_1(0)=f(0,2)=\frac{8}{3}\,,\qquad \min_{L_1}f=F_1(2)=f(2,2)=\frac{2}{3}\,.$$

The values of f along L_2 are

$$F_2(x)=f(x,0)=\frac{1}{2}x^2\,.$$

Therefore the maximum and minimum values of f on L_2 are

$$\max_{L_2}f=F_2(2)=f(2,0)=2\,,\qquad \min_{L_1}f=F_2(0)=f(0,0)=0\,.$$

The values of f along L_3 are

$$F_3(y)=f(2,y)=\frac{1}{3}y^3-2y+2\,.$$

The derivative $F_3'(y)=y^2-2$ vanishes at $y=\sqrt{2}$ in the interval $[0,2]$. Since $F_3(0)=2$, $F_3(2)=\frac{2}{3}$, and $F_3(\sqrt{2})=2-4\sqrt{2}/3$, the maximum and minimum values of f on L_3 are

$$\max_{L_3} f = F_3(0) = f(2,0) = 2\,, \qquad \min_{L_3} f = F(\sqrt{2}) = f(2,\sqrt{2}) = 2 - \frac{4\sqrt{2}}{3}\,.$$

Finally, the values f along L_4 are $F_4(y) = f(0,y) = \frac{1}{3}y^3$, which is monotonically increasing function. So the maximum and minimum values of f on L_4 are

$$\max_{L_4} f = F_4(2) = f(0,2) = \frac{8}{3}\,, \qquad \min_{L_4} f = F_4(0) = f(0,0) = 0\,.$$

Thus, the extreme values of f on the boundary C of D are

$$\max_{C} f = f(0,2) = \frac{8}{3}\,, \quad \min_{C} f = f(0,0) = 0\,.$$

Step 3. The maximum and minimum values of f on D are, respectively,

$$\max_{D} f = \max\{0,\ -1/6,\ 8/3\} = 8/3 = f(0,2)\,,$$
$$\min_{D} f = \min\{0,\ -1/6,\ 0\} = -1/6 = f(1,1)\,.$$

The absolute maximum occurs on the boundary of D. □

26.6. Exercises.

1–10. For each of the following functions, find all critical points and determine whether the function has a relative maximum, a relative minimum, or a saddle point at each critical point:

1. $f(x,y,z) = x^2 + y^2 + z^2 + 2x + 4y - 8z$;
2. $f(x,y,z) = x^2 + y^3 + z^2 + 12xy - 2z$;
3. $f(x,y,z) = x^2 + y^3 - z^2 + 12xy + 2z$;
4. $f(x,y,z) = \sin x + z \sin y$;
5. $f(x,y,z) = x^2 + \frac{5}{3}y^3 + z^2 - 2xy - 4zy$;
6. $f(x,y,z) = x + y^2/(4x) + z^2/y + 2/z$;
7. $f(x,y,z) = a^2/x + x^2/y + y^2/z + z^2/b,\ \ x > 0,\ y > 0,\ z > 0,\ b > 0$;
8. $f(x,y,z) = \sin x + \sin y + \sin z + \sin(x+y+z)$, where (x,y,z) is in the rectangular box $[0,\pi]\times[0,\pi]\times[0,\pi]$;
9. $f(x_1,...,x_m) = \sum_{k=1}^{m} \sin x_k$;
10. $f(\mathbf{r}) = (R^2 - \|\mathbf{r}\|^2)^2$ where $\mathbf{r} = \langle x_1,...,x_m\rangle$ and R is a constant;
11. $f(x_1,...,x_m) = x_1 + x_2/x_1 + x_3/x_2 + \cdots + x_m/x_{m-1} + 2/x_m,\ \ x_i > 0,$ $i = 1,2,...,m$.

12. Given two positive numbers a and b, find m numbers x_i, $i = 1,2,...,m$, between a and b so that the ratio

$$\frac{x_1 x_2 \cdots x_m}{(a + x_1)(x_1 + x_2)\cdots(x_m + b)}$$

is maximal.

13–21. Use multivariable Taylor polynomials to show that the origin is a critical point of each of the following functions. Determine whether the function has a local extreme value at the critical point:

13. $f(x,y) = x^2 + xy^2 + y^4$;
14. $f(x,y) = \ln(1+x^2y^2)$;
15. $f(x,y) = x^2\ln(1+x^2y^2)$;
16. $f(x,y) = xy(\cos(x^2y)-1)$;
17. $f(x,y) = (x^2+2y^2)\tan^{-1}(x+y)$;
18. $f(x,y) = \cos(e^{xy}-1)$;
19. $f(x,y) = \ln(y^2\sin^2 x + 1)$;
20. $f(x,y) = e^{x+y^2} - 1 - \sin(x-y^2)$;
21. $f(x,y,z) = \sin(xy+z^2)/(xy+z^2)$, where f is defined on the set $xy+z^2 = 0$ by the continuous extension (show that this is possible);
22. $f(x,y,z) = 2 - 2\cos(x+y+z) - x^2 - y^2 - z^2$.

23. Let $f(x,y,z) = xy^2z^3(a-x-2y-3z)$, $a > 0$. Find relative extreme values of f.

24–26. Give examples of a function f of two variables that attains its extreme values in the specified set D and has the following properties:

24. f is continuous on D and D is not closed;
25. f is not continuous on D and D is bounded and closed;
26. f is not continuous on D and D is not bounded and not closed.

27–31. For each of the following functions find the extreme values on the specified set D:

27. $f(x,y) = 1 + 2x - 3y$, D is the closed triangle with vertices $(0,0)$, $(1,2)$, and $(2,1)$;
28. $f(x,y) = x^2 + y^2 + xy^2 - 1$, $D = \{(x,y)|\,|x| \le 1,\ |y| \le 2\}$;
29. $f(x,y) = yx^2$, $D = \{(x,y)|x \ge 0,\ y \ge 0,\ x^2+y^2 \le 4\}$;
30. $f(x,y,z) = xy^2 + z$, $D = \{(x,y,z)|\,|x| \le 1,\ |y| \le 1,\ |z| \le 1\}$;
31. $f(x,y,z) = xy^2 + z$, $D = \{(x,y,z)|\,1 \le x^2+y^2 \le 4,\ -2+x \le z \le 2-x\}$.

32. Find the point on the plane $x+y-z = 1$ that is closest to the point $(1,2,3)$. Hint: Let the point in question have the coordinates (x,y,z). Then the squared distance between it and $(1,2,3)$ is $f(x,y) = (x-1)^2 + (y-2)^2 + (z-3)^2$ where $z = x+y-1$ because (x,y,z) is in the plane.

33. Find the point on the cone $z^2 = x^2 + y^2$ that is closest to the point $(1,2,3)$

34. Find an equation of the plane that passes through the point $(3,2,1)$ and cuts off the smallest volume in the first octant.

35. Find the extreme values of $f(x,y) = ax^2 + 2bxy + cy^2$ on the circle $x^2+y^2 = 1$.

36. Find the extreme values of $f(x,y,z) = x^2/a^2 + y^2/b^2 + z^2/c^2$ on the sphere $x^2+y^2+z^2 = 1$.

37. Find 2 positive numbers whose product is fixed, while the sum of their reciprocals is minimal.

38. Find m positive numbers whose product is fixed, while the sum of their reciprocals is minimal.

39. A solid object consists of a solid cylinder and a solid circular cone so that the base of the cone coincides with the base of the cylinder. If the total surface area of the object is fixed, what are dimensions of the cone and cylinder at which the object has maximal volume?

40. Find a linear approximation $y = mx+b$ to the parabola $y = x^2$ such that the deviation $\Delta = \max|x^2 - mx - b|$ is minimal in the interval $1 \leq x \leq 3$.

41. Let N points of masses m_j, $j = 1, 2, ..., N$, be positioned in a plane at $P_j = (x_j, y_j)$. Recall from Calculus II that the moment of inertia of this system about a point $P = (x, y)$ is

$$I(x, y) = \sum_{j=1}^{N} m_j |PP_j|^2$$

Find P about which the moment of inertia is minimal.

27. Lagrange Multipliers

Let $f(x, y)$ be the height of a hill at a point (x, y) in the base of the hill. A hiker walks along a path $\mathbf{r}(t) = \langle x(t), y(t)\rangle$. What are the local maxima and minima along the path? What are the maximum and minimum heights along the path? These questions are easy to answer if parametric equations of the path are explicitly known. Indeed, the height along the path is the single-variable function $F(t) = f(\mathbf{r}(t))$ and the problem is reduced to the standard extreme value problem for $F(t)$ on an interval $a \le t \le b$.

EXAMPLE **27.1**. *Find extreme values of the function $f(x, y) = -xy$ on the circle $x^2 + y^2 = 4$.*

SOLUTION: The parametric equation of the circle can be taken in the form

$$\mathbf{r}(t) = \langle 2\cos t,\ 2\sin t\rangle\,, \quad 0 \le t \le 2\pi\,.$$

The values of f on the circle are

$$F(t) = f(2\cos t, 2\sin t) = -4\cos t \sin t = -2\sin(2t)\,.$$

On the interval $[0, 2\pi]$, the function $-\sin(2t)$ attains its absolute minimum value at $t = \pi/4$ and $t = \pi/4 + \pi$ and its absolute maximum value at $t = 3\pi/4$ and $t = 3\pi/4 + \pi$. So, on the circle, the function f attains the absolute maximum value 2 at $(\sqrt{2}, -\sqrt{2})$ and $(-\sqrt{2}, \sqrt{2})$ and the absolute minimum value -2 at $(\sqrt{2}, \sqrt{2})$ and $(-\sqrt{2}, -\sqrt{2})$. The solution is illustrated in the right panel of Fig. **26.1**. □

The problem considered may be generalized in the following way:

Find extreme values of $f(x, y)$ on the set defined by the equation $g(x, y) = 0$.

In other words, only the points (x, y) that satisfy the condition $g(x, y) = 0$ are permitted in the argument of f; that is, the variables x and y are no longer independent in the extreme value problem. The condition $g(x, y) = 0$ is called a *constraint*. Problems of this type occur for functions of more than two variables. For example, let $f(x, y, z)$ be the temperature as a function of position in space. A reasonable question to ask is: What are the maximum and minimum temperatures on a surface? A surface may be described by imposing one constraint $g(x, y, z) = 0$ on the variables x, y, and z. Nothing precludes us from asking about the maximum and minimum temperatures along a curve defined as an intersection of two surfaces $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 0$. So the variables x, y, and z are now subject to two constraints.

In general, what are the extreme values of a multivariable function $f(\mathbf{r})$ whose arguments are subject to several constraints $g_a(\mathbf{r}) = 0$, $a = 1, 2, ..., M$ (assuming, of course, that the set defined by all constraints is not empty)? Evidently, the constraints cannot always be solved explicitly to obtain the values f on the set defined by the constraints. It is therefore desirable to develop a technique to find extreme values of f without solving the constraints. If constraints are level sets of differentiable functions, then this technique is known the *method of Lagrange multipliers*.

DEFINITION **27.1**. (Local Maxima and Minima Subject to Constraints). *A function f has a local maximum (or minimum) at $\mathbf{r}_0$ on the set defined by constraints $g_a(\mathbf{r}) = 0$, $a = 1, 2, ..., M$, if $f(\mathbf{r}) \le f(\mathbf{r}_0)$ (or $f(\mathbf{r}) \ge f(\mathbf{r}_0)$) for all $\mathbf{r}$ in some neighborhood of $\mathbf{r}_0$ that satisfy the constraints, that is, $g_a(\mathbf{r}) = 0$.*

The extreme values of a function subject to constraints are defined similarly.

DEFINITION **27.2**. (Extreme Values of a Function Subject to Constraints) *A function f has a maximum (or minimum) value at a point $\mathbf{r}_0$ on the set defined by constraints $g_a(\mathbf{r}) = 0$, $a = 1, 2, ..., M$, if $f(\mathbf{r}) \le f(\mathbf{r}_0)$ (or $f(\mathbf{r}) \ge f(\mathbf{r}_0)$) for all $\mathbf{r}$ such that $g_a(\mathbf{r}) = 0$.*

Note that a function f may not have local maxima or minima in its domain. However, when its arguments become subject to constraints, it may well have local maxima and minima on the set defined by the constraints. In the example considered, $f(x, y) = -xy$ has no local maxima or minima, but, when it is restricted on the circle by imposing the constraint $g(x, y) = x^2 + y^2 - 4 = 0$, it happens to have two local minima and maxima. In this case, the local maxima and minima also determine the extreme values of the function f on the circle.

27.1. Critical Points of a Function Subject to a Constraint. The extreme value problem with constraints amounts to finding the critical points of a function whose arguments are subject to constraints. Suppose that f is a differentiable function. The example discussed above shows that the equation $\boldsymbol{\nabla} f = \mathbf{0}$ no longer determines the critical points of f in the set defined by constraints. A new condition has to be found.

Consider first the case of a single constraint for two variables $\mathbf{r} = \langle x, y\rangle$. Let $\mathbf{r}_0$ be a point at which a differentiable function f has a local extremum on the set S defined by the constraint $g(\mathbf{r}) = 0$. Let us further assume that g has continuous partial derivatives in a neighborhood of $\mathbf{r}_0$ and $\boldsymbol{\nabla} g(\mathbf{r}_0) \neq \mathbf{0}$. Then by the analysis of Section **24.2** the equation $g(\mathbf{r}) = 0$ defines a *smooth* curve through the point $\mathbf{r}_0$. Let $\mathbf{r}(t)$ be parametric equations of this curve in a neighborhood of $\mathbf{r}_0$, that is, for some $t = t_0$, $\mathbf{r}(t_0) = \mathbf{r}_0$ and $\mathbf{r}'(t_0) \neq \mathbf{0}$ (a smooth curve has a tangent vector at any point). The function $F(t) = f(\mathbf{r}(t))$ defines values of f along the curve and has a local extremum at t_0. Since the curve is smooth, the vector function $\mathbf{r}(t)$ is differentiable and it is concluded that F has no rate of change at $t = t_0$, $F'(t_0) = 0$. Since f is a differentiable function, the derivative $F'(t_0)$ can also be computed by the

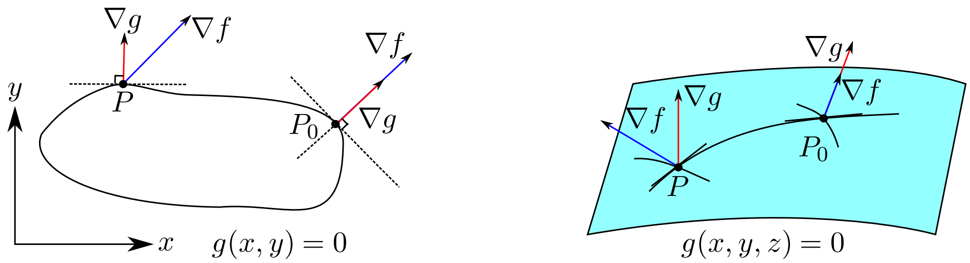


FIGURE **27.1**. **Left**: Relative orientations of the gradients ∇f and ∇g along the curve $g(x, y) = 0$. At the point P_0, the function f has a local extreme value along the curve $g = 0$. At this point, the gradients are parallel and *the level curve of f through P_0 and the curve $g = 0$ have a common tangent line*. **Right**: Relative orientations of the gradients ∇f and ∇g along any curve in the constraint surface $g(x, y, z) = 0$. At the point P_0, the function f has a local extreme value on the surface $g = 0$. At this point, the gradients are parallel and *the level surface of f through P_0 and the surface $g = 0$ have a common tangent plane*.

chain rule:

$$\begin{aligned} F'(t_0) &= f'_x(\mathbf{r}_0)x'(t_0) + f'_y(\mathbf{r}_0)y'(t_0) \\ &= \boldsymbol{\nabla} f(\mathbf{r}_0) \cdot \mathbf{r}'(t_0) = 0 \quad \Rightarrow \quad \boldsymbol{\nabla} f(\mathbf{r}_0) \perp \mathbf{r}'(t_0)\,. \end{aligned}$$

The gradient $\boldsymbol{\nabla} f(\mathbf{r}_0)$ is orthogonal to a tangent vector to the curve at *the point where f has a local extremum on the curve*. By Theorem **24.2**, the gradient $\boldsymbol{\nabla} g(\mathbf{r})$ at *any* point is normal to the level curve $g(\mathbf{r}) = 0$, that is, $\boldsymbol{\nabla} g(\mathbf{r}(t)) \perp \mathbf{r}'(t)$ for any t, *provided* $\boldsymbol{\nabla} g(\mathbf{r}_0) \neq \mathbf{0}$. Therefore, the gradients $\boldsymbol{\nabla} f(\mathbf{r}_0)$ and $\boldsymbol{\nabla} g(\mathbf{r}_0)$ must be parallel at $\mathbf{r}_0$ (see Fig. **27.1**):

$$\left.\begin{aligned} \boldsymbol{\nabla} f(\mathbf{r}_0) &\perp \mathbf{r}'(t_0) \\ \boldsymbol{\nabla} g(\mathbf{r}_0) &\perp \mathbf{r}'(t_0) \end{aligned}\right\} \quad \Rightarrow \quad \boldsymbol{\nabla} f(\mathbf{r}_0) = \lambda \boldsymbol{\nabla} g(\mathbf{r}_0)\,,$$

for some number λ. This is a characteristic property of a critical point of f subject to a constraint. This algebraic condition also has a simple geometrical interpretation. Suppose that f has continuous partial derivatives in a neighborhood of $\mathbf{r}_0$ and $\boldsymbol{\nabla} f(\mathbf{r}_0) \neq \mathbf{0}$. Then the level set $f(\mathbf{r}) = f(\mathbf{r}_0)$ is a smooth curve through $\mathbf{r}_0$. The characteristic geometrical property of the point $\mathbf{r}_0$ is that *the level curve of f and the curve $g(x, y) = 0$ are intersecting at $\mathbf{r}_0$ and tangential to one another (they share the same tangent line through $\mathbf{r}_0$)*.

The conclusion is readily extended to functions of three or more variables. Let $\mathbf{r} = \langle x, y, z\rangle$. Then by the analysis of Section **24.2**, the equation $g(\mathbf{r}) = 0$, where g has continuous partial derivatives and $\boldsymbol{\nabla} g(\mathbf{r}_0) \neq \mathbf{0}$, defines a smooth surface S through the point $\mathbf{r}_0$ in a neighborhood of $\mathbf{r}_0$. If f has a local extreme value at $\mathbf{r}_0$ on the surface $g(\mathbf{r}) = 0$, then f attains a local extreme value along any smooth curve $\mathbf{r} = \mathbf{r}(t)$ through $\mathbf{r}_0$ in that surface. Therefore the derivative

$$F'(t) = \frac{d}{dt}f(\mathbf{r}(t)) = f'_x x' + f'_y y' + f'_z z' = \boldsymbol{\nabla} f(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$$

must vanish at t_0 for which $\mathbf{r}(t_0) = \mathbf{r}_0$. Therefore, the gradient $\boldsymbol{\nabla} f(\mathbf{r}_0)$ is orthogonal to a tangent vector of *any* smooth curve in the surface S at $\mathbf{r}_0$:

$$F'(t_0) = \boldsymbol{\nabla} f(\mathbf{r}_0) \cdot \mathbf{r}'(t_0) = 0 \quad \Rightarrow \quad \boldsymbol{\nabla} f(\mathbf{r}_0) \perp \mathbf{r}'(t_0)\,.$$

On the other hand, by the properties of the gradient (Theorem **24.2**), the vector $\boldsymbol{\nabla} g(\mathbf{r}_0)$ is orthogonal to $\mathbf{r}'(t_0)$ for every such curve. Therefore, at the point $\mathbf{r}_0$, the gradients of f and g are orthogonal to the tangent plane of S and, hence, must be parallel. A similar line of reasoning proves the following theorem for any number of variables.

THEOREM **27.1**. (Critical Points Subject to a Constraint). *Suppose that f has a local extreme value at a point $\mathbf{r}_0$ on the set defined by a constraint $g(\mathbf{r}) = 0$. Suppose that g has continuous partial derivatives in a neighborhood of $\mathbf{r}_0$ and $\boldsymbol{\nabla} g(\mathbf{r}_0) \neq \mathbf{0}$. If f is differentiable at $\mathbf{r}_0$, then there exists a number λ such that*

$$\boldsymbol{\nabla} f(\mathbf{r}_0) = \lambda \boldsymbol{\nabla} g(\mathbf{r}_0).$$

This theorem provides a powerful method to find critical points of f subject to a constraint $g = 0$. It is called the *method of Lagrange multipliers.* To find the critical points of f, the following system of equations must be solved:

$$\boldsymbol{\nabla} f(\mathbf{r}) = \lambda \boldsymbol{\nabla} g(\mathbf{r})\,, \quad g(\mathbf{r}) = 0. \tag{27.1}$$

If $\mathbf{r} = \langle x, y\rangle$, this is a system of three equations:

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

for three variables (x, y, λ). For each solution (x_0, y_0, λ_0), the corresponding critical point of f is (x_0, y_0). The numerical value of λ is not relevant; only its existence must be established by solving the system. In the case of functions of three variables, the system contains four equations for four variables (x, y, z, λ):

$$\begin{cases} f'_x(x,y,z) = \lambda g'_x(x,y,z) \\ f'_y(x,y,z) = \lambda g'_y(x,y,z) \\ f'_z(x,y,z) = \lambda g'_z(x,y,z) \\ g(x,y,z) = 0 \end{cases}$$

For each solution $(x_0, y_0, z_0, \lambda_0)$, the corresponding critical point of f is (x_0, y_0, z_0).

On applicability of the Lagrange method. Suppose that f has a local extremum at a point $\mathbf{r}_0$ on a curve $g(\mathbf{r}) = 0$. The hypothesis $\boldsymbol{\nabla} g(\mathbf{r}_0) \neq \mathbf{0}$ in Theorem **27.1** is crucial for the method of Lagrange multipliers to work. If $\boldsymbol{\nabla} f(\mathbf{r}_0) \neq \mathbf{0}$ (i.e. $\mathbf{r}_0$ is not a critical point of f without the constraint), then Eqs. (**27.1**) have no solution when $\boldsymbol{\nabla} g(\mathbf{r}_0) = \mathbf{0}$, and the method of Lagrange multiplers fails to detect a local extremum of f. Recall that the derivation of Eq. (**27.1**) requires that a curve defined by the equation $g(\mathbf{r}) = 0$ is smooth near $\mathbf{r}_0$, which may no longer be the case if $\boldsymbol{\nabla} g(\mathbf{r}_0) = \mathbf{0}$. So, if f attains its local extreme value at a point where the curve $g(x, y) = 0$ is not smooth, then this point cannot be determined by Eq. (**27.1**). For example, the equation

$$g(x, y) = x^3 - y^2 = 0 \quad \Rightarrow \quad x = y^{2/3}$$

defines a curve that has a cusp at $(0, 0)$ and $\boldsymbol{\nabla} g(0, 0) = \mathbf{0}$. Since $x = y^{2/3} \geq 0$ on the curve, the function

$$f(x, y) = x \quad \Rightarrow \quad f(y^{2/3}, y) = y^{2/3}\,, \quad -\infty < y < \infty\,,$$

attains its absolute minimum value 0 along this curve at the origin. However, $\boldsymbol{\nabla} f(0, 0) = \langle 1, 0\rangle \neq \mathbf{0}$ and the method of Lagrange multipliers fails to detect this point because there is no λ at which Eqs. (**27.1**) are satisfied.

On the other hand, the function $f(x, y) = x^2$ also attains its absolute minimum value at the origin. Equations (**27.1**) do have the solution $(x, y, \lambda) = (0, 0, 0)$. The difference between the two cases is that in the latter case $\boldsymbol{\nabla} f(0, 0) = \mathbf{0}$, i.e., $(0, 0)$ is also a critical point of f without the constraint. Thus, *if the gradient $\boldsymbol{\nabla} g$ vanishes at some points in the set defined by the constraint $g = 0$, then these points should also be studied as critical points of a function f subject to the constraint $g = 0$.*

27.2. Extreme Value Problem Subject to a Constraint. A differentiable function does not necessarily attains its extreme values on a set defined by a constraint $g = 0$. For example, $f(x, y) = xy$ has no maximum value of the line $g(x, y) = y - x = 0$, but it attains its minimum value because the values of f on the line are $f(x, x) = x^2$ where x is any real. On the line $g(x, y) = y + x = 0$, the function f has no minimum value, but attains its maximum value: $f(x, -x) = -x^2$. On the parabola $g(x, y) = y - x^2 = 0$, the function f has no maximum and minimum values because $f(x, x^2) = x^3$. So, the mere existence of critical points as solutions to Eqs. (**27.1**) does not guarantee that f has extreme values at a critical point.

If, however, a differentiable function f is known to attain its extreme values on a set defined by a constraint $g = 0$ where g is also a differentiable function, then finding the extreme values becomes a two-steps procedure:

Step 1. Find all solutions $(\mathbf{r}, \lambda) = (\mathbf{r}_0, \lambda_0)$ of the system (**27.1**); find all points in the set $g = 0$ where $\boldsymbol{\nabla} g = \mathbf{0}$.

Step 2. Evaluate f at all the critical points that result from Step 1. The largest of these values is the maximum value of f; the smallest is the minimum value of f.

The extreme value theorem (Theorem **26.2**) provides sufficient conditions under which the above two-steps procedure allows us to find the extreme values of a function subject to constraints. *If the constraints define a closed and bounded set, then a differentiable function attains its extreme values on the set at some of the critical points determined by the method of Lagrange multipliers.*

If $g(x, y) = 0$ define a piecewise smooth curve that is a boundary of a bounded region in a plane, then that curve is a closed bounded set. A line in a plane is a closed, but not bounded set because it is not contained in any disk. Similarly, if $g(x, y, z) = 0$ define a piecewise smooth surface that is a boundary of a bounded region in space, then that surface is a closed bounded set. A plane or a line in space is not a bounded set as these sets are not contained in any ball.

EXAMPLE **27.2**. *Use the method of Lagrange multipliers to solve the the problem in Example* **27.1**.

SOLUTION: Put $g(x, y) = x^2 + y^2 - 4$. The functions $f(x, y) = -xy$ and g are differentiable everywhere as they are polynomials.
Step 1: Then

$$\begin{cases} f'_x = \lambda g'_x \\ f'_y = \lambda g'_y \\ g = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} -y = 2\lambda x \\ -x = 2\lambda y \\ x^2 + y^2 = 4 \end{cases} .$$

The substitution of the first equation into the second one gives $x = 4\lambda^2 x$. This means that either $x = 0$ or $\lambda = \pm 1/2$. If $x = 0$, then $y = 0$ by the first equation, which contradicts the constraint. For $\lambda = 1/2$, $x = -y$ and the constraint gives $2x^2 = 4$ or $x = \pm\sqrt{2}$. The critical points corresponding to $\lambda = 1/2$ are

$$P_1 = (\sqrt{2}, -\sqrt{2}) , \quad P_2 = (-\sqrt{2}, \sqrt{2}) .$$

If $\lambda = -1/2$, $x = y$ and the constraint gives $2x^2 = 4$ or $x = \pm\sqrt{2}$. The critical points corresponding to $\lambda = -1/2$ are

$$P_3 = (\sqrt{2}, \sqrt{2}) , \quad P_4 = (-\sqrt{2}, -\sqrt{2}) .$$

The gradient $\nabla g(x, y) = \langle 2x, 2y \rangle$ vanishes only at the origin which is not on the circle $x^2 + y^2 = 4$. Alternatively, one could just notice that a circle is a smooth curve.
Step 2: The circle $x^2 + y^2 = 4$ is is a closed, bounded set in the plane as a boundary of a disk. So f attains its extreme values at (some of) the found critical points: $f(\pm\sqrt{2}, \mp\sqrt{2}) = 2$ is the maximum value and $f(\pm\sqrt{2}, \pm\sqrt{2}) = -2$ is the minimum one. $\square$

27.3. The Nature of Critical Points Subject to a Constraint. The mere existence of a solution to (**27.1**) does not guarantee that f has a local extremum at the point found. The situation is fully analogous to finding local extrema of a differentiable function without constraints: the function in question does not necessarily attains a local extremum value at each critical point. An additional study is needed to determine the nature of a critical point (e.g., the second derivative test). For example, in the case of a function f of two variables subject to a constraint $g = 0$, Eq. (**27.1**) means that the function f has no rate of change *along the curve* $g = 0$ at a particular point of the curve. This can happen not only when f has a local extremum at this point, but also when f has an *inflection point* along the curve. Here is an example of this kind.

Consider the following function and constraint

$$f(x, y) = xy\,, \quad g(x, y) = y - x^2 = 0\,.$$

It is easy to see that the constraint defines the parabola $y = x^2$ in the plane. The function f has no extreme values on the parabola: $f(x, x^2) = x^3$ where $-\infty < x < \infty$. However, the system (**27.1**) has one solution:

$$\begin{cases} f'_x = \lambda g'_x \\ f'_y = \lambda g' y \\ \;g = 0 \end{cases} \quad\Rightarrow\quad \begin{cases} y = -2\lambda x \\ x = \lambda \\ y = x^2 \end{cases} \quad\Rightarrow\quad (x, y, \lambda) = (0, 0, 0)\,.$$

The substitution of the first and second equations into the third one yields $-2\lambda^2 = \lambda^2$ or $\lambda = 0$, and, hence, $x = y = 0$. Note that $f(x, x^2) = x^3$ has an inflection point at $x = y = 0$.

How does one determine the nature of a critical point in the case of constrains? There is an analog of the second derivative test in the case of a constrained extreme value problem. It will be discussed in Section **27.6**. However, this test is generally far more difficult to use than the second derivative test for non-constrained extreme value problems. So, a direct use of Definition **27.1** often works better. Yet, in the case when constraints define a *curve*, there is a simple way to determine the nature of critical points.

Let $g(x, y) = 0$ define a smooth curve so that all critical points P_1, P_2, ..., P_n are solutions to (**27.1**). Suppose that it is known that f has a local maximum at P_1 on the curve. For instance, if the curve is closed, then by the extreme value theorem f attains its maximum value at one of the critical points, say, P_1. In particular, it is also a local maximum. The critical points can always be *ordered* along the curve, that is, if a curve is traversed in a particular direction, then the point P_k follows the point P_{k-1}, $k = 2, 3, ...$. Since f has a local maximum at P_1, it should have either a local minimum or an inflection at the neighboring critical point P_2 along the curve. Let P_3 be the critical point next to P_2 along the curve. Then f has a local minimum at P_2 if $f(P_2) < f(P_3)$ and an inflection if $f(P_2) > f(P_3)$. In other words,

- Only a local minimum (maximum) can occur at the only critical point between two neighboring local maxima (minima);
- One or more consecutive inflection points can only occur between two neighboring local maximum and minimum.

This procedure may be continued until all critical points are exhausted. Compare this pattern of critical points with the behavior of a height along a hiking path.

EXAMPLE **27.3**. *Find all critical points of the function $f(x, y) = x^2 y$ on the circle $x^2 + y^2 = 1$ and determine the nature of each critical point.*

SOLUTION: The function f is a polynomial and hence differentiable everywhere. The constraint $g(x, y) = x^2 + y^2 - 1 = 0$ is defined by a differentiable function $g(x, y)$ whose gradient $\boldsymbol{\nabla} g = \langle 2x, 2y\rangle$ does not vanish anywhere on the circle. Therefore all critical point of f are determined by solutions of the system (**27.1**):

$$\begin{cases} f'_x = \lambda g'_x \\ f'_y = \lambda g'_y \\ g = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} 2xy = 2\lambda x \\ x^2 = 2\lambda y \\ x^2 + y^2 = 1 \end{cases}$$

It follow from the first equation that either $x = 0$ or $y = \lambda$. If $x = 0$, then $y = \pm 1$ (by the constraint) and $\lambda = 0$ (by the second equation). So, $(0, \pm 1)$ are critical points corresponding to $\lambda = 0$. If $y = \lambda$, then by the second equation $x^2 = 2\lambda^2$. Then by the constraint $x^2 + y^2 = 3\lambda^2 = 1$ and, hence, $\lambda = \pm 1/\sqrt{3}$. Thus, $(\pm\sqrt{(2/3)}, 1/\sqrt{3})$ are critical points corresponding to $\lambda = 1/\sqrt{3}$ and $(\pm\sqrt{(2/3)}, -1/\sqrt{3})$ are critical points corresponding to $\lambda = -1/\sqrt{3}$. Let us order the critical points along the circle counterclockwise starting from the point in the positive quadrant:

$$P_1 = (\sqrt{(2/3)}, 1/\sqrt{3}), \quad P_2 = (0, 1), \quad P_3 = (-\sqrt{(2/3)}, 1/\sqrt{3}),$$
$$P_4 = (-\sqrt{2/3}, -1/\sqrt{3}), \quad P_5 = (0, -1), \quad P_6 = (\sqrt{(2/3)}, -1/\sqrt{3}).$$

By examining the sequence of values of f at the critical points

$$\begin{array}{ccccccccccc} f(P_1) & \to & f(P_2) & \to & f(P_3) & \to & f(P_4) & \to & f(P_5) & \to & f(P_6) \\ \frac{\sqrt{2}}{3\sqrt{3}} & \to & 0 & \to & \frac{\sqrt{2}}{3\sqrt{3}} & \to & -\frac{\sqrt{2}}{3\sqrt{3}} & \to & 0 & \to & -\frac{\sqrt{2}}{3\sqrt{3}} \\ \max & \to & \min & \to & \max & \to & \min & \to & \max & \to & \min \end{array}$$

it is concluded that the function f has local maxima at P_1, P_3 and P_5 and it has local minima at P_2, P_4, and P_6. □

27.4. Applications of the method of Lagrange multipliers. The method of Lagrange multipliers is used in various optimization problems where some of the variables are not independent and subject to constraints.

EXAMPLE **27.4**. *A rectangular box without a lid is to be made from cardboard. Find the dimensions of the box of a given volume V such that the cost of material is minimal.*

SOLUTION: Let the dimensions of the box be x, y, and z, where z is the height. The amount of cardboard needed is determined by the surface area

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

The question is to find the minimal value of f subject to constraint that the volume of the box has a given value V:

$$g(x,y,z) = xyz - V = 0\,.$$

Since f and g have continuous partial derivatives, let us apply the method of Lagrange multipliers:

$$\begin{cases} f'_x &= \lambda g'_x \\ f'_y &= \lambda g'_y \\ f'_z &= \lambda g'_z \\ g &= 0 \end{cases} \Rightarrow \begin{cases} y + 2z &= \lambda yz \\ x + 2z &= \lambda xz \\ 2x + 2y &= \lambda xy \\ xyz &= V \end{cases} \Rightarrow \begin{cases} xy + 2xz &= \lambda V \\ xy + 2zy &= \lambda V \\ 2xz + 2yz &= \lambda V \\ xyz &= V \end{cases},$$

where the last system has been obtained by multiplying the first equation by x, the second one by y, and the third one by z with the subsequent use of the constraint. Combining the first two equations, one infers $2z(y-x) = 0$. Since $z \neq 0$ ($V \neq 0$), one has $y = x$. Combining the first and third equations, one infers $y(x-2z) = 0$ and hence $x = 2z$. The substitution of $y = x = 2z$ into the constraint yields $4z^3 = V$. Hence, the optimal dimensions are

$$x = y = (2V)^{1/3}\,, \quad z = \frac{1}{2}\,(2V)^{1/3}\,.$$

The amount of cardboard minimizing the cost is $3(2V)^{2/3}$ (the value of f at the critical point). From the geometry of the problem, it is clear that f attains its minimum value at the only critical point. □

Extreme values on a set. The method of Lagrange multipliers can be used to determine extreme values of a function on a set D. Recall that the extreme values may occur on the boundary of D. In Example **26.3**, explicit parametric equations of the boundary of D have been used (Step 2 in the solution). Instead, an algebraic equation of the boundary,

$$g(x,y) = x^2 + y^2 - 4 = 0$$

can be used in combination with the method of Lagrange multipliers. Indeed, if

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

then its critical points along the boundary circle satisfy the system of equations:

$$\begin{cases} f'_x &= \lambda g'_x \\ f'_y &= \lambda g'_y \\ g &= 0 \end{cases} \quad \Longrightarrow \quad \begin{cases} 2x + y &= 2\lambda x \\ 2y + x &= 2\lambda y \\ x^2 + y^2 &= 4 \end{cases}.$$

By subtracting the second equation from the first one, it follows that

$$x - y = 2\lambda(x - y)\,.$$

Hence, either $x = y$ or $\lambda = 1/2$. In the former case, the constraint yields $2x^2 = 4$ or $x = \pm\sqrt{2}$. The corresponding critical points are $(\pm\sqrt{2}, \pm\sqrt{2})$. If $\lambda = 1/2$, then from the first two equations in the system, one infers that $x = -y$. The constraint becomes $2x^2 = 4$ or $x = \pm\sqrt{2}$, and the critical points are $(\pm\sqrt{2}, \mp\sqrt{2})$.

27.5. The Case of Two or More Constraints. Let a differentiable function of three variables f have a local extreme value at point $\mathbf{r}_0$ on the set defined by two constraints $g_1(\mathbf{r}) = 0$ and $g_2(\mathbf{r}) = 0$. Suppose that g_1 and g_2 have continuous partial derivatives in a neighborhood of $\mathbf{r}_0$ and the gradients of g_1 and g_2 do not vanish at $\mathbf{r}_0$, $\boldsymbol{\nabla} g_1(\mathbf{r}_0) \neq \mathbf{0}$ and $\boldsymbol{\nabla} g_2(\mathbf{r}_0) \neq \mathbf{0}$. Then each constraint defines a smooth surface in a neighborhood of $\mathbf{r}_0$. Suppose that *the vectors* $\boldsymbol{\nabla} g_1(\mathbf{r}_0)$ *and* $\boldsymbol{\nabla} g_2(\mathbf{r}_0)$ *are not parallel or, equivalently,* $\boldsymbol{\nabla} g_1(\mathbf{r}_0)$ *is not proportional to* $\boldsymbol{\nabla} g_2(\mathbf{r}_0)$. Then the set defined by the constraints is a smooth curve of intersection of two surfaces. Let $\mathbf{v}$ be a tangent vector to the curve at $\mathbf{r}_0$. Since the curve lies in the level surface $g_1 = 0$, by the earlier arguments, $\boldsymbol{\nabla} f(\mathbf{r}_0) \perp \mathbf{v}$ and $\boldsymbol{\nabla} g_1(\mathbf{r}_0) \perp \mathbf{v}$. On the other hand, the curve also lies in the level surface $g_2 = 0$ and hence $\boldsymbol{\nabla} g_2(\mathbf{r}_0) \perp \mathbf{v}$. It follows that the gradients $\boldsymbol{\nabla} f$, $\boldsymbol{\nabla} g_1$, and $\boldsymbol{\nabla} g_2$ become *coplanar* at the point $\mathbf{r}_0$ as they lie in the plane normal to $\mathbf{v}$. Any vector in the plane normal to $\mathbf{v}$ is a linear combination of two non-parallel vectors in it. Therefore there exist numbers λ_1 and λ_2 such that

$$\boldsymbol{\nabla} f(\mathbf{r}) = \lambda_1 \boldsymbol{\nabla} g_1(\mathbf{r}) + \lambda_2 \boldsymbol{\nabla} g_2(\mathbf{r}), \qquad g_1(\mathbf{r}) = g_2(\mathbf{r}) = 0$$

when $\mathbf{r} = \mathbf{r}_0$. This is a system of five equations for five variables $(x, y, z, \lambda_1, \lambda_2)$. For any solution $(x_0, y_0, z_0, \lambda_{10}, \lambda_{20})$, the point (x_0, y_0, z_0) is a critical point of f on the set defined by the constraints. In general, the following result can be proved by a similar line of reasoning.

THEOREM **27.2**. (Critical Points Subject to Constraints).
Suppose that functions g_a, $a = 1, 2, ..., M$, *of* m *variables,* $m > M$, *have continuous partial derivatives in a neighborhood of a point* $\mathbf{r}_0$ *and a function* f *has a local extreme value at* $\mathbf{r}_0$ *in the set defined by the constraints* $g_a(\mathbf{r}) = 0$. *Suppose that* $\boldsymbol{\nabla} g_a(\mathbf{r}_0)$ *are non-zero vectors any of which cannot be expressed as a linear combination of the others and* f *is differentiable at* $\mathbf{r}_0$. *Then there exist numbers* λ_a *such that*

$$\boldsymbol{\nabla} f(\mathbf{r}_0) = \lambda_1 \boldsymbol{\nabla} g_1(\mathbf{r}_0) + \lambda_2 \boldsymbol{\nabla} g_2(\mathbf{r}_0) + \cdots + \lambda_M \boldsymbol{\nabla} g_M(\mathbf{r}_0).$$

EXAMPLE **27.5**. *Find extreme values of the functions* $f(x, y, z) = xyz$ *on the curve that is an intersection of the sphere* $x^2 + y^2 + z^2 = 1$ *and the plane* $x + y + z = 0$.

SOLUTION: Put $g_1(x, y, z) = x^2 + y^2 + z^2 - 1$ and $g_2(x, y, z) = x + y + z$. One has $\boldsymbol{\nabla} g_1 = \langle 2x, 2y, 2z \rangle$ which can only vanish at $(0, 0, 0)$ and, hence, $\boldsymbol{\nabla} g_1 \neq \mathbf{0}$ on the sphere. Also, $\boldsymbol{\nabla} g_2 = \langle 1, 1, 1 \rangle \neq \mathbf{0}$. Therefore critical points of f on the surface of constraints are determined by the equations:

$$\begin{array}{ll} f'_x = \lambda_1 g'_x + \lambda_2 g'_x & \qquad yz = 2\lambda_1 x + \lambda_2 \\ f'_y = \lambda_1 g'_y + \lambda_2 g'_y & \qquad xz = 2\lambda_1 y + \lambda_2 \\ f'_z = \lambda_1 g'_z + \lambda_2 g'_z \quad \Rightarrow & \qquad xy = 2\lambda_1 z + \lambda_2 \\ g_1 = 0 & \qquad 1 = x^2 + y^2 + z^2 \\ g_2 = 0 & \qquad 0 = x + y + z \end{array}$$

Subtract the second equation from the first one to obtain that $(y-x)z = 2\lambda_1(x-y)$. It follows then that either $x = y$ or $z = -2\lambda_1$. Suppose first that $y = x$. Then $z = -2x$ by the fifth equation. The substitution of $x = y$ and $z = -2x$ into the fourth equation yields $6x^2 = 1$ or $x = \pm 1/\sqrt{6}$. Therefore the points

$$\mathbf{r}_1 = \langle 1/\sqrt{6}, 1/\sqrt{6}, -2/\sqrt{6}\rangle\,, \quad \mathbf{r}_2 = \langle -1/\sqrt{6}, -1/\sqrt{6}, 2/\sqrt{6}\rangle$$

are critical points, *provided there exist the corresponding values* λ_1 *and* λ_2 *such that all equations are satisfied.* For example, take $\mathbf{r}_1$. Then the second and third equations become

$$\begin{cases} -\frac{2}{6} = \frac{2}{\sqrt{6}}\lambda_1 + \lambda_2 \\ \frac{1}{6} = -\frac{4}{\sqrt{6}}\lambda_1 + \lambda_2 \end{cases} \quad \Leftrightarrow \quad \begin{cases} -\frac{3}{6} = \frac{6}{\sqrt{6}}\lambda_1 \\ \frac{1}{6} = -\frac{4}{\sqrt{6}}\lambda_1 + \lambda_2 \end{cases}$$

So $\lambda_1 = -1/(2\sqrt{6})$ and $\lambda_2 = -1/6$. The existence of λ_1 and λ_2 for the point $\mathbf{r}_2$ is verified similarly. Next suppose that $z = -2\lambda_1$. Subtract the third equation from the second one to obtain that $(z-y)x = 2\lambda_1(y-z)$. It follows that either $y = z$ or $x = -2\lambda_1$. Let $y = z$. The fifth equation yields $x = -2y$ and the fourth equation is reduced to $6y^2 = 1$. Therefore there are two more critical points

$$\mathbf{r}_3 = \langle -2/\sqrt{6}, 1/\sqrt{6}, 1/\sqrt{6}\rangle\,, \quad \mathbf{r}_4 = \langle 2/\sqrt{6}, -1/\sqrt{6}, -1/\sqrt{6}\rangle\,.$$

The reader is to verify the existence of λ_1 and λ_2 in these cases (note that $\lambda_1 = -z/2$). Finally, let $x = -2\lambda_1$ and $z = -2\lambda_1$. These conditions imply that $x = z$ and, by the fifth equation, $y = -2x$. The fourth equation yields $6x^2 = 1$ so that there is another pair of critical points:

$$\mathbf{r}_5 = \langle 1/\sqrt{6}, -2/\sqrt{6}, 1/\sqrt{6}\rangle\,, \quad \mathbf{r}_6 = \langle -1/\sqrt{6}, 2/\sqrt{6}, -1/\sqrt{6}\rangle$$

(the reader is to verify the existence of λ_1 and λ_2). The intersection of the sphere and the plane is a circle, which is a closed and bounded set. So, f attains its extreme values at some of the critical points found. By examining the largest and smallest values of f at the critical points, it is concluded that f attains the maximum value $2/\sqrt{6}$ at $\mathbf{r}_2$, $\mathbf{r}_4$, and $\mathbf{r}_6$ and the minimum value $-2/\sqrt{6}$ at $\mathbf{r}_1$, $\mathbf{r}_3$, and $\mathbf{r}_5$. $\square$

Let $f(\mathbf{r})$ be a function subject to a constraint $g(\mathbf{r})$. Define the function

$$F(\mathbf{r}, \lambda) = f(\mathbf{r}) - \lambda g(\mathbf{r}),$$

where λ is viewed as an additional independent variable. Then critical points of F are determined by (**27.1**). Indeed, if $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$, then

$$F'_\lambda(\mathbf{r}, \lambda) = 0 \;\Rightarrow\; g(\mathbf{r}) = 0$$
$$F'_{x_j}(\mathbf{r}, \lambda) = 0 \;\Rightarrow\; f'_{x_j}(\mathbf{r}, \lambda) - \lambda g'_{x_j}(\mathbf{r}, \lambda) = 0 \;\Rightarrow\; \boldsymbol{\nabla} f = \lambda \boldsymbol{\nabla} g\,.$$

Similarly, if there are several constraints, critical points of the function with additional variables λ_a, $a = 1, 2, ..., M$,

$$F(\mathbf{r}, \lambda_1, \lambda_2, ..., \lambda_M) = f(\mathbf{r}) - \sum_{a=1}^{M} \lambda_a g_a(\mathbf{r}) \tag{27.2}$$

coincide with the critical points of f subject to the constraints $g_a = 0$ as stated in Theorem **27.2**. The functions F and f have the same values on the set defined by the constraints $g_a = 0$ because they differ by a linear combination of constraint functions with the coefficients being the *Lagrange multipliers*. The above observation provides a simple way to formulate the equations for critical points subject to constraints.

27.6. Finding Local Maxima and Minima. In the simplest case of a function f of two variables subject to a constraint, the nature of critical points has been determined by ordering critical points along the curve and by examining the values of the function at the critical points. Unfortunately, this method is limited in applications. For example, if the constraint defines a surface, then it is not possible to order critical points. It is possible to develop an analog of the second-derivative test for critical points subject to constraints.

The second-derivative test. If the constraints can be solved, then an explicit form of f on the set defined by the constraints can be found, and the standard second-derivative test applies! For instance, in Example **27.4**, the constraint can be solved $z = V/(xy)$. The values of the function f on the constraint surface are

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

The equations $F'_x = 0$ and $F'_y = 0$ determine the critical point $x = y = (2V)^{1/3}$ (and $z = V/(xy) = (2V)^{1/3}/2$). So the second-derivative test can be applied to the function $F(x, y)$ at the critical point $x = y = (2V)^{1/3}$ to show that indeed F has a minimum and hence f has a minimum on the constraint surface.

Even if the constraint cannot be solved explicitly, the implicit function theorem may be used to obtain an analog of the second-derivative test for critical points of functions subject to constraints. Its general formulation is not simple. So the discussion is limited to the simplest case of a function of two variables subject to a constraint (see also Study Problem **27.2** where the case of three variables and one constraint is studied).

Suppose that g and f have continuous second order partial derivatives and $\nabla g(\mathbf{r}_0) \neq \mathbf{0}$ where $\mathbf{r}_0$ is a solution of the system (**27.1**) for some λ. Then g'_x and g'_y cannot simultaneously vanish at the critical point. Without loss of generality, assume that $g'_y \neq 0$ at $\mathbf{r}_0$. By the implicit function theorem, there is a neighborhood of $\mathbf{r}_0$ in which the equation $g(x, y) = 0$ has a unique solution $y = h(x)$. The values of f on the level curve $g = 0$ near the critical point are $F(x) = f(x, h(x))$. By the chain rule, one infers that $F' = f'_x + f'_y h'$ and

$$F'' = (d/dx)(f'_x + f'_y h') = f''_{xx} + 2f''_{xy}h' + f''_{yy}(h')^2 + f'_y h''. \tag{27.3}$$

So, in order to find $F''(x_0)$, one has to calculate $h'(x_0)$ and $h''(x_0)$. This latter task is accomplished by the implicit differentiation.

By the definition of $h(x)$, $G(x) = g(x, h(x)) = 0$ for all x in an open interval containing x_0. Therefore, $G'(x) = 0$, which defines h':

$$G(x) = 0 \quad \Rightarrow \quad 0 = G'(x) = g'_x(x, h) + g'_y(x, h)h'(x) \quad \Rightarrow \quad h'(x) = -\frac{g'_x}{g'_y}\,.$$

Similarly, $G''(x) = 0$ yields

$$G'' = g''_{xx} + 2g''_{xy}h' + g''_{yy}(h')^2 + g'_y h'' = 0, \tag{27.4}$$

which can be solved for h'':

$$h'' = -\frac{1}{g'_y}\left(g''_{xx} + 2g''_{xy}h' + g''_{yy}(h')^2\right), \quad h' = -\frac{g'_x}{g'_y}$$

So the derivatives $h'(x_0)$ and $h''(x_0)$ can be computed in terms of the values of the partial derivatives of g at the critical point (x_0, y_0). Their substitution into (**27.3**) gives the value $F''(x_0)$. If $F''(x_0) > 0$ (or $F''(x_0) < 0$), then f has a local minimum (or maximum) at (x_0, y_0) along the curve $g = 0$. Note also that $F'(x_0) = 0$ as required owing to the conditions $f'_x = \lambda g'_x$ and $f'_y = \lambda g'_y$ satisfied at the critical point.

If $g'_y(\mathbf{r}_0) = 0$, then $g'_x(\mathbf{r}_0) \neq 0$, and there is a function $x = h(y)$ that solves the equation $g(x, y) = 0$. So, by swapping x and y in the above arguments, the same conclusion is proved to hold.

EXAMPLE **27.6**. *Show that the point $\mathbf{r}_0 = \mathbf{0}$ is a critical point of the function $f(x, y) = x^2y + y + x$ subject to the constraint $e^{xy} = x + y + 1$ and determine whether f has a local minimum or maximum at it.*

SOLUTION:
Critical point. Put $g(x, y) = e^{xy} - x - y - 1$. Then $g(0, 0) = 0$; that is, the point $(0, 0)$ satisfies the constraint. The first partial derivatives of f and g are

$$f'_x = 2xy + 1\,, \quad f'_y = x^2 + 1\,, \quad g'_x = ye^{xy} - 1\,, \quad g'_y = xe^{xy} - 1\,.$$

Therefore, both equations $f'_x(0, 0) = \lambda g'_x(0, 0)$ and $f'_y(0, 0) = \lambda g'_y(0, 0)$ are satisfied at $\lambda = -1$. Thus, the point $(0, 0)$ is a critical point of f subject to

the constraint $g = 0$.
Second-derivative test. Since $g'_y(0,0) = -1 \neq 0$, there is a function $y = h(x)$ near $x = 0$ such that $G(x) = g(x, h(x)) = 0$. By the implicit differentiation,

$$h'(0) = -g'_x(0,0)/g'_y(0,0) = -1\,.$$

The second partial derivatives of g are

$$g''_{xx} = y^2 e^{xy}\,,\quad g''_{yy} = x^2 e^{xy}\,,\quad g''_{xy} = e^{xy} + xye^{xy}\,.$$

The derivative $h''(0)$ is found from (**27.4**), where $g''_{xx}(0,0) = g''_{yy}(0,0) = 0$, $g''_{xy}(0,0) = 1$, $h'(0) = -1$, and $g'_y(0,0) = -1$:

$$h''(0) = -[g''_{xx}(0,0) + 2g''_{xy}(0,0)h'(0) + g''_{yy}(0,0)(h'(0))^2]/g'_y(0,0) = -2.$$

The second partial derivatives of f are

$$f''_{xx} = 2y\,,\quad f''_{yy} = 0\,,\quad f''_{xy} = 2x.$$

The substitution of $f''_{xx}(0,0) = f''_{yy}(0,0) = f''_{xy}(0,0) = 0$, $h'(0) = -1$, $f'_y(0,0) = 1$, and $h''(0) = -2$ into (**27.3**) gives $F''(0) = -2 < 0$. Therefore, f attains a local maximum at $(0,0)$ along the curve $g = 0$. Note also that $F'(0) = f'_x(0,0) + f'_y(0,0)h'(0) = 1 - 1 = 0$ as required. □

The implicit differentiation and the implicit function theorem can be used to establish the second-derivative test for the multivariable case with constraints (see another example in Study Problem **27.2**).

27.7. Study Problems.

Problem **27.1**. *An axially symmetric solid consists of a circular cylinder and a right-angled circular cone attached to one of the cylinder's bases. What are the dimensions of the solid at which it has a maximal volume if the surface area of the solid has a fixed value S?*

SOLUTION: Let r and h be the radius and height of the cylinder. Since the cone is right-angled, its height is r. The surface area is the sum of three terms: the area of the base (disk) πr^2, the area of the side of the cylinder $2\pi rh$, and the surface area S_c of the cone. A cone of height a and with the radius of the base r is obtained by rotation of a straight line $y = mx$, where $m = r/a$ and $0 \leq x \leq a$, about the $x-$axis. In the present case $r = a$ and $m = 1$. Recall from Calculus II that the area of a surface obtained by rotation about the x axis is

$$S_c = \int_0^a 2\pi y\sqrt{1 + (dy/dx)^2}dx = \sqrt{2}\int_0^r 2\pi x dx = \pi\sqrt{2}\,r^2\,.$$

Similarly, the volume of the cone is

$$V_c = \int_0^a \pi y^2 dx = \int_0^r \pi x^2 dx = \frac{\pi r^3}{3}\,.$$

Therefore the problem is reduced to finding the maximal value of the function (volume) $V(r,h) = \pi r^2 h + \pi r^3/3$ subject to the constraint

$$g(r,h) = 2\pi r h + (1+\sqrt{2})\pi r^2 - S = 0\,.$$

The critical points of V satisfy the equations:

$$\begin{cases} V'_r = \lambda g'_r \\ V'_h = \lambda g'_h \\ \;0 = g(r,h) \end{cases} \quad \Rightarrow \quad \begin{cases} 2\pi r h + \pi r^2 h = \lambda(2\pi h + 2(1+\sqrt{2})\pi r) \\ \pi r^2 = 2\pi\lambda r \\ S = 2\pi r h + (1+\sqrt{2})\pi r^2 \end{cases}$$

Since $r \neq 0$ (the third equation is not satisfied if $r = 0$), the second equation implies that $\lambda = r/2$. The substitution of the latter into the first equation yields $rh = r^2\sqrt{2}$ or $h = r\sqrt{2}$. Then it follows from the third equation that $r = (S/[\pi(1+2\sqrt{2})])^{1/2}$. By the geometrical nature of the problem it is clear that the found critical point corresponds to the maximum value of the volume. $\square$

Problem **27**.**2**. *Let functions f and g of three variables $\mathbf{r} = \langle x,y,z\rangle$ have continuous partial derivatives up to order 2. Use the implicit differentiation to establish the second-derivative test for critical points of f on the surface $g(\mathbf{r}) = 0$.*

SOLUTION: Suppose that $\boldsymbol{\nabla} g(\mathbf{r}_0) \neq \mathbf{0}$ at a critical point $\mathbf{r}_0$. Without loss of generality, one can assume that $g'_z(\mathbf{r}_0) \neq 0$. By the implicit function theorem, there exists a function $z = h(x,y)$ such that $G(x,y) = g(x,y,h(x,y)) = 0$ in some neighborhood of the critical point. Then the equations $G'_x(x,y) = 0$ and $G'_y(x,y) = 0$ determine the first partial derivatives of h:

$$g'_x + g'_z h'_x = 0 \;\Rightarrow\; h'_x = -g'_x/g'_z\,; \quad g'_y + g'_z h'_y = 0 \;\Rightarrow\; h'_y = -g'_y/g'_z\,.$$

The second partial derivatives h''_{xx}, h''_{xy}, and h''_{yy} are found from the equations

$$\begin{aligned} G''_{xx} = 0 \quad &\Rightarrow \quad g''_{xx} + 2g''_{xz}h'_x + g''_{zz}(h'_x)^2 + g'_z h''_{xx} = 0, \\ G''_{yy} = 0 \quad &\Rightarrow \quad g''_{yy} + 2g''_{yz}h'_y + g''_{zz}(h'_y)^2 + g'_z h''_{yy} = 0, \\ G''_{xy} = 0 \quad &\Rightarrow \quad g''_{xy} + g''_{xz}h'_x + g''_{yz}h'_y + g''_{zz}h'_x h'_y + g'_z h''_{xy} = 0. \end{aligned}$$

The values of the function f on the level surface $g(x,y,z) = 0$ near the critical point are $F(x,y) = f(x,y,h(x,y))$. To apply the second-derivative test to the function F, its second partial derivatives have to be computed at the critical point. By the chain rule

$$\begin{aligned} F''_{xx} &= (f'_x + f'_z h'_x)'_x = f''_{xx} + 2f''_{xz}h'_x + f''_{zz}(h'_x)^2 + f'_z h''_{xx}, \\ F''_{yy} &= (f'_y + f'_z h'_y)'_y = f''_{yy} + 2f''_{yz}h'_y + f''_{zz}(h'_y)^2 + f'_z h''_{yy}, \\ F''_{xy} &= (f'_x + f'_z h'_x)'_y = f''_{xy} + f''_{xz}h'_x + f''_{yz}h'_y + f''_{zz}h'_x h'_y + f'_z h''_{xy}, \end{aligned}$$

where the first and second partial derivatives of h have been found earlier. If (x_0,y_0,z_0) is the critical point found by solving the system (**27**.**1**), then

$a = F''_{xx}(x_0, y_0)$, $b = F''_{yy}(x_0, y_0)$, and $c = F''_{xy}(x_0, y_0)$ in the second-derivative test for the two-variable function F. □

27.8. Exercises.
1–18. Use Lagrange multipliers to find the maximum and minimum values of each of the following functions subject to the specified constraints:

1. $f(x, y) = xy$, $x + y = 1$;
2. $f(x, y) = x^2 + y^2$, $x/a + y/b = 1$;
3. $f(x, y) = xy^2$, $2x^2 + y^2 = 6$;
4. $f(x, y) = y^2$, $x^2 + y^2 = 4$;
5. $f(x, y) = x + y$, $x^2/16 + y^2/9 = 1$;
6. $f(x, y) = 2x^2 - 2y^2$, $x^4 + y^4 = 16$;
7. $f(x, y) = Ax^2 + 2Bxy + Cy^2$, $x^2 + y^2 = 1$;
8. $f(x, y, z) = xyz$, $3x^2 + 2y^2 + z^2 = 6$;
9. $f(x, y, z) = x - 2y + 2z$, $x^2 + y^2 + z^2 = 1$;
10. $f(x, y, z) = x^2 + y^2 + z^2$, $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$;
11. $f(x, y, z) = -x + 3y - 3z$, $x + y - z = 0$, $y^2 + 2z^2 = 1$;
12. $f(x, y, z) = xy + yz$, $xy = 1$, $y^2 + 2z^2 = 1$;
13. $f(x, y, z) = xy + yz$, $x^2 + y^2 = 2$, $y + z = 2$ ($x > 0$, $y > 0$, $z > 0$);
14. $f(x, y, z) = \sin(x)\sin(y)\sin(z)$, $x + y + z = \pi/2$ ($x > 0$, $y > 0$, $z > 0$);
15. $f(x, y, z) = x^2/a^2 + y^2/b^2 + z^2/c^2$, $x^2 + y^2 + z^2 = 1$, $n_1x + n_2y + n_3z = 0$, where $\hat{\mathbf{n}} = \langle n_1, n_2, n_3\rangle$ is a unit vector;
16. $f(\mathbf{r}) = \hat{\mathbf{u}} \cdot \mathbf{r}$, $\|\mathbf{r}\| = R$, where $\mathbf{r} = \langle x_1, ..., x_m\rangle$, $\hat{\mathbf{u}}$ is a constant unit vector, and $R > 0$ is a constant;
17. $f(\mathbf{r}) = \mathbf{r} \cdot \mathbf{r}$, $\mathbf{n} \cdot \mathbf{r} = 1$ where $\mathbf{n}$ has strictly positive components and $\mathbf{r} = \langle x_1, x_2, ..., x_m\rangle$;
18. $f(\mathbf{r}) = x_1^n + x_2^n + \cdots + x_m^n$, $x_1 + x_2 + \cdots + x_m = a$ where $n > 0$ and $a > 0$.

19. Prove the inequality

$$\frac{x^n + y^n}{2} \geq \left(\frac{x + y}{2}\right)^n$$

if $n \geq 1$, $x \geq 0$, and $y \geq 0$. Hint: Minimize the function $f = (x^n + y^n)/2$ under the condition $x + y = s$.
20. Find the minimal value of the function $f(x, y) = y$ on the curve $x^2 + y^4 - y^3 = 0$. Explain why the method of Lagrange multipliers fails. Hint: Sketch the curve near the origin.
21. Use the method of Lagrange multipliers to maximize the function $f(x, y) = 3x + 2y$ on the curve $\sqrt{x} + \sqrt{y} = 5$. Compare the obtained value with $f(0, 25)$. Explain why the method of Lagrange multipliers fails.
22. Find three positive numbers whose sum is a fixed number $c > 0$ and whose product is maximal.

23–27. Use the method of Lagrange multipliers to solve the following Problems from Section **26.6**:

23. Exercise **35** ;
24. Exercise **36** ;
25. Exercise **37** ;
26. Exercise **38** ;
27. Exercise **39** .

28. The cross section of a cylindrical tab is a half-disk. If the tab has total area S, what are the dimensions at which the tab has maximal volume?
29. Find a rectangle with a fixed perimeter $2p$ that forms a solid of the maximal volume under rotation about one of its sides.
30. Find a triangle with a fixed perimeter $2p$ that form a solid of the maximal volume under rotation about one of its sides.
31. Find a rectangular box with the maximal volume that is contained in a half-ball of radius R.
32. Find a rectangular box with the maximal volume that is contained in an ellipsoid $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$.
33. Consider a circular cone obtained by rotation of a straight line segment of length l about the axis through an endpoint of the segment. If the angle between the segment and the axis is θ, find a rectangular box within the cone that has a maximal volume.
34. The solid consists of a rectangular box and two identical pyramids whose bases are opposite faces of the box. The edges of the pyramid adjacent at the vertex opposite to its base have equal lengths. If the solid has a fixed volume V, at what angle between the edges of the pyramid and its base is the surface area of the solid minimal?
35. Use Lagrange multipliers to find the distance between the parabola $y = x^2$ and the line $x - y = 2$.
36. Find the maximum value of the function $f(\mathbf{r}) = \sqrt[m]{x_1 x_2 \cdots x_m}$ given that $x_1 + x_2 + \cdots + x_m = c$, where c is a positive constant. Deduce from the result that if $x_i > 0$, $i = 1, 2, ..., m$, then

$$\sqrt[m]{x_1 x_2 \cdots x_m} \leq \frac{1}{m}(x_1 + x_2 + \cdots + x_m)$$

that is, the *geometrical mean* of m numbers is no larger than the *arithmetic mean*. When is the equality reached?
37. Give an alternative proof of the Cauchy-Schwarz inequality in a Euclidean space (Theorem **8.1**) using the method of Lagrange multipliers to maximize the function of $2m$ variables $f(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ subject to the contraints $\mathbf{x} \cdot \mathbf{x} = 1$ and $\mathbf{y} \cdot \mathbf{y} = 1$, where $\mathbf{x} = \langle x_1, ..., x_m \rangle$ and $\mathbf{y} = \langle y_1, ..., y_m \rangle$. Hint: After maximizing the function, put $\mathbf{x} = \mathbf{a}/\|\mathbf{a}\|$ and $\mathbf{y} = \mathbf{b}/\|\mathbf{b}\|$ for any two non-zero vectors $\mathbf{a}$ and $\mathbf{b}$.

Selected Answers and Hints to Exercises

Section 16.6. **1**. $D = \{(x, y)|y \neq 0\}$, $R = (-\infty, \infty)$. **2**. $D = \{(x, y)|(x, y) \neq (0, 0)\}$, $R = (-\infty, \infty)$. **3**. $D = \{(x, y)|y \neq \pm 2x\}$, $R = (-\infty, \infty)$. **4**. $D = \{(x, y)|(x/3)^2 + (y/6)^2 < 1\}$ (an interior of an ellipse, the boundary is not included), $R = (-\infty, \ln 9]$. **5**. $D = \{(x, y)|x^2 + (y/3)^2 \leq 1\}$ (an interior of an ellipse, the boundary is included), $R = [0, 1]$. **6**. $D = \{(x, y)|x^2 + y^2 \leq 4,\ y > 0\}$ (the half-disk with the circular boundary included), $R = (-\infty, \infty)$. **7**. $D = \{(x, y)|x^2 + y^2 \leq 4,\ y \neq 0\}$ (the disk with the circular boundary included and with the diameter along the x axis excluded), $R = (-\infty, \infty)$. **8**. $D = \{(x, y)|x^2 + y^2 \leq 4,\ x^2 + y^2 > 1\}$ (the region between the circles, the inner circle is excluded), $R = (-\infty, \ln 3]$ **9**. $D = \{(x, y, z)|yz \neq 0\}$ (the whole space with the planes $y = 0$ and $z = 0$ excluded), $R = (-\infty, \infty)$. **10**. $D = \{(x, y, z)|x \neq y^2 + z^2\}$ (the whole space with the paraboloid $x = y^2 + z^2$ excluded), $R = (-\infty, \infty)$. **11**. $D = \{(x, y, z)|z - 1 < x^2 + y^2\}$ (the solid region below the circular paraboloid $z - 1 = x^2 + y^2$, the paraboloid is not included), $R = (-\infty, \infty)$. **12**. $D = \{(x, y, z)|x^2 \geq y^2 + z^2\ ,\ x^2 + y^2 + z^2 < 1\}$ (the part of the ball that lies inside the double cone whose axis is the x axis, the boundary of the ball is not included). **14**. The contour map consists of ellipses $(x/a)^2 + (y/b)^2 = 1$ where $a = \sqrt{k}$, $b = \sqrt{k}/2$, and $k > 0$; the graph is the elliptic paraboloid, $z = x^2 + 4y^2$. **15**. The contour map consists of hyperbolas $y = k/x$, $k \neq 0$ and the coordinate axes (level curves $f(x, y) = 0$); the graph is the hyperbolic paraboloid with axes rotated by $45°$ (the axes are $y = \pm x$) or a "saddle". **17**. The contour map consists of ellipses $(x/a)^2 + (y/b)^2 = 1$ where $b = a/3$, the graph is the elliptic cone $z^2 = x^2 + 9y^2$, $z \geq 0$. **18**. The contour map consists of coordinate lines $x = \sin^{-1}(k)$, the graph is a cylindrical surface swept by the graph $z = \sin x$ by moving the latter parallel to the y axis. **20**. Parallel planes perpendicular to $\mathbf{n} = \langle 1, 2, 3\rangle$. **21**. Ellipsoids $(x/a)^2 + (y/b)^2 + (z/c)^2 = 1$ where $b = a/2$ and $c = a/3$, $a > 0$. **22**. Paraboloids $z - k = -(x^2 + y^2)$ where k is real. **24**. Hyperboloids of one sheet $x^2 + y^2 - z^2 = e^k > 0$, where k is real, with the axis parallel to the z axis. **33**. The surface $z = f(x, y)$ is shifted along the z axis by k units. **34**. The surface $z = f(x, y)$ is stretched along the z axis if $|m| > 1$ and compressed if $0 < |m| < 1$, a negative m corresponds to the reflection about the xy plane. **35**. The surface $z = f(x, y)$ is rigidly shifted so that each point is along the vector $(a, b, 0)$. **36**. The surface $z = f(x, y)$ is stretched ($|p| < 1$, $|q| < 1$) and/or compressed ($|p| > 1$, $|q| > 1$) along the x and y axes; negative values of p and/or q correspond to reflections of the surface about yz and/or xz planes, respectively. **41**. $f(x, y) = x^2(1 - y)/(1 + y)$. **45**. The surface is obtained by rotation of the graph $z = f(u)$, $u \geq 0$, about the z axis

Section 17.7. **1.** $0 < \delta \leq \varepsilon/10$. **2.** $0 < \delta \leq 3\varepsilon/10$. **3.** $0 < \delta \leq \delta_1$ where δ_1 is the smallest number of $3\varepsilon/10$ and 1. **4.** $0 < \delta \leq 3\varepsilon/10$. **5.** $0 < \delta \leq \delta_1$ where δ_1 is the smallest number of $\varepsilon/12$ and 1. **6.** $-|x|\sqrt{y} \leq y \sin(x/\sqrt{y}) \leq |x|\sqrt{y}$.

7. $-\frac{1}{2}y^2 \le [1-\cos(y/x)]x^2 \le \frac{1}{2}y^2$. **9.** The value $f(\mathbf{r}_0)$ is unknown. If f is continuous, then $f(\mathbf{r}_0) = 2$. **12.** f is continuous on its domain. **17.** f is not continuous at $y = x$, $x \neq 0$. **20.** No such c exists. **21.** $c = 0$. **22.** $c = 0$. **23.** $\frac{1}{2}$. **24.** 0. **25.** 0. **26.** -1. **27.** 0.

Section 18.7. **4.** $\frac{1}{4}$. **5.** $-\frac{1}{6}$. **6.** $\frac{1}{2}$. **7.** The limit does not exist. **8.** 0. **9.** $\frac{1}{2}$. **10.** The limit does not exist. **11.** 0. **12.** $\ln 2$. **13.** 1. **14.** 1. **15.** 0. **16.** 0. **17.** 0. **19.** $f(t, at^2) = a/(1+a^2) \neq 0$ (limits along parabolas do not vanish), $a \neq 0$. **21.** The limit along the lines is zero. But the limit at infinity does not exist because the limit along the parabola does not vanish: $f(t,t^2) = t^2 \to \infty$ as $t \to \infty$. **22.** ∞. **23.** ∞. **24.** The limit does not exist (consider $x = y = z = t$ as $t \to \infty$ and $x = y = t$, $z = e^{-at^2}$ as $t \to \infty$). **25.** ∞. **26.** The limit does not exist. **27.** The limit does not exist. Hint: compare the limits along $\langle t, t, -t\rangle$ and $\langle t, t, -t^2\rangle$ as $t \to 0^+$. **28.** The limit does not exist. **30.** The limit does not exist.

Section 19.4. **1.** $f'_x(1,2) = \frac{4}{9}$, $f'_y(1,2) = -\frac{2}{9}$. **2.** $f'_x(1,2,3) = \frac{2}{5}$, $f'_y(1,2,3) = f'_z(1,2,3) = 0$. **3.** $f'_{x_k}(\mathbf{0}) = k$, $k = 1,2,...,n$. **4.** $f'_x(1,2,\frac{\pi}{2}) = 0$, $f'_y(1,2,\frac{\pi}{2}) = -\frac{\pi}{2}$, $f'_z(1,2,\frac{\pi}{2}) = -2$. **5.** $f'_x(1,1) = 1$, $f'_y(1,1) = \frac{\pi}{2}$. **6.** $f'_x(0,0) = f'_y(0,0) = 1$. **7.** $f'_x(0,0) = f'_y(0,0) = 0$. **8.** $f'_x = n(x+y^2)^{n-1}$, $f'_y = 2ny(x+y^2)^{n-1}$. **9.** $f'_x = yx^{y-1}$, $f'_y = \ln(x)x^y$. **10.** $f'_x = (1+2x^2+4xy)e^{(x+2y)^2}$, $f'_y = 4x(x+2y)e^{(x+2y)^2}$. **11.** $f'_x = y\cos(xy)\cos(x^2+y^2) - 2x\sin(xy)\sin(x^2+y^2)$, $f'_y = x\cos(xy)\cos(x^2+y^2) - 2y\sin(xy)\sin(x^2+y^2)$. **12.** $f'_x = 1/(x+y^2+z^3)$, $f'_y = 2y/(x+y^2+z^3)$, $f'_z = 3z^2/(x+y^2+z^3)$. **13.** $f'_x = y^2\cos(xz^2) - xy^2z^2\sin(xz^2)$, $f'_y = 2xy\cos(xz^2)$, $f'_z = -2x^2y^2z\sin(xz^2)$. **14.** $f'_{x_k} = ma_k(\mathbf{a}\cdot\mathbf{r})^{m-1}$, $k = 1,2,3$. **15.** $f'_x = -y/(x^2+y^2)$, $f'_y = x/(x^2+y^2)$. **16.** $f'_x = y/(x^2+y^2)$, $f'_y = -x/(x^2+y^2)$. **17.** $f'_x = y^z x^{y^z-1}$, $f'_y = z\ln(x)x^{y^z}y^{z-1}$, $f'_z = x^{y^z}y\ln(x)$. **22.** $f'_{x_j} = (\mathbf{a}\times\mathbf{b})_j$, $j = 1,2,3$. **23.** $f'_{x_j} = \frac{1}{\|\mathbf{a}\times\mathbf{r}\|}[(\mathbf{a}\cdot\mathbf{a})x_j - (\mathbf{a}\cdot\mathbf{r})a_j]$, $j = 1,2,3$, where $\mathbf{a} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{r} = \langle x_1, x_2, x_3\rangle$. **24.** $f'_x(1,2) = \frac{4}{9} > 0$ (increasing), $f'_y(1,2) = \frac{1}{9} > 0$ (increasing). **25.** $f'_x(1,1) = \frac{2}{3} > 0$ (increasing), $f'_y(1,1) = -\frac{4}{3} < 0$ (decreasing). **26.** $f'_x(-1,\pi) = -\pi < 0$ (decreasing), $f'_y(-1,\pi) = 1 > 0$ (increasing).

Section 20.4. **1.** $f''_{xx} = -2xy^3/(1+x^2y^2)^2$, $f''_{yy} = -2x^3y/(1+x^2y^2)^2$, $f''_{xy} = f''_{yx} = (1-x^2y^2)/(1+x^2y^2)^2$. **2.** $f''_{xx} = 0$, $f''_{yy} = 2xz\cos(zy^2) - 4xy^2z^2\sin(zy^2)$, $f''_{zz} = -xy^4\sin(zy^2)$, $f''_{zx} = f''_{xz} = y^2\cos(zy^2)$, $f''_{xy} = f''_{yx} = 2zy\cos(zy^2)$, $f''_{yz} = f''_{zy} = 2xy\cos(zy^2) - 2xyz^3\sin(zy^2)$. **3.** $f''_{xx} = 6x$, $f''_{xy} = f''_{yx} = 0$, $f''_{yy} = 0$, $f''_{zy} = f''_{yz} = 1$, $f''_{zz} = 2$. **7.** $f''_{x_ix_j} = a_ib_j + a_jb_i$, $i,j = 1,2,...,m$. **13.** $f'''_{xxy} = f'''_{xyx} = f'''_{yxx} = 0$. **14.** $f'''_{xyz} = 0$ $f'''_{xxz} = 0$, $f'''_{yzz} = 0$ (the order of differentiation does not matter). **15.** $\partial^5 f/\partial z^5 = \sin(xy)e^z$, $f^{(4)}_{xyzz} = e^z[\cos(xy) - xy\sin(xy)]$ (the order of differentiation does not matter). **16.** $-4!\sin(x+2y+3z-4t)$. **21.** $f = x^3y + y^3 + c$. **22.** $f = xyz + x^3 + 2y^2 + z + c$.

23. $f = \sum_{k=1}^{m} \frac{k}{2} x_k^2 + c$. **24.** f does not exist. **25.** $f = x\sin(xy) + c$. **33.** $c^2 = a^2 + b^2$ (a double circular cone). **34.** $F(t) = t^2u'''(t) + 3tu''(t) + u'(t)$. **39.** $f''_{xy}(0,0)$ does not exist, $f''_{yx}(0,0) = 0$ (note $f'_x(0,0) = f'_y(0,0) = 0$, $f'_x(0,y) = 1$ if $y \neq 0$, whereas $f'_y(x,0) = 0$ if $x \neq 0$). **40.** $f(x,y) = g(y)x + h(y)$ for some $g(y)$ and $h(y)$. **41.** $f(x,y) = g(x) + h(y)$ for some $g(x)$ and $h(y)$. **42.** $f(x,y) = g_n(x)y^{n-1} + g_{n-1}(x)y^{n-1} + \cdots + g_2(x)y + g_1(x)$ for some $g_k(x)$, $k = 1, 2, ..., n$. **43.** $g(x,y,z) = f(x,y) + h(y,z) + p(x,z)$ for some $f(x,y)$, $h(y,z)$, and $p(x,z)$. **44.** $f(x,y) = x^2y + y^2 - 2x^4 + 1$. **45.** $f(x,y) = 2y^2 + xy + 2$. **46.** $f(x,y) = \frac{1}{2}(x^2y + y^2x) + x + y^2$.

Section 21.6. **3.** No; $f'_x(0,0)$ and $f'_y(0,0)$ do not exist. **4.** No; since $f'_x(0,0) = f'_y(0,0) = 0$, there should be $L(x,y) = 0$, but $[f(x,y) - L(x,y)]/\sqrt{x^2+y^2}$ does not have a limit as $(x,y) \to (0,0)$. **5.** $f'_x(0,0) = f'_y(0,0) = 0$, so $L(x,y) = 0$, but $[f(x,y) - L(x,y)]/\sqrt{x^2+y^2}$ does not have a limit as $(x,y) \to (0,0)$ and, hence, f is not differentiable at $(0,0)$. **7.** $0 < x < \infty$, $-\infty < y < \infty$. **8.** $2x + y \neq 0$. **9.** All (x,y,z). **10.** $x^2 + y^2 > z^2$ (solid region outside the double cone). **14.** The plane is $6x + 9y + z = 23$, the line is $(x-1)/6 = (y-2)/9 = z+1$. **15.** The plane is $3(x-1) + \frac{1}{4}(y-4) - (z-2) = 0$, the line is $(x-1)/3 = 4(y-4) = 2-z$. **16.** The plane is $4(x-2) - 3(y-1) - z = 0$, the line is $\frac{1}{4}(x-2) = \frac{1}{3}(1-y) = -z$. **17.** The plane is $\frac{1}{2}(x-1) - (y - \frac{\pi}{4}) - (z+1) = 0$, the line is $2(x-1) = \frac{\pi}{4} - y = -z-1$. **18.** The plane is $z - x = 0$, the line is $x - 1 = 1 - z$, $y = 1$. **19.** The plane is $2z - x - y = 0$, the line is $x - 1 = y - 1 = (1-z)/2$ (consider the graph $y = z + \ln z - \ln x = f(x,z)$ and the tangent plane to it at $(1,1,1)$). **20.** $L(x,y) = 3 - 12x + 2y$. **21.** $L(x,y,z) = 1 + \frac{1}{2}x + \frac{1}{3}(z-1)$. **22.** $L(\mathbf{r}) = \mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0)$. **23.** $3 - \frac{7}{3}\Delta x - \frac{2}{3}\Delta y$, where $\Delta x = 0.08$ and $\Delta y = -0.05$. **24.** $108(1 + \Delta x + \Delta y + \Delta z)$ where $\Delta x = 0.002$, $\Delta y = 0.003$, and $\Delta z = 0.004$. **25.** $1 + 2\Delta x - \frac{1}{3}\Delta y - \frac{1}{6}\Delta z$ where $\Delta x = 0.03$, $\Delta y = -0.02$, and $\Delta z = 0.05$ (linearize $f(x,y,z) = x^2y^{-1/3}z^{-1/6}$ at $(1,1,1)$). **26.** $1 + \Delta x$ where $\Delta x = -0.03$ (linearize $f(x,y) = x^y$ at $(1,1)$).

Section 22.7. **2.** $dz/dt = (12t^5 + 2\ln t/t)[1 + 4t^6 + 2(\ln t)^2]^{-1/2}$. **3.** $z'_t = z'_x s + z'_y t/\sqrt{t^2+s^2}$ and $z'_s = z'_x t + z'_y s/\sqrt{t^2+s^2}$, where $z'_x = e^{-x}[y\cos(xy) - \sin(xy)]$ and $z'_y = e^{-x}x\cos(xy)$. **6.** $z'_u = 23$, $z'_v = 32$. **9.** $z''_{uu} = 2f'_x + 8uvf''_{xy} + 4u^2f''_{xx} + 4v^2f''_{yy}$, $z''_{vv} = 2f'_x + 8uvf''_{xy} + 4u^2f''_{yy} + 4v^2f''_{xx}$, $z''_{uv} = z''_{vu} = 2f'_y + 2y(f''_{xx} + f''_{yy}) + 4xf''_{xy}$. **11.** $g'_x = 2xf'(u)$, $g'_y = 2yf'(u)$, $g'_z = 2zf'(u)$, $g''_{xx} = 2f'(u) + 4x^2f''(u)$, $g''_{yy} = 2f'(u) + 4y^2f''(u)$, $g''_{zz} = 2f'(u) + 4z^2f''(u)$, $g''_{xy} = g''_{yx} = 4xyf''(u)$, $g''_{xz} = g''_{zx} = 4xzf''(u)$, $g''_{zy} = g''_{yz} = 4zyf''(u)$ where $u = x^2 + y^2 + z^2$. **12.** If $u = x$ and $v = x/y$ so that $g(x,y) = f(u,v)$, then $g'_x = f'_u + \frac{1}{y}f'_v$, $g'_y = -\frac{x}{y^2}f'_v$, $g''_{xx} = f''_{uu} + \frac{2}{y}f''_{uv} + \frac{1}{y^2}f''_{vv}$, $g''_{yy} = \frac{2x}{y^3}f'_v + \frac{x^2}{y^4}f''_{vv}$, $g''_{xy} = g''_{yx} = -\frac{1}{y^3}[yf'_v + xyf''_{uv} + xf''_{vv}]$. **25.** $z'_x = (e^z - 3)^{-1}$, $z'_y = 2(e^z - 3)^{-1}$, $z''_{xx} = -e^z(e^z - 3)^{-3}$, $z''_{xy} = z''_{yx} = -2e^z(e^z - 3)^{-3}$, $z''_{yy} = -4e^z(e^z - 3)^{-3}$

where $z \neq \ln 3$. **31.** $z(2.8, -2.3) \approx 2 + \frac{2}{9}\Delta x - \frac{1}{9}\Delta y$ where $\Delta x = -0.2$ and $\Delta y = -0.3$. **36.** At the point $(2, 6, 8)$, the temperature is increasing at the rate $dT/dt = 4.5$ degrees Celsius per second. **39.** The volume is increasing at the rate 6 m^3/s, the surface area is increasing at the rate 10 m^2/s, and the diagonal is not changing. **40.** Examples: $x + 3y$ if $n = 1$, $x^2 + xy - y^2$ if $n = 2$, $x^3 - 4y^3 + 2xy^2$ if $n = 3$, etc.

Section 23.7. **1.** $df = (3x^2 - 6xy + 3y^2)dx + (3y^2 - 3x^2 + 6xy)dy$. **2.** $df = -2xy^2 \sin(x^2 y)dx + [\cos(x^2 y) - yx^2 \sin(x^2 y)]dy$. **3.** $df = 2\cos(x^2 + y^2)(xdx + ydy)$. **4.** $df = (1 + y^2 z e^{xyz})dx + (z + xzye^{xyz} + e^{xyz})dy + (y + xy^2 e^{xyz})dz$. **8.** Let $f(x, y, z, u) = xyzu$; the rounding error is bounded by 0.5 so that the differentials dx, dy, dz, and du, representing the rounding errors, cannot exceed 0.5, and, hence, df cannot exceed $4 \cdot 10^6 \cdot 0.5 = 2 \cdot 10^6$. **9.** 30 m^2, the error bound of the estimate is 0.04 m^2. **10.** The area is decreased by 0.14 m^2 (the error of the estimate is approximately 10^{-3} m^2), the diagonal is decreased by 2.8 cm (the error of the estimate is approximately 0.01 cm). **11.** Put $S = R^2\theta/2$, $dS = 0$, and $d\theta = \pi/180$ so that $dR = -1/6$ cm (the error of the estimate is approximately $\frac{3}{8}R(d\theta/\theta)^2 = 1/480 \approx 0.002$ cm). **13.** The volume is $\pi(15.0 \pm 2.3)$ and the relative error is 15.5%. **14.** The third side is 173 ± 6 m and the relative error is 3.5%. **15.** 0.5%. **19.** $df = (1 + 2xy)dx + (-1 + x^2)dy$, $d^2 f = 2y(dx)^2 + 4xdxdy$, $d^3 f = 6(dx)^2 dy$, and $d^n f = 0$ for $n \geq 4$. **22.** $df = yzdx + xzdy + xydz$, $d^2 f = 2(zdxdy + ydxdz + xdydz)$, $d^3 f = 6dxdydz$, and $d^n f = 0$ for $n \geq 4$. **23.** If $u = xyz$, then $df = -(1+u)^{-2}du$ and $d^2 f = 2(1+u)^{-3}(du)^2 - (1+u)^{-2}d^2 u$ (see the previous answer for du and $d^2 u$). **27.** $T_2 = 3 + 3(x-1) + 3(y-1) + 2(x-1)^2 + (y-1)^2$. **28.** $T_2 = 1 - \frac{1}{2}(x - \frac{\pi}{2})^2 - \frac{\pi^2}{8}(y-1)^2 - \frac{\pi}{2}(x - \frac{\pi}{2})(y-1)$. **29.** $T_2 = 1 + (x - 1) + (x-1)(y-1)$. **31.** $T_0 = 1$, $T_1 = 1 + \frac{1}{2}(x + 2y)$, $T_2 = T_1 - \frac{1}{8}(x + 2y)^2$. **32.** $T_0 = T_1 = 0$, $T_2 = T_3 = xy$, $T_4 = xy(1 + x^2 + y^2)$; **33.** $T_0 = 0$, $T_1 = x + 2y$, $T_2 = x + 2y + z^2$, $T_3 = T_2 - \frac{1}{6}(x + 2y)^3$. **38.** $P_2 = -xy - yz - xz$ **39.** $T_0 = f(x_0, y_0)$, $T_1(r) = T_0$, $T_2(r) = T_0 + (f''_{xx}(x_0, y_0) + f''_{yy}(x_0, y_0))r^2/4$, $T_3(r) = T_2(r)$.

Section 24.5. **1.** If $\nabla f \neq \mathbf{0}$ at a point, then the function has no rate of change at that point in all direction orthogonal to ∇f (in the two-variable case, two such directions exist and in the three-variable case, any direction in the plane orthogonal to ∇f. **3.** $\nabla f(1, 2) = \langle 4, 1\rangle$, $D_{\mathbf{v}} f(1, 2) = \frac{19}{5} > 0$ (increasing). **4.** $\nabla f(1, 1) = \langle \frac{1}{4}, -\frac{1}{4}\rangle$, $D_{\mathbf{v}} f(1, 1) = \frac{1}{4\sqrt{5}} > 0$ (increasing). **5.** $\nabla f(1, 2, -1) = \langle 5, 5, -6\rangle$, $D_{\mathbf{v}} f(1, 2, -1) = -\frac{17}{3} < 0$ (decreasing) **6.** $\nabla f(1, -1, 1) = \frac{1}{17}\langle 1, -2, 3\rangle$, $D_{\mathbf{v}} f(1, 2, -1) = \frac{2}{17\sqrt{3}} > 0$ (increasing). **7.** $\nabla f(1, 1, 3) = \frac{1}{4}\langle 1, 3, 1\rangle$, $D_{\mathbf{v}} f(1, 1, 3) = \frac{23}{28} > 0$ (increasing), **8.** $\nabla f(2, 1, 1) = \langle 1, 1, -3\rangle$, $D_{\mathbf{v}} f(2, 1, 1) = \frac{7}{3} > 0$ (increasing). **9.** The maximal rate $\sqrt{17}$ occurs in the direction $\langle 1, -4\rangle$, the minimal rate $-\sqrt{17}$ occurs in the direction $\langle -1, 4\rangle$, the rate vanishes in the directions $\langle 4, 1\rangle$

and $\langle -4, -1\rangle$. **10.** The maximal rate $\sqrt{1+4(\ln 2)^2}$ occurs in the direction $\langle 1, 2\ln 2\rangle$, the minimal rate $-\sqrt{1+4(\ln 2)^2}$ occurs in the direction $\langle -1, -2\ln 2\rangle$, the rate vanishes in the directions $\langle 2\ln 2, -1\rangle$ and $\langle -2\ln 2, 1\rangle$. **11.** The maximal rate $\sqrt{523}/49$ occurs in the direction $\langle 21, -9, 1\rangle$, the minimal rate $-\sqrt{523}/49$ occurs in the direction $\langle -21, 9, -1\rangle$, the rate vanishes along any direction orthogonal to $\mathbf{n} = \langle 21, -9, 1\rangle$ (along any vector in the plane orthogonal to $\mathbf{n}$). **14.** Define the angle ϕ by $\cos\phi = a/\sqrt{a^2+b^2}$ and $\sin\phi = b/\sqrt{a^2+b^2}$ where $a = 3$ and $b = 5/4$. Let $\hat{\mathbf{u}} = (\cos\theta, \sin\theta)$. Then $\cos(\theta - \phi) = p$ or $\theta = \cos^{-1} p + \phi$. **15.** (i) $\sqrt{46}$, $\langle -1, -3, 6\rangle$; (ii) all vectors that make the angle $\pi/3$ with the vector $\mathbf{n} = \langle -1, -3, 6\rangle$ (they form a cone about $\mathbf{n}$); (iii) 13/3. **23.** The direction toward P_2; f increases most rapidly in the direction of $\mathbf{v} = \langle 2, 1, 1\rangle$. **24.** (i) $\sqrt{2}$; (ii) 1; (iii) $\pi/4$. **25.** $-8/(27\sqrt{33})$. **26.** (i) all points on the surface $z^2 = xy$; (ii) all points $(0, 0, a)$ where $a \neq 0$; (iii) all points (a, a, a) where a is real. **28.** The plane: $3(x-3) + 4(y-4) + 12(z-12) = 0$; the line: $(x-3)/3 = (y-4)/4 = (z-12)/12$. **29.** the plane: $4(x-2) - 5(y-1) - (z+1) = 0$; the line: $(2-x)/4 = (y-1)/5 = z+1$. **30.** The plane: $2(x-\pi/4) - y - z = 0$; the line: $z - 1 = 1 - y = (x - \pi/4)/2$. **34.** The tangent plane is parallel to the yz plane at the points $(4, -2, 0)$ and $(-4, 2, 0)$, to the zx plane at the points $(-2, 4, -2)$ and $(2, -4, 2)$, and to the xy plane at the points $(0, 2\sqrt{2}, -2\sqrt{2})$ and $(0, -2\sqrt{2}, 2\sqrt{2})$. **35.** The planes through the points $(1, 2, 2)$ and $(-1, -2, -2)$ and orthogonal to $\mathbf{n} = \langle 1, 4, 6\rangle$. **41.** Toward P_1. **42.** Shoot in the direction $\mathbf{v} = \langle -2, -6, -3\rangle$; the escape trajectory is traced out by the vector function $\mathbf{r}(t) = \langle 2t, 3t^2, t^3\rangle$ where $t \geq 1$. **44.** $f'_x(P_0) = (b\sin\theta - a\sin\phi)/\sin(\theta-\phi)$ and $f'_y(P_0) = (a\cos\phi - b\cos\theta)/\sin(\theta - \phi)$. **45.** Drill in the direction of $\mathbf{a} = \langle 1, 3, 1\rangle$; the concentration is increasing at the rate $\sqrt{11}$ g/m in this direction. **46.** Yes, the tangent plane is $12(x-2) - 7(y-1) + 9(z-3) = 0$. **52.** The focus is at $(0, 0, a/4)$.

Section 25.6. **1.** (i) A local maximum; (ii) a local minimum; (iii) a saddle point; (iv) no information. **2**. A local minimum at $(0, 2)$. **3**. A saddle at $(0, 2)$. **6.** A local minimum at $(1, 0)$. **7.** A local minimum at $(3, \frac{1}{2})$, a saddle at $(-1, \frac{1}{2})$. **8.** A local minimum at $(1, 1)$, saddle at $(0, 0)$. **9.** A local minimum at $(5, 2)$. **10.** Local minima at $(1, \pm 1)$ and $(\pm 1, 1)$. **11.** Local minima at $(\pi n, \pi m)$ if $n + m$ is odd, local maxima at $(\pi n, \pi m)$ if $n + m$ is even, and saddles at $(\frac{\pi}{2} + \pi n, \frac{\pi}{2} + \pi m)$, where n and m are integers. **12.** Local minima at $(\pi n, 0)$ if n is odd, saddles at $(\pi n, 0)$ if n is even, where n is an integer. **13.** A saddle at $(0, 0)$, a local maximum at $(-\frac{1}{2}, -1)$. **14.** A saddle at $(0, 0)$, a local minimum at $(\frac{2}{3}, \frac{2}{3})$. **15.** Saddles at $(0, 0)$, $(1, 0)$, and $(0, 1)$; a local maximum at $(\frac{1}{3}, \frac{1}{3})$. **16.** Saddles at $(0, \frac{\pi}{2} + \pi n)$ where n is an integer. **28.** There are two solutions $z = z(x, y)$, one attains a maximum value 6 at $(1, -1)$ and the other attains a minimum value -2 at $(1, -1)$.

Section 26.6. **1.** A local minimum at $(-1,-2,4)$. **2.** A saddle at $(0,0,1)$; a local minimum at $(-144,24,1)$. **3.** Saddles at $(0,0,1)$ and $(-144,24,1)$. **4.** Saddles at $(\frac{\pi}{2}+\pi n, \pi m, 0)$ where n and m are integers. **5.** A saddle at $(0,0,0)$; a local minimum at $(2,2,4)$. **13.** A local minimum (complete the squares: $(x+\frac{1}{2}y^2)^2+\frac{3}{4}y^4$). **14.** A local minimum. **15.** A local minimum. **16.** No local extremum. **17.** no local extremum. **18.** A local maximum. **19.** A local minimum. **20.** A local minimum. **21.** A local maximum. **22.** No local extremum. **24.** Let $0<a<u<b$ for some a and b. Take a continuous function $g(u)$ on (a,b) such that it attains maximum and minimum values in (a,b). Put $f(x,y)=g(\sqrt{x^2+y^2})$. The function f attains extreme values in the open region $a^2<x^2+y^2<b^2$. **25.** Take $g(u)$ on $[a,b]$, $a>0$, that is not continuous and attains its extreme values in $[a,b]$. Put $f(x,y)=g(\sqrt{x^2+y^2})$. **26.** A similar construction as in **24** and **25** on (a,∞), $a>0$. **27.** $\max_D f=f(2,1)=2$ and $\min_D f=f(1,2)=-3$. **28.** $\max_D f=f(1,\pm 2)=8$ and $\min_D f=f(0,0)=-1$. **29.** $\min_D f=f(a,0)=f(0,b)=0$, where a and b are real, and $\max_D f=f(2\sqrt{2}/\sqrt{3},2/\sqrt{3})=16/3\sqrt{3}$. **30.** $\max_D f=f(1,\pm 1,1)=2$ and $\min_D f=f(-1,\pm 1,-1)=-2$.

Section 27.6. **1.** $\max f=\frac{1}{4}$. **3.** $\min f=-4$, $\max f=4$. **4.** $\min f=0$, $\max f=4$. **5.** $\min f=-5$, $\max f=5$. **6.** $\min f=-8$, $\max f=8$. **8.** $\min f=-2/\sqrt{3}$, $\max f=2/\sqrt{3}$. **9.** $\min f=-3$, $\max f=3$. **11.** $\min f=-2\sqrt{6}$, $\max f=2\sqrt{6}$. **12.** $\min f=1-2^{-3/2}$, $\max f=1+2^{-3/2}$. **25.** If the numbers are x and y, and $xy=p$, then $x=y=\sqrt{p}$. **26.** If the numbers are x_j, $j=1,2,...,m$, and $x_1x_2\ldots x_m=p$, then $x_j=p^{1/m}$, $j=1,2,...,m$. **28.** If R and h are the radius and the length of the tab, then $h=2R$ and $R=\sqrt{S/(3\pi)}$. **31.** The base of the box is a square whose side is $2R/\sqrt{3}$, and the height of the box is $R/\sqrt{3}$. **32.** $|x|\le a/\sqrt{3}$, $|y|\le b/\sqrt{3}$, $|z|\le c/\sqrt{3}$.