CHAPTER 4

Multiple Integrals

28. Double Integrals

28.1. The Volume Problem. Suppose one needs to determine the volume of a hill whose height $f(\mathbf{r})$ as a function of position $\mathbf{r} = \langle x, y\rangle$ in the base of the hill is known. For example, the hill must be leveled to construct a highway. Its volume is required to estimate the number of truck loads needed to move the soil away. The following procedure can be used to estimate the volume. The base D of the hill is first partitioned into small pieces D_p of area ΔA_p, where $p = 1, 2, ..., N$ enumerates the pieces; that is, the union of all the pieces D_p is the region D. The partition elements should be small enough so that the height $f(\mathbf{r})$ has no significant variation when $\mathbf{r}$ ranges over D_p. The volume of the portion of the hill above each partition element D_p is approximately $\Delta V_p \approx f(\mathbf{r}_p)\,\Delta A_p$, where $\mathbf{r}_p$ is a point in D_p (see the left panel of Fig. **28.1**). The volume of the hill can therefore be estimated as

$$V \approx \sum_{p=1}^{N} f(\mathbf{r}_p)\,\Delta A_p.$$

For practical purposes, the values $f(\mathbf{r}_p)$ can be found, for example, from a contour map of f.

The above approximation neglects variations of values of f within a partition element D_p. Therefore it is expected to become more accurate with decreasing the size of the partition elements (naturally, their number N has to increase). If R_p is the smallest radius of a disk that contains D_p, then put $R_N^* = \max_p R_p$, which determines the size of the largest partition element. One says that the partition is *refined* if R_N^* is decreasing with increasing the number N of partition elements. Note that the reduction of the maximal area $\max_p \Delta A_p$ versus the maximal size R_N^* may not be good enough to improve the accuracy of the estimate. If D_p looks like a narrow strip, its area is small, but the variations f along the strip may be significant and the accuracy of the approximation $\Delta V_p \approx f(\mathbf{r}_p)\,\Delta A_p$ is poor. One can therefore expect that the exact value of the volume is obtained in the limit

$$V = \lim_{\substack{N\to\infty \\ (R_N^*\to 0)}} \sum_{p=1}^{N} f(\mathbf{r}_p)\,\Delta A_p\,. \qquad (\mathbf{28.1})$$

The volume V may be viewed as the volume of a solid bounded from above by the surface $z = f(x, y)$, which is the graph of f, and by the portion D of

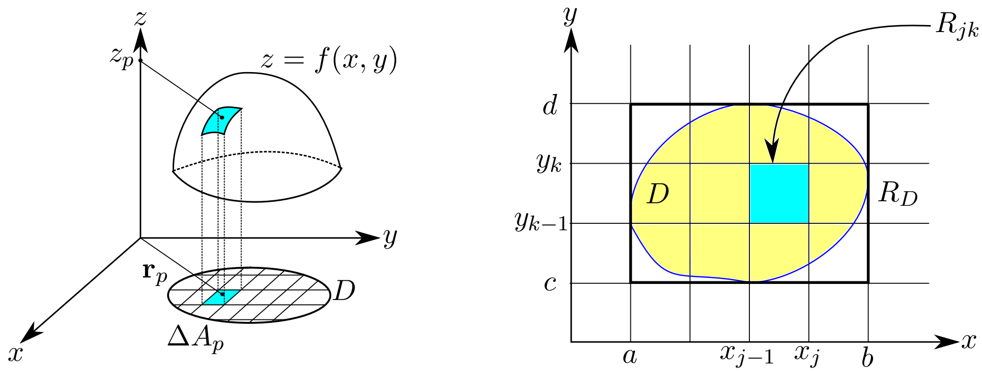


FIGURE **28.1**. **Left**: The volume of a solid region bounded from above by the graph $z = f(x, y)$ and from below by a portion D of the xy plane is approximated by the sum of volumes $\Delta V_p = z_p \Delta A_p$ of columns with the base area ΔA_p and the height $z_p = f(\mathbf{r}_p)$ where $\mathbf{r}_p$ is a sample point within the base and p enumerates the columns. **Right**: A rectangular partition of a region D is obtained by embedding D into a rectangle R_D. Then the rectangle R_D is partitioned into smaller rectangles R_{kj}.

the xy plane. Naturally, it is not expected to depend on the way the region D is partitioned, neither should it depend on the choice of *sample points* $\mathbf{r}_p$ in each partition element.

The limit (**28.1**) resembles the limit of a Riemann sum for a single-variable function $f(x)$ on an interval $[a, b]$ used to determine the area under the graph of f. Indeed, if x_k, $k = 0, 1, ..., N$, $x_0 = a < x_1 < \cdots < x_{N-1} < x_N = b$ is the partition of $[a, b]$, then ΔA_p is the analog of $\Delta x_j = x_j - x_{j-1}$, $j = 1, 2, ..., N$, the number R_N^* is the analog of $\Delta_N = \max_j \Delta x_j$, and the values $f(\mathbf{r}_p)$ are analogous to $f(x_j^*)$, where x_j^* is in $[x_{j-1}, x_j]$. The area under the graph is then

$$A = \lim_{\substack{N\to\infty \\ (\Delta_N \to 0)}} \sum_{j=1}^{N} f(x_j^*)\, \Delta x_j = \int_a^b f(x)\, dx\,.$$

So, the limit (**28.1**) seems to define an integral over a two-dimensional region D (i.e., with respect to both variables x and y used to label points in D). This observation leads to the concept of a *double integral*. However, the qualitative construction used to analyze the volume problem still lacks the level of rigor used to define the single-variable integration. For example, how does one choose the "shape" of the partition elements D_p, or how does one calculate their areas? These kinds of questions were not even present in the single-variable case and have to be addressed.

28.2. Preliminaries. The *closure* of a set D in a Euclidean space is the set obtained from D by adding all its limit points to D. The closure of D is denoted $\bar{D}$. For example, let D be the open disk $x^2 + y^2 < 1$. Every point of D is a limit point and every point of the circle $x^2 + y^2 = 1$ is also a limit point. Therefore the closure $\bar{D}$ is the (closed) disk $x^2 + y^2 \leq 1$.

Definition **28.1**. (A Region in a Euclidean Space)
An open connected set in a Euclidean space is called an open region*. The closure of an open connected set is called a* closed region*. A set D is a region in a Euclidean space if there is an open region G that is contained in D while the closure of G contains D.*

The whole idea of introducing the notion of a region is to give a name to sets in a plane that have a non-zero area and to sets in space that have a non-zero volume. Note that a region in a plane always contains an open set and this open set has a disk that lies in it. As any disk has a non-zero area, a region is expected to have a non-zero area. In particular, the volume problem considered above makes sense if D is a region. But in order to make the notion of the area (or volume) of a region precise, some additional conditions on the *boundary* of the region have to be imposed. If all points of an open region D are removed from its closure $\bar{D}$, then the obtained set is called *the boundary of D*. In other words,

The boundary of an open region D is the difference of its closure $\bar{D}$ and D itself.

For example, if D is the open disk $x^2 + y^2 < 1$, then its closure is the closed disk $x^2 + y^2 \leq 1$, and the difference between the two sets is the circle $x^2 + y^2 = 1$ which is the boundary of D. Now recall that a point of a set is an interior point of the set if there is an open ball of sufficiently small radius that contains the point and lies in the set. So,

The boundary of a closed region D is obtained from D by removing all interior points of D.

Clearly, if D is an open region, then the interior of $\bar{D}$ is D. Let G be a region. Then by definition there exists an open region D that it lies in G, while G is contained in the closure $\bar{D}$:

$$D \subset G \subset \bar{D}\,.$$

Then D is nothing but the collection of all interior points of G. It follows that the boundary of G coincides with the boundary of D or the boundary of $\bar{D}$ (since D and $\bar{D}$ have the same boundary). Thus, *the difference between an open region and a region is that the region may contain its boundary or a part of it, while an open region contains no point of its boundary.*

For example, let G be the set in the xy plane defined by the conditions $x^2 + y^2 < 1$ if $y \geq 0$ and $x^2 + y^2 \leq 1$ if $y < 0$. So, G is the disk of radius 1. The upper part of its boundary circle ($y \geq 0$) does not belong to G, while its lower part lies in G. The largest open set D that is contained in G is the open disk $x^2 + y^2 < 1$. It is an open region. Its closure $\bar{D}$ is the closed disk $x^2 + y^2 \leq 1$. Evidently, $D \subset G \subset \bar{D}$. So, the boundary of G is the circle $x^2 + y^2 = 1$. Note that G contains a part of its boundary.

Definition **28.2**. (Smooth Boundary of a Region)
The boundary of a region is called smooth *if in a neighborhood of every point it coincides with a level set of a function that has continuous partial derivatives and whose gradient does not vanish. The boundary is called* piecewise smooth *if it consists of finitely many smooth pieces.*

Let D be a region in the plane. The boundary of D is smooth if in a neighborhood of each point (x_0, y_0) of the boundary there is a function g of two variables such that the boundary is the level set $g(x, y) = g(x_0, y_0)$, where the function g has continuous partial derivatives, and $\boldsymbol{\nabla} g \neq \mathbf{0}$. Recall from Section **24.2** that under these conditions on g, the level set is a smooth curve. Similarly, a smooth boundary of a region in space is a smooth surface. For example, the disk $x^2 + y^2 < 1$ has the boundary $x^2 + y^2 = 1$ which is the level curve of the function $g(x, y) = x^2 + y^2$. The boundary of the ball in space $x^2 + y^2 + z^2 < 1$ is the sphere $x^2 + y^2 + z^2 = 1$, which is the level set of the function $g(x, y, z) = x^2 + y^2 + z^2$. In both the cases, g has continuous partial derivatives and $\boldsymbol{\nabla} g \neq \mathbf{0}$ near any point of the level set $g = 1$. By the properties of the gradient, $\boldsymbol{\nabla} g$ is normal to the boundary and its components are continuous functions. An open rectangle in the plane, $a_1 < x < b_1$ and $a_2 < y < b_2$, is a region whose boundary is piecewise smooth as it consists of four straight line segments and each segment is a smooth curve. Similarly, an open rectangular box in space, $a_1 < x < b_1$, $a_2 < y < b_2$, and $a_3 < z < b_3$, is a region whose boundary is piecewise smooth as it consists of six (coordinate) planes that are smooth surfaces.

Definition **28.3**. (Bounded Functions)
A function f is called bounded on a set D if there are numbers m and M such that $m \leq f(\mathbf{r}) \leq M$ for all $\mathbf{r}$ in D. The numbers m and M are called, respectively, lower and upper bounds of f on D.

Evidently, upper and lower bounds are not unique because any number smaller than m is also a lower bound, and, similarly, any number greater than M is an upper bound. However, the smallest upper bound and the largest lower bound are unique.

Definition **28.4**. (Supremum and Infimum).
Let f be bounded on a set D. The smallest upper bound of f on D is called the supremum *of f on D and denoted by $\sup_D f$. The largest lower bound of f on D is called the* infimum *of f on D and denoted by $\inf_D f$.*

In other words, $\sup_D f$ is an upper bound of f on D such that the number $\sup_D f - a$ is not an upper bound for any positive number $a > 0$. Similarly, $\inf_D f$ is a lower bound of f on D such that the number $\inf_D f + a$ is not a lower bound for any positive number $a > 0$. If f is continuous and the set D is closed and bounded, then by the extreme value theorem (Theorem **26.2**) the function f attains it absolute maximum and minimum values on D, and in this case

$$\inf_D f = \min_D f\,, \qquad \sup_D f = \max_D f\,.$$

Let

$$f(x,y) = x^2 + y^2$$

and D be the open rectangle:

$$D: \quad 0 < x < 1\,, \qquad 0 < y < 1\,.$$

Then f does not attain its extreme values on D despite that it is continuous on D. But

$$\sup_D f = 2\,, \qquad \inf_D f = 0\,.$$

Indeed, $f(x,y) = x^2 + y^2 < 2$ for all (x,y) in D so that 2 is an upper bound. For any number $a > 0$, one can find points in D such that $f(x,y) > 2 - a$ and hence $2 - a$ is not an upper bound. So, 2 is the smallest upper bound. Similarly, $f(x,y) = x^2 + y^2 > 0$ for all (x,y) in D so that 0 is a lower bound of f on D. For any number $a > 0$, one can find a point (x,y) in D such that $f(x,y) < 0 + a = a$ and hence a is not a lower bound. Therefore 0 is the greatest lower bound of f on D.

28.3. Double Integral. Suppose D is a bounded, closed region in the plane and f is a bounded function of two variables (x,y) on D. The function f is *extended* to the whole plane by setting $f(x,y) = 0$ if (x,y) is not in D. Since D is bounded, it can always be embedded into a closed rectangle

$$R_D = \{(x,y)\,|\, a \le x \le b,\ c \le y \le d\} = [a,b] \times [c,d]$$

The latter equality defines a short notation for a rectangle in a plane. Consider a partition x_j of the interval $[a,b]$ and a partition y_k of the interval $[c,d]$ where

$$\begin{aligned} x_j &= a + j\,\Delta x\,, \quad \Delta x = (b-a)/N_1\,, \quad j = 0,1,...,N_1\,,\\ y_k &= c + k\,\Delta y\,, \quad \Delta y = (d-c)/N_2\,, \quad k = 0,1,...,N_2\,. \end{aligned}$$

These partitions induce a partition of the rectangle R_D by rectangles

$$R_{jk} = [x_{j-1}, x_j] \times [y_{k-1}, y_k]\,, \quad j = 1,2,...,N_1\,, \quad k = 1,2,...,N_2\,.$$

The area of each partition rectangle R_{kj} is $\Delta A = \Delta x\,\Delta y$. This partition is called a *rectangular partition* of R_D. It is depicted in the right panel of

Fig. **28.1**. For every pair (j,k) put

$$M_{jk} = \sup_{R_{jk}} f\,, \qquad m_{jk} = \inf_{R_{jk}} f$$

which are the supremum and infimum of f on R_{jk}, respectively.

Definition **28.5**. (Upper and Lower Sums).
Let f be a bounded function on a bounded, closed region D. Let R_D be a rectangle that contains D and let the function f be defined to have zero value for all points of R_D that do not belong to D. Given a rectangular partition R_{jk}, the sums

$$U(f,N_1,N_2) = \sum_{j=1}^{N_1}\sum_{k=1}^{N_2} M_{jk}\,\Delta A\,, \quad L(f,N_1,N_2) = \sum_{j=1}^{N_1}\sum_{k=1}^{N_2} m_{jk}\,\Delta A$$

are called the upper and lower sums *of f over D.*

The upper and lower sums are examples of double sequences.

Definition **28.6**. (Double Sequence)
A double sequence is a rule that assigns a real number a_{nm} to an ordered pair of integers (n,m), $n,m = 1,2,\ldots.$

In other words, a double sequence is a function f of two variables (x,y) whose domain consists of points with positive integer-valued coordinates, $a_{nm} = f(n,m)$. Similarly to ordinary numerical sequences, the limit a double sequence can be defined.

Definition **28.7**. (The Limit of a Double Sequence)
If for any positive number $\varepsilon > 0$ there exists an integer N such that

$$|a_{nm} - a| < \varepsilon \quad \text{for all } n,m > N\,,$$

then the sequence is said to converge to a and the number a is called the limit of the sequence and denoted

$$\lim_{n,m\to\infty} a_{nm} = a\,.$$

In other words, the number a is the limit of a double sequence a_{nm}, if the deviation $|a - a_{nm}|$ of values of a_{nm} from a can be made arbitrary small for *all* sufficiently large integers n and m. One can also say that the number a is the limit of the sequence a_{nm} if there are only *finitely many* terms of the sequence a_{nm} outside any interval $(a-\varepsilon, a+\varepsilon)$, where $\varepsilon > 0$, no matter how small is ε.

The limit of a double sequence is analogous to the limit of a function of two variables. It can be found by studying the corresponding limit of a function of two variables whose range contains the double sequence. Suppose $a_{nm} = f(1/n, 1/m)$ and $f(x,y) \to a$ as $(x,y) \to (0,0)$. The latter means that for any $\varepsilon > 0$ there is a number $\delta > 0$ such that $|f(x,y) - a| < \varepsilon$ for all $\|\mathbf{r}\| < \delta$ where $\mathbf{r} = \langle x, y\rangle$. In particular, for $\mathbf{r} = \langle 1/n, 1/m\rangle$, the condition

$\|\mathbf{r}\|^2 = 1/n^2 + 1/m^2 < \delta^2$ is satisfied for all $n, m > N > 2/\delta$. Hence, for all such n, m, $|a_{nm} - a| < \varepsilon$, which means that $a_{nm} \to a$ as $n, m \to \infty$.

Definition **28.8**. (Double Integral). *If the limits of the upper and lower sums exist as $(N_1, N_2) \to \infty$ and coincide, then f is said to be Riemann integrable on D, and the limit of the upper and lower sums*

$$\lim_{N_{1,2}\to\infty} U(f, N_1, N_2) = \lim_{N_{1,2}\to\infty} L(f, N_1, N_2) = \iint_D f(x, y)\, dA$$

is called the double integral *of f over the region D.*

Let us discuss Definition **28.8** from the point of view of the volume problem. If $f(x, y) \geq 0$ in D, then for a given partition *the upper and lower sums represent the smallest upper estimate and the greatest lower estimate of the volume* of the solid region under the graph $z = f(x, y)$ above the region D. The values of the sums should become closer to the volume as the partition becomes finer, that is, the limits of $L(f, N_1, N_2)$ and $U(f, N_1, N_2)$ as $N_1, N_2 \to \infty$ exist and coincide with the volume under the graph $z = f(x, y)$ over D. However, a specific partition of D by rectangles has been used in the definition of the double integral. In this way, the area ΔA_p of the partition element has been given a precise meaning as the area of a rectangle. Later, it will be shown that if the double integral exists in the sense of the above definition, then it exists if the rectangular partition is replaced by any partition of D by elements D_p of an arbitrary shape subject to certain conditions that allow for a precise evaluation of their areas.

Example **28.1**. *Determine whether the function $f(x, y) = xy$ is integrable on the region $D = [0, 1] \times [0, 1]$ and find the double integral if it exists.*

Solution: The region D is embedded into the closed rectangle $R_D = [0, 1] \times [0, 1]$. Put

$$\Delta x = \frac{1}{N_1}\,, \quad \Delta y = \frac{1}{N_2}\,, \quad x_j = j\Delta x = \frac{j}{N_1}\,, \quad y_k = k\Delta y = \frac{k}{N_2}\,,$$

where $j = 1, 2, ..., N_1$ and $k = 1, 2, ..., N_2$. Since $x_j > x_{j-1} \geq 0$ and $y_k > y_{k-1} \geq 0$, the numbers $f(x_j, y_k) = x_j y_k$ and $f(x_{j-1}, y_{k-1}) = x_{j-1} y_{k-1}$ are the maximum and minimum values of the function f on the partition rectangle $R_{jk} = [x_{j-1}, x_j] \times [y_{k-1}, y_k]$:

$$M_{jk} = x_j y_k = \frac{jk}{N_1 N_2}\,, \quad m_{jk} = x_{j-1} y_{k-1} = \frac{(j-1)(k-1)}{N_1 N_2}\,.$$

Now recall that

$$1 + 2 + \cdots + N = \sum_{k=1}^{N} k = \frac{1}{2} N(N+1)\,.$$

The following relation is useful to convert a double sum into the product of sums:

$$\begin{aligned}\Big(\sum_{j=1}^{N_1} j\Big)\Big(\sum_{k=1}^{N_2} k\Big) &= \Big(1+2+\cdots+N_1\Big)\Big(1+2+\cdots+N_2\Big)\\ &= 1\cdot 1+1\cdot 2+2\cdot 1+2\cdot 2+\cdots+N_1\cdot N_2\\ &= \sum_{j=1}^{N_1}\sum_{k=1}^{N_2} jk\,.\end{aligned}$$

The upper and lower sums are

$$\begin{aligned}U(f,N_1,N_2) &= \frac{\Delta x\Delta y}{N_1N_2}\sum_{j=1}^{N_1}\sum_{k=1}^{N_2} jk = \frac{1}{N_1^2N_2^2}\Big(\sum_{j=1}^{N_1} j\Big)\Big(\sum_{k=1}^{N_2} k\Big)\\ &= \frac{1}{4}\,\frac{N_1(N_1+1)N_2(N_2+1)}{N_1^2N_2^2}\\ &= \frac{1}{4}+\frac{1}{N_1}+\frac{1}{N_2}+\frac{1}{N_1N_2}\\ L(f,N_1,N_2) &= \frac{\Delta x\Delta y}{N_1N_2}\Big(\sum_{j=1}^{N_1}(j-1)\Big)\Big(\sum_{k=1}^{N_2}(k-1)\Big)\\ &= \frac{1}{4}\,\frac{N_1(N_1-1)N_2(N_2-1)}{N_1^2N_2^2}\\ &= \frac{1}{4}-\frac{1}{N_1}-\frac{1}{N_2}+\frac{1}{N_1N_2}\end{aligned}$$

The function f is integrable on D because

$$\lim_{N_{1,2}\to\infty} U(f,N_1,N_2) = \lim_{N_{1,2}\to\infty} L(f,N_1,N_2) = \frac{1}{4} = \iint_D f(x,y)dA$$

□

28.4. Riemann Sums.

Definition **28.9**. (Riemann Sum).
Let f be a bounded function on a region D that is contained in a rectangle R_D. Let f be defined by zero values outside of D in R_D. Let $\mathbf{r}_{jk}^*$ be a point in a rectangle R_{jk}, where R_{jk} form a rectangular partition of R_D. The sum

$$R(f,N_1,N_2) = \sum_{j=1}^{N_1}\sum_{k=1}^{N_2} f(\mathbf{r}_{jk}^*)\,\Delta A$$

is called a Riemann sum.

If $f(x,y)\geq 0$, then Riemann sums of f approximate the volume under the graph of f. Since the volume is also given by the double integral of f

over a region D, the sequence of Riemann sums is expected to converge to the double integral *independently of the choice of sample points* $\mathbf{r}^*_{jk}$. This is indeed so.

THEOREM **28.1**. (Convergence of Riemann Sums).
If a function f is integrable on D, then its Riemann sums for any choice of sample points $\mathbf{r}^*_{jk}$ converge to the double integral:

$$\lim_{N_{1,2}\to\infty} R(f,N_1,N_2) = \iint_D f\,dA.$$

PROOF. For any partition rectangle R_{jk} and any sample point $\mathbf{r}^*_{jk}$ in it,

$$m_{jk} \le f(\mathbf{r}^*_{jk}) \le M_{jk}\,.$$

Multiplying this inequality by ΔA and taking the sums over j and k, it is concluded that

$$L(f,N_1,N_2) \le R(f,N_1,N_2) \le U(f,N_1,N_2)\,.$$

Since f is integrable, the limits of the upper and lower sums exist and coincide. The conclusion of the theorem follows from the squeeze principle for limits. □

Approximations of a Double Integral. If f is integrable, its double integral can be approximated by a suitable Riemann sum. A commonly used choice of sample points is to take $\mathbf{r}^*_{jk}$ to be the intersection of the diagonals of partition rectangles R_{jk}, that is, $\mathbf{r}^*_{jk} = \langle \bar{x}_j, \bar{y}_k\rangle$, where $\bar{x}_j$ and $\bar{y}_k$ are the midpoints of the intervals $[x_{j-1}, x_j]$ and $[y_{k-1}, y_k]$, respectively. This rule is called the *midpoint rule*. In general, the accuracy of the approximation of a double integral by Riemann sums can be assessed by finding the upper and lower sums; their difference determines the upper bound on the absolute error of the approximation. Note first that, given a partition of D, the value of the double integral of f over D and the value of the Riemann sum lie between the values of the lower and upper sums. Therefore the following chain of inequalities holds:

$$-(U-L) = L-U \le L-R \le \iint_D f dA - R \le U-R \le U-L$$

where for the sake of brevity the arguments (f, N_1, N_2) in the upper, lower, and Riemann sums are omitted. This inequality can be written in the form

$$\left|\iint_D f dA - R(f,N_1,N_2)\right| \le U(f,N_1,N_2) - L(f,N_1,N_2)$$

for any choice of sample points in the Riemann sum. In Example **28.3**, the error of the approximation of the integral by a Riemann sum does not exceed $2/N_1 + 2/N_2$ for any choice of sample points.

Alternatively, if the double integral of an *integrable* function f over a region D is to be evaluated up to some significant decimals, the partition in the Riemann sum has to be refined until the value of the sum does not change in the significant digits.

28.5. Continuity and Integrability. Not every bounded function is integrable. There are functions whose behavior is so irregular that one cannot give any meaning to the volume under their graph by converging upper and lower sums.

An Example of a Nonintegrable Function. Consider the function

$$f(x,y) = \begin{cases} 1 & \text{if } x \text{ and } y \text{ are rational} \\ 2 & \text{if } x \text{ and } y \text{ are irrational} \\ 0 & \text{otherwise} \end{cases}$$

Let $D = [0,1] \times [0,1]$. Since D is a rectangle, one can take $R_D = D$ when constructing the lower and upper sums. Recall that any interval $[a,b]$ contains both rational and irrational numbers. Therefore, any partition rectangle R_{jk} contains points whose coordinates are both rational, or both irrational, or pairs of rational and irrational numbers. Hence,

$$M_{jk} = 2\,, \qquad m_{jk} = 0\,.$$

The lower sum vanishes

$$L = \sum_{j,k} m_{jk}\Delta A = 0$$

for any partition and therefore its limit is 0, whereas the upper sum is

$$U = \sum_{j,k} M_{jk}\Delta A = 2\sum_{jk} \Delta A = 2A = 2$$

for any partition, where $A = 1$ is the total area of all partition elements or the area of the square. The limits of the upper and lower sums do not coincide, $2 \neq 0$, and the double integral of f does not exist. The Riemann sum for this function can converge to any number between 2 and 0, depending on the choice of sample points. For example, if the sample points have rational coordinates, then the Riemann sum equals 1. If the sample points have irrational coordinates, then the Riemann sum equals 2. If the sample points are such that one coordinate is rational while the other is irrational, then the Riemann sum vanishes. Clearly, the conclusion about non-integrability of this function can be extended to any bounded region.

The following theorem describes a class of integrable functions that is sufficient in many practical applications.

THEOREM **28.2**. (Integrability of Continuous Functions).
Let D be a bounded, closed region whose boundaries are piecewise-smooth curves. If a function f is continuous on D, then it is integrable on D.

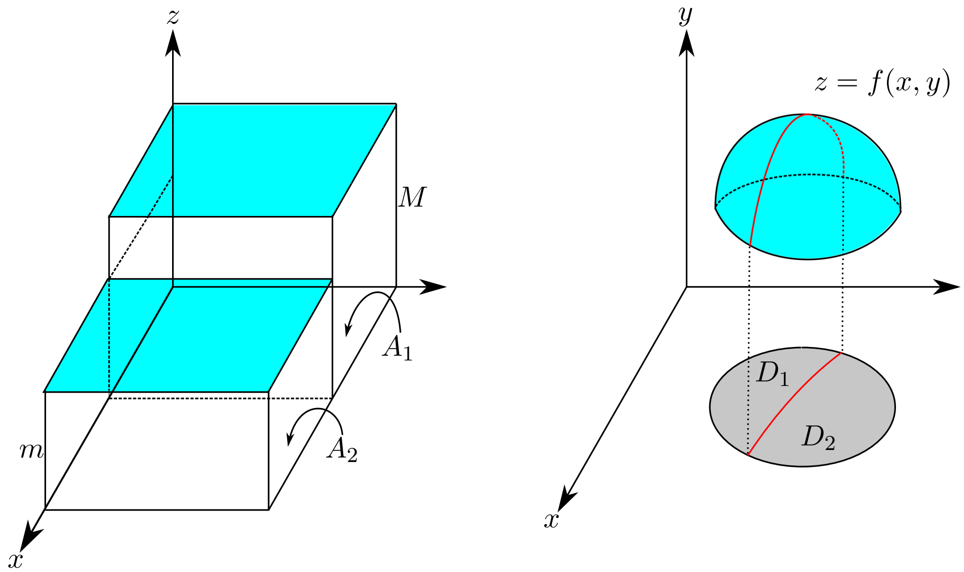


FIGURE **28.2**. **Left**: The graph of a piecewise constant function. The function is not continuous on a straight line. The volume under the graph is $V = MA_1 + mA_2$. Despite the discontinuity, the function is integrable and the value of the double integral coincides with the volume V. **Right**: Additivity of the double integral (see Section **29**). If a region D is split by a smooth curve into two regions D_1 and D_2, then the double integral of f over D is the sum of integrals over D_1 and D_2. The additivity of the double integral is analogous to the additivity of the volume: The volume under the graph $z = f(x, y)$ and above D is the sum of volumes above D_1 and D_2.

Note that the converse is not true; that is, the class of integrable functions is wider than the class of all continuous functions. This is a rather natural conclusion in view of the analogy between the double integral and the volume. Let $f(x, y)$ be defined on $D = [0, 2] \times [0, 1]$ as

$$f(x,y) = \begin{cases} M\,, & 0 \le x \le 1 \\ m\,, & 1 < x \le 2 \end{cases}$$

where for definitiveness $0 < m < M$. The function is not continuous along the line $x = 1$ in D. Its graph is shown in the left panel of Fig. **28.2**. The volume below the graph $z = f(x, y)$ and above D is easy to find; it is the sum of volumes of two rectangular boxes with the bases $[0, 1] \times [0, 1]$ and $[1, 2] \times [0, 1]$ and the corresponding heights M and m. So, the volume is

$$V = MA_1 + mA_2$$

where A_1 and A_2 are the areas of the bases, $A_1 = A_2 = 1$. The double integral of f exists and also equals the volume V. Indeed, for a partition $x_j = j\Delta x$, $\Delta x = 2/N_1$, of the interval $[0, 2]$, there exists j' such that

$$x_{j'-1} \le 1 < x_{j'}\,.$$

The numbers M_{jk} and m_{jk} differs only for partition rectangles intersected by the line $x = 1$, that is, in the rectangles $R_{j'k}$:

$$M_{jk} = \begin{cases} M\,, & j < j' \\ M\,, & j = j' \\ m\,, & j > j' \end{cases} \qquad m_{jk} = \begin{cases} M\,, & j < j' \\ m\,, & j = j' \\ m\,, & j > j' \end{cases}$$

for all $k = 1, 2, ..., N_2$. The length of the curve along which f is discontinuous is $l = 1$. If $\Delta y = 1/N_2$, then $l = \sum_k \Delta y$. To find the upper sum, the summation over j is split into three terms, $j < j'$, $j = j'$, and $j > j'$:

$$\begin{aligned} U(f, N_1, N_2) &= \sum_{k=1}^{N_2} \left(\sum_{j=1}^{j'-1} M_{jk} + M_{j'k} + \sum_{j=j'+1}^{N_1} M_{jk} \right) \Delta x \Delta y \\ &= Mlx_{j'-1} + Ml\Delta x + ml(2 - x_{j'})\,, \\ L(f, N_1, N_2) &= Mlx_{j'-1} + ml\Delta x + ml(2 - x_{j'}) \end{aligned}$$

because M_{jk} and m_{jk} differs only for $j = j'$. Put $1 - x_{j'-1} = p\Delta x$ for some $0 \le p < 1$ so that $x_{j'} - 1 = (1 - p)\Delta x$. Then the areas of the bases of the rectangles on which the function has a constant values can be written in the form

$$\begin{aligned} A_1 &= l \cdot (1 - 0) = l(x_{j'-1} + p\Delta x) \\ A_2 &= l \cdot (2 - 1) = l[(2 - x_{j'}) + (1 - p)\Delta x] \end{aligned}$$

Using these relations, it is not difficult to express the upper and lower sums in terms of the volume V:

$$\begin{aligned} U(f, N_1, N_2) &= MA_1 + mA_2 + (M - m)(1 - p)l\Delta x \\ &= V + (M - m)(1 - p)l\Delta x\,, \\ L(f, N_1, N_2) &= MA_1 + mA_2 - (M - m)pl\Delta x \\ &= V - (M - m)pl\Delta x\,. \end{aligned}$$

Therefore the upper and lower sums converge to V as $\Delta x \to 0$ (or $N_1 \to \infty$) and the double integral of f over D exists and is equal to V. Note that $M-m$ is the value of the jump discontinuity of f across the line $x = 1$ and $l\Delta x$ is the total area of the partition rectangles intersecting the "discontinuity" curve.

Furthermore, if a bounded function g coincides with f for all $x \ne 1$, but $g(1, y) \ne f(1, y)$, then the function g is also integrable and its double integral is also equal to V. Note that only the coefficient $M - m$ at the term $l\Delta x$ has to be changed to obtained the upper and lower sums of g, but this term vanishes in the limit for any value of the coefficient. This observation resembles a similar property of the ordinary integral: The value of the integral does not change if the integrand is changed at a single point.

In general, if a bounded function f is not continuous on a smooth curve, then the contribution of partition rectangles intersecting the curve to the upper and lower sums tends to zero as $N_{1,2} \to \infty$. This can be shown by a similar line of arguments as in the above example and the following assertion holds.

COROLLARY **28.1**. *Let D be a closed, bounded region whose boundaries are piecewise smooth curves. If a function f is bounded on D and not continuous on a finite number of smooth curves, then it is integrable on D.*

Note that by this Corollary, a bounded function that is not continuous only on the boundary of D is integrable. So a continuous bounded function f on an open region D with piecewise smooth boundaries is integrable on the closure $\bar{D}$, and the value of the double integral does not depend on the values of f on the boundary of D. Similarly, the value of the double integral does not depend on the values of f on a smooth curve where f is not continuous in D.

28.6. Exercises.

1–6. For each of the following functions and the specified rectangular domain D, find the upper and lower sums, investigate their convergence, and find the double integral or show that the function is not integrable.

1. $f(x,y) = k = \text{const}, \quad D = [a,b] \times [c,d]$;

2. $f(x,y) = k_1 = \text{const}$ if $y > 0$ and $f(x,y) = k_2 = \text{const}$ if $y \le 0$, $D = [0,1] \times [-1,1]$;

3. $f(x,y) = xy^2, \quad D = [0,1] \times [0,1]$ Hint: $1 + 2^2 + \cdots + n^2 = \frac{1}{6}n(n+1)(2n+1)$;

4. $f(x,y) = 1 - x - y, \quad D = \{(x,y) | \, 0 \le y \le 1 - x \, , 0 \le x \le 1\}$;

5. $f(x,y) = 1$ if one of the variables is rational, and otherwise $f(x,y) = xy$, $D = [0,1] \times [0,1]$

6. $f(x,y) = x^2 + y^2, \quad D = [1,2] \times [1,3]$.

7-8. For each of the following functions use a Riemann sum with specified N_1 and N_2 and sample points at lower right corners to estimate the double integral over a given region D.

7. $f(x,y) = x + y^2, \quad (N_1, N_2) = (2,2)$, $D = [0,2] \times [0,4]$;

8. $f(x,y) = \sin(x+y), \quad (N_1, N_2) = (3,3)$, $D = [0,\pi] \times [0,\pi]$.

9. Approximate the integral of $f(x,y) = (24 + x^2 + y^2)^{-1/2}$ over the disk $x^2 + y^2 \le 25$ by a Riemann sum. Use a partition by squares whose vertices have integer-valued coordinates and sample points at vertices of the squares that are farthest from the origin. Assess the accuracy of the approximation by calculating the difference of the upper and lower sums.

10-18. Evaluate each of the following double integrals by identifying it as the volume of a solid, e.g., by sketching the graph of the integrand.

10. $\iint_D k dA$ if D is the disk $x^2 + y^2 \leq 1$ and k is a constant;
11. $\iint_D \sqrt{1 - x^2 - y^2} dA$ if D is the disk $x^2 + y^2 \leq 1$;
12. $\iint_D (1 - x - y) dA$ if D is the triangle with vertices $(0, 0)$, $(0, 1)$, and $(1, 0)$;
13. $\iint_D (c - \frac{c}{a} x - \frac{c}{b} y) dA$ if D is the triangle with vertices $(0, 0)$, $(0, b)$, and $(a, 0)$ where a, b, and c are positive numbers;
14. $\iint_D (k - x) dA$ if D is the rectangle $0 \leq x \leq k$ and $0 \leq y \leq a$;
15. $\iint_D (2 - \sqrt{x^2 + y^2}) dA$ if D is the part of the disk $x^2 + y^2 \leq 1$ in the first quadrant. Hint: the volume of a circular solid cone with the base being the disk of radius R and the height h is $\pi R^2 h/3$;
16. $\iint_D (2 - \sqrt{x^2 + y^2}) dA$ if D is the ring $1 \leq x^2 + y^2 \leq 2$;
17. $\iint_D (\sqrt{1 - x^2 - y^2} - \sqrt{x^2 + y^2} + 1) dA$ if D is the disk $x^2 + y^2 \leq 1$;
18. $\iint_D (\sqrt{1 - x^2 - y^2} + \sqrt{x^2 + y^2} - 1) dA$ if D is the disk $x^2 + y^2 \leq 1$.

29. Properties of the Double Integral

The properties of the double integral are similar to those of an ordinary integral and can be established directly from the definition using the basic limit laws.

Linearity. Let f and g be functions integrable on D and let c be a number. Then

$$\iint_D (f+g)\,dA = \iint_D f\,dA + \iint_D g\,dA\,,$$
$$\iint_D cf\,dA = c\iint_D f\,dA\,.$$

Area. The double integral

$$A(D) = \iint_D dA \tag{29.1}$$

is called the *area* of D (if it exists). If D is bounded by piecewise smooth curves, then it exists because the unit function $f = 1$ is continuous on D. By the geometrical interpretation of the double integral, the number $A(D)$ is the volume of the solid cylinder with the cross section D and the *unit* height ($f = 1$). Intuitively, the region D can always be covered by the union of adjacent rectangles of area $\Delta A = \Delta x\,\Delta y$. In the limit $(\Delta x, \Delta y) \to (0,0)$, the total area of these rectangles converges to the area of D. Let D be a region with a piecewise smooth boundary. By Corollary **28.1** and the remark following it, it is natural to define the area of D as the area of its closure $\bar{D}$ given by the integral (**29.1**). It will be shown that the value of (**29.1**) for a disk of radius a is πa^2. Furthermore, *a set D in a plane is said to have zero area if it is contained in the union of open disks which can be chosen so that their total area is less than any preassigned positive number.* For example, the piecewise smooth boundary of a bounded region has zero area. Indeed, let l be the arclength of the boundary curve (it exists as the curve is piecewise smooth). Suppose that the curve is partition into N pieces of length $\Delta s = l/N$. Then the curve is covered by the union of N disks of radius Δs (centered at the mid-points of each partition segment of the curve) so that their total area is $\pi N(\Delta s)^2 = \pi l^2/N \to 0$ as $N \to \infty$.

Additivity. Suppose that D is the union of D_1 and D_2 such that the area of their intersection is 0; that is, D_1 and D_2 may only have common points at their boundaries or no common points at all. If f is integrable on D, then

$$\iint_D f\,dA = \iint_{D_1} f\,dA + \iint_{D_2} f\,dA\,.$$

This property is difficult to prove directly from the definition. However, it appears rather natural when making the analogy of the double integral and the volume. If the region D is cut into two regions D_1 and D_2 by a piecewise

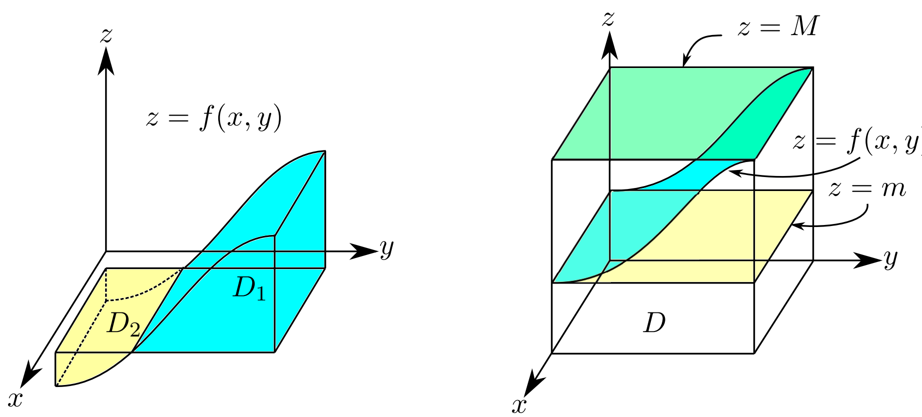


FIGURE **29.1**. **Left**: A function f is nonnegative on the region D_1 and nonpositive on D_2. The double integral of $f(x, y)$ over the union of regions D_1 and D_2 is the difference of the indicated volumes. The volume below the xy plane and above the graph of f contributes to the double integral with the negative sign. **Right**: An illustration to the upper and lower bounds of the double integral of a function f over a region D. If $A(D)$ is the area of D and $m \leq f(x, y) \leq M$ in D, then the volume under the graph of f is no less than the volume $mA(D)$ and no larger than $MA(D)$.

smooth curve, then the solid above D is also cut into two solids, one above D_1 and the other above D_2. Naturally, the volume is additive (see the right panel of Fig. **28.2** in the previous section).

Suppose that f is nonnegative on D_1 and nonpositive on D_2. The double integral over D_1 is the volume V_1 of the solid above D_1 and below the graph of f. Since $-f \geq 0$ on D_2, the double integral over D_2 is $-V_2$ where V_2 is the volume of the solid *below* D_2 and *above* the graph of f. When f becomes negative, its graph goes below the plane $z = 0$ (the xy plane). So, the double integral is the difference of the volumes above and below the xy plane:

$$\iint_D f dA = V_1 - V_2$$

Therefore it may vanish or take negative values, depending on which volume is larger. This property is analogous to the familiar relation between the ordinary integral and the area under the graph. It is illustrated in Fig. **29.1** (the left panel).

Positivity. If $f(\mathbf{r}) \geq 0$ for all $\mathbf{r}$ in D, then

$$\iint_D f\, dA \geq 0,$$

and, as a consequence of the linearity,

$$\iint_D f\,dA \geq \iint_D g\,dA$$

if $f(\mathbf{r}) \geq g(\mathbf{r})$ for all $\mathbf{r}$ in D.

Upper and Lower Bounds. Let $m = \inf_D f$ and $M = \sup_D f$. Then $m \leq f(\mathbf{r}) \leq M$ for all $\mathbf{r}$ in D. From the positivity property for the double integrals of $f(x,y) - m \geq 0$ and $M - f(x,y) \geq 0$ over D and Eq. (**29.1**), it follows that

$$mA(D) \leq \iint_D f\,dA \leq MA(D)\,.$$

This inequality is easy to visualize. If f is positive, then the double integral is the volume of the solid below the graph of f. The solid lies in the cylinder with the cross section D. The graph of f lies between the planes $z = m$ and $z = M$. Therefore, the volume of the cylinder of height m cannot exceed the volume of the solid, whereas the latter cannot exceed the volume of the cylinder of height M as shown in the right panel of Fig. **29.1**.

THEOREM **29.1**. (Integral Mean Value Theorem).
If f is continuous on a bounded closed region D, then there exists a point $\mathbf{r}_0$ in D such that

$$\iint_D f\,dA = f(\mathbf{r}_0)A(D)\,.$$

PROOF. Let h be a number. Put

$$g(h) = \iint_D (f-h)\,dA = \iint_D f\,dA - hA(D)\,.$$

From the upper and lower bounds for the double integral, it follows that $g(M) \leq 0$ and $g(m) \geq 0$. Since $g(h)$ is linear in h, there exists $h = h_0$ in $[m, M]$ such that $g(h_0) = 0$. On the other hand, a continuous function on a closed, bounded region D takes its maximal and minimal values as well as all the values between them (Theorem **26.3**). Therefore, for any $m \leq h_0 \leq M$, there is $\mathbf{r}_0$ in D such that $f(\mathbf{r}_0) = h_0$. $\square$

A geometrical interpretation of the integral mean value theorem is rather simple. Imagine that the solid below the graph of f is made of clay (see the left panel of Fig. **29.2**). The shape of a piece of clay may be deformed while the volume is preserved under deformation. The nonflat top of the solid can be deformed so that it becomes flat, turning the solid into a cylinder of height h_0, which, by volume preservation, should be between the smallest and the largest heights of the original solid. The integral mean value theorem merely states the existence of such an *average* height at which the volume of the cylinder coincides with the volume of the solid with a nonflat top. The continuity of the function is sufficient (but not necessary) to establish that there is a point at which the average height coincides with the value of the function.

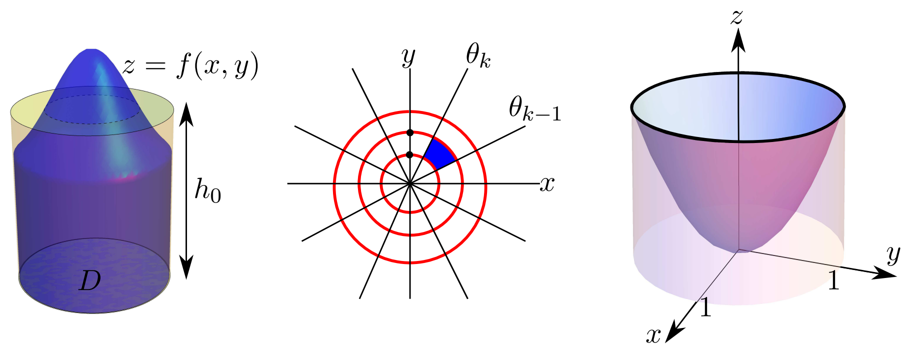


Figure **29.2**. **Left**: A clay solid with a non-flat top (the graph of a continuous function f) may be deformed to the solid of the same volume and with the same horizontal cross section D, but with a flat top $z = h_0$. The function f takes the value h_0 at some point of D. This illustrates the integral mean value theorem.
Middle: A partition of a disk by concentric circles of radii $r = r_p$ and rays $\theta = \theta_k$ as described in Example **29.1**. A partition element is the region $r_{p-1} \le r \le r_p$ and $\theta_{k-1} \le \theta_k$.
Right: The volume below the graph $z = x^2 + y^2$ and above the disk D, $x^2 + y^2 \le 1$. The corresponding double integral is evaluated in Example **29.1** by taking the limit of Riemann sums for the partition of D shown in the middle panel.

Definition **29.1**. (Average value of a function)
Let f be integrable on D and let $A(D)$ be the area of D. The average value *of f on D is*

$$\frac{1}{A(D)} \iint_D f dA$$

If f is continuous on D, then the integral mean value theorem asserts that f attains its average value at some point in D. The continuity hypothesis is crucial here. For example, the function depicted in the left panel of Fig. **28.2** is not continuous. Its average value is $(MA_1 + mA_2)/(A_1 + A_2)$ which generally coincides neither with M nor m.

Integrability of the Absolute Value. Suppose that f is integrable on a bounded, closed region D. Then its absolute value $|f|$ is also integrable and

$$\left|\iint_D f\, dA\right| \le \iint_D |f|\, dA.$$

If f is continuous on D, then $|f|$ is also continuous on D and, hence, integrable by Theorem **28.2**. For a generic integrable function f, a proof of the integrability of $|f|$ is rather technical and omitted. Once the integrability of

$|f|$ is established, the inequality is a simple consequence of $|a+b| \le |a| + |b|$ applied to a Riemann sum of f. Making the analogy between the double integral and the volume, suppose that $f \ge 0$ on D_1 and $f \le 0$ on D_2, where $D_{1,2}$ are two portions of D. If V_1 and V_2 stand for the volumes of the solids bounded by the graph of f and D_1 and D_2, respectively, then the double integral of f over D is $V_1 - V_2$, while the double integral of $|f|$ is $V_1 + V_2$. Naturally, $|V_1 - V_2| \le V_1 + V_2$ for positive $V_{1,2}$.

Independence of Partition. Suppose f is continuous and nonnegative on a closed bounded region D. Then the volume under the graph of f is given by the double integral of f over D. On the other hand, the volume can be computed by (**28.1**) in which the Riemann sum is defined for an *arbitrary* (nonrectangular) partition of D. It seems natural to require that the numerical value of the volume should not depend on the choice of partitions in the Riemann sums (**28.1**). This observation leads to a conjecture that the double integral, if it exists, may also be computed as the limit of Riemann sums with arbitrary partitions. The analysis is limited to the case when f is continuous.

A continuous function on a closed bounded region has the following remarkable property called the *uniform continuity.*

THEOREM **29.2**. (Uniform Continuity)
Suppose f is a continuous function on a bounded closed region D in a Euclidean space. Then for any number $\varepsilon > 0$ there exists a number $\delta > 0$ such that

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

for any $\mathbf{r}$ and $\mathbf{r}'$ in D.

The assertion can be understood as follows. Fix a point $\mathbf{r}'$ in D. By continuity of f, for any $\varepsilon > 0$, one can find a ball (or disk) of sufficiently small radius centered at the point $\mathbf{r}'$ such that the values of f in this ball deviates from $f(\mathbf{r}')$ no more than ε (recall $\lim_{\mathbf{r}\to\mathbf{r}'} f(\mathbf{r}) = f(\mathbf{r}')$ for a continuous f). Note that the radius δ depends on both the number ε and the point $\mathbf{r}'$, in general. The uniform continuity implies a stronger condition. Namely, the radius δ does not depends on the point. A ball of radius δ can be centered at *any* point in D and the values of the function at any two points in this ball differ by no more than ε. In other words, as soon as *the distance between any two points in D in less than δ, the difference between the values of the function becomes less than ε*. This is why this property is called the *uniform* continuity. Variations of values of f in any ball in D of a fixed radius are uniformly bounded.

A continuous function on a non-closed (or non-bounded) set may not have this property. For example, put $f(x, y) = 1/x$ which is continuous in the rectangle $D = (0, 1] \times [0, 1]$. Note D is not closed. Then in a disk whose center is sufficiently close to the line $x = 0$, the values of f can have variations as large as desired within this disk because $1/x$ diverges as x approaches zero. Take an interval (x_1, x_2) of a length $\delta = x_2 - x_1$. Then $1/x_1 - 1/x_2$ can be made arbitrary large by taking x_1 closer to 0 for any choice of $\delta > 0$. So the variations of f in any disk in D of some non-zero radius cannot be bounded by a fixed number ε. Similarly, take $f(x, y) = x^2$ which is continuous in the unbounded rectangle $D = [0, \infty) \times [0, 1]$. Then in a disk whose center is sufficiently far from the line $x = 0$, the values of f can have variations as large as desired within this disk. For an interval (x_1, x_2) of a length $\delta > 0$, the variation $x_2^2 - x_1^2 = \delta(x_2 + x_1)$ can be made as large as desired by taking x_2 large enough no matter how small δ is.

Let f be a continuous function on a closed bounded region D. Let D be partitioned by piecewise smooth curves into partition elements D_p, $p = 1, 2, ..., N$, so that the union of D_p is D and $A(D) = \sum_{p=1}^{N} \Delta A_p$, where ΔA_p is the area of D_p defined by Eq. (**29.1**). If R_p is the smallest radius of a disk that contains D_p, put $R_N^* = \max_p R_p$; that is, R_p characterizes the size of the partition element D_p and R_N^* is the size of the largest partition element. Suppose that $R_N^* \to 0$ as $N \to \infty$. Under the aforementioned conditions the following theorem holds.

THEOREM **29.3**. (Independence of the Partition)
For any choice of sample points $\mathbf{r}_p^*$ and any choice of partition elements D_p,

$$\iint_D f\, dA = \lim_{\substack{N\to\infty \\ (R_N^* \to 0)}} \sum_{p=1}^{N} f(\mathbf{r}_p^*)\, \Delta A_p \tag{29.2}$$

PROOF. As f is continuous on D, in each D_p there is a point $\mathbf{r}_p$ such that

$$\iint_D f\, dA = \sum_{p=1}^{N} \iint_{D_p} f\, dA = \sum_{p=1}^{N} f(\mathbf{r}_p)\, \Delta A_p\,.$$

The first equality follows from the additivity of the double integral, and the second one holds by the integral mean value theorem. Consider the Riemann sum

$$R(f, N) = \sum_{p=1}^{N} f(\mathbf{r}_p^*)\, \Delta A_p,$$

where $\mathbf{r}_p^*$ is a sample point in D_p. If $\mathbf{r}_p^* \neq \mathbf{r}_p$, then the Riemann sum does not coincide with the double integral. However, its limit as $N \to \infty$ equals the double integral. Indeed, put $c_p = |f(\mathbf{r}_p^*) - f(\mathbf{r}_p)|$ and $c_N = \max c_p$, $p = 1, 2, ..., N$. Fix a number $\varepsilon > 0$. By Theorem **29.2**, there is $\delta > 0$ such that variations of f in any disk of radius δ in D do not exceed ε. Since $R_N^* \to 0$ as $N \to \infty$, $R_N^* < \delta$ for all N larger than some N_0. Hence, $c_N < \varepsilon$ because any partition element D_p is contained in a disk of radius $R_p \leq R_N^* < \delta$, which implies that $c_N \to 0$ as $N \to \infty$. Therefore, the deviation of the Riemann sum from the double integral converges to 0:

$$\begin{aligned}\left|\iint_D f\,dA - R(f,N)\right| &= \left|\sum_{p=1}^{N}(f(\mathbf{r}_p) - f(\mathbf{r}_p^*))\,\Delta A_p\right| \\ &\le \sum_{p=1}^{N}|f(\mathbf{r}_p) - f(\mathbf{r}_p^*)|\Delta A_p \\ &= \sum_{p=1}^{N} c_p\,\Delta A_p \le c_N \sum_{p=1}^{N}\Delta A_p = c_N A(D) \to 0\end{aligned}$$

as $N \to \infty$. □

A practical significance of this theorem is that the double integral can be approximated by Riemann sums for *any convenient partition* of the integration region. Note that the region D is no longer required to be embedded in a rectangle and f does not have to be extended outside of D. This property is useful for evaluating double integrals by means of *change of variables* discussed later in this chapter. It is also useful to simplify calculations of Riemann sums.

EXAMPLE **29.1**. *Find the double integral of $f(x,y) = x^2 + y^2$ over the disk D, $x^2 + y^2 \le 1$, using the partition of D by concentric circles and rays from the origin.*

SOLUTION: Consider circles $x^2 + y^2 = r_p^2$, where

$$r_p = p\,\Delta r\,, \quad \Delta r = \frac{1}{N}\,, \quad p = 0, 1, 2, ..., N\,.$$

If θ is the polar angle in the plane, then points with a fixed value of θ form a ray from the origin. Let the disk D be partitioned by circles of radii r_p and rays

$$\theta = \theta_k = k\,\Delta\theta\,, \quad \Delta\theta = \frac{2\pi}{n}\,, \quad k = 1, 2, ..., n\,.$$

Each partition element lies in the sector of angle $\Delta\theta$ and is bounded by two circles whose radii differ by Δr (see the middle panel of Fig. **29.2**). The area of a sector of radius r_p is $r_p^2\,\Delta\theta/2$. Therefore, the area of a partition element between circles of radii r_p and r_{p-1} is

$$\Delta A_p = \frac{1}{2}r_p^2\,\Delta\theta - \frac{1}{2}r_{p-1}^2\,\Delta\theta = \frac{1}{2}(r_p^2 - r_{p-1}^2)\,\Delta\theta = \frac{1}{2}(r_p + r_{p-1})\,\Delta r\,\Delta\theta$$

where $p = 1, 2, ..., N$. In the Riemann sum, use the midpoint rule; that is, the sample points are intersections of the circles of radius $\bar{r}_p = (r_p + r_{p-1})/2$ and the rays with angles $\bar{\theta}_k = (\theta_{k+1} + \theta_k)/2$. The values of f at the sample points are $f(\mathbf{r}_p^*) = \bar{r}_p^2$, the area elements are $\Delta A_p = \bar{r}_p\,\Delta r\,\Delta\theta$, and the corresponding Riemann sum reads

$$R(f, N, n) = \sum_{k=1}^{n} \sum_{p=1}^{N} \bar{r}_p^3 \, \Delta r \, \Delta\theta = 2\pi \sum_{p=1}^{N} \bar{r}_p^3 \, \Delta r$$

because $\sum_{k=1}^{n} \Delta\theta = 2\pi$, the total range of θ in the disk D. The sum over p is the Riemann sum for the single-variable function $g(r) = r^3$ on the interval $0 \le r \le 1$. In the limit $N \to \infty$, this sum converges to the integral of g over the interval $[0, 1]$, that is,

$$\iint_D (x^2 + y^2)\, dA = 2\pi \lim_{N\to\infty} \sum_{p=1}^{N} \bar{r}_p^3 \, \Delta r = 2\pi \int_0^1 r^3 \, dr = \pi/2\,.$$

So, by choosing the partition according to the shape of D, the double Riemann sum has been reduced to a Riemann sum for a single-variable function. $\square$

The numerical value of the double integral in this example is the volume of the solid that lies below the paraboloid $z = x^2 + y^2$ and above the disk D of unit radius in the xy plane. It can also be represented as the volume of the cylinder with height $h = 1/2$, $V = hA(D) = \pi h = \pi/2$. This observation illustrates the integral mean value theorem. The function f takes the value $h = 1/2$ on the circle $x^2 + y^2 = 1/2$ of radius $1/\sqrt{2}$ in D.

29.1. Exercises.

1–5. Evaluate each of the following double integrals by using the properties of the double integral and its interpretation as the volume of a solid.

1. $\iint_D k dA$ where k is a constant and D is the square $[-2, 2] \times [-2, 2]$ with a circular hole of radius 1, i.e., $x^2 + y^2 \ge 1$ in D;

2. $\iint_D f dA$ where D is a disk $x^2 + y^2 \le 4$ and f is piecewise constant function: $f(x, y) = 2$ if $1 \le x^2 + y^2 \le 4$ and $f(x, y) = -3$ if $0 \le x^2 + y^2 < 1$;

3. $\iint_D (4 + 3\sqrt{x^2 + y^2}) dA$ where D is the disk $x^2 + y^2 \le 1$;

4. $\iint_D (\sqrt{4 - x^2 - y^2} - 2) dA$ if D is the part of the disk $x^2 + y^2 \le 4$ in the first quadrant;

5. $\iint_D (4 - x - y) dA$ where D is the triangle with vertices $(0, 0)$, $(1, 0)$, and $(0, 1)$. Hint: Use the identity $4 - x - y = 3 + (1 - x - y)$ and the linearity of the double integral.

6–7. Use the positivity of the double integral to prove the following inequalities.

6. $\iint_D \sin(xy)/(xy)\, dA \le A(D)$ where D is a bounded region in which $x > 0$ and $y > 0$;

7. $\iint_D (ax^2 + by^2) dA \le (a + b)\pi/2$ where D is the disk $x^2 + y^2 \le 1$. Hint: Put $r^2 = x^2 + y^2$. Then use $x^2 \le r^2$ and $y^2 \le r^2$ and apply the result of Example **29.1**.

8-11. Find the lower and upper bounds for each of the following integrals.

8. $\iint_D xy^3 dA$ where $D = [1,2] \times [1,2]$;

9. $\iint_D \sqrt{1 + xe^{-y}} dA$ where $D = [0,1] \times [0,1]$;

10. $\iint_D \sin(x+y) dA$ where D is the triangle with vertices $(0,0)$, $(0,\pi)$, and $(\pi/4, 0)$. Hint: Graph D and determine the set of point at which $\sin(x+y)$ attains its maximum value;

11. $\iint_D (100 + \cos^2 x + \cos^2 y)^{-1} dA$ where D is defined by $|x| + |y| \le 10$.

12. Let f be continuous on a bounded region D with a non-zero area. If the double integral of f over D vanishes, prove that there is a point in D at which f vanishes.

13. Use the method of Example **29.1** to find $\iint_D e^{x^2+y^2} dA$ where D is the part of the disk $x^2 + y^2 \le 1$.

14. Use a Riemann sum to approximate the double integral of $f(x,y) = \sqrt{x+y}$ over the triangle bounded by the lines $x = 0$, $y = 0$, and $x + y = 1$. Partition the integration region into four equal triangles by the lines $x =$ const, $y =$ const, and $x + y =$ const. Choose sample points to be centroids of the triangle.

15–17. Determine the sign of each of the following integrals.

15. $\iint_D \ln(x^2 + y^2) dA$ where D is defined by $|x| + |y| \le 1$;

16. $\iint_D \sqrt[3]{1 - x^2 - y^2} dA$ where D is defined by $x^2 + y^2 \le 4$;

17. $\iint_D \sin^{-1}(x+y) dA$ where D is defined by $0 \le x \le 1$ and $0 \le y \le 1 - x$.

30. Iterated Integrals

Here a practical method for evaluating double integrals will be developed. To simplify the technicalities, the derivation of the method is given for continuous functions. In combination with the properties of the double integral, it is sufficient for many applications.

Recall that if a multivariable limit exists, then the *repeated limits* exist and coincide. A similar statement is true for double sequences.

THEOREM **30.1**. *Suppose that a double sequence a_{nm} converges to a as $n, m \to \infty$. Then*

$$\lim_{n\to\infty}\left(\lim_{m\to\infty} a_{nm}\right) = \lim_{m\to\infty}\left(\lim_{n\to\infty} a_{nm}\right) = a$$

A proof of this simple theorem is left to the reader as an exercise. The limit $\lim_{m\to\infty} a_{nm} = b_n$ is taken for a fixed value n. Similarly, the limit $\lim_{n\to\infty} a_{nm} = c_m$ is taken for a fixed m. The theorem states that the limits of two generally *different* sequences b_n and c_m coincide and are equal to the limit of the double sequence. This property of double sequences will be applied to Riemann sums of an integrable function to reduce a double integral to ordinary *iterated* integrals.

30.1. Rectangular Domains. The simplest case of a rectangular domain is considered first. The double integral over general domains is studied in the next section. Let a function f be continuous on a rectangle $D = [a, b] \times [c, d]$. Let R_{jk} be a rectangular partition of D as defined earlier. For any choice of sample points (x_j^*, y_k^*), where $x_{j-1} \le x_j^* \le x_j$ and $y_{k-1} \le y_k^* \le y_k$, the Riemann sum $R(f, N_1, N_2)$ converges to the double integral of f over D by Theorem **28.1**. Since the limit of the double sequence $R(f, N_1, N_2)$ exists, it should not depend on the order in which the limits $N_1 \to \infty$ (or $\Delta x \to 0$) and $N_2 \to \infty$ (or $\Delta y \to 0$) are computed (Theorem **30.1**). Suppose the limit $N_2 \to \infty$ is to be evaluated first:

$$\begin{aligned}\iint_D f\,dA &= \lim_{N_{1,2}\to\infty} R(f, N_1, N_2) \\ &= \lim_{N_1\to\infty} \sum_{j=1}^{N_1} \left(\lim_{N_2\to\infty} \sum_{k=1}^{N_2} f(x_j^*, y_k^*)\,\Delta y \right) \Delta x\,.\end{aligned}$$

For each j, the expression in parentheses is nothing but the Riemann sum for the single-variable function $g_j(y) = f(x_j^*, y)$ on the interval $c \le y \le d$. So, if the functions $g_j(y)$ are integrable on $[c, d]$, then the limit of their Riemann sums is the integral of g_j over the interval. If f is continuous on D, then it must also be continuous along the lines $x = x_j^*$ in D; that is, $g_j(y) = f(x_j^*, y)$ is continuous for every j and hence integrable on $[c, d]$. Thus,

$$\lim_{N_2\to\infty} \sum_{k=1}^{N_2} f(x_j^*, y_k^*)\,\Delta y = \int_c^d f(x_j^*, y)\,dy\,. \tag{30.1}$$

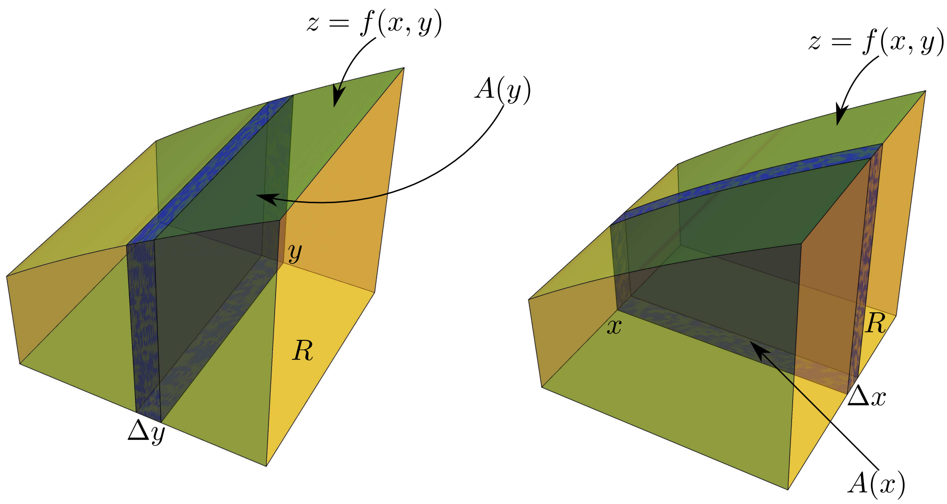


FIGURE **30.1**. An illustration to Fubini's theorem. The volume of a solid below the graph $z = f(x, y)$ and above a rectangle R is the sum of the volumes of the slices. **Left**: The slicing is done parallel to the x axis so that the volume of each slice is $\Delta y A(y)$ where $A(y)$ is the area of the cross section by a plane with a fixed value of y. **Right**: The slicing is done parallel to the y axis so that the volume of each slice is $\Delta x A(x)$ where $A(x)$ is the area of the cross section by a plane with a fixed value of x as given in (30.**2**).

Define a function $A(x)$ by

$$A(x) = \int_c^d f(x, y)\, dy\,. \tag{30.2}$$

The value of A at x is given by the integral of f with respect to y; the integration with respect to y is carried out as if x were a fixed number. For example, put $f(x, y) = x^2y + e^{xy}$ and $[c, d] = [0, 1]$. Then an antiderivative $F(x, y)$ of $f(x, y)$ with respect to y is $F(x, y) = x^2y^2/2 + e^{xy}/x$, which means that $F'_y(x, y) = f(x, y)$. Therefore,

$$A(x) = \int_0^1 (x^2y + e^{xy})\, dy = \frac{1}{2}x^2y^2 + \frac{e^{xy}}{x}\Big|_0^1 = \frac{1}{2}x^2 + \frac{e^x}{x} - \frac{1}{x}\,.$$

A geometrical interpretation of $A(x)$ is simple. If $f \geq 0$, then $A(x_j^*)$ is the area of the cross section of the solid below the graph $z = f(x, y)$ by the plane $x = x_j^*$. If variations of $A(x)$ within the interval $[x_{j-1}, x_j]$ are small (or Δx is small enough), then $A(x_j^*)\, \Delta x$ is the volume of the slice of the solid of width Δx (see the right panel of Fig. **30.1**).

One can prove that the function A is continuous on $[a,b]$ (see Study Problem **30.1**). The second sum in the Riemann sum for the double integral is the Riemann sum of $A(x)$ on the interval $[a,b]$ which converges to the integral of A over $[a,b]$:

$$\begin{aligned}\iint_D f\,dA &= \lim_{N_1\to\infty}\sum_{j=1}^{N_1} A(x_j^*)\,\Delta x = \int_a^b A(x)\,dx\\ &= \int_a^b\left(\int_c^d f(x,y)\,dy\right)dx\,.\end{aligned}$$

The integral on the right side of this equality is called the *iterated integral*. In what follows, the parentheses in the iterated integral will be omitted. The order in which the integrals are evaluated is specified by the order of the differentials in it; for example, $dy\,dx$ means that the integration with respect to y is to be carried out first.

In a similar fashion, by computing the limit $N_1\to\infty$ first, the double integral can be expressed as an iterated integral in which the integration is carried out with respect to x and then with respect to y. So the following result has been established.

THEOREM **30.2**. (Fubini's Theorem).
If f is continuous on the rectangle $D=[a,b]\times[c,d]$, then

$$\iint_D f(x,y)\,dA = \int_c^d\int_a^b f(x,y)\,dx\,dy = \int_a^b\int_c^d f(x,y)\,dy\,dx.$$

Think of a loaf of bread with a rectangular base and with a top having the shape of the graph $z=f(x,y)\geq 0$. It can be sliced along either of the two directions parallel to adjacent sides of its base. Fubini's theorem implies that the volume of the loaf is the sum of the volumes of the slices and is independent of how the slicing is done (see Fig. **30.1**).

EXAMPLE **30.1**. *Verify Fubini's theorem for the double integral*

$$\iint_D \frac{dA}{\sqrt{2x-y}}\,,\qquad D=[1,2]\times[0,1]\,.$$

SOLUTION: The integrand is the composition of two continuous functions $u^{-1/2}$ and $u=2x-y$. Therefore it continuous for all (x,y) for which $u>0$. In the rectangle $[2,1]\times[0,1]$, $u\geq 1$. So, the function $(2x-y)^{-1/2}$ is continuous on D and, hence, integrable on D, and Fubini's theorem applies to evaluate the double integral. Let us first integrate with respect to x. Since

$$\frac{\partial}{\partial x}\sqrt{2x-y} = \frac{1}{\sqrt{2x-y}}$$

the integration with respect to x yields:

$$\begin{aligned}\int_0^1\int_1^2 \frac{dx}{\sqrt{2x-y}}\,dy &= \int_0^1 \sqrt{2x-y}\,\Big|_1^2\,dy = \int_0^1\left(\sqrt{4-y}-\sqrt{2-y}\right)\,dy\\ &= \frac{2}{3}\left[(2-y)^{3/2}-(4-y)^{3/2}\right]\Big|_0^1\\ &= \frac{2}{3}\left(9-3\sqrt{3}-2\sqrt{2}\right)\end{aligned}$$

Let us integrate with respect to y first. Since

$$\frac{\partial}{\partial y}\left(-2\sqrt{2x-y}\right) = \frac{1}{\sqrt{2x-y}}\,,$$

the integration with respect to y yields:

$$\begin{aligned}\int_1^2\int_0^1 \frac{dy}{\sqrt{2x-y}}\,dx &= -2\int_1^2 \sqrt{2x-y}\,\Big|_0^1 dx = 2\int_1^2\left(\sqrt{2x}-\sqrt{2x-1}\right)\,dx\\ &= \frac{2}{3}\left[(2x)^{3/2}-(2x-1)^{3/2}\right]\Big|_1^2\\ &= \frac{2}{3}\left(9-3\sqrt{3}-2\sqrt{2}\right)\,.\end{aligned}$$

The values of the iterated integrals coincide in accord with Fubini's theorem. □

EXAMPLE **30.2**. *Find the volume of the solid bounded from above by the portion of the paraboloid $z = 4 - x^2 - 2y^2$ and from below by the portion of the paraboloid $z = -4 + x^2 + 2y^2$, where (x, y) in $[0, 1] \times [0, 1]$.*

SOLUTION: Let $h(x, y)$ be the height of the solid at a point (x, y). Given (x, y), the point (x, y, z) lies in the solid if $z_{\text{bot}}(x, y) \le z \le z_{\text{top}}(x, y)$ where the graphs $z = z_{\text{top}}(x, y)$ and $z = z_{\text{bot}}(x, y)$ are the top and bottom boundaries of the solid. Then the height of the solid at any (x, y) in D is

$$h(x,y) = z_{\text{top}}(x,y) - z_{\text{bot}}(x,y)\,.$$

Therefore the volume is

$$\begin{aligned}V &= \iint_D h(x,y)\,dA = \iint_D [z_{\text{top}}(x,y) - z_{\text{bot}}(x,y)]\,dA\\ &= \iint_D (8-2x^2-4y^2)\,dA = \int_0^1\int_0^1 (8-2x^2-4y^2)\,dy\,dx\\ &= \int_0^1 [(8-2x^2)y - 4y^3/3]\Big|_0^1 dx = \int_0^1 (8-2x^2-4/3)\,dx = 6.\end{aligned}$$

□

If a function of two variables happens to be the product of two functions of a single variable, then Fubini's theorem allows one to convert the double integral into the product of ordinary integrals.

COROLLARY **30.1**. (Factorization of Iterated Integrals).
Let D be a rectangle $[a,b]\times[c,d]$. Suppose $f(x,y)=g(x)h(y)$, where the functions g and h are integrable on $[a,b]$ and $[c,d]$, respectively. Then

$$\iint_D f(x,y)\,dA=\int_a^b g(x)\,dx\int_c^d h(y)\,dy\,.$$

Indeed, integrating first with respect to x by Fubini's theorem

$$\begin{aligned}\iint_D f\,dA &= \int_a^b\int_c^d g(x)h(y)\,dxdy=\int_c^d h(y)\left(\int_a^b g(x)\,dx\right)\,dy\\ &=\int_a^b g(x)\,dx\int_c^d h(y)\,dy\end{aligned}$$

where the constant equal to the integral of g over $[a,b]$ has been pulled from the integral with respect to y using the basic properties (linearity) of the integral. This simple consequence of Fubini's theorem is quite useful.

EXAMPLE **30.3**. *Evaluate the double integral of $f(x,y)=\sin(x+y)$ over the rectangle $[0,\pi]\times[-\pi/2,\pi/2]$.*

SOLUTION: One has $\sin(x+y)=\sin x\cos y+\cos x\sin y$. The integral of $\sin y$ over $[-\pi/2,\pi/2]$ vanishes by symmetry. So, by the factorization property of the iterated integral, only the first term contributes to the double integral:

$$\iint_D \sin(x+y)\,dA=\int_0^\pi \sin x\,dx\int_{-\pi/2}^{\pi/2}\cos y\,dy=2\cdot 2=4\,.$$

$\square$

The following example illustrates the use of the additivity of double integrals.

EXAMPLE **30.4**. *Evaluate the double integral of $f(x,y)=15x^4y^2$ over the region D, which is the rectangle $[-2,2]\times[-2,2]$ with the rectangular hole $[-1,1]\times[-1,1]$.*

SOLUTION: Let $D_1=[-2,2]\times[-2,2]$ and let $D_2=[-1,1]\times[-1,1]$. The rectangle D_1 is the union of D and D_2 such that their intersection has no area. Hence,

$$\begin{aligned}\iint_{D_1} f\,dA &= \iint_D f\,dA+\iint_{D_2} f\,dA \quad\Rightarrow\\ \iint_D f\,dA &= \iint_{D_1} f\,dA-\iint_{D_2} f\,dA.\end{aligned}$$

By evaluating the double integrals over $D_{1,2}$,

$$\begin{aligned}\iint_{D_1} 15x^4y^2\,dA &= 15\int_{-2}^2 x^4\,dx\int_{-2}^2 y^2\,dy=2^{10},\\ \iint_{D_2} 15x^4y^2\,dA &= 15\int_{-1}^1 x^4\,dx\int_{-1}^1 y^2\,dy=4.\end{aligned}$$

the double integral over D is obtained, $1024 - 4 = 1020$. □

30.2. Study Problems.

Problem **30.1**. *Prove that the function $A(x)$ defined in (**30.2**) is continuous in $[a, b]$.*

Solution: The function f is continuous on a closed rectangle and, hence, *uniformly continuous* by Theorem **29.2**. This implies that for any number $\varepsilon > 0$ one can find a number $\delta > 0$ such that

$$|f(x,y) - f(x',y')| < \frac{\varepsilon}{d-c} \quad \text{whenever} \quad \sqrt{(x-x')^2 + (y-y')^2} < \delta$$

for any choice of such pairs (x, y) and (x', y'). In particular, set $y = y'$ so that

$$|f(x,y) - f(x',y)| < \frac{\varepsilon}{d-c} \quad \text{whenever} \quad |x - x'| < \delta\,.$$

Then the following chain of inequalities holds

$$\begin{aligned} |A(x) - A(x')| &= \Big|\int_c^d \Big(f(x,y) - f(x',y)\Big)dy\Big| \leq \int_c^d \Big|f(x,y) - f(x',y)\Big|dy \\ &< \frac{\varepsilon}{d-c}\int_c^d dy = \varepsilon \quad \text{whenever} \quad |x - x'| < \delta\,. \end{aligned}$$

By the definition of a limit, the latter implies that

$$\lim_{x \to x'} A(x) = A(x') \quad \text{for any} \quad x' \in [a, b]\,,$$

and, hence, $A(x)$ is continuous in $[a, b]$. □

Problem **30.2**. *Suppose a function f has continuous second partial derivatives on the rectangle $D = [0, 1] \times [0, 1]$. Find $\iint_D f''_{xy}\, dA$ if $f(0, 0) = 1$, $f(0, 1) = 2$, $f(1, 0) = 3$, and $f(1, 1) = 5$.*

Solution: By Fubini's theorem,

$$\begin{aligned} \iint_D f''_{xy}\, dA &= \int_0^1 \int_0^1 \frac{\partial}{\partial x} f'_y(x,y)\, dx\, dy = \int_0^1 f'_y(x,y)\Big|_0^1 dy \\ &= \int_0^1 [f'_y(1,y) - f'_y(0,y)]\, dy = \int_0^1 \frac{d}{dy}[f(1,y) - f(0,y)]\, dy \\ &= [f(1,y) - f(0,y)]\Big|_0^1 \\ &= [f(1,1) - f(0,1)] - [f(1,0) - f(0,0)] = 1. \end{aligned}$$

By Clairaut's theorem $f''_{xy} = f''_{yx}$ and the value of the integral is independent of the order of integration. □

30.3. Exercises.

1–15. Evaluate each of the following double integrals over the specified rectangular region.

1. $\iint_D (x+y)dA, \quad D=[0,1]\times[0,2]\,;$

2. $\iint_D xy^2\,dA, \quad D=[0,1]\times[-1,1]\,;$

3. $\iint_D \sqrt{x+2y}\,dA, \quad D=[1,2]\times[0,1]\,;$

4. $\iint_D (1+3x^2y)dA, \quad D=[0,1]\times[0,2]\,;$

5. $\iint_D xe^{yx}dA, \quad D=[0,1]\times[0,1]\,;$

6. $\iint_D \cos(x+2y)dA, \quad D=[0,\pi]\times[0,\pi/4]\,;$

7. $\iint_D \frac{1+2x}{1+y^2}dA, \quad D=[0,1]\times[0,1]\,;$

8. $\iint_D \frac{y}{x^2+y^2}dA, \quad D=[0,1]\times[1,2]\,;$

9. $\iint_D (x-y)^n dA, \quad D=[0,1]\times[0,1]$, where n is a positive integer;

10. $\iint_D e^x\sqrt{y+e^x}dA, \quad D=[0,1]\times[0,2]\,;$

11. $\iint_D \sin^2(x)\sin^2(y)dA, \quad D=[0,\pi]\times[0,\pi]\,;$

12. $\iint_D \ln(x+y)dA,\ D=[1,2]\times[1,2]\,;$

13. $\iint_D \frac{1}{2x+y}\,dA, \quad D[0,1]\times[1,2]\,;$

14. $\iint_D x2^{x-y}\,dA, \quad D=[0,1]\times[0,1]\,;$

15. $\iint_D \frac{1-(xy)^3}{1-xy}\,dA, \quad D=[0,\frac{1}{2}]\times[0,\frac{1}{2}]$.

16–18. Find the volume of each of the following solids E.

16. E lies under the paraboloid $z=1+3x^2+6y^2$ and above the rectangle $[-1,1]\times[0,2]\,;$

17. E lies in the first octant and is bounded by the cylinder $z=4-y^2$ and the plane $x=3\,;$

18. E lies in the first octant and is bounded by the planes $x+y-z=0$, $y=2$, and $x=1$.

19. Evaluate $\iint_D xydA$ where D is the part of the square $[-1,1]\times[-1,1]$ that does not lie in the first quadrant.

20. Let f be continuous on $[a,b]\times[c,d]$ and $g(u,v)=\iint_{D_{uv}} f(x,y)dA$ where $D_{uv}=[a,u]\times[c,v]$ for $a<u<b$ and $c<v<d$. Show that $g''_{uv}=g''_{vu}=f(u,v)$.

21. Let f be a continuous function on $[a,b]$. Prove that

$$\left(\int_a^b f(x)dx\right)^2 \le (b-a)\int_a^b \Big(f(x)\Big)^2 dx$$

where the equality is reached only if $f(x)=\text{const}$. Hint: Consider the iterated integral

$$\int_a^b\int_a^b \Big[f(x)-f(y)\Big]^2 dy\,dx$$

22. Find the average value of the squared distance from the origin to a point of the disk $(x-a)^2+(y-b)^2\le R^2$. Hint: Set up the coordinate system at the center of the disk, consider the average value of the squared distance from the point (a, b) to a point of the disk, and use the method of Example **29.1** to convert the double integral to an iterated integral.

31. Double Integrals Over General Regions

The concept of the iterated integral can be extended to general regions subject to the following conditions.

31.1. Simple Regions.

Definition **31.1**. (Simple and Convex Regions). *A region D is said to be* simple *in the direction* $\mathbf{u}$ *if any line parallel to the vector* $\mathbf{u}$ *intersects D along at most one straight line segment. A region D is called* convex *(or simple) if it is simple in any direction.*

This definition is illustrated in Fig. **31.1**. Suppose D is bounded and simple in the direction of the y axis. It will be referred to as *y simple* or *vertically simple*. Since D is bounded, there is an interval $[a, b]$ such that vertical lines $x = x_0$ intersect D if x_0 in $[a, b]$. In other words, the region D lies within the vertical strip $a \le x \le b$ where a and b are the minimal and maximal values of the x coordinate for all points (x, y) in D.

Take a vertical line $x = x_0$ and consider all points of D that also belong to the line, that is, pairs (x_0, y) in D, where the first coordinate is fixed. Since the line intersects D along a segment, the variable y ranges over an interval. The endpoints of this interval depend on the line or the value of x_0; that is, for every x_0 in $[a, b]$, $y_{\rm bot} \le y \le y_{\rm top}$, where the numbers $y_{\rm bot}$ and $y_{\rm top}$ depend on x_0. So, all vertically simple regions admit the following algebraic description.

Algebraic Description of Vertically Simple Regions. If D is bounded and vertically simple, then it lies in the vertical strip $a \le x \le b$ and is bounded from below by the graph $y = y_{\rm bot}(x)$ and from above by the graph $y = y_{\rm top}(x)$:

$$D = \{(x, y) \,|\, y_{\rm bot}(x) \le y \le y_{\rm top}(x)\,, \quad a \le x \le b\}. \tag{31.1}$$

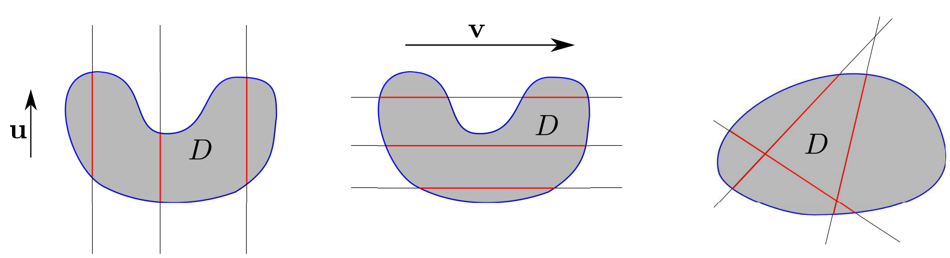


Figure **31.1**. **Left**: A region D is simple in the direction $\mathbf{u}$. **Middle**: A region D is not simple in the direction $\mathbf{v}$. **Right**: A region D is simple or a convex. Any straight line intersects it along at most one segment, or a straight line segment connecting any two points of D lies in D.

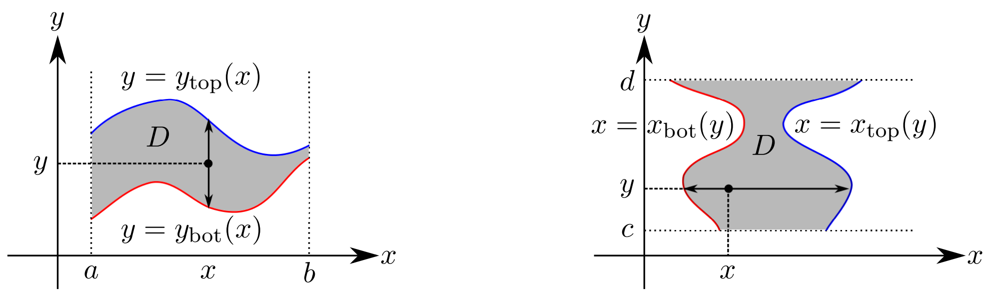


FIGURE **31.2**. **Left**: An algebraic description of a vertically simple region as given in Eq. (**31.1**): for every $a \leq x \leq b$, the y coordinate ranges over the interval $y_{\rm bot}(x) \leq y \leq y_{\rm top}(x)$. **Right**: An algebraic description of a horizontally simple region D as given in Eq. (**31.2**): for every $c \leq y \leq d$, the x coordinate ranges over the interval $x_{\rm bot}(y) \leq x \leq x_{\rm top}(y)$.

The numbers a and b are, respectively, the smallest and the largest values of the x coordinate of points of D. The graphs $y = y_{\rm bot}(x)$ and $y = y_{\rm top}(x)$ are *top* and *bottom* boundaries of D, respectively, relative to the direction of the y axis. The algebraic description of a vertically simple region is illustrated in the left panel of Figure **31.2**.

EXAMPLE **31.1**. *Give an algebraic description of the half-disk* $x^2+y^2 \leq 1$, $y \geq 0$, *as a vertically simple region.*

SOLUTION: The x coordinate of any point in the disk lies in the interval $[a, b] = [-1, 1]$ (see Fig. **31.3** (left panel)). Take a vertical line corresponding to a fixed value of x in this interval. This line intersects the half-disk along the segment whose one endpoint lies on the x axis; that is, $y = 0 = y_{\rm bot}(x)$. The other endpoint lies on the circle. Solving the equation of the circle for y, one finds $y = \pm\sqrt{1-x^2}$. Since $y \geq 0$ in the half-disk, the positive solution has to be taken, $y = \sqrt{1-x^2} = y_{\rm top}(x)$. So the region is bounded by two graphs $y = 0$ and $y = \sqrt{1-x^2}$:

$$D = \{(x, y)\,|\, 0 \leq y \leq \sqrt{1-x^2}\,,\ -1 \leq x \leq 1\}\,.$$

□

Suppose D is simple in the direction of the x axis. It will be referred to as x *simple* or *horizontally simple*. Since D is bounded, there is an interval $[c, d]$ such that horizontal lines $y = y_0$ intersect D if y_0 in $[c, d]$. In other words, the region D lies within the horizontal strip $c \leq y \leq d$. Take a horizontal line $y = y_0$ and consider all points of D that also belong to the line, that is, pairs (x, y_0) in D, where the second coordinate is fixed. Since the line intersects D along a segment, the variable x ranges over an interval. The endpoints of this interval depend on the line or the value of y_0; that is, for every y_0 in $[c, d]$, $x_{\rm bot} \leq x \leq x_{\rm top}$, where the numbers $x_{\rm bot}$ and

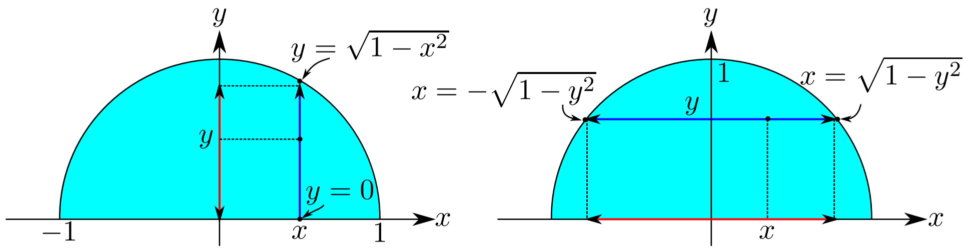


FIGURE **31.3**. The half-disk D, $x^2 + y^2 \leq 1$, $y \geq 0$, is a simple region. **Left**: An algebraic description of D as a vertically simple region as given in (**31.1**). The maximal range of x in D is $[-1, 1]$. For every such x, the y coordinate in D has the range $0 \leq y \leq \sqrt{1-x^2}$.
Right: An algebraic description of D as a horizontally simple region as given in (**31.2**). The maximal range of y in D is $[0, 1]$. For every such y, the x coordinate in D has the range $-\sqrt{1-y^2} \leq x \leq \sqrt{1-y^2}$.

x_{top} depend on y_0. So, all horizontally simple regions admit the following algebraic description.

Algebraic Description of Horizontally Simple Regions. If D is bounded and horizontally simple, then it lies in a horizontal strip $c \leq y \leq d$, where c and d are the minimal and maximal values of the y coordinates for all points of D. Furthermore, the region D is bounded from below by the graph $x = x_{\text{bot}}(y)$ and from above by the graph $x = x_{\text{top}}(y)$:

$$D = \{(x, y) \,|\, x_{\text{bot}}(y) \leq x \leq x_{\text{top}}(y)\,, \quad c \leq y \leq d\}. \tag{31.2}$$

The terms "*top*" and "*bottom*" boundaries are now defined relative to the line of sight in the direction of the x axis. The algebraic description of a horizontally simple region is illustrated in the right panel of Figure **31.2**.

EXAMPLE **31.2**. *Give an algebraic description of the half-disk* $x^2+y^2 \leq 1$, $y \geq 0$, *as a horizontally simple region.*

SOLUTION: The y coordinate of any point in the disk lies in the interval $[c, d] = [0, 1]$. Take a horizontal line corresponding to a fixed value of y from this interval. The line intersects the half-disk along a segment whose endpoints lie on the circle. Solving the equation of the circle for x, the x coordinates of the endpoints are obtained: $x = \pm\sqrt{1-y^2}$. So,

$$D = \{(x, y) \,|\, -\sqrt{1-y^2} \leq x \leq \sqrt{1-y^2}\,,\ 0 \leq y \leq 1\}\,.$$

When viewed in the horizontal direction, the top boundary of the region is the graph $x = \sqrt{1-y^2} = x_{\text{top}}(y)$ and the bottom boundary is the graph $x = -\sqrt{1-y^2} = x_{\text{bot}}(y)$ (see Fig. **31.3** (right panel)). $\square$

31.2. Iterated Integrals for Simple Regions. Suppose D is vertically simple and bounded by piecewise smooth curves. Then it should have an algebraic description according to (**31**.**1**) where the graphs $y = y_{\rm bot}(x)$ and $y = y_{\rm top}(x)$ are smooth curves (or a finite collection of them). Note that the functions $y_{\rm bot}$ and $y_{\rm top}$ are not necessarily continuous. For example, if D is the union of the rectangles $[0,1]\times[0,1]$ and $[1,2]\times[0,2]$, then D is vertically simple and $y_{\rm top}(x) = 2$ if $1 \le x \le 2$, while $y_{\rm top}(x) = 1$ if $0 \le x < 1$ so that the graph $y = y_{\rm bot}(x)$ consists of two smooth curves.

Let $R_D = [a,b]\times[c,d]$ be an embedding rectangle for the region D, where $c \le y_{\rm bot}(x) \le y_{\rm top}(x) \le d$ for all x in $[a,b]$. Suppose that a function f is continuous on D and defined by zero values outside D:

$$f(x,y) = 0 \quad \text{if} \quad c \le y < y_{\rm bot}(x) \quad \text{or} \quad y_{\rm top}(x) < y \le d\,, \quad a \le x \le b\,.$$

Consider a Riemann sum for a rectangular partition of R_D with sample points (x_j^*, y_k^*) just like in Section **30.1**. Since f is integrable, the double integral exists, and the double limit of the Riemann sum should not depend on the order in which the limits $N_1 \to \infty$ and $N_2 \to \infty$ are taken (Theorem **30**.**1**). For a vertically simple D, the limit $N_2 \to \infty$ (or $\Delta y \to 0$) is taken first. Similarly to Eq. (**30**.**1**), one infers that

$$\lim_{N_2\to\infty} \sum_{k=1}^{N_2} f(x_j^*, y_k^*)\,\Delta y = \int_c^d f(x_j^*, y)\,dy = \int_{y_{\rm bot}(x_j^*)}^{y_{\rm top}(x_j^*)} f(x_j^*, y)\,dy$$

because the function f vanishes outside the interval $y_{\rm bot}(x) \le y \le y_{\rm top}(x)$ for any x in $[a,b]$.

Suppose that $f(x,y) \ge 0$ and consider the solid bounded from above by the graph $z = f(x,y)$ and from below by the region D. The area of the cross section of the solid by the coordinate plane corresponding to a fixed value of x is given by Eq. (**30**.**2**):

$$A(x) = \int_c^d f(x,y)\,dy = \int_{y_{\rm bot}(x)}^{y_{\rm top}(x)} f(x,y)\,dy\,.$$

So just like in the case of rectangular domains, the above limit equals $A(x_j^*)$. That the area of the cross section is given by an integral over a *single* interval is only possible for a vertically simple base D of the solid. If D were not vertically simple, then such a slice would not have been a single slice but rather a few disjoint slices, depending on how many disjoint intervals are in the intersection of a vertical line with D. In this case, the integration with respect to y would have yielded a sum of integrals over all such intervals. The reason the integration with respect to y is to be carried out first only for vertically simple regions is exactly to avoid the necessity to integrate over a union of disjoint intervals.

Finally, the value of the double integral is given by the integral of $A(x)$ over the interval $[a,b]$. Recall that the volume of a slice of width dx and the cross section area $A(x)$ is $dV = A(x)dx$ so that the total volume of the solid

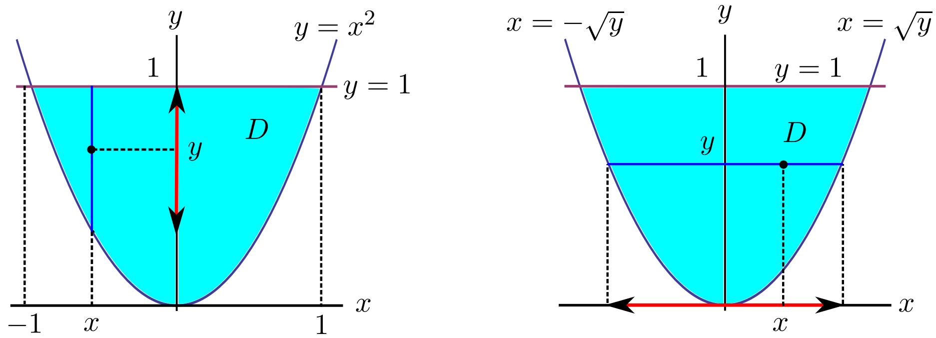


FIGURE **31.4**. Illustration to Example **31.3**. **Left**: The integration region as a vertically simple region: $-1 \le x \le 1$ and, for every such x, $x^2 \le y \le 1$. **Right**: The integration region as a horizontally simple region: $0 \le y \le 1$ and, for every such y, $-\sqrt{y} \le x \le \sqrt{y}$.

is given by the integral $V = \int_a^b A(x)dx$ (as the sum of volumes of all slices in the solid).

Iterated Integral for Vertically Simple regions. Let D be a vertically simple region; that is, it admits the algebraic description (**31.1**). The double integral of f over D is then given by the iterated integral

$$(\mathbf{31.3}) \qquad \iint_D f(x,y)\,dA = \int_a^b \int_{y_{\rm bot}(x)}^{y_{\rm top}(x)} f(x,y)\,dy\,dx.$$

Iterated Integral for Horizontally Simple Regions. Naturally, for horizontally simple regions, the integration with respect to x should be carried out first. Therefore the limit $N_1 \to \infty$ (or $\Delta x \to 0$) should be taken first in the Riemann sum. The technicalities are similar to the case of vertically simple regions. Let D be a horizontally simple region; that is, it admits the algebraic description (**31.2**). The double integral of f over D is then given by the iterated integral

$$(\mathbf{31.4}) \qquad \iint_D f(x,y)\,dA = \int_c^d \int_{x_{\rm bot}(y)}^{x_{\rm top}(y)} f(x,y)\,dx\,dy.$$

EXAMPLE **31.3**. *Evaluate the double integral of $f(x,y) = 6yx^2$ over the region D bounded by the line $y = 1$ and the parabola $y = x^2$.*

SOLUTION: The region D is both horizontally and vertically simple. It is therefore possible to use either (**31.3**) or (**31.4**). To find an algebraic description of D as a vertically simple region, one has to first specify the maximal range of the x coordinate in D. It is determined by the intersection of the line $y = 1$ and the parabola $y = x^2$, that is, $1 = x^2$, and hence x in $[a,b] = [-1,1]$ for all points of D (see the left panel of Figure **31.4**) For any x in $[-1,1]$, the y coordinate of points of D attains the smallest value on the parabola (i.e., $y_{\rm bot}(x) = x^2$), and the largest value on the line (i.e., $y_{\rm top}(x) = 1$). One has

$$D = \{(x,y)\,|\,x^2 \le y \le 1\,,\ -1 \le x \le 1\}$$
$$\iint_D 6yx^2\,dA = 6\int_{-1}^{1} x^2 \int_{x^2}^{1} y\,dy\,dx = 3\int_{-1}^{1} x^2(1-x^4)\,dx = \frac{8}{7}\,.$$

It is also instructive to obtain this result using the reverse order of integration. To find an algebraic description of D as a horizontally simple region, one has to first specify the maximal range of the y coordinate in D. The smallest value of y is 0 and the largest value is 1; that is, y in $[c,d]=[0,1]$ for all points of D. For any fixed y in $[0,1]$, the x coordinate of points of D attains the smallest and largest values on the parabola $y = x^2$ or $x = \pm\sqrt{y}$, that is, $x_{\rm bot}(y) = -\sqrt{y}$ and $x_{\rm top}(y) = \sqrt{y}$ (see the right panel of Figure **31.4**). One has

$$D = \{(x,y)\,|\,-\sqrt{y} \le x \le \sqrt{y}\,,\ 0 \le y \le 1\}$$
$$\begin{aligned}\iint_D 6yx^2\,dA &= 6\int_0^1 y\int_{-\sqrt{y}}^{\sqrt{y}} x^2\,dx\,dy = 2\int_0^1 y(2y^{3/2})\,dy\\ &= 4\int_0^1 y^{5/2}\,dy = \frac{8}{7}\,.\end{aligned}$$

□

Iterated Integrals for Nonsimple Regions. If the integration region D is not simple, how can one evaluate the double integral? A region bounded by a piecewise smooth curve can be cut by suitable smooth curves into simple regions D_p, $p = 1, 2, ..., n$. The double integral over simple regions can then be evaluated. The double integral over D is then the sum of the double integrals over D_p by the additivity property. When evaluating a double integral, it is sometimes technically convenient to cut the integration region into two or more pieces even if the region is simple (see Example **31.5**).

Integrals of non-continuous functions. If a function f is not continuous in D on a smooth curve C, then by Corollary **28.1** f is integrable on D. How to evaluate the double integral over D? One can show that the representation of the double integral over D by an iterated integral also holds in this case. The latter can be established by cutting the region D by a smooth curve that contains the curve on which f is not continuous into two regions D_1 and D_2, convert the double integrals over D_1 and D_2 into the corresponding iterated integrals, and then add the results. The procedure is illustrated by an example.

EXAMPLE **31.4**. *Evaluate the double integral of*

$$f(x,y) = \begin{cases} x\,, & y \le x^2 \\ y\,, & y > x^2 \end{cases}$$

over $D = [0,1] \times [0,1]$.

SOLUTION: Clearly f is not continuous along the parabola $y = x^2$ in D. The parabola cuts D into two simple closed regions:

$$\begin{aligned} D_1 &= \{(x,y)\,|\,x^2 \le y \le 1\,,\ 0 \le x \le 1\} \\ D_2 &= \{(x,y)\,|\,0 \le y \le x^2\,,\ 0 \le x \le 1\} \end{aligned}$$

The function f is continuous on D_2 by the definition of f so that

$$\iint_{D_2} f(x,y)dA = \int_0^1 x \int_0^{x^2} dydx = \int_0^1 x^3 dx = \frac{1}{4}\,.$$

However, f is not continuous on D_1. Note that on the boundary $y = x^2$ of D_1, $f(x, x^2) = x$, but $f(x, x^2 + a) = x^2 + a$ for any small positive $a > 0$ so that $f(x, x^2 + a) \to x^2$ as $a \to 0^+$. Nevertheless

$$\iint_{D_1} f(x,y)dA = \int_0^1 \int_{x^2}^1 f(x,y)dydx = \int_0^1 \int_{x^2}^1 ydydx = \frac{1}{2} \int_0^1 (1 - x^4)dx = \frac{2}{5}\,,$$

because $f(x,y) = y$ in the interval $[x^2 + a, 1]$ so the integral of $f(x,y)$ with respect to y over $[x^2 + a, 1]$ converges to the integral of y over $[x^2, 1]$ as $a \to 0^+$. Therefore

$$\int_0^1 f(x,y)dy = \int_0^{x^2} xdy + \int_{x^2}^1 ydy$$

and

$$\iint_D fdA = \int_0^1 \int_0^1 f(x,y)dydx = \iint_{D_1} fdA + \iint_{D_2} fdA = \frac{13}{20}\,.$$

□

31.3. Reversing the Order of Integration. Suppose f is integrable on a simple region D. Then either Eq. (**31.3**) or Eq. (**31.4**) can be used to evaluate the double integral of f over D. However, the technicalities involved in each case may be quite different. In fact, it may happen that an analytic evaluation of the integral is impossible in one order of integration, whereas it is possible in the other order. The following two examples illustrate these observations.

EXAMPLE **31.5**. *Evaluate the double integral of* $f(x,y) = 2x$ *over the region* D *bounded by the line* $x = 2y + 2$ *and the parabola* $x = y^2 - 1$.

SOLUTION: The region D is both vertically and horizontally simple. However, an evaluation of the iterated integral based on the algebraic description of D as a vertically simple region is more involved. Indeed, the largest value

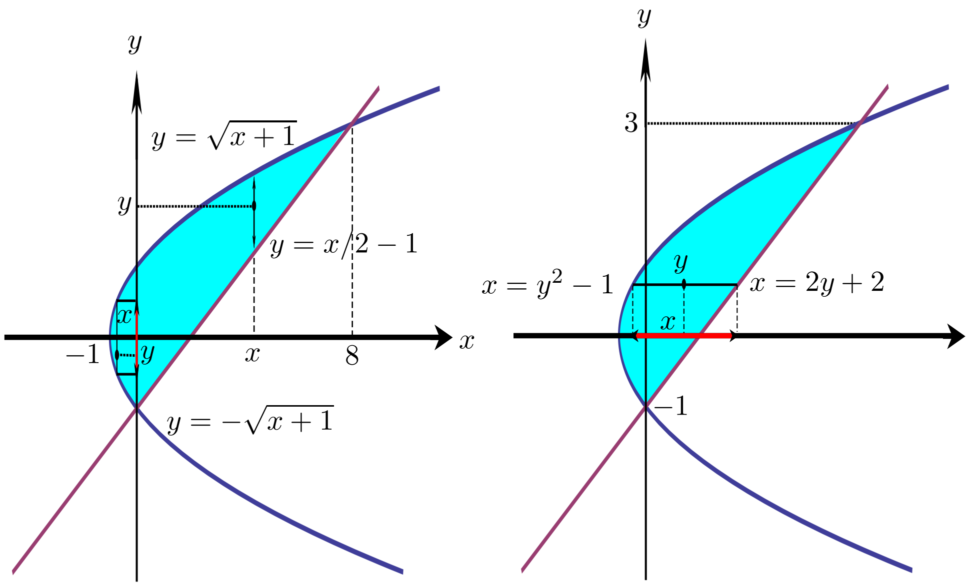


FIGURE **31.5**. Illustration to Example **31.5**. **Left**: The integration region D as a vertically simple region. An algebraic description requires to split the maximal range of x into two intervals. For every $-1 \le x \le 0$, the y coordinate ranges over the interval $-\sqrt{x+1} \le y \le \sqrt{x+1}$, whereas for every $0 \le x \le 8$, $x/2 - 1 \le y \le \sqrt{x+1}$. Accordingly, when converting the double integral to the iterated integral, the region D has to be split into two parts in which $x \le 0$ and $x \ge 0$. **Right**: The integration region D as a horizontally simple region. For every $-1 \le y \le 3$, the x coordinate ranges the interval $y^2 - 1 \le x \le 2y + 2$. So, the double integral can be converted to a single iterated integral.

of the x coordinate in D occurs at one of the points of intersection of the line and the parabola, $2y+2 = y^2 - 1$ or $(y-1)^2 = 4$, and hence, $y = -1$, 3. The largest value of x in D is $x = 3^2 - 1 = 8$. The smallest value of x occurs at the point of intersection of the parabola with the x axis, $x = -1$. So $[a, b] = [-1, 8]$. For any fixed $x \in [-1, 0]$, the range of the y coordinate is determined by the parabola $x = y^2 - 1$. Solutions of this equation are $y = \pm\sqrt{x+1}$ and the range of the y coordinate is $-\sqrt{x+1} \le y \le \sqrt{x+1}$. For any fixed $x \in [0, 8]$, the largest value of y still occurs on the parabola, $y = \sqrt{x+1}$, while the smallest value occurs on the line, $x = 2y + 2$ or $y = (x-2)/2$ so that $-\sqrt{x+1} \le y \le (x-2)/2$. The boundaries of D are

$$y = y_{\rm top}(x) = \sqrt{x+1}, \quad y = y_{\rm bot}(x) = \begin{cases} -\sqrt{x+1} & \text{if} \quad -1 \le x \le 0 \\ x/2 - 1 & \text{if} \quad 0 \le x \le 8 \end{cases}$$

That the bottom boundary consists of two graphs dictates the necessity to split the region D into two regions D_1 and D_2 such that x in $[-1,0]$ for all points in D_1 and $x\in[0,8]$ for all points in D_2. The corresponding iterated integral reads

$$\begin{aligned}\iint\limits_{D} 2x\,dA &= \iint\limits_{D_1} 2x\,dA + \iint\limits_{D_2} 2x\,dA\\ &= 2\int\limits_{-1}^{0} x \int\limits_{-\sqrt{x+1}}^{\sqrt{x+1}} dy\,dx + 2\int\limits_{0}^{8} x\int\limits_{-\sqrt{x+1}}^{x/2-1} dy\,dx.\end{aligned}$$

On the other hand, if the iterated integral corresponding to the algebraic description of D as a horizontally simple region is used, the technicalities are greatly simplified. The smallest and largest values of y in D occur at the points of intersection of the line and the parabola found above, $y=-1,3$, that is, $[c,d]=[-1,3]$. For any fixed $y\in[-1,3]$, the x coordinate ranges from its value on the parabola to its value on the line, $x_{\rm bot}(y)=y^2-1\leq x\leq 2y+2=x_{\rm top}(y)$. The corresponding iterated integral reads

$$\iint_D 2x\,dA = 2\int_{-1}^{3}\int_{y^2-1}^{2y+2} x\,dx\,dy = \int_{-1}^{3}(-y^4+6y^2+8y+3)\,dy = \frac{256}{5}\,,$$

which is simpler to evaluate than the previous one. □

Sometimes the iterated integration cannot even be carried out in one order, but it can still be done in the other order.

Example **31.6**. *Evaluate the double integral of $f(x,y)=\sin(y^2)$ over the region D, which is the triangle bounded by the lines $x=0$, $y=x$, and $y=\sqrt{\pi}$.*

Solution: Suppose that the iterated integral for vertically simple regions is used. The range of the x coordinate in D is the interval $[0,\sqrt{\pi}]=[a,b]$, and, for every fixed $x\in[0,\sqrt{\pi}]$, the range of the y coordinate is $y_{\rm bot}(x)=x\leq y\leq\sqrt{\pi}=y_{\rm top}(x)$ in D:

$$D=\{\ (x,y)\mid x\leq y\leq\sqrt{\pi}\,,\ x\in[0,\sqrt{\pi}]\}\,.$$

The iterated integral reads

$$\iint_D \sin(y^2)\,dA = \int_0^{\sqrt{\pi}}\int_x^{\sqrt{\pi}}\sin(y^2)\,dy\,dx\,.$$

However, the antiderivative of $\sin(y^2)$ cannot be expressed in elementary functions! Let us reverse the order of integration. The maximal range of the y coordinate in D is $[0,\sqrt{\pi}]=[c,d]$. For every fixed y in $[0,\sqrt{\pi}]$, the range of the x coordinate is $x_{\rm bot}(y)=0\leq x\leq y=x_{\rm top}(y)$ in D:

$$D=\{\ (x,y)\mid 0\leq x\leq y\,,\ y\in[0,\sqrt{\pi}]\}\,.$$

Therefore, the iterated integral reads

$$\iint_D \sin(y^2)\,dA = \int_0^{\sqrt{\pi}} \sin(y^2) \int_0^y dx\,dy$$
$$= \int_0^{\sqrt{\pi}} \sin(y^2) y\,dy = -\frac{1}{2}\cos(y^2)\Big|_0^{\sqrt{\pi}} = 1\,,$$

where the last integral is evaluated by the substitution $u = y^2$. □

31.4. The Use of Symmetry. The symmetry property has been established in single-variable integration:

$$f(-x) = -f(x) \quad \Rightarrow \quad \int_{-a}^{a} f(x)\,dx = 0,$$

which is quite useful. For example, an indefinite integral of $\sin(x^{2011})$ cannot be expressed in elementary functions. Nevertheless, to find its definite integral over any *symmetric* interval $[-a, a]$, an explicit form of the indefinite integral is not necessary. Indeed, the function $\sin(x^{2011})$ is antisymmetric and, hence, its integral over any symmetric interval vanishes. A similar property can be established for double integrals.

Consider a transformation T that maps each point (x, y) of the plane to another point (x_s, y_s) so that a region D is mapped to a region D^s. One writes

$$T:\ D \ \rightarrow\ D^s \quad \text{and} \quad T(D) = D^s\,.$$

A region D is said to be *symmetric* under a transformation $T:\ (x, y) \rightarrow (x_s, y_s)$ if the image D^s of D coincides with D (i.e., $T(D) = D$). For example, let D be bounded by an ellipse:

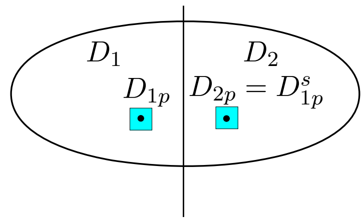
$$D = \{\ (x, y)\mid x^2/a^2 + y^2/b^2 \le 1\}\,.$$

Then D is symmetric under reflections about the x axis, the y axis, or their combination:

$$\begin{aligned}
T_x:&\quad (x, y) \rightarrow (x_s, y_s) = (-x, y)\,, \quad T_x(D) = D\,,\\
T_y:&\quad (x, y) \rightarrow (x_s, y_s) = (x, -y)\,, \quad T_y(D) = D\,,\\
T_{xy}:&\quad (x, y) \rightarrow (x_s, y_s) = (-x, -y)\,, \quad T_{xy}(D) = D\,.
\end{aligned}$$

A transformation of the plane $(x, y) \rightarrow (x_s, y_s)$ is said to be *area preserving* if the image D^s of any region D under this transformation has the same area, that is, $A(D) = A(D^s)$. For example, translations, rotations, reflections about lines, and their combinations are area-preserving transformations.

THEOREM **31.1**. (Symmetry Property)
Let f be integrable on a region D which is symmetric under an area-preserving transformation $(x, y) \rightarrow (x_s, y_s)$. If the function f is skew symmetric under this transformation, $f(x_s, y_s) = -f(x, y)$, then the integral of f over D



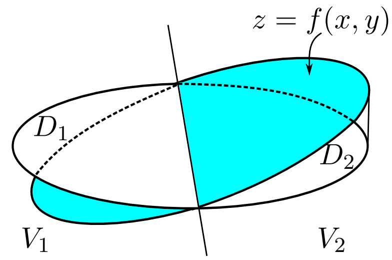


FIGURE **31.6**. **Left**: The region D is symmetric relative the reflection about the line. Under this reflection $D_1 \to D_2$ and $D_2 \to D_1$. Any partition of D_1 by elements D_{1p} induces the partition of D_2 by taking the images of D_{1p} under the reflection. **Right**: The graph of a function f that is skew symmetric under the reflection. If f is positive in D_2, then it is negative in D_1. The volume V_2 of the solid below the graph and above D_2 is exactly the same as the volume $V_1 = V_2$ of the solid above the graph and below D_1. But the latter solid lies below the xy plane and, hence, the double integral over D is $V_2 - V_1 = 0$.

vanishes:

$$\iint_D f(x, y)\, dA = 0\,.$$

A proof is postponed until the change of variables in double integrals is discussed. Here the simplest case of a reflection about a line is considered. If D is symmetric under this reflection, then the line cuts D into two equal-area regions D_1 and D_2 so that $D_1^s = D_2$ and $D_2^s = D_1$. The double integral is independent of the choice of partition (see (**29.2**)). Consider a partition of D_1 by elements D_{1p}, $p = 1, 2, ..., N$. By symmetry, the images D_{1p}^s of the partition elements D_{1p} form a partition of D_2 such that $\Delta A_p = A(D_{1p}) = A(D_{1p}^s)$ by area preservation. Choose elements D_{1p} and D_{1p}^s to partition the region D as shown in the left panel of Fig. **31.6**. Now recall that the double integral is also independent of the choice of sample points. Suppose (x_p, y_p) are sample points in D_{1p}. Choose sample points in D_{1p}^s to be the images (x_{ps}, y_{ps}) of (x_p, y_p) under the reflection. With these choices of the partition of D and sample points, the Riemann sum (**29.2**) vanishes:

$$\iint_D f\, dA = \lim_{N\to\infty} \sum_{p=1}^{N} \Big(f(x_p, y_p)\, \Delta A_p + f(x_{ps}, y_{ps})\, \Delta A_p \Big) = 0\,,$$

where the two terms in the sum correspond to partitions of D_1 and D_2 in D; by the hypothesis, the function f is antisymmetric under the reflection and therefore $f(x_{ps}, y_{ps}) = -f(x_p, y_p)$ for all p. From a geometrical point of view, the portion of the solid bounded by the graph $z = f(x, y)$ that lies

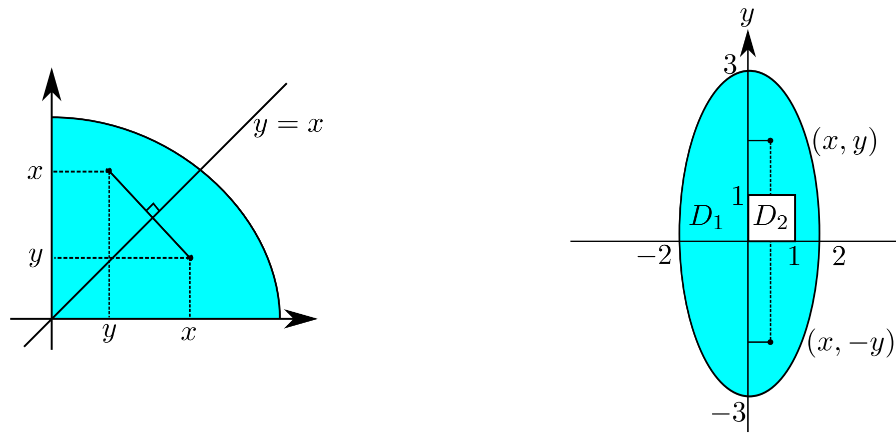


FIGURE **31.7**. **Left**: Illustration to Example **31.7**. The region is symmetric under the reflection about the line $y = x$. **Right**: The integration region D in Example **31.8**. It can be viewed as the difference of the elliptic region D_1 and the square D_2. The elliptic region is symmetric under the reflection about the x axis, whereas the function $f(x, y) = x^2y^3$ is skew-symmetric, $f(x, -y) = -f(x, y)$. So the integral over D_1 must vanish and the double integral over D is the negative of the integral over D_2.

above the xy plane has exactly the same shape as that below the xy plane, and therefore their volumes contribute with opposite signs to the double integral and cancel each other (see the right panel of **31.6**).

EXAMPLE **31.7**. *Evaluate the double integral of* $\sin[(x-y)^3]$ *over the portion* D *of the disk* $x^2 + y^2 \le 1$ *that lies in the first quadrant* $(x, y \ge 0)$*.*

SOLUTION: The region D is symmetric under the reflection about the line $y = x$ (see the left panel of Fig. **31.7**):

$$T: \quad (x, y) \to (x_s, y_s) = (y, x)\,, \quad T(D) = D\,,$$

whereas the function is skew-symmetric,

$$\begin{aligned} f(x_s, y_s) &= f(y, x) = \sin[(y-x)^3] = \sin[-(x-y)^3] = -\sin[(x-y)^3] \\ &= -f(x, y)\,, \end{aligned}$$

By the symmetry property (Theorem **31.1**), the double integral vanishes. □

EXAMPLE **31.8**. *Evaluate the double integral of* $f(x, y) = x^2y^3$ *over the region* D*, which is obtained from the elliptic region* $x^2/4 + y^2/9 \le 1$ *by removing the square* $[0, 1] \times [0, 1]$*.*

SOLUTION: Let D_1 and D_2 be the elliptic and square regions, respectively. The elliptic region D_1 is large enough to include the square D_2 as shown in the right panel of Fig. **31.7**. Therefore, the additivity of the double integral can be used (compare Example **30.4**) to transform the double integral over a non-simple region D into two double integrals over simple regions:

$$\begin{aligned}\iint_D x^2y^3\,dA &= \iint_{D_1} x^2y^3\,dA - \iint_{D_2} x^2y^3\,dA \\ &= -\iint_{D_2} x^2y^3\,dA = -\int_0^1 x^2\,dx\int_0^1 y^3\,dy = -\frac{1}{12}\,;\end{aligned}$$

the integral over D_1 vanishes by Theorem **31.1** because the elliptic region D_1 is symmetric under the reflection

$$T:\quad (x,y)\to(x_s,y_s)=(x,-y)\,,\quad T(D_1)=D_1\,,$$

whereas the integrand is skew-symmetric,

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

□

31.5. Study Problems.

Problem **31.1**. *Prove the Dirichlet formula*

$$\int_0^a\int_0^x f(x,y)dydx=\int_0^a\int_y^a f(x,y)dxdy\,,\qquad a>0\,.$$

SOLUTION: The left side of the equation is an iterated integral for the double integral $\iint_D f dA$. Let us find the shape of D. According to the limits of integration, D admits the following algebraic description (as a vertically simple region). For every $0\le x\le a$, the y coordinate changes in the interval $0\le y\le x$. So the region D is the triangle bounded by the lines $y=0$, $y=x$, and $x=a$. To reverse the order of integration, let us find an algebraic description of D as a horizontally simple region. The maximal range of y in D is the interval $[0,a]$. For every fixed $0\le y\le a$, the x coordinate spans the interval $y\le x\le a$ in D. So the two sides of the Dirichlet formula represent the same double integral as iterated integrals in different orders and, hence, are equal. □

Problem **31.2**. *Reverse the order of integration*

$$\int_1^2\int_{2-x}^{\sqrt{2x-x^2}} f(x,y)\,dydx$$

SOLUTION: The given iterated integral represents a double integral $\iint_D f dA$ where the integration region admits the following description (as a vertically simple region). For every fixed $1\le x\le 2$, the y coordinates spans the interval $2-x\le y\le\sqrt{2x-x^2}$. So D is bounded by the graphs:

$$y=2-x\quad\text{and}\quad y=\sqrt{2x-x^2}\ \text{ or }\ y^2=2x-x^2\ \text{ or }\ (x-1)^2+y^2=1\,,$$

where the squares have been completed to obtain the last equation. The boundaries of D contain the line and the circle of radius 1 centered at $(1,0)$. The circle and the line intersect at the points $(1,1)$ and $(2,0)$. Thus, the region D is the part of the disk $(x-1)^2+y^2 \le 1$ that lies above the line $y = 2-x$:

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

The reader is advised to sketch it. To reverse the order of integration, let us find an algebraic description of D as a horizontally simple region. The maximal range of y is the interval $[0,1]$, which is determined by the points of intersection of the circle and the line. Viewing the region D along the x axis, one can see that for every fixed $0 \le y \le 1$, the smallest value of x in D is attained on the line $y = 2-x$ or $x = 2-y = x_{\rm bot}(y)$, while its greatest value in D is attained on the circle $(x-1)^2+y^2 = 1$ or $x-1 = \pm\sqrt{1-y^2}$ or $x = 1+\sqrt{1-y^2} = x_{\rm top}(y)$ because the solution with the plus sign corresponds to the part of the circle that lies above the line. Hence, the integral in the reversed order reads

$$\int_0^1 \int_{2-y}^{1+\sqrt{1-y^2}} f(x,y)\,dxdy$$

$\square$

31.6. Exercises.

1–5. For each of the two orders of integration, specify the limits in the iterated integrals for $\iint_D f(x,y)dA$, splitting the integration region when necessary.

1. D is the triangle with vertices $(0,0)$, $(2,1)$, and $(-2,1)$;
2. D is a trapezoid with vertices $(0,0)$, $(1,0)$, $(1,2)$, and $(0,1)$;
3. D is the disk $x^2+y^2 \le 1$;
4. D is the disk $x^2+y^2 \le y$;
5. D is the ring $1 \le x^2+y^2 \le 4$.

6–17. Evaluate each of the following double integrals over the specified region.

6. $\iint_D xy dA$ where D is bounded by the curves $y = x^2$ and $y = x$;
7. $\iint_D (2+y)dA$ where D is the region bounded by the graphs of $x = 3$ and $x = 4-y^2$;
8. $\int\int_D dxdy\,(x+y)$ where D is bounded by the curves $x = y^4$ and $x = y$;
9. $\iint_D (2+y)dA$ where D is the region bounded by the three lines of $x = 3$, $y+x = 0$ and $y-x = 0$. Find the value of the integral by geometric means;
10. $\iint_D x^2y dA$ where D is the region bounded by the graphs of $y = 2+x^2$ and $y = 4-x^2$;

11. $\iint_D \sqrt{1-y^2}dA$ where D is the triangle with vertices $(0,0)$, $(0,1)$, and $(1,1)$;

12. $\iint_D xydA$ where D is bounded by the lines $y=1$, $x=-3y$ and $x=2y$;

13. $\iint_D y\sqrt{x^2-y^2}dA$ where where D is the triangle with vertices $(0,0)$, $(1,0)$, and $(1,1)$;

14. $\iint_D (2a-x)^{-1/2}dA$ where D is bounded by the coordinate axes and by the shortest arc of the circle of radius a and centered at (a,a);

15. $\iint_D |xy|dA$ where D is the disk of radius a centered at the origin;

16. $\iint_D (x^2+y^2)dA$ where D is the parallelogram with the sides $y=x$, $y=x+a$, $y=a$, and $y=3a$ $(a>0)$;

17. $\iint_D y^2dA$ where D is bounded by the x axis and by one arc of the cycloid $x=a(t-\sin t)$, $y=a(1-\cos t)$, $0\le t\le 2\pi$.

18–23. Sketch the solid region whose volume is given by each of the following integrals.

18. $\int_0^1\int_0^{1-x}(x^2+y^2)dydx$;

19. $\iint_D (x+y)dA$, where D is defined by the inequalities $0\le x+y\le 1$, $x\ge 0$, and $y\ge 0$;

20. $\iint_D \sqrt{x^2+y^2}dA$, where D is defined by the inequality $x^2+y^2\le x$;

21. $\iint_D (x^2+y^2)dA$, where D is defined by the inequality $|x|+|y|\le 1$;

22. $\iint_D \sqrt{1-(x/2)^2-(y/3)^2}dA$, where D is defined by the inequality $(x/2)^2+(y/3)^2\le 1$.

23–27. Use the double integral to find the volume of the specified solid region E.

23. E is bounded by the plane $x+y+z=1$ and the coordinate planes;

24. E lies under the paraboloid $z=3x^2+y^2$ and above the region in the xy-plane bounded by the curves $x=y^2$ and $x=1$;

25. E is bounded by the cylinder $x^2+y^2=1$ and the planes $y=z$, $z=0$ in the first octant;

26. E is bounded the cylinders $x^2+y^2=a^2$ and $y^2+z^2=a^2$;

27. E is enclosed by the parabolic cylinders $y=1-x^2$, $y=x^2-1$ and the planes $x+y+z=2$, $2x+2y-z=10$.

28–38. Sketch the region of integration and reverse the order of integration in each of the following iterated integrals. Evaluate the integral if the integrand is specified.

28. $\int_0^{\sqrt{\pi}}\int_y^{\sqrt{\pi}}\cos(x^2)dxdy$ *Hint*: after reversing the integration order, make the substitution $u=x^2$ to do the integral;

29. $\int_0^1\int_{x^3}^{\sqrt{x}} f(x,y)dydx$;

30. $\int_0^1\int_{y^2}^{y} f(x,y)dxdy$;

31. $\int_0^1\int_1^{e^x} f(x,y)dydx$;

32. $\int_1^4\int_{\sqrt{y}}^{2} f(x,y)dxdy$;

33. $\int_0^3 \int_0^y f(x,y)dxdy + \int_3^6 \int_0^{6-y} f(x,y)dxdy$;
34. $\int_0^4 \int_{\sqrt{x}}^2 (1+y^3)^{-1}dydx$;
35. $\int_{-6}^2 \int_{(x^2/4)-1}^{2-x} f(x,y)\, dydx$;
36. $\int_{-1}^1 \int_{-\sqrt{1-x^2}}^{1-x^2} f(x,y)\, dydx$;
37. $\int_0^{2a} \int_{\sqrt{2ax-x^2}}^{\sqrt{2ax}} f(x,y)\, dydx \quad (a>0)$;
38. $\int_0^{\pi} \int_0^{\sin x} f(x,y)\, dydx$.

39–41. Use the symmetry and the properties of the double integral to find each of the following integrals.

39. $\iint_D e^{x^2}\sin(y^3)dA$ where D is the triangle with vertices $(0,1)$, $(0,-1)$, and $(1,0)$;
40. $\iint_D (y^2-x^2)^9 dA$ where $D = \{(x,y)|1 \le |x|+|y| \le 2\}$;
41. $\iint_D x dA$ where D is bounded by the ellipse $x^2/a^2 + y^2/b^2 = 1$ and has the triangular hole with vertices $(0,b)$, $(0,-b)$, and $(a,0)$;
42. $\iint_D (\cos(x^2) - \cos(y^2))dA$ where D is the disk $x^2+y^2 \le a^2$.

43–45. Use the double integral to find the area of each of the following regions.

43. D is bounded by the curves $xy = a^2$ and $x+y = 5a/2$, $a>0$;
44. D is bounded by the curves $y^2 = 2px + p^2$ and $y^2 = -2qx + q^2$ where p and q are positive numbers;
45. D is bounded by $(x-y)^2 + x^2 = a^2$.

32. Double Integrals in Polar Coordinates

32.1. Polar coordinates. Polar coordinates (r,θ) are defined by the following relations:

$$x = r\cos\theta\,,\qquad y = r\sin\theta\,,$$

where (x,y) are rectangular coordinates in a plane. Polar coordinates used for integration are a restricted version of polar coordinates introduced in Calculus 2. For purposes of integration, the range of r and θ is restricted. The variable r is always required to be non-negative, $r \geq 0$, so that $r = \sqrt{x^2+y^2}$ is the distance from the origin to the point (x,y). The polar angle θ is usually required to take its values in an interval of length 2π, the most commonly-used intervals being $[0,2\pi)$, $[-\pi,\pi)$, $[0,2\pi]$, and $[-\pi,\pi]$. If the either of the first two intervals is used, then every point in the xy plane, other than the origin, has exactly one pair of polar coordinates (r,θ). If $0 \leq \theta \leq 2\pi$, then each point $(x,0)$ on the positive x axis has two pairs of polar coordinates: $(r,\theta) = (x,0)$ and $(r,\theta) = (x,2\pi)$; all other points besides the origin have exactly one pair of polar coordinates. If the interval is $-\pi \leq \theta \leq \pi$, then each point $(x,0)$ on the *negative* x axis has two pairs of polar coordinates: $(r,\theta) = (|x|,-\pi)$ and $(r,\theta) = (|x|,\pi)$; all other points besides the origin have exactly one pair of polar coordinates. No matter what interval is chosen, the origin always has infinitely many pairs of polar coordinates: $(r = 0, \theta = \text{any value in the chosen interval})$. Fortunately, *none of these coordinate-duplications has any effect on integration*, because the set of points with more than one pair of polar coordinates has zero area. Therefore there is never any harm in using the closed interval $[0,2\pi]$ or $[-\pi,\pi]$ for θ in integration, and these choices make formulas look less strange than if half-open intervals were used.

However, if one wishes to assign every point of the xy plane, other than the origin, a *unique* pair of polar coordinates, a half-open interval has to be used such as $[0,2\pi)$, or $[-\pi,\pi)$. To express θ in terms of $(x,y) \neq (0,0)$, one can use the geometrical interpretation of θ as the angle between the positive x axis and the ray from the origin through the point (x,y) counted counterclockwise if θ is positive and clockwise if θ is negative. Recall that the function $\tan^{-1} u$ has the domain $(-\infty,\infty)$ and the range $(-\pi/2,\pi/2)$; it is monotonic and $\lim_{u\to\infty} = \pi/2$ and $\lim_{u\to-\infty} = -\pi/2$. If one takes the interval $-\pi \leq \theta < \pi$, then

$$r = \sqrt{x^2+y^2}\,,\quad \theta = \begin{cases} \tan^{-1}(y/x) & \text{if } x>0 \\ \pi/2 & \text{if } x=0,\ y>0 \\ -\pi/2 & \text{if } x=0,\ y<0 \\ \tan^{-1}(y/x)+\pi & \text{if } x<0,\ y>0 \\ \tan^{-1}(y/x)-\pi & \text{if } x<0,\ y\leq 0 \end{cases}$$

The first relation defines the rotation angle θ in in quadrants I and IV, with the y-axis excluded. The second and third relations define the rotation angle on the y axis (except the origin). The fourth and fifth relations define

θ, respectively, in the interior of quadrant II and in quadrant III, with the negative y axis excluded. Similarly, if $0 \leq \theta < 2\pi$, then

$$\textbf{(32.1)} \qquad r = \sqrt{x^2 + y^2}\,, \quad \theta = \begin{cases} \tan^{-1}(y/x) & \text{if } x > 0,\ y \geq 0 \\ \pi/2 & \text{if } x = 0,\ y > 0 \\ 3\pi/2 & \text{if } x = 0,\ y < 0 \\ \tan^{-1}(y/x) + \pi & \text{if } x < 0 \\ \tan^{-1}(y/x) + 2\pi & \text{if } x > 0,\ y < 0 \end{cases}$$

Note that none of the cases above define θ at the origin. As already mentioned, at the origin, θ is allowed to have any value from the chosen interval.

An ordered pair of numbers (r, θ) is considered as a point in a *polar plane*, just like an ordered pair (x, y) represents a point in a plane in space. Suppose D' is a region that lies in the part of the polar plane in which $r \geq 0$ and θ lies in one of the standard intervals described above, e.g., $0 \leq \theta \leq 2\pi$. The relations $x = r\cos\theta$, $y = r\sin\theta$ define a *transformation* T of any region D' in the polar plane to a region D in the xy plane:

$$T:\ D' \to D\,,$$

that is, to every ordered pair (r, θ) corresponding to a point of D', an ordered pair (x, y) corresponding to a point of D is assigned. If D' is a region in the open rectangle $(0, \infty) \times (0, 2\pi)$, then the transformation T is *one-to-one* according to the analysis given above. Therefore the transformation T is one-to-one for any region D' in $[0, \infty) \times [0, 2\pi]$ except possibly on the boundary of D'. Furthermore, if D is bounded, then it is contained in a disk $x^2 + y^2 \leq R^2$ of a sufficiently large radius R. The region D' whose image is a bounded region D is also bounded because $0 \leq r \leq R$ for all r in D' (recall that the variable θ always ranges over an interval of length 2π). In particular,

$$T:\ \text{boundary of } D' \ \to\ \text{boundary of } D\,.$$

This observation allows us to reconstruct the region D' for a given bounded region D. First note that, if D' is a rectangle:

$$D' = [a, b] \times [\theta_1, \theta_2]\,,$$

then

$$\begin{array}{lll} T:\ r = a & \to & x^2 + y^2 = a^2 \\ T:\ r = b & \to & x^2 + y^2 = b^2 \\ T:\ \theta = \theta_1 & \to & (x, y) = (t\cos\theta_1, t\sin\theta_2)\,,\ t > 0 \\ T:\ \theta = \theta_2 & \to & (x, y) = (t\cos\theta_1, t\sin\theta_2)\,,\ t > 0 \end{array}$$

Therefore D is bounded by two concentric circles of radii a and b, and by two rays from the origin that make the angles θ_1 and θ_2 with the positive x axis that are counted counterclockwise ($0 \leq \theta_1 < \theta_2 \leq 2\pi$). If $a = 0$, then the boundary $r = 0$ of D' is collapsed to a single point $(x, y) = (0, 0)$ under the transformation T. In this case, D is a wedge of the disk of radius b that lies between the two rays (provided $\theta_1 \neq 0$ or $\theta_2 \neq 2\pi$). If $\theta_1 = 0$ and $\theta_2 = 2\pi$, then D is bounded only by the circles of radii $a \neq 0$ and b, and just by the circle of radius b if $a = 0$. It is also clear that if the origin is an interior point of a region D, then θ takes its full range (e.g., $[0, 2\pi]$) in D'.

EXAMPLE **32.1**. *Find a closed region D' in the part $[0, \infty) \times [0, 2\pi]$ of the polar plane whose image under the transformation T is*

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

SOLUTION: The boundary of D contains the origin. Therefore the boundary of D' has to contain a part of the line $r = 0$ in the polar plane because $T: \; r = 0 \;\to\; (x, y) = (0, 0)$. To find the other part of the boundary of D', let us write the equation of the boundary of D with the origin excluded in the polar coordinates:

$$\begin{aligned} x^2 + (y-1)^2 = 1 \quad &\Rightarrow \quad x^2 + y^2 - 2y = 0 \quad \Rightarrow \quad r^2 = 2r\sin\theta \\ &\Rightarrow \quad r = 2\sin\theta \end{aligned}$$

because $r \ne 0$. Since $r \ge 0$ in any D', the range of θ is determined by the intersection of the line $r = 0$ and the graph $r = 2\sin\theta$ in the polar plane, which gives the interval $[0, \pi]$. Thus,

$$T: \; D' = \{(r, \theta) \,|\, 0 \le r \le 2\sin\theta\,,\; 0 \le \theta \le \pi\} \;\to\; D\,.$$

□

Let D be the disk:

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

Its boundary $(x-1)^2 + y^2 = 1$ or $x^2 + y^2 = 2x$ in the polar coordinates is $r^2 = 2r\cos\theta$. The region D lies in the first and fourth quadrants of the xy plane, while none of its parts is either in the second or third quadrants. To describe the region D' in the polar plane whose image is D, it is therefore more convenient to choose the interval $[-\pi, \pi]$ as the full range of θ. Using a similar line of argument, it is concluded that

$$T: \; D' = \{(r, \theta) \,|\, 0 \le r \le 2\cos\theta\,,\; -\pi/2 \le \theta \le \pi/2\} \;\to\; D\,.$$

32.2. The double integral in polar coordinates. In what follows, the range of polar coordinates r and θ in the polar plane is always restricted so that $r \ge 0$ and θ takes values in a closed interval of length 2π, and a region D' is always assumed to lie in the part of the polar plane defined by the restricted range of the polar coordinates.

Let D' be a closed bounded region in the polar plane and D its image in the xy plane. Let $R'_D = [a, b] \times [c, d]$ be a closed rectangle containing D' in the polar plane so that the image of R'_D contains D. According to the above analysis, D lies between two concentric circles of radii a and b and between two rays extended from the origin. Consider partitions of the intervals $r \in [a, b]$ and $\theta \in [c, d]$:

$$\begin{aligned} &r_0 = a\,, \quad r_j = r_{j-1} + \Delta r\,, \quad \Delta r = (b-a)/N_1\,, \quad j = 1, 2, ..., N_1 \\ &\theta_0 = c\,, \quad \theta_k = \theta_{k-1} + \Delta\theta\,, \quad \Delta\theta = (d-c)/N_2\,, \quad k = 1, 2, ..., N_2\,. \end{aligned}$$

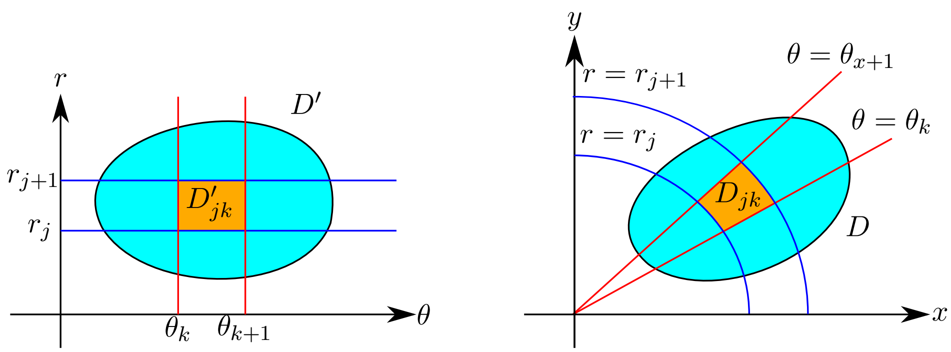


FIGURE **32.1**. **Left**: A partition of D' by the coordinate lines $r = r_j$ and $\theta = \theta_k$ where $r_j - r_{j-1} = \Delta r$ and $\theta_k - \theta_{k-1} = \Delta\theta$. A patrition element is a rectangle D'_{jk}. Its area is $\Delta A'_{jk} = \Delta r \Delta\theta$. **Right**: A partition of D by the images of the coordinate curves $r = r_j$ (concentric circles) and $\theta = \theta_k$ (rays extended from the origin). A patrition element D_{jk} is the image of the rectangle D'_{jk}. Its area is $\Delta A_{jk} = \frac{1}{2}(r_j^2 - r_{j-1}^2)\Delta\theta = \frac{1}{2}(r_j + r_{j-1})\Delta A'_{jk}$.

The partitions of the above interval generate a rectangular partition of R'_D such that each partition rectangle D'_{jk} is bounded by the coordinate lines

$$r = r_{j-1}\,, \quad r = r_j\,, \quad \theta = \theta_{k-1}\,, \quad \theta = \theta_k$$

in the polar plane as shown in Fig. **32.1** (left panel). Each partition rectangle has the area

$$\Delta A' = \Delta r\,\Delta\theta\,.$$

The image of the coordinate line $r = r_k$ in the xy plane is the circle of radius r_k centered at the origin. The image of the coordinate line $\theta = \theta_k$ in the xy plane is the ray from the origin that makes the angle θ_k with the positive x axis (as defined above). The rays and circles are called *coordinate curves* of the polar coordinate system, that is, the curves along which either the coordinate r or the coordinate θ remains constant (concentric circles and rays, respectively). A rectangular partition of D' induces a partition of D by coordinate curves of the polar coordinates. Each partition element D_{jk} is the image of the rectangle D'_{jk} and is bounded by two circles and two rays (if the origin is viewed as the circle of zero radius).

Let f be an integrable function on D that is extended outside D by setting its values to 0. By Theorem **29.3** the double integral of f over D can be computed as the limit of Riemann sums (**29.2**) and the limit does not depend on either the choice of partition elements or sample points in them. Let ΔA_{jk} be the area of D_{jk}. The area of the sector of the disk of radius r_j

that has the angle $\Delta\theta$ is $r_j^2\Delta\theta/2$. Therefore,

$$\Delta A_{jk} = \frac{1}{2}(r_j^2 - r_{j-1}^2)\,\Delta\theta = \frac{1}{2}(r_j + r_{j-1})\,\Delta r\,\Delta\theta = \frac{1}{2}(r_j + r_{j-1})\,\Delta A' .$$

In (**29.2**), put $\Delta A_p = \Delta A_{jk}$ and choose the sample points $\mathbf{r}_p$ being the images of sample points (r_j^*, θ_k^*) in D'_{jk} so that $f(\mathbf{r}_p) = f(r_j^*\cos\theta_k^*, r_j^*\sin\theta_k^*)$. The limit in (**29.2**) is understood as the double limit $(N_1, N_2) \to \infty$ (or $(\Delta r, \Delta\theta) \to (0,0)$). Owing to the independence of the limit of the choice of sample points, let us use the midpoint rule

$$r_j^* = \frac{1}{2}(r_j + r_{j-1})\,.$$

With this choice,

$$\Delta A_{jk} = r_j^*\Delta A' .$$

By taking the limit of the Riemann sum (**29.2**) it is concluded that

$$\iint_D f(x,y)dA = \lim_{N_{1,2}\to\infty}\sum_{j=1}^{N_1}\sum_{k=1}^{N_2} f(r_j^*\cos\theta_k^*, r_j^*\sin\theta_k^*)r_j^*\,\Delta A' .$$

The right side of this equation is a Riemann sum of the function

$$g(r,\theta) = f(r\cos\theta, r\sin\theta)J(r)$$

over the region D', where $J(r) = r$ is called the *Jacobian* of the polar coordinates. The Jacobian defines the area element transformation

$$dA = J\,dA' = r\,dA'.$$

The Riemann sum converges to the integral of g over D', provided g is integrable. One can prove that *if f is integrable on D, then g is integrable on D'*. For example, if f is continuous on a disk D centered at the origin, $x^2 + y^2 \le b^2$, then g is continuous on the rectangle $D' = [0,b] \times [0, 2\pi]$ because g is a composition of continuous functions $f(x,y)$ and $x = r\cos\theta$, $y = r\sin\theta$. Therefore g is integrable on D' as any continuous function on a region with a piecewise smooth boundary (Theorem **28.2**).

THEOREM **32.1**. (Double Integral in Polar Coordinates). *Let a closed bounded region D be the image of a closed bounded region D' in the polar plane spanned by ordered pairs (r,θ) of polar coordinates. Let $f(x,y)$ be continuous on D. Then $f(r\cos\theta, r\sin\theta)J(r)$ is integrable on D' and*

$$\iint\limits_D f(x,y)\,dA = \iint\limits_{D'} f(r\cos\theta, r\sin\theta)\,J(r)\,dA' , \quad J(r) = r\,.$$

The area of a planar region. By setting $f(x,y) = 1$ in the double integral in polar coordinates, it is concluded that the area of a region D is given by

$$A(D) = \iint_D dA = \iint_{D'} r\, dA'\,.$$

A similarity between the double integral in rectangular and polar coordinates is that they both use partitions by corresponding coordinate curves. Note that horizontal and vertical lines are coordinate curves of the rectangular coordinates. So the very term "a double integral in polar coordinates" refers to a specific partition of D in the Riemann sum, namely, by *coordinate curves of polar coordinates* (by circles and rays).

32.3. Evaluation of double integrals in polar coordinates. *The double integral over D' can be evaluated by the standard means, that is, by converting it to a suitable iterated integral with respect to r and θ.* Suppose that D' is a vertically simple region as shown in Fig. **32.2** (right panel):

$$D' = \{\ (r,\theta)\ |\ r_{\rm bot}(\theta) \le r \le r_{\rm top}(\theta), \quad \theta_1 \le \theta \le \theta_2\ \}$$

Then D is bounded by the *polar graphs* $r = r_{\rm bot}(\theta)$, $r = r_{\rm top}(\theta)$ and by the lines $y\cos\theta_1 = x\sin\theta_1$ and $y\cos\theta_2 = x\sin\theta_2$ (see the left panel of Fig, **32.2**). Recall from Calculus 2 that curves defined by the equation $r = g(\theta)$ are called polar graphs. They can be visualized by means of a simple geometrical procedure. Take a ray corresponding to a fixed value of the polar angle θ. On this ray, mark the point at a distance $r = g(\theta)$ from the origin. All such points obtained for all values of θ in a specified interval form the polar graph. The double integral over D can be written as the iterated integral over D':

$$\iint_D f(x,y)dA = \int_{\theta_1}^{\theta_2}\int_{r_{\rm bot}(\theta)}^{r_{\rm top}(\theta)} f(r\cos\theta, r\sin\theta)\, r\, dr d\theta\,.$$

The iterated integral over a horizontally simple region D' is obtained similarly. Thus, the evaluation of a double integral in polar coordinates includes three essential steps:

Step 1. Find the region D' in the polar plane whose image is the given region D under the transformation $x = r\cos\theta$, $y = r\sin\theta$;

Step 2. Write the integrand as a function of polar coordinates $f(x,y) = f(r\cos\theta, r\sin\theta)$;

Step 3. Evaluate the double integral of f *multiplied by the Jacobian* $J(r) = r$ over D' by converting it to a suitable iterated integral.

EXAMPLE **32.2**. *Use polar coordinates to evaluate the double integral of $f(x,y) = xy^2\sqrt{x^2+y^2}$ over D, which is the portion of the disk $x^2+y^2 \le 1$ that lies in the first quadrant.*

SOLUTION: Step 1. Since $0 \le x^2 + y^2 \le 1$, the region D' is bounded by the lines $r = 0$ and $r = 1$. The boundary $(x,0)$, $x > 0$, of D is the ray

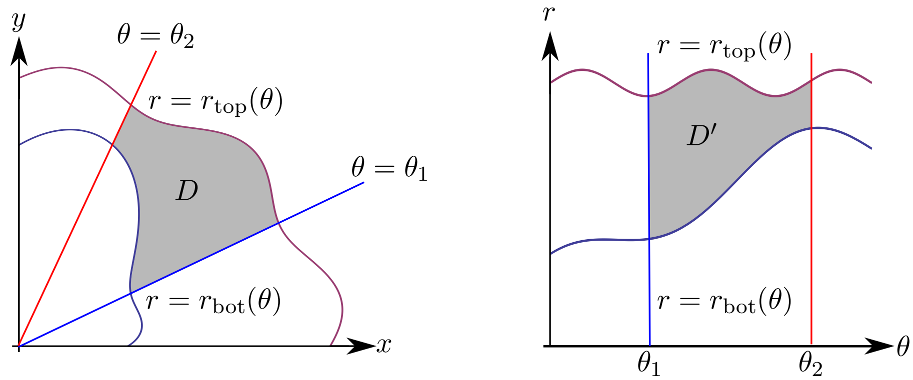


FIGURE **32.2**. **Left**: In polar coordinates, the boundary of a region D, which is the image of a vertically simple region D' in the polar plane, can be viewed as polar graphs and lines through the origin. **Right**: A vertically simple region D' in the polar plane.

$\theta = 0$ and the boundary $(0, y)$, $y > 0$, is the ray $\theta = \pi/2$. Therefore $D' = [0, 1] \times [0, \pi/2]$.
Step 2. $f(r\cos\theta, r\sin\theta) = xy^2\sqrt{x^2+y^2} = r^4\cos\theta\sin^2\theta$.
Step 3. The double integral over the rectangle D' can be evaluated by Fubini's theorem:

$$\begin{aligned}\iint_D f(x,y)\,dA &= \iint_{D'} r^5\cos\theta\sin^2\theta dA' = \int_0^{\pi/2}\sin^2\theta\cos\theta\,d\theta\int_0^1 r^5 dr\\ &= \frac{1}{3}\sin^3\theta\Big|_0^{\pi/2}\cdot\frac{1}{6}\,r^6\Big|_0^1 = \frac{1}{18}\,,\end{aligned}$$

where the area transformation law $dA = rdA'$ has been taken into account. □

This example shows that the technicalities involved in evaluating the double integral have been substantially simplified by passing to polar coordinates. The simplification is twofold. First, the domain of integration has been simplified; the new domain is a rectangle, which is much simpler to handle in the iterated integral than a portion of a disk. Second, the evaluation of ordinary integrals with respect to r and θ appears to be simpler than the integration of f with respect to either x or y needed in the iterated integral. However, these simplifications cannot always be achieved by converting the double integral to polar coordinates. The region D and the integrand f should have some particular properties that guarantee the observed simplifications and thereby justify the use of polar coordinates. Here are some guiding principles to decide whether the conversion of a double integral to polar coordinates could be helpful:

- The domain D is bounded by circles, lines through the origin, and polar graphs.

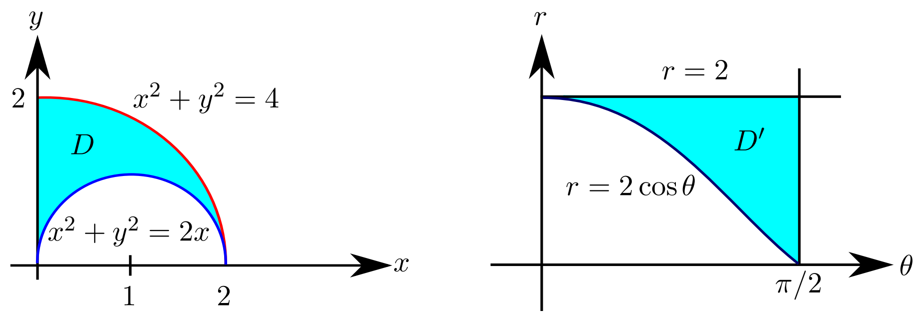


FIGURE **32.3**. Illustration to Example **32.3**.

- The function $f(x, y)$ depends on either the combination $x^2+y^2 = r^2$ or $y/x = \tan\theta$.

Indeed, if D is bounded only by circles centered at the origin and lines through the origin, then D' is a rectangle because the boundaries of D are *coordinate curves* of polar coordinates. If the boundaries of D contain circles not centered at the origin or, generally, polar graphs, that is, curves defined by the relations $r = g(\theta)$, then an algebraic description of the boundaries of D' is simpler than that of the boundaries of D. If $f(x, y) = h(u)$, where $u = x^2+y^2 = r^2$ or $u = y/x = \tan\theta$, then in the iterated integral one of the integrations, either with respect to θ or r, becomes trivial.

EXAMPLE **32.3**. *Evaluate the double integral of $f(x, y) = xy$ over the region D that lies in the first quadrant and is bounded by the circles $x^2+y^2 = 4$ and $x^2 + y^2 = 2x$.*

SOLUTION: Step 1. Using the principle that the boundary of D' is mapped into the boundary of D, the boundary of D' is obtained by converting the equations for the boundary of D to polar coordinates. The boundary of the region D contains three curves:

$$\begin{array}{lclcl} x^2 + y^2 = 4 & \Rightarrow & r^2 = 4 & \Rightarrow & r = 2, \\ x^2 + y^2 = 2x & \Rightarrow & r^2 = 2r\cos\theta & \Rightarrow & r = 2\cos\theta\,, \\ x = 0,\ y \geq 0 & \Rightarrow & \theta = \pi/2 & & \end{array}$$

So, in the polar plane, the region D' is bounded by the horizontal line $r = 2$, the graph $r = 2\cos\theta$, and the vertical line $\theta = \pi/2$ (see Fig. **32.3**).
Step 2. $f(x, y) = xy = r^2\cos\theta\sin\theta$.
Step 3. It is convenient to use an algebraic description of D' as a vertically simple region:

$$D' = \{(r, \theta)\,|\, r_{\rm bot}(\theta) = 2\cos\theta \leq r \leq 2 = r_{\rm top}(\theta)\,,\quad \theta_1 = 0 \leq \theta \leq \pi/2 = \theta_2\}$$

because $r_{\text{top}}(0) = r_{\text{bot}}(0)$. Multiplying f by the Jacobian $J = r$, the double integral in question is converted to polar coordinates and then evaluated:

$$\begin{aligned}\iint_D xy\,dA &= \iint_{D'} r^3 \sin\theta\cos\theta\,dA' = \int_{\theta_1}^{\theta_2} \sin\theta\cos\theta \int_{r_{\text{bot}}(\theta)}^{r_{\text{top}}(\theta)} r^3\,dr\,d\theta \\ &= \int_0^{\pi/2} \sin\theta\cos\theta \int_{2\cos\theta}^{2} r^3\,dr\,d\theta \\ &= 4\int_0^{\pi/2} (1-\cos^4\theta)\cos\theta\sin\theta\,d\theta \\ &= 4\int_0^1 (1-u^4)u\,du = 4\Big(\frac{1}{2}-\frac{1}{6}\Big) = \frac{4}{3},\end{aligned}$$

where the change of variables $u = \cos\theta$ has been used. □

EXAMPLE **32.4**. *Find the area of the region D that is bounded by two spirals $r = \theta$ and $r = 2\theta$, where θ in $[0, 2\pi]$, and the positive x axis.*

Before solving the problem, let us make a few comments about the shape of D. The boundaries $r = \theta$ and $r = 2\theta$ are polar graphs. Given a value of θ, $r = \theta$ (or $r = 2\theta$) is the distance from the point on the graph to the origin. As this distance increases monotonically with increasing θ, the polar graphs are spirals winding about the origin. The region D lies between two spirals; it is not simple in any direction (see the left panel of Fig. **32.4**). Let us write the equations of the boundary of D in the rectangular coordinates. For example, in the first quadrant with $x \neq 0$ (the y axis is excluded)

$$r = \theta \quad\Rightarrow\quad \tan r = \tan\theta \quad\Rightarrow\quad x\tan\sqrt{x^2+y^2} = y\,.$$

There is no way to find an analytic solution of this equation to express y as a function of x or vice versa. Therefore, had one tried to evaluate the double integral in the rectangular coordinates by cutting the region D into simple pieces with a subsequent conversion of the double integrals into iterated integrals, one would have faced an *unsolvable* problem of finding the equations for the boundaries of D in the form $y = y_{\text{top}}(x)$ and $y = y_{\text{bot}}(x)$ or $x = x_{\text{top}}(y)$ and $x = x_{\text{bot}}(y)$!

SOLUTION: The region D is bounded by three curves, two spirals (polar graphs), and the line $y = 0$, $x > 0$. They are the images of the lines $r = \theta$, $r = 2\theta$, and $\theta = 2\pi$ in the polar plane as shown in the right panel of Fig. **32.4**. These lines form the boundary of D'. An algebraic description of D' as a vertically simple region is convenient to use:

$$D' = \{(r,\theta)\,|\,r_{\text{bot}}(\theta) = \theta \le r \le 2\theta = r_{\text{top}}(\theta)\,,\quad \theta_1 = 0 \le \theta \le 2\pi = \theta_2\}\,.$$

Hence,

$$A(D) = \iint_D dA = \iint_{D'} r\,dA' = \int_0^{2\pi}\int_\theta^{2\theta} r\,dr\,d\theta = \frac{3}{2}\int_0^{2\pi}\theta^2\,d\theta = 4\pi^3.$$

□

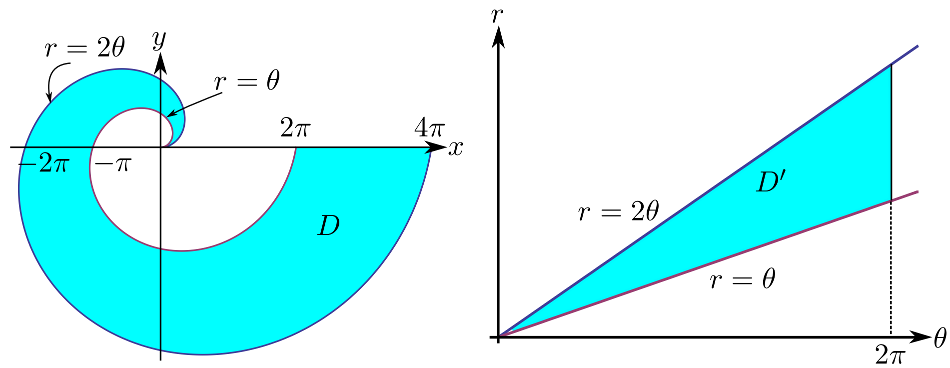


FIGURE **32.4**. An illustration to Example **32.4**. **Left**: The integration region D lies between two spirals. It is not simple in any direction. **Right**: The region D' in the polar plane whose image is D. The region D' is simple and is bounded by straight lines.

EXAMPLE **32.5**. *Find the volume of the part of the solid bounded by the cone $z = \sqrt{x^2+y^2}$ and the paraboloid $z = 2 - x^2 - y^2$ that lies in the first octant.*

SOLUTION: The solid is shown in the left panel of Fig. **32.5**. The intersection of the cone (bottom boundary) and paraboloid (top boundary),

$$z = z_{\text{top}}(x,y) = 2 - x^2 - y^2\,, \quad z = z_{\text{bot}}(x,y) = \sqrt{x^2+y^2}\,,$$

is a circle of unit radius. Indeed, put $r = \sqrt{x^2+y^2}$. Then the points of intersection satisfy the condition

$$z_{\text{top}} = z_{\text{bot}} \quad \Rightarrow \quad \sqrt{x^2+y^2} = 2 - x^2 - y^2 \quad \Rightarrow \quad r = 2 - r^2 \quad \Rightarrow \quad r = 1\,.$$

If the point (x, y, z) is in the solid, then the point $(x, y, 0)$ is said to lie in the projection D of the solid onto the xy plane along the z axis. Clearly, D is the part of the disk $r \le 1$ in the first quadrant. The volume in question is

$$V = \iint_D h(x,y)dA\,,$$

where $h(x, y)$ is the height of the solid at (x, y) in D. Since a line parallel to the z axis through a point (x, y) in D intersects the solid along a segment, the height of the solid at (x, y) is

$$h(x,y) = z_{\text{top}}(x,y) - z_{\text{bot}}(x,y) = 2 - x^2 - y^2 - \sqrt{x^2+y^2} = 2 - r^2 - r\,.$$

The height does not depend on the polar angle θ and the region D is bounded by the circle and two straight lines through the origin. Therefore, the conversion of the double integral to polar coordinates can simplify its evaluation. The region D is the image of the rectangle $D' = [0, 1] \times [0, \pi/2]$ in the polar

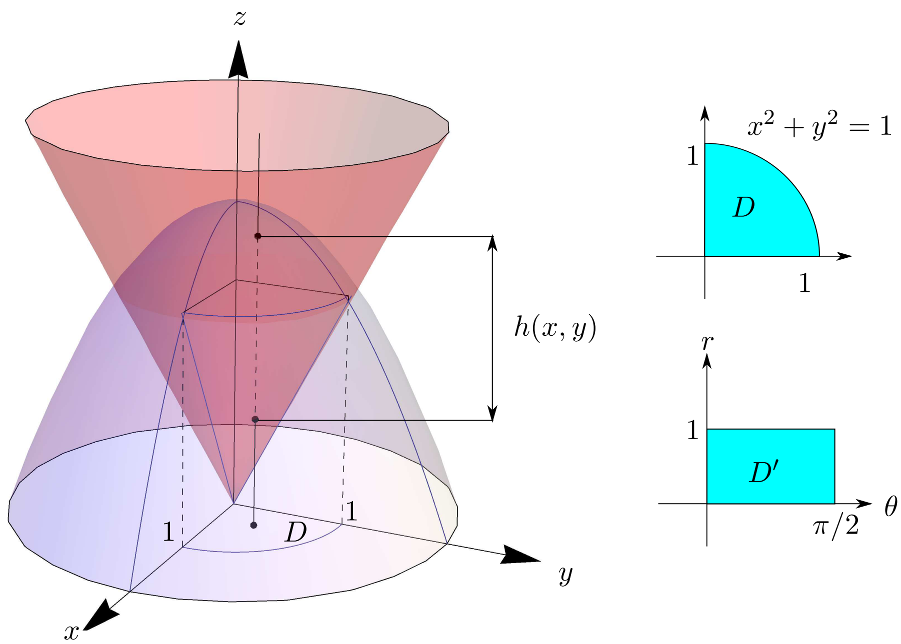


FIGURE **32.5**. An illustration to Example **32.5**. **Left**: The solid whose volume is sought. Its vertical projection onto the xy plane is D which is the part of the disk $r \le 1$ in the first quadrant. At a point (x, y) in D, the height $h(x, y)$ of the solid is the difference between the values of the z coordinate on the top and bottom boundaries (the paraboloid and the cone, respectively). **Right**: The region D' in the polar plane whose image is D .

plane. The volume is

$$
\begin{aligned}
V &= \iint_D h(x,y)\,dA = \iint_{D'} (2 - r^2 - r)r\,dA' \\
&= \int_0^{\pi/2} d\theta \int_0^1 (2r - r^3 - r^2)\,dr = \frac{\pi}{2}\left(1 - \frac{1}{4} - \frac{1}{3}\right) = \frac{5\pi}{24}.
\end{aligned}
$$

□

32.4. Study problems.

Problem **32.1**. *Find the area of the four-leaved rose bounded by the polar graph* $r = \cos(2\theta)$.

SOLUTION: The polar graph comes through the origin $r = 0$ four times when $\theta = \pi/4$, $\theta = \pi/4 + \pi/2$, $\theta = \pi/4 + \pi$, and $\theta = \pi/4 + 3\pi/2$. These angles may be changed by adding an integer multiple of π, owing to the periodicity of $\cos(2\theta)$. Therefore each leaf of the rose corresponds to the range of θ between two neighboring zeros of $\cos(2\theta)$. Since all leaves have the same area, it is sufficient to find the area of one leaf, say, for $-\pi/4 \le \theta \le \pi/4$. With this choice, the leaf is the image of the vertically simple region

$$D' = \{(r,\theta)|0 \le r \le \cos(2\theta),\ -\pi/4 \le \theta \le \pi/4\}$$

in the polar plane. Therefore, its area is given by the double integral

$$\begin{aligned} A(D) &= \iint_D dA = \iint_{D'} r\,dA' = \int_{-\pi/4}^{\pi/4}\int_0^{\cos(2\theta)} r\,dr\,d\theta \\ &= \frac{1}{2}\int_{-\pi/4}^{\pi/4}\cos^2(2\theta)d\theta = \frac{1}{4}\int_{-\pi/4}^{\pi/4}(1+\cos(4\theta))d\theta \\ &= \frac{1}{4}\Big(\theta + \frac{1}{4}\sin(4\theta)\Big)\Big|_{-\pi/4}^{\pi/4} = \frac{\pi}{8}\,. \end{aligned}$$

Thus, the total area is $4A(D) = \pi/2$. □

32.5. Exercises.

1-4. Sketch the region in the xy plane whose area is given by each of the following iterated integrals in polar coordinates and find the area of the region.

1. $\int_0^{\pi}\int_1^2 r\,dr\,d\theta\,;$
2. $\int_{-\pi/2}^{\pi/2}\int_0^{2a\cos\theta} r\,dr\,d\theta\,;$
3. $\int_{-\pi/4}^{\pi/4}\int_0^{1/\cos\theta} r\,dr\,d\theta\,;$
4. $\int_{-\pi}^{\pi}\int_0^{1+\cos\theta} r\,dr\,d\theta\,.$

5-8. Convert the double integral $\iint_D f(x,y)dA$ to an iterated integral in polar coordinates for each of the following regions.

5. D is the disk $x^2+y^2 \le a^2\,;$
6. D is the disk $x^2+y^2 \le ax$, $a>0\,;$
7. D is the ring $a^2 \le x^2+y^2 \le b^2\,;$
8. D is the parabolic segment $-a \le x \le a$, $x^2/a \le y \le a$, $a>0$.

9–13. Evaluate each of the following double integrals by changing to polar coordinates.

9. $\iint_D xy\,dA$ where D is the part of the ring $a^2 \le x^2+y^2 \le b^2$ in the first quadrant;
10. $\iint_D \sin(x^2+y^2)dA$ where D is the disk $x^2+y^2 \le a^2\,;$
11. $\iint_D \arctan(y/x)dA$ where D is the part of the ring $0 < a^2 \le x^2 + y^2 \le b^2$ between the lines $y=\sqrt{3}x$ and $y = x/\sqrt{3}$ in the first quadrant;
12. $\iint_D \ln(x^2+y^2)dA$ where D is the portion of the ring $0 < a^2 \le x^2+y^2 \le b^2$ between two half-lines $x = \pm y$, $y>0\,;$
13. $\iint_D \sin(\sqrt{x^2+y^2})\,dA$ where D is the ring $\pi^2 \le x^2+y^2 \le 4\pi^2\,.$

14–16. If r and θ are polar coordinates, reverse the order of integration in each of the following iterated integrals and sketch the integration region in the xy plane.

14. $\int_{-\pi/2}^{\pi/2}\int_0^{\cos\theta} f(r,\theta)rdrd\theta$;

15. $\int_0^{\pi/2}\int_0^{a\sqrt{\sin(2\theta)}} f(r,\theta)rdrd\theta$, $a>0$;

16. $\int_0^{a}\int_0^{\theta} f(r,\theta)rdrd\theta$, $0<a<2\pi$.

17–20. Sketch the region of integration and evaluate each of the following integrals by converting it to polar coordinates.

17. $\int_{-1}^{1}\int_0^{\sqrt{1-y^2}} e^{x^2+y^2}\,dxdy$;

18. $\int_{-1}^{0}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}(x+y)dydx$;

19. $\int_0^{2}\int_0^{\sqrt{2y-y^2}} \sqrt{x^2+y^2}dxdy$;

20. $\int_{1/\sqrt{2}}^{1}\int_{\sqrt{1-x^2}}^{x} xydydx+\int_1^{\sqrt{2}}\int_0^x xydydx+\int_{\sqrt{2}}^{2}\int_0^{\sqrt{4-x^2}} xydydx$.

21–26. Convert each of the following integrals in rectangular coordinates to an iterated integral in polar coordinates.

21. $\int_0^2\int_x^{x\sqrt{3}} f(\sqrt{x^2+y^2})\,dydx$;

22. $\int_0^1\int_0^{x^2} f(x,y)\,dydx$;

23. $\iint_D f(\sqrt{x^2+y^2})dA$ where D is the disk $x^2+y^2\le 1$;

24. $\iint_D f(\sqrt{x^2+y^2})dA$ where $D=\{(x,y)||y|\le|x|,|x|\le 1\}$;

25. $\iint_D f(y/x)dA$ where D is the disk $x^2+y^2\le x$;

26. $\iint_D f(\sqrt{x^2+y^2})dA$ where D is bounded by the curve $(x^2+y^2)^2=a^2(x^2-y^2)$.

27–33. Find the area of the specified region D.

27. D is enclosed by the polar graph $r=1+\cos\theta$;

28. D is bounded by two spirals $r=\theta/4$ and $r=\theta/2$, where $0\le\theta\le 2\pi$, and the positive x axis;

29. D is the part of the region enclosed by the cardioid $r=1+\sin\theta$ that lies outside the disk $x^2+y^2\le 9/4$;

30. D is bounded by the curve $(x^2+y^2)^2=2a^2(x^2-y^2)$ and $x^2+y^2\ge a^2$ if (x,y) is in D;

31. D is bounded by the curve $(x^3+y^3)^2=x^2+y^2$ and lies in the first quadrant;

32. D is bounded by the curve $(x^2+y^2)^2=a(x^3-3xy^2)$, $a>0$;

33. D is bounded by the curve $(x^2+y^2)^2=8a^2xy$ and $(x-a)^2+(y-a)^2\le a^2$, $a>0$, if (x,y) is in D.

34–38. Find the volume of the specified solid E.

34. E is bounded by the cones $z=3\sqrt{x^2+y^2}$ and $z=4-\sqrt{x^2+y^2}$;

35. E is bounded by the cone $z=\sqrt{x^2+y^2}$, the plane $z=0$, and the cylinders $x^2+y^2=1$, $x^2+y^2=4$;

36. E is bounded by the paraboloid $z = 1 - x^2 - y^2$ and the plane $z = -3$;

37. E is bounded by the hyperboloid $x^2 + y^2 - z^2 = -1$ and the plane $z = 2$;

38. E lies under the paraboloid $z = x^2 + y^2$, above the xy plane, and inside the cylinder $x^2 + y^2 = 2x$.

39. Find

$$\lim_{a\to 0} \frac{1}{\pi a^2} \iint_D f(x,y)dA, \quad D = \{(x,y)\,|\,x^2 + y^2 \le a^2\}$$

if f is a continuous function.

33. Change of Variables in Double Integrals

With an example of polar coordinates, it is quite clear that a smart choice of integration variables can significantly simplify the technicalities involved when evaluating double integrals. The simplification is twofold: simplifying the shape of the integration region (a rectangular shape is most desirable) and finding antiderivatives when calculating the iterated integral. It is therefore of interest to develop a technique for a general change of variables in double integrals so that one would be able to *design* new variables specific to the double integral in question in which the sought-for simplification can be achieved.

33.1. Change of Variables. Let the functions $x(u, v)$ and $y(u, v)$ be defined on an open region D'. Then, for every pair (u, v) in D', one can find a pair (x, y), where $x = x(u, v)$ and $y = y(u, v)$. All such pairs form a region in the xy plane that is denoted D. In other words, the functions $x(u, v)$ and $y(u, v)$ define a *transformation* T of a region D' in the uv plane onto a region D in the xy plane:

$$T: \quad D' \rightarrow D\,; \qquad T: \quad (u, v) \rightarrow (x,\, y) = (x(u, v),\, y(u, v))\,.$$

If no two points in D' have the same image point in D, then the transformation is called *one-to-one*. For a one-to-one transformation, one can define the *inverse transformation* T^{-1}, that is, the functions $u(x, y)$ and $v(x, y)$ that assign a pair (u, v) in D' to a pair (x, y) in D, where $u = u(x, y)$ and $v = v(x, y)$:

$$T^{-1}: \quad D \rightarrow D'\,; \qquad T^{-1}: \quad (x, y) \rightarrow (u,\, v) = (u(x, y),\, v(x, y))\,.$$

Owing to this one-to-one correspondence between rectangular coordinates (x, y) and pairs (u, v), one can describe points in a plane by *new coordinates* (u, v). For example, the relations $x = x(r, \theta) = r\cos\theta$ and $y = y(r, \theta) = r\sin\theta$ define polar coordinates. In any open set D' of pairs (r, θ) that lie within the half-strip $[0, \infty) \times [0, 2\pi)$, the transformation is one-to-one. The corresponding inverse functions $r = r(x, y)$ and $\theta = \theta(x, y)$ have been found in the previous section.

Definition **33.1**. (Change of Variables in a Plane).
A one-to-one transformation of an open region D' defined by $x = x(u, v)$ and $y = y(u, v)$ is called a change of variables *if the functions $x(u, v)$ and $y(u, v)$ have continuous first-order partial derivatives on D'.*

The pairs (u, v) are often called *curvilinear coordinates*. Recall that, in a rectangular coordinate system, a point of a plane can be described as a point of intersection of two coordinate lines $x = x_p$ and $y = y_p$. The point (x_p, y_p) in D is a unique image of a point (u_p, v_p) in D'. Consider the inverse transformation $u = u(x, y)$ and $v = v(x, y)$. Since $u(x_p, y_p) = u_p$ and $v(x_p, y_p) = v_p$, the point (x_p, y_p) in D can be viewed as the point of intersection of two curves $u(x, y) = u_p$ and $v(x, y) = v_p$. The curves $u(x, y) = u_p$

and $v(x,y) = v_p$ are called *coordinate curves* of the new coordinates u and v; that is, the coordinate u has a fixed value along its coordinate curve $u(x,y) = u_p$, and, similarly, the coordinate v has a fixed value along its coordinate curve $v(x,y) = v_p$. The coordinate curves are images of the straight lines $u = u_p$ and $v = v_p$ in D' under the transformation:

$$\begin{array}{lllll} T: & \text{coordinate lines in } D' & \rightarrow & \text{coordinate curves in } D \\ T: & u = u_p & \rightarrow & u(x,y) = u_p \\ T: & v = v_p & \rightarrow & v(x,y) = v_p \end{array}$$

If the coordinate curves are not straight lines (as in a rectangular coordinate system), then such coordinates are naturally curvilinear. A collection of level curves of the functions $u(x,y)$ and $v(x,y)$ is called a *coordinate grid* of curvilinear coordinates (u,v), just like a rectangular coordinate grid in a plane. The coordinate curves through a point (x_p, y_p) can also be defined as parametric curves:

$$u(x,y) = u_p \quad \Leftrightarrow \quad \begin{cases} x = x(u_p, v) \\ y = y(u_p, v) \end{cases}$$

$$v(x,y) = v_p \quad \Leftrightarrow \quad \begin{cases} x = x(u, v_p) \\ y = y(u, v_p) \end{cases}$$

For example, if the radial variable is fixed, $r = r_p$, in $x = r\cos\theta$, $y = r\sin\theta$, then parametric equations of the circle are obtained:

$$r(x,y) = \sqrt{x^2+y^2} = r_p \quad \Leftrightarrow \quad \begin{cases} x = r_p\cos\theta \\ y = r_p\sin\theta \end{cases}, \quad 0 \le \theta \le 2\pi\,.$$

Similarly, by fixing $\theta = \theta_p$, parametric equations of a ray are obtained:

$$\theta = \theta_p \quad \Leftrightarrow \quad \begin{cases} x = r\cos\theta_p \\ y = r\sin\theta_p \end{cases}, \quad 0 \le r < \infty\,.$$

A collection of concentric circles and rays from the origin is a coordinate grid of polar coordinates.

33.2. A criterion for a transformation to be a change of variables. Suppose that the functions $x(u,v)$ and $y(u,v)$ have continuous partial derivatives. How can one verify whether the transformation $x = x(u,v)$, $y = y(u,v)$ is one-to-one and, hence, defines a change of variables? Of course, one might just try to find the inverse transformation but the latter implies solving a system of non-linear equations, which might be a technically formidable problem. It turns out that the question can be answered by much simpler means.

Definition **33.2**. (Jacobian of a Transformation).
The Jacobian of a transformation defined by differentiable functions $x = x(u,v)$ and $y = y(u,v)$ is

$$\frac{\partial(x,y)}{\partial(u,v)} = \det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{pmatrix} = x'_u y'_v - x'_v y'_u\,.$$

The left side of this relation is a convenient notation of the Jacobian. The matrix whose determinant is evaluated has the *first* row composed of the partial derivatives of the variables in the numerator with respect to the *first variable* in the denominator, while the *second* row contains the partial derivatives of the variables in the numerator with respect to the *second variable* in the denominator. This rule is easy to remember.

The Jacobian of the polar coordinates

$$J = \frac{\partial(x, y)}{\partial(r, \theta)} = x'_r y'_\theta - x'_\theta y'_r = r(\cos^2\theta + \sin^2\theta) = r$$

vanishes at $r = 0$ and the transformation $x = r\cos\theta$, $y = r\sin\theta$ is not one-to-one on the line $r = 0$ of the polar plane. But in a neighborhood of any point (r_0, θ_0), $r_0 \neq 0$, where the Jacobian is not zero, the transformation $x = r\cos\theta$, $y = r\sin\theta$ is one-to-one and has the inverse constructed in Section **32.1**.

Example **33.1**. *Find the Jacobian of the transformation $x = u(1-v)$, $y = uv$. Find all zeros of the Jacobian and compare them with the set of points at which the transformation is not one-to-one.*

Solution: The Jacobian of the transformation is

$$\frac{\partial(x, y)}{\partial(u, v)} = \det\begin{pmatrix} x'_u & y'_u \\ x'_v & y'_v \end{pmatrix} = \det\begin{pmatrix} 1-v & v \\ -u & u \end{pmatrix} = u(1-v) + uv = u$$

The zero of the Jacobian form a line $u = 0$ (the v axis). The transformation is not one-to-one on the line $u = 0$. Indeed, if $u \neq 0$, then the inverse transformation is not difficult to find

$$u = u(x, y) = x + y\,, \quad v = v(x, y) = \frac{y}{x+y}\,.$$

For example, the first relation follow from $x/u = 1 - v$ and $y/u = v$ and, hence, $x/u = 1 - y/u$ (which is valid if $u \neq 0$), while the second relation is obtained by substituting the first one into $v = y/u$. Note that $u(x, y) = 0$ on the line $y = -x$ and $v(x, y)$ does not exist. Thus, the set of zeros of the Jacobian coincides with the set on which the transformation is not one-to-one. □

That a transformation is one-to-one and hence has the inverse on the set where the Jacobian does not vanish is not coincidental and specific to the two considered examples. It is, in fact, true in general.

Theorem **33.1**. (Inverse Function Theorem for Two Variables) *Let the transformation $(u, v) \to (x, y)$ be defined on an open set U' containing a point (u_0, v_0). Suppose that the functions $x(u, v)$ and $y(u, v)$ have continuous partial derivatives in U' and the Jacobian of the transformation does not vanish at the point (u_0, v_0). Then there exists an inverse transformation $u = u(x, y)$, $v = v(x, y)$ in an open set U containing the image point $(x_0, y_0) = (x(u_0, v_0), y(u_0, v_0))$ and the functions $u(x, y)$ and $v(x, y)$ have continuous partial derivatives in U.*

Let $P_0 = (x_0, y_0)$ be the image of $P_0' = (u_0, v_0)$. Consider two coordinate curves through the point P_0:

$$\begin{aligned} C_1 : &\quad x = x(u, v_0)\,,\ y = y(u, v_0)\,, \quad u_1 < u < u_2\,, \quad u_0 \in (u_1, u_2)\,,\\ C_2 : &\quad x = x(u_0, v)\,,\ y = y(u_0, v)\,, \quad v_1 < v < v_2\,, \quad v_0 \in (v_1, v_2)\,. \end{aligned}$$

The curve C_1 is the coordinate curve of the variable u. It can also be viewed as the level curve $v(x, y) = v_0$. Similarly, the curve C_2 is the coordinate curve of the variable v and can also be represented as the level curve $u(x, u) = u_0$. The curves are intersecting at the point $P_0 = (x_0, y_0)$. The tangent vectors to the curves at the point of intersection are

$$\mathbf{T}_1 = \langle x_u'(u_0, v_0)\,,\ y_u'(u_0, v_0)\rangle\,, \quad \mathbf{T}_2 = \langle x_v'(u_0, v_0)\,,\ y_v'(u_0, v_0)\rangle\,.$$

Without loss of generality, suppose that $J(u_0, v_0) > 0$. Then it is not difficult to see that

$$J(u_0, v_0) = \|\mathbf{T}_1 \times \mathbf{T}_2\| \neq 0$$

Therefore the coordinate curves are intersecting at P_0 at some non-zero angle (because the tangent vectors $\mathbf{T}_1$ and $\mathbf{T}_2$ are not parallel). The partial derivatives of $x(u, v)$ and $y(u, v)$ are continuous near (u_0, v_0) and so must be the Jacobian $J(u, v)$. Since the Jacobian is continuous and does not vanish at (u_0, v_0), it does not vanish in a neighborhood of (u_0, v_0). This implies that coordinate curves through any other point (x_p, y_p) in a neighborhood of (x_0, y_0) are also intersecting at a non-zero angle. Thus, is a neighborhood of (x_0, y_0), the coordinate curves of the new variables look like a continuously deformed rectangular grid; each small "parallelogram" bounded by *four* neighboring coordinate curves. Each point near (x_0, y_0) is *uniquely* represented as the point of intersection of two coordinate curves and corresponds to the point of intersection of two intersecting perpendicular lines in the uv plane. So, the transformation is one-to-one.

33.3. Change of Variables in a Double Integral. Let f be an integrable function on a region D. Let $x = x(u, v)$ and $y = y(u, v)$ define a transformation of a region D' to D. Suppose that the transformation is defined on a rectangle $R_D' = [a, b] \times [c, d]$ and is a change of variables in its interior $(a, b) \times (c, d)$ which contains the region D'. Then there is an inverse transformation, that is, the transformation of D to D', which is defined by the functions $u = u(x, y)$ and $v = v(x, y)$. Therefore D is contained in the image R_D of the rectangle R_D':

$$T : \quad R_D' \subset D' \ \to \ R_D \subset D\,.$$

The function f is extended to R_D by setting its values to 0 for all points that are not in D. According to (**29.2**), the double integral of f over D is the limit of Riemann sums. The limit depends neither on a partition of D by area elements nor on sample points in the partition elements. Following the analogy with polar coordinates, consider a partition of D (or R_D) by coordinate curves

$$u(x,y)=u_i\,,\quad i=0,1,2,...,N_1\,,\quad v(x,y)=v_j\,,\quad j=0,1,2,...,N_2\,,$$

where

$$\begin{aligned} u_0 &= a\,, \quad u_i = u_{i-1}+\Delta u\,, \quad \Delta u = (b-a)/N_1\,,\\ v_0 &= c\,, \quad v_j = v_{j-1}+\Delta v\,, \quad \Delta v = (d-c)/N_2\,. \end{aligned}$$

This partition of D is induced by a rectangular partition of D' by horizontal lines $v=v_k$ and vertical lines $u=u_k$ in the uv plane. Each partition element D'_{ij} of D' has the area

$$\Delta A' = \Delta u\,\Delta v\,.$$

The image of D'_{ij} is a partition element D_{ij} of D (see Figure **33.1**). If (u_i^*, v_j^*) is a sample point in D'_{ij}, then the corresponding sample point in D_{ij} is $\mathbf{r}^*_{ij}=(x(u_i^*,v_j^*),\, y(u_i^*,v_j^*))$, and (**29.2**) becomes

$$\iint_D f\,dA = \lim_{N_1,N_2\to\infty}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} f(\mathbf{r}^*_{ij})\,\Delta A_{ij},$$

where ΔA_{ij} is the area of the partition element D_{ij}. The limit N_1, $N_2\to\infty$ is understood in the sense of a double limit $(\Delta u,\Delta v)\to(0,0)$. As before, the values of $f(x(u,v),y(u,v))$ outside D' are set to 0 when calculating the value of f in a partition rectangle that intersects the boundary of D'.

As in the case of polar coordinates, the aim is to convert this limit into a double integral of $f(x(u,v),y(u,v))$ over the region D'. This can be accomplished by finding a relation between ΔA_{ij} and $\Delta A'_{ij}$, that is, the rule of the area element transformation under a change of variables. Consider a rectangle D'_{ij} in the uv plane bounded by the lines

$$u=u_{i-1}\,,\quad u=u_i\,,\quad v=v_{j-1}\,,\quad v=v_j\,.$$

Let us mark three vertices of the partition rectangle D'_{ij} (see Fig. **33.1**):

$$A'=(u_{i-1},v_{j-1})\,,\quad B'=(u_i+\Delta u, v_{j-1})\,,\quad C'=(u_{i-1},v_{j-1}+\Delta v)\,,$$

and their images under the transformation of D'_{ij} to D_{ij}:

$$A=(x(A'),y(A'))\,,\quad B=(x(B'),y(B'))\,,\quad C=(x(C'),y(C'))\,.$$

Since the transformation $(u,v)\to(x,y)$ is a change of variables (see Definition **33.1**), the functions $x(u,v)$ and $y(u,v)$ have continuous partial derivatives and, hence, differentiable. Therefore their variations in D'_{ij} can be well approximated by their linearization (recall Definition **21.1**). The distances $|A'B'|=\Delta u$ and $|A'C'|=\Delta v$ are small. So, when calculating the area ΔA of D_{ij}, it is sufficient to consider variations of x and y within D_{ij} *linear* in variations of u and v within D'_{ij}. In the limit $(\Delta u,\Delta v)\to(0,0)$, their higher powers can be neglected (as they would not contribute to the limit of the Riemann sum), and the area transformation law should have the form

$$\Delta A = J\,\Delta u\,\Delta v = J\,\Delta A',$$

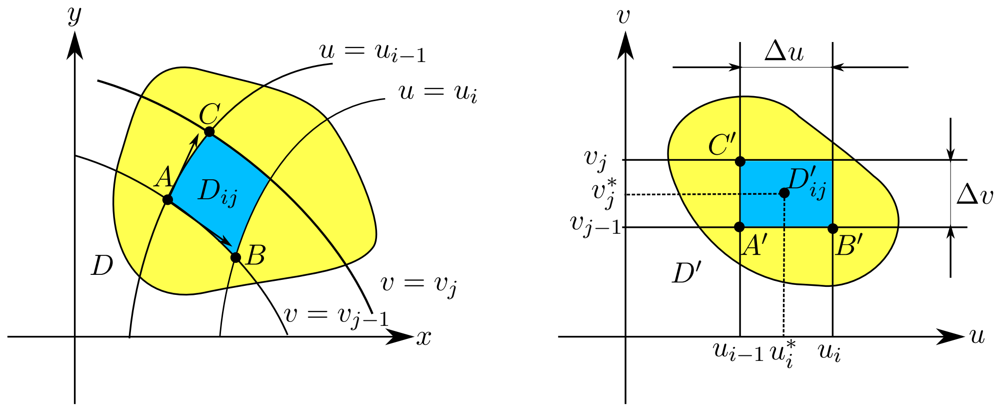


FIGURE **33.1**. **Left**: A partition of a region D by the coordinate curves of the new variables $u(x, y) = u_i$ and $v(x, y) = v_j$ which are the images of the straight lines $u = u_i$ and $v = v_j$ in the uv plane. A partition element D_{ij} is bounded by the coordinate curves for which $u_i - u_{i-1} = \Delta u$ and $v_j - v_{j-1} = \Delta v$. **Right**: The region D', whose image is the integration region D under the coordinate transformation, is partitioned by the coordinate lines $u = u_i$ and $v = v_j$. A partition element is the rectangle D'_{ij} whose area is $\Delta u \Delta v$. The change of variables establishes a one-to-one correspondence between points of D and D'. In particular, A, B, and C in D correspond to A', B', and C' in D', respectively.

where the coefficient J is to be found. Recall that $J = r$ for polar coordinates, that is, J coincides with the Jacobian of the transformation. It turns out that J is the absolute value of the Jacobian for a general change of variables.

To find J, consider the coordinate curve $x = x(u, v_{j-1})$, $y = y(u, v_{j-1})$, where $u_{i-1} \leq u \leq u_i$, that goes from A to B, and the coordinate curve $x = x(u_{i-1}, v)$, $y = y(u_{i-1}, v)$, where $v_{j-1} \leq v \leq v_i$, that connects A and C. Owing to differentiability of $x(u, v)$ and $y(u, v)$, the curves are smooth. Therefore the arclength from A to B and from A to C along the corresponding coordinate curves are well approximated by the length of the corresponding secant line segments, $\|\overrightarrow{AB}\|$ and $\|\overrightarrow{AC}\|$, respectively (recall Section **13**). The error of the approximation decreases to zero faster than Δu and Δv. This suggests that the area of D_{ij} can be approximated by the area of the parallelogram with adjacent sides being the vectors **b** and **c**:

$$\Delta A_{ij} = \|\overrightarrow{AB} \times \overrightarrow{AC}\| \,.$$

The error of the approximation should be decreasing faster than $\Delta u \Delta v$ in the limit $(\Delta u, \Delta v) \to (0, 0)$. Linearizing the functions $x(u, v)$ and $y(u, v)$ at the point A', one infers that

$$\begin{aligned}
\overrightarrow{AB} &= \langle x(B') - x(A')\,,\ y(B') - y(A')\,,\ 0\rangle \\
x(B') - x(A') &= x(u_{i-1} + \Delta u, v_{j-1}) - x(u_{i-1}, v_{j-1}) = x'_u(u_{i-1}, v_{j-1})\,\Delta u \\
y(B') - y(A') &= y(u_{i-1} + \Delta u, v_{j-1}) - y(u_{i-1}, v_{j-1}) = y'_u(u_{i-1}, v_{j-1})\,\Delta u \\
\overrightarrow{AB} &= \Delta u \left\langle x'_u(u_{i-1}, v_{j-1}),\ y'_u(u_{i-1}, v_{j-1}),\ 0\right\rangle .
\end{aligned}$$

The third component of $\overrightarrow{AB}$ is set to 0 as the vector is planar. An analogous calculation of the components of $\overrightarrow{AC}$ yields

$$\begin{aligned}
\overrightarrow{AC} &= \langle x(C') - x(A')\,,\ y(C') - y(A')\,,\ 0\rangle \\
&= \Delta v \left\langle x'_v(u_{i-1}, v_{j-1}),\ y'_v(u_{i-1}, v_{j-1}),\ 0\right\rangle .
\end{aligned}$$

The cross product of the vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$ in the xy plane is parallel to the z axis so that that the area of the parallelogram is equal to the absolute value of the z component of the cross product:

$$\Delta A_{ij} = \left|\det \begin{pmatrix} x'_u & y'_u \\ x'_v & y'_v \end{pmatrix}\right| \Delta u\,\Delta v = J(u_{i-1}, v_{j-1})\,\Delta u\,\Delta v\,. \tag{33.1}$$

The absolute value is needed because the z component of the cross product may be negative, $\|(0,0,z)\| = \sqrt{z^2} = |z|$. The determinant that appears in Eq. (**33**.**1**) plays a significant role in the theory of transformations. So it has a special name.

Furthermore, the coefficient J in (**33**.**1**) is the *absolute value* of the Jacobian. If the partial derivatives of x and y with respect to u and v are continuous on D', J is continuous on D', too. Therefore, for any sample point (u_i^*, v_j^*) in D'_{ij}, the difference

$$\frac{\Delta A_{ij} - J(u_i^*, v_j^*)\Delta A'}{\Delta A'} = J(u_{i-1}, v_{j-1}) - J(u_i^*, v_j^*)$$

vanishes in the limit $(\Delta u, \Delta v) \to (0,0)$. So, if in (**33**.**1**) the value of the Jacobian is taken at a sample point other than A', then the corresponding change in the value of ΔA_{ij} should decrease to zero faster than $\Delta u \Delta v$ in the limit $(\Delta u, \Delta v) \to (0,0)$. Thus, with the same accuracy used in the approximation of ΔA_{ij}, one can always put

$$\Delta A_{ij} = J(u_i^*, v_j^*)\,\Delta u\,\Delta v$$

in the Riemann sum for any choice of sample points. The limit of the Riemann sum

$$\iint_D f\,dA = \lim_{N_1, N_2 \to \infty} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} f(x(u_i^*, v_j^*), y(u_i^*, v_j^*))\,\Delta A_{ij},$$

defines the double integral of the function $f(x(u,v), y(u,v))J(u,v)$ over the region D', provided that the latter function is integrable on D'. One can prove that *if f is integrable on D and $x = x(u,v)$, $y = y(u,v)$ is a change of* *variables, then $f(x(u,v), y(u,v))J(u,v)$ is integrable on D'.* The foregoing arguments suggest that the following theorem is true (a full proof is given in advanced calculus courses).

THEOREM **33.2**. (Change of Variables in a Double Integral). *Suppose a transformation $x = x(u,v)$, $y = y(u,v)$ has continuous first-order partial derivatives and maps a region D' bounded by piecewise-smooth curves onto a region D. Suppose that this transformation is one-to-one and has a nonvanishing Jacobian, except perhaps on the boundary of D'. Then*

$$\iint_D f(x,y)\,dA = \iint_{D'} f(x(u,v), y(u,v))J(u,v)\,dA',$$

$$J(u,v) = \left|\frac{\partial(x,y)}{\partial(u,v)}\right|.$$

In the case of polar coordinates, the boundary of D' may contain the line $r = 0$ on which the Jacobian $J = r$ vanishes. This entire line collapses into a single point, the origin $(x,y) = (0,0)$ in the xy plane, upon the transformation $x = r\cos\theta$ and $y = r\sin\theta$; that is, this transformation is not one-to-one on this line. A full proof of the theorem requires an analysis of such subtleties in a general change of variables as well as a rigorous justification of the linear approximation in the area transformation law, which were excluded in the above analysis.

In practice, the change of variables in a double integral entails the following steps:

Step 1. Finding the region D' whose image under the transformation $x = x(u,v)$, $y = y(u,v)$ is the region of integration D. A useful rule to remember here is:

$$T: \text{ boundary of } D' \longrightarrow \text{boundary of } D$$

under the transformation T. In particular, if equations of boundaries of D are given, then equations of the corresponding boundaries of D' can be obtained by expressing the former in terms of the new variables by the substitution $x = x(u,v)$ and $y = y(u,v)$.

Step 2. Transformation of the function to new variables

$$f(x,y) = f(x(u,v), y(u,v)).$$

Step 3. Calculation of the Jacobian that defines the area element transformation:

$$dA = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv = J\,dA', \qquad J = \left|\frac{\partial(x,y)}{\partial(u,v)}\right|.$$

Step 4. Evaluation of the double integral of fJ over D' by converting it to a suitable iterated integral. The choice of new variables should be motivated by simplifying the shape of D' (a rectangular shape is the most desirable).

EXAMPLE **33.2**. *Use the change of variables $x = u(1-v)$, $y = uv$ (see Example* **33.1***) to evaluate the integral $\iint_D (x+y)^5 y^5 dA$ where D is the triangle bounded by the lines $y = 0$, $x = 0$, and $x + y = 1$.*

SOLUTION: Step 1. Note first that the line $u = 0$ is mapped to a single point, the origin, in the xy plane. So a part of the line $u = 0$ must be in the boundary of D'. The equation $x = 0$ in the new variables becomes $u(1-v) = 0$ which means that either $u = 0$ or $v = 1$. Therefore a part of the line $v = 1$ is in the boundary of D' as it is mapped to the boundary line $x = 0$. The equation $y = 0$ in the new variables reads $uv = 0$. Therefore a part of the line $v = 0$ is also in the boundary of D'. The equation $x + y = 1$ in the new variables has the form $u = 1$. So,

$$\begin{aligned}
T: &\quad u = 0 \ \text{ or } \ v = 0 \ \rightarrow \ y = 0\,,\\
T: &\quad u = 0 \ \text{ or } \ v = 1 \ \rightarrow \ x = 0\,,\\
T: &\quad u = 1 \ \rightarrow \ x + y = 1
\end{aligned}$$

This suggests that D' is bounded by four lines $u = 0$, $u = 1$, $v = 0$, and $v = 1$ because the boundary of D' is mapped onto the boundary of D. Therefore D' is the square

$$D' = [0,1] \times [0,1]\,.$$

Note that the transformation is not one-to-one on the line $u = 0$ because the line $u = 0$ is mapped to a single point, the origin.

Step 2. Since $x + y = u$, the integrand in the new variables is

$$f(x,y) = (x+y)^5 y^5 = u^5 (uv)^5 = u^{10} v^5\,.$$

Step 3. By Example **33.1** the Jacobian of the transformation is $J = u$. Therefore the area element transformation is $dA = |u| dA'$. The absolute value may be omitted because $u \geq 0$ in D'. Note that the Jacobian vanishes only on the boundary of D' and, hence, the hypotheses of Theorem **33.2** are fulfilled.

Step 4. The double integral in the new variables is evaluated by Fubini's theorem:

$$\iint_D (x+y)^5 y^5 dA = \iint_{D'} u^{11} v^5 dA' = \int_0^1 u^{11} du \int_0^1 v^5 dv = \frac{1}{12} \cdot \frac{1}{6} = \frac{1}{72}$$

□

Let D be a region of integration. Suppose new variables are defined by a transformation.

If the Jacobian of the transformation does not vanish in the interior of D, then by Theorem **33.1** *the transformation defines a genuine change of variables in the double integral and the conclusion of Theorem* **33.2** *holds.*

Note that zeros of the Jacobian on the boundary of the region of integration do not affect the conclusion of Theorem **33.2**. Furthermore, suppose that the Jacobian vanishes at a single point of the interior of D. For example, let D be a disk centered at the origin. The Jacobian of polar coordinates vanishes at the origin. Does the conclusion of Theorem **33.2** hold in this case? Let us cut the disk along its diameter and represent a double integral over the disk as the sum of integrals over two half-disks. The conclusion of the Theorem **33.2** holds for each of the two integrals because the zero of the Jacobian lies on the boundary of each half-disk. Therefore it holds for the whole disk. This observation can be generalized:

If zeros of the Jacobian lie on a smooth curve in the region of integration, then the conclusion of Theorem **33.2** *still holds.*

Indeed, the region of integration can be cut into two regions along a smooth curve that contain zeros of the Jacobian. The Jacobian does not vanish in the interiors of two new regions of integration and, hence, the conclusion of Theorem **33.2** holds for them. Evidently, the procedure may be repeated for finitely many smooth curves on which the Jacobian has zeros.

33.4. Jacobian of the inverse transformation. By Theorem **33.1**, the Jacobian of the *inverse transformation* can be calculated as $\partial(u,v)/\partial(x,y)$ so that the area transformation law is

$$dudv = \left|\frac{\partial(u,v)}{\partial(x,y)}\right| dxdy$$

and the following statement holds:

COROLLARY **33.1**. *If $u = u(x,y)$ and $v = v(x,y)$ is the inverse of the transformation $x = x(u,v)$ and $y = y(u,v)$, then*

$$\frac{\partial(x,y)}{\partial(u,v)} = \frac{1}{\frac{\partial(u,v)}{\partial(x,y)}} = \frac{1}{\det\begin{pmatrix} u'_x & v'_x \\ u'_y & v'_y \end{pmatrix}}. \qquad (\mathbf{33.2})$$

The analogy with a change of variables in the one-dimensional case can be made. If $x = f(u)$ where f has continuous derivative $f'(u)$ that does not vanish, then by the inverse function theorem for functions of one variable (Theorem **13.2**) there is an inverse function $u = g(x)$ whose derivative is continuous and $g'(x) = 1/f'(u)$ where $u = g(x)$. Then the transformation of the differential dx can be written in two equivalent forms, just like the transformation of the area element $dA = dxdy$:

$$dx = f'(u)du = \frac{du}{g'(x)} \quad \longleftrightarrow \quad dxdy = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| dudv = \frac{dudv}{\left|\frac{\partial(u,v)}{\partial(x,y)}\right|}.$$

Note the absence of the absolute value bars in the one-variable case. If $f'(u) > 0$, then $f(a') = a < b = f(b')$ if $a' < b'$. If $f'(u) < 0$, then

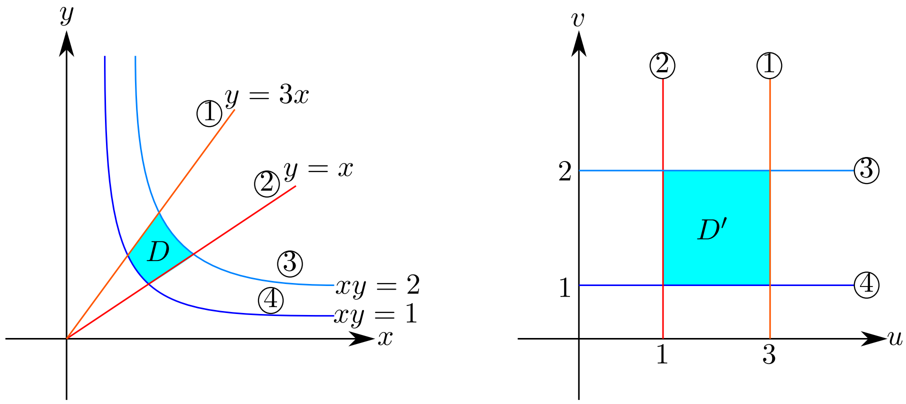


FIGURE **33.2**. An illustration to Example **33.3**. The transformation of the integration region D. Equations of the boundaries of D, $y = 3x$, $y = x$, $xy = 2$, and $xy = 1$, are written in the new variables $u = y/x$ and $v = xy$ to obtain the equations of the boundaries of D', $u = 3$, $u = 1$, $v = 2$, and $v = 1$, respectively. The correspondence between the boundaries of D and D' is indicated by encircled numbers enumerating the boundary curves.

$f(a') = a < b = f(b')$ if $a' > b'$. In the latter case,

$$\int_a^b h(x)dx = \int_{a'}^{b'} h(f(u))f'(u)du = -\int_{b'}^{a'} h(f(u))f'(u)du$$
$$= \int_{b'}^{a'} h(f(u))|f'(u)|du\,.$$

Therefore the full analogy with the two-variable case is achieved, that is, $dx = |f'(u)|du$ if one agrees that the new lower integration limit is always set to be smaller than the upper one. In other words, the length of a segment is always given by

$$b - a = \int_a^b dx = \int_{a'}^{b'} |f'(u)|du\,,$$

where $a' < b'$.

Equation (**33.2**) defines the Jacobian as a function of (x, y). Sometimes it is technically simpler to express the product $f(x, y)J(x, y)$ in the new variables rather than doing so for f and J separately. This is illustrated by the following example.

EXAMPLE **33.3**. *Use a suitable change of variables to evaluate the double integral of $f(x, y) = xy^3$ over the region D that lies in the first quadrant and is bounded by the lines $y = x$ and $y = 3x$ and by the hyperbolas $yx = 1$ and $yx = 2$.*

SOLUTION: The equations of the lines can be written in the form $y/x = 1$ and $y/x = 3$ because $y, x > 0$ in D (see Fig. **33.2**). Note that the equations of boundaries of D depend on just two particular combinations y/x and yx that take constant values on the boundaries of D. Consider the transformation T defined by the functions

$$u = u(x, y) = \frac{y}{x}\,, \quad v = v(x, y) = xy\,.$$

Under this transformation the boundary curves of D becomes the straight lines:

$$\begin{array}{llll} T: \; y \;\;= 3x & \to \quad u = 3\,; & T: \; y \;\;= x & \to \quad u = 1\,; \\ T: \; yx = 2 & \to \quad v = 2\,; & T: \; yx = 1 & \to \quad v = 1\,. \end{array}$$

This suggests that the image D' of D is the rectangle $[1, 3] \times [1, 2]$ in the uv plane. To verify that the defined transformation is a change of variables and, hence, can be used to simplify the region of integration, the Jacobian of the transformation should not vanish in the interior of D. By means of Eq. (**33**.**2**) the Jacobian as a function of (x, y) is obtained:

$$J = \left|\det\begin{pmatrix} u'_x & v'_x \\ u'_y & v'_y \end{pmatrix}\right|^{-1} = \left|\det\begin{pmatrix} -y/x^2 & y \\ 1/x & x \end{pmatrix}\right|^{-1} = \left|-\frac{2y}{x}\right|^{-1} = \frac{x}{2y}\,.$$

The absolute value bars may be omitted as x and y are strictly positive in D. Thus, $J \neq 0$ in D and the transformation is indeed a change of variables. Let us put aside for a moment the problem of expressing x and y as functions of new variables, which is needed to express f and J as functions of u and v, and find first the product fJ as a function of (x, y) and then express it in terms of the new variables (u, v):

$$f(x, y)J(x, y) = \frac{1}{2}x^2y^2 = \frac{1}{2}v^2\,.$$

So finding the functions $x = x(u, v)$ and $y = y(u, v)$ happens to be unnecessary in this example! Hence,

$$\iint_D xy^3\,dA = \frac{1}{2}\iint_{D'} v^2\,dA' = \frac{1}{2}\int_1^3 du \int_1^2 v^2\,dv = \frac{7}{3}\,.$$

The reader is advised to evaluate the double integral in the original rectangular coordinates to compare the amount of work needed with this solution. $\square$

The following example illustrates how a change of variables can be used to simplify the integrand of a double integral.

EXAMPLE **33**.**4**. *Evaluate the double integral of the function* $f(x, y) = \cos[(y - x)/(y + x)]$ *over the trapezoidal region with vertices* $(1, 0)$, $(2, 0)$, $(0, 1)$, *and* $(0, 2)$.

SOLUTION: An iterated integral in the rectangular coordinates would contain the integral of the cosine function of a rational argument (either with respect to x or y), which is difficult to evaluate. So a change of variables

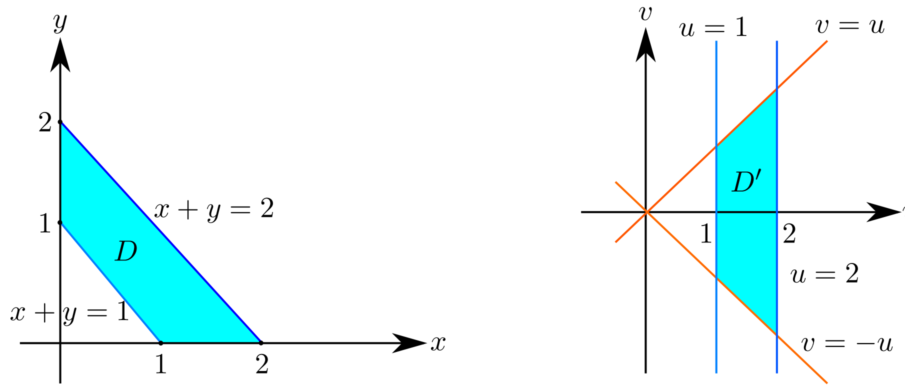


FIGURE **33.3**. **Left**: The integration region D in Example **33.4** is bounded by the lines $x+y=1$, $x+y=2$, $x=0$ and $y=0$. **Right**: The image D' of D under the change of variables $u=x+y$ and $v=y-x$. The boundaries of D' are obtained by substituting the new variables into the equations for boundaries of D so that $x+y=1 \to u=1$, $x+y=2 \to u=2$, $x=0 \to v=u$, and $y=0 \to v=-u$.

should be used to simplify the argument of the cosine function. The region D is bounded by the lines $x+y=1$, $x+y=2$, $x=0$, and $y=0$. Consider the transformation T defined by the functions

$$u = u(x,y) = x+y\,, \quad v = v(x,y) = y-x\,,$$

so that the function in the new variables becomes

$$f(x,y) = \cos\left(\frac{y-x}{y+x}\right) = \cos\left(\frac{v}{u}\right).$$

The transformation is a change of variables because the Jacobian

$$J = \frac{1}{\left|\frac{\partial(u,v)}{\partial(x,y)}\right|} = \frac{1}{\left|\det\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}\right|} = \frac{1}{2}$$

does not vanish anywhere. Under this transformation, the boundary of D is mapped onto the boundary of $D' = T(D)$:

$$\begin{array}{llllllll} T: & x+y=1 & \to & u=1\,; & T: & x+y=2 & \to & u=2\,;\\ T: & x=0 & \to & v=u\,; & T: & y=0 & \to & v=-u\,; \end{array}$$

The last two relations follow from the inverse transformation

$$x = \frac{1}{2}(u-v)\,, \quad y = \frac{1}{2}(u+v)\,,$$

so that the line $x=0$ is mapped onto the line $v=u$, while the line $y=0$ is mapped onto the line $v=-u$. Thus, the new integration region is

$$D' = \{(u,v)| -u \le v \le u,\ 1 \le u \le 2\}\,.$$

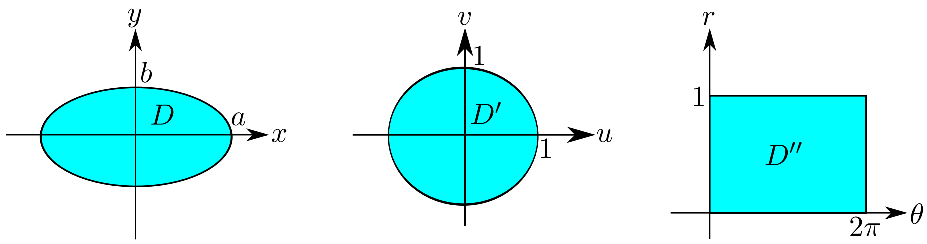


FIGURE **33.4**. The transformation of the integration region D in Example **33.5** (Area of an ellipse). The region D, $x^2/a^2 + y^2/b^2 \le 1$, is first transformed into the disk D', $u^2 + v^2 \le 1$, by $x = au$, $y = bv$, and then D' is transformed into the rectangle D'' by $u = r\cos\theta$, $v = r\sin\theta$.

Hence, using $dA = JdA' = \frac{1}{2}dA'$,

$$
\begin{aligned}
\iint_D \cos\Big(\frac{y-x}{y+x}\Big)dA &= \iint_{D'} \cos\Big(\frac{v}{u}\Big)JdA' = \frac{1}{2}\int_1^2\int_{-u}^{u}\cos\Big(\frac{v}{u}\Big)dv\,du \\
&= \frac{1}{2}\int_1^2 u\sin\Big(\frac{v}{u}\Big)\Big|_{-u}^{u}du \\
&= \sin(1)\int_1^2 udu = \frac{3}{2}\sin(1)\,.
\end{aligned}
$$

□

EXAMPLE **33.5**. (Area of an Ellipse).
Find the area of the region D bounded by the ellipse $x^2/a^2 + y^2/b^2 = 1$.

SOLUTION: Under the change of variables $u = x/a$, $v = y/b$, the ellipse is transformed into the circle $u^2 + v^2 = 1$ of unit radius. The Jacobian is

$$J = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \left|\det\begin{pmatrix} a & 0 \\ 0 & b\end{pmatrix}\right| = ab\,.$$

Therefore

$$A(D) = \iint_D dA = \iint_{D'} J\,dA' = ab\iint_{D'} dA' = abA(D') = \pi ab\,.$$

Of course the area $A(D')$ of the disk $u^2 + v^2 \le 1$ can also be evaluated by converting the integral over D' to polar coordinates $u = r\cos\theta$, $v = r\sin\theta$. The disk D' is the image of the rectangle $D'' = [0,1]\times[0,2\pi]$ and the Jacobian is r. The corresponding transformations of the region of integration are shown in Fig. **33.4**. □

When $a = b$, the ellipse becomes a circle of radius $R = a = b$, and the area of the ellipse becomes the area of the disk, $A = \pi R^2$.

33.5. Symmetries and a Change of Variables. Using the concept of a change of variables in a double integral one can give an algebraic criterion of area-preserving transformations introduced in Section **31.4** and prove Theorem **31.1**. *A transformation $x = x(u, v)$, $y = y(u, v)$ is said to be area preserving if the absolute value of its Jacobian is 1, that is, $dA = dA'$.* Indeed, since the Jacobian does not vanish, the transformation is a change of variables. For any closed bounded region D' whose image under such transformation is D,

$$A(D) = \iint_D dA = \iint_{D'} J dA' = \iint_{D'} dA' = A(D')\,.$$

For example, rotations, translations, and reflections are area-preserving transformations for obvious geometrical reasons (they preserve the distance between any two points of a region). The following theorem holds.

THEOREM **33.3**. (Symmetry of Double Integrals)
Suppose that an area-preserving transformation $x = x(u, v)$, $y = y(u, v)$ maps a region D onto itself. Suppose that a function f is skew-symmetric under this transformation, that is, $f(x(u, v), y(u, v)) = -f(u, v)$. Then the double integral of f over D vanishes.

PROOF. Since $D' = D$ and $dA = dA'$, the change of variables yields

$$\begin{aligned} I = \iint_D f(x, y)\, dA &= \iint_D f(x(u, v), y(u, v))\, dA' \\ &= -\iint_D f(u, v)\, dA' = -I, \end{aligned}$$

that is, $I = -I$, or $I = 0$. □

33.6. Study Problems.

Problem **33.1**. (Generalized Polar Coordinates)
Generalized polar coordinates are defined by the transformation

$$x = ar\cos^n\theta\,, \qquad y = br\sin^n\theta$$

where a, b, and n are parameters. Find the Jacobian of the transformation. Use the generalized polar coordinates with a suitable choice of parameters to find the area of the region in the first quadrant that is bounded by the curve $\sqrt[4]{x/a} + \sqrt[4]{y/b} = 1$.

SOLUTION: The Jacobian of the generalized polar coordinates is

$$\begin{aligned} \frac{\partial(x, y)}{\partial(r, \theta)} &= \det\begin{pmatrix} x'_r & y'_r \\ x'_\theta & y'_\theta \end{pmatrix} = \det\begin{pmatrix} a\cos^n\theta & b\sin^n\theta \\ -nar\sin\theta\cos^{n-1}\theta & nbr\cos\theta\sin^{n-1}\theta \end{pmatrix} \\ &= nabr\left(\cos^{n+1}\theta\sin^{n-1}\theta + \cos^{n-1}\theta\sin^{n+1}\theta\right) \\ &= nabr\cos^{n-1}\theta\sin^{n-1}\theta(\cos^2\theta + \sin^2\theta) \\ &= nabr\cos^{n-1}\theta\sin^{n-1}\theta \end{aligned}$$

Choosing the parameter $n = 8$, the equation of the curve $\sqrt[4]{x/a} + \sqrt[4]{y/b} = 1$ becomes $\sqrt[4]{r} = 1$ or $r = 1$. Since the region in question lies in the first quadrant, it is also bounded by the lines $y = 0$ and $x = 0$ which are the images of the lines $\theta = \pi/2$ and $\theta = 0$ in the (r, θ) plane. Therefore the rectangle $D' = [0, 1] \times [0, \pi/2]$ is mapped onto the region D in question. The Jacobian of the transformation is *positive* in the interior of D' so the absolute value of the Jacobian in the area element transformation may be omitted. The area of D is

$$\begin{aligned}
A(D) &= \iint_D dA = \iint_{D'} J dA' = 8ab \int_0^{\pi/2} \cos^7\theta \sin^7\theta d\theta \int_0^1 r dr \\
&= \frac{ab}{32} \int_0^{\pi/2} \Big(\sin(2\theta)\Big)^7 d\theta = -\frac{ab}{64} \int_0^{\pi/2} \Big(\sin(2\theta)\Big)^6 d\cos(2\theta) \\
&= \frac{ab}{64} \int_{-1}^1 \Big(1 - u^2\Big)^3 du = \frac{ab}{70}
\end{aligned}$$

where first the double angle formula $\cos\theta \sin\theta = \frac{1}{2}\sin(2\theta)$ has been used and then the integration has been carried out with the help of the substitution $u = \cos(2\theta)$. $\square$

33.7. Exercises.

1–4. Find the Jacobian of each of the following transformations.

1. $x = 3u - 2v$, $y = u + 3v$;

2. $x = e^r \cos\theta$, $y = e^r \sin\theta$;

3. $x = uv$, $y = u^2 - v^2$;

4. $x = u \cosh v$, $y = u \sinh v$.

5. Consider *hyperbolic coordinates* in the first quadrant $x > 0$, $y > 0$ defined by the transformation $x = ve^u$, $y = ve^{-u}$. Calculate the Jacobian. Determine the range of (u, v) in which the transformation is one-to-one. Find the inverse transformation and sketch coordinate curves of hyperbolic coordinates.

6. Find the conditions on the parameters of a linear transformation $x = a_1u+b_1v+c_1$, $y = a_2u+b_2v+c_2$ so that the transformation is area-preserving. In particular, prove that rotations discussed in Study Problem **1.2** are area-preserving.

7–9. Find the image D of the specified region D' under the given transformation.

7. $D' = [0, 1] \times [0, 1]$ and the transformation is $x = u$, $y = v(1 - u^2)$;

8. D' is the triangle with vertices $(0, 0)$, $(1, 0)$, and $(1, 1)$, and the transformation is $x = v^2$, $y = u$;

9. D' is the region defined by the inequality $|u| + |v| \leq 1$, and the transformation is $x = u + v$, $y = u - v$.

10. Find a linear transformation that maps the triangle D' with vertices $(0, 0)$, $(0, 1)$, and $(1, 0)$ onto the triangle D with vertices $(0, 0)$, (a, b), and

(b, a) where a and b are positive, non-equal numbers. Use this transformation to evaluate the integral of $f(x, y) = bx - ay$ over the triangle D.

11–14. Evaluate each of the following double integrals using the specified change of variables.

11. $\iint_D (8x+4y)dA$, where D is the parallelogram with vertices $(3, -1)$, $(-3, 1)$, $(-1, 3)$, and $(5, 1)$; the change of variables is $x = (v-3u)/4$, $y = (u+v)/4$;

12. $\iint_D (x^2 - xy + y^2)dA$, where D is the region bounded by the ellipse $x^2 - xy + y^2 = 1$; the change of variables is $x = u - v/\sqrt{3}$, $y = u + v/\sqrt{3}$;

13. $\iint_D (x^2 - y^2)^{-1/2}dA$, where D is in the first quadrant and bounded by hyperbolas $x^2 - y^2 = 1$, $x^2 - y^2 = 4$ and by the lines $x = 2y$, $x = 4y$; the change of variables is $x = u\cosh v$, $y = u\sinh v$;

14. $\int\int_D e^{(x/y)}\,(x + y)^3/y^2 dA$, where D is bounded by the lines $y = x, y = 2x, x + y = 1$ and $x + y = 2$; the change of variables $u = x/y, v = x + y$. Hint: Follow the procedure based on Eq. (**33**.**2**) as illustrated in Example **33**.**3**.

15. Find the image D' of the square $a < x < a + h$, $b < y < b + h$, where a, b, and h are positive numbers, under the transformation $u = y^2/x$, $v = \sqrt{xy}$. Find the ratio of the area $A(D')$ to the area $A(D)$. What is the limit of the ratio when $h \to 0$?

16–17. Use the specified change of variables to convert the iterated integral to an iterated integral in the new variables.

16. $\int_a^b \int_{\alpha x}^{\beta x} f(x, y)\,dydx$ where $0 < a < b$ and $0 < \alpha < \beta$ if $u = x$ and $v = y/x$;

17. $\int_0^2 \int_{1-x}^{2-x} f(x, y)\,dydx$ if $u = x + y$ and $v = x - y$.

18. Convert the double integral $\iint_D f(x, y)dA$ to an iterated integral in the new variables where D is bounded by the curve $\sqrt{x} + \sqrt{y} = \sqrt{a}$ $(a > 0)$ and the lines $x = 0$, $y = 0$ if $x = u\cos^4 v$ and $y = u\sin^4 v$.

19–28. Evaluate each of the following double integrals by making a suitable change of variables.

19. $\int\int_D yx^2\,dA$ where D is in the first quadrant and bounded by the curves $xy = 1$, $xy = 2$, $yx^2 = 1$ and $yx^2 = 2$;

20. $\iint_D e^{x-y}\,dA$ where D is given by the inequality $|x| + |y| \le 1$;

21. $\iint_D (1+3x^2)dA$ where D is bounded by the lines $x+y = 1$, $x+y = 2$ and by the curves $y - x^3 = 0$, $y - x^3 = 1$;

22. $\iint_D (y+2x^2)dA$ where the domain D is bounded by two parabolas, $y = x^2 + 1$, $y = x^2 + 2$ and by two hyperbolas $xy = -1$ $(x < 0)$, $xy = 1$ $(x > 0)$;

23. $\iint_D (x+y)^2/x^2 dA$ where D is bounded by four lines $y = x$, $y = 2x$, $y + x = 1$ and $y + x = 2$;

24. $\iint_D \sqrt{y-x}/(x+y)$ where D is the square with vertices $(0,2a)$, (a,a), $(2a,2a)$, and $(a,3a)$ with $a>0$;

25. $\iint_D \cos(x^2/a^2+y^2/b^2)dA$ where D is bounded by the ellipse $x^2/a^2+y^2/b^2=1$;

26. $\iint_D (x+y)dA$ where D is bounded by $x^2+y^2=x+y$;

27. $\iint_D (|x|+|y|)dA$ where D is defined by $|x|+|y|\le 1$;

28. $\iint_D (1-\frac{x^2}{a^2}-\frac{y^2}{b^2})^{1/2}dA$ where D is bounded by the ellipse $x^2/a^2+y^2/b^2=1$.

29. Let f be continuous on $[0,1]$. Show that $\iint_D f(x+y)dA=\int_0^1 uf(u)du$ if D is the triangle with vertices $(0,0)$, $(0,1)$, and $(1,0)$.

30–32. Use a suitable change of variables to reduce the double integral to a single integral.

30. $\iint_D f(x+y)dA$ where D is defined by $|x|+|y|\le 1$;

31. $\iint_D f(ax+by+c)dA$ where D is the disk $x^2+y^2\le 1$ and $a^2+b^2\ne 0$;

32. $\iint_D f(xy)dA$ where D lies in the first quadrant and is bounded by the curves $xy=1$, $xy=2$, $y=x$, and $y=4x$.

33. Let n and m be positive integers. Prove that if $\iint_D x^n y^m dA=0$, where D is bounded by an ellipse $x^2/a^2+y^2/b^2=1$, then at least one of the numbers n and m is odd.

34. Suppose that level curves of a function $f(x,y)$ are simple, closed, and smooth. Let a region D be bounded by two level curves $f(x,y)=a$ and $f(x,y)=b$. Prove that

$$\iint_D f(x,y)dA=\int_a^b uF'(u)du$$

where $F(u)$ is the area of the region between the curves $f(x,y)=a$ and $f(x,y)=u$. Hint: partition the region D by level curves of the function f.

35–37. Use the generalized polar coordinates with a suitable choice of parameters to find the area of the given region D.

35. D is bounded by the curves $x^3/a^2+y^3/b^3=x^2+y^2$ and lies in the first quadrant ;

36. D is bounded by the curves $x^3/a^2+y^3/b^3=x^2/c^2-y^2/k^2$ and lies in the first quadrant ;

37. D is bounded by the curve $(x/a+y/b)^5=x^2y^2/c^4$.

38–42. Use the double integral and a suitable change of variables to find the area of the given region D.

38. D is bounded by the curves $x+y=a$, $x+y=b$, $y=mx$, $y=nx$ and lies in the first quadrant;

39. D is bounded by the curves $y^2=2ax$, $y^2=2bx$, $x^2=2cy$, $x^2=2ky$ where $0<a<b$ and $0<c<k$;

40. D is bounded by the curves $(x/a)^{1/2}+(y/b)^{1/2}=1$, $(x/a)^{1/2}+(y/b)^{1/2}=2$, $x/a=y/b$, $4x/a=y/b$, where $a>0$ and $b>0$;

41. D is bounded by the curves $(x/a)^{2/3} + (y/b)^{2/3} = 1$, $(x/a)^{2/3} + (y/b)^{2/3} = 4$, $x/a = y/b$, $8x/a = y/b$ and lies in the first quadrant;

42. D is bounded by the ellipses $x^2/\cosh^2 u + y^2/\sinh^2 u = 1$, where $u = u_1$ and $u = u_2 > u_1$, and by the hyperbolas $x^2/\cos^2 v - y^2/\sin^2 v = 1$, where $v = v_1$ and $v = v_2 > v_1$. Hint: Consider the transformation $x = \cosh u \cos v$, $y = \sinh u \sin v$.

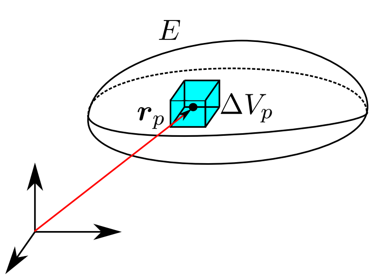
34. Triple Integrals

Suppose a solid region E is filled with an inhomogeneous material. The latter means that, if a small volume ΔV of the material is taken at two distinct points of E, then the masses of these two pieces are different, despite the equality of their volumes. The inhomogeneity of the material can be characterized by the *mass density* as a function of position. Let $\Delta m(\mathbf{r})$ be the mass of a small piece of material of volume ΔV cut out around a point $\mathbf{r}$. Then the mass density is defined by

$$\sigma(\mathbf{r}) = \lim_{\Delta V \to 0} \frac{\Delta m(\mathbf{r})}{\Delta V}.$$

The limit is understood in the following sense. If R is the radius of the smallest ball that contains the region of volume ΔV, then the limit means that $R \to 0$ (i.e., roughly speaking, all the dimensions of the piece decrease uniformly in the limit). The mass density is measured in units of mass per unit volume. For example, the value $\sigma(\mathbf{r}) = 5$ g/cm^3 means that a piece of material of volume 1 cm^3 cut out around the point $\mathbf{r}$ has a mass of 5 gr.

Suppose that the mass density of the material in a region E is known. The question is: What is the total mass of the material in E? A practical answer to this question is to partition the region E so that each partition element E_p, $p = 1, 2, ..., N$, has a mass Δm_p. The total mass is $M = \sum_p \Delta m_p$. If a partition element E_p has a volume ΔV_p, then $\Delta m_p \approx \sigma(\mathbf{r}_p)\,\Delta V_p$ for some



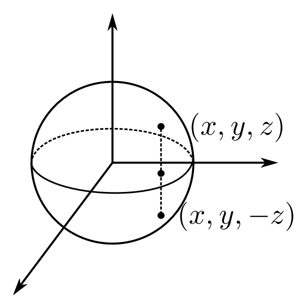


FIGURE **34.1**. **Left**: A partition element of a solid region and $\mathbf{r}_p$ is the position vector of a sample point in it. If $\sigma(\mathbf{r})$ is the mass density, then the mass of the partition element is $\Delta m(\mathbf{r}_p) \approx \sigma(\mathbf{r}_p)\Delta V_p$ where ΔV_p is the volume of the partition element. The total mass is the sum of $\Delta m(\mathbf{r}_p)$ over the partition of the solid E as given in Eq. (**34.1**).
Right: An illustration to Example **34.1**. A ball is symmetric under the reflection about the xy plane: $(x, y, z) \to (x, y, -z)$. If the function f is skew-symmetric under this reflection, $f(x, y, -z) = -f(x, y, z)$, then the triple integral of f over the ball vanishes.

$\mathbf{r}_p$ in E_p (see the left panel of Fig. **34.1**). If R_p is the radius of the smallest ball that contains E_p, put $R_N^* = \max\{R_1, R_2, ..., R_N\}$. Then, by increasing the number N of partition elements so that $R_p \leq R_N^* \to 0$ as $N \to \infty$, the approximation $\Delta m_p \approx \sigma(\mathbf{r}_p)\,\Delta V_p$ becomes more and more accurate by the definition of the mass density because $\Delta V_p \to 0$ for all p. So the total mass is

$$(\mathbf{34.1}) \qquad M = \lim_{\substack{N\to\infty \\ (R_N^*\to 0)}} \sum_{p=1}^{N} \sigma(\mathbf{r}_p)\,\Delta V_p\,,$$

which is to be compared with (**28.1**). In contrast to (**28.1**), the summation over the partition should include a triple sum, one sum per each direction in space. This gives an intuitive idea of a triple integral. Its abstract mathematical construction follows exactly the footsteps of the double-integral construction.

34.1. Definition of a Triple Integral. Suppose E is a closed bounded region in space (recall Definition **28.1**). Let f be a bounded function on E, that is $m \leq f(\mathbf{r}) \leq M$ for all $\mathbf{r}$ in E. The function f is extended to the whole space by setting its values to zero for all points that are not in E.

Rectangular Partition. Since E is bounded it can be embedded into a rectangular box

$$R_E = \{(x,y,z)\,|\, a \leq x \leq b\,,\ c \leq y \leq d\,,\ s \leq z \leq q\} = [a,b]\times[c,d]\times[s,q]\,.$$

The rectangle R_E is partitioned by the coordinate planes

$$\begin{aligned} x &= x_i = a + i\,\Delta x\,, \quad i = 0,1,...,N_1\,, \quad \Delta x = (b-a)/N_1\,,\\ y &= y_j = c + j\,\Delta y\,, \quad j = 0,1,...,N_2\,, \quad \Delta y = (d-c)/N_2\,,\\ z &= z_k = s + k\,\Delta z\,, \quad k = 0,1,...,N_3\,, \quad \Delta z = (q-s)/N_3\,. \end{aligned}$$

Each partition element is a rectangular box

$$R_{ijk} = [x_{i-1}, x_i]\times[y_{j-1}, y_j]\times[z_{k-1}, z_k]\,, \quad i,j,k \geq 1\,,$$

Its volume is $\Delta V = \Delta x\,\Delta y\,\Delta z$. The total number of partition elements is $N = N_1 N_2 N_3$.

Upper and Lower Sums. By analogy with Definition **28.5**, the lower and upper sums are defined. Put

$$M_{ijk} = \sup_{R_{ijk}} f(\mathbf{r})\,, \quad m_{ijk} = \inf_{R_{ijk}} f(\mathbf{r})\,,$$

where the supremum and infimum are taken over the partition element R_{ijk}. Then the upper and lower sums are

$$U(f,\mathbf{N}) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{k=1}^{N_3} M_{ijk}\,\Delta V\,, \quad L(f,\mathbf{N}) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{k=1}^{N_3} m_{ijk}\,\Delta V,$$

where $\mathbf{N} = \langle N_1, N_2, N_3\rangle$. So the upper and lower sums are *triple* sequences. *A rule that assigns a unique number $a_{\mathbf{n}}$ to an ordered triple of integers $\mathbf{n} = \langle n_1, n_2, n_3\rangle$ is called a triple sequence.* The limit of a triple sequence is defined similarly to the limit of a double sequence (see Definition **28.6**). *A number a is the limit of a triple sequence,*

$$\lim_{\mathbf{n}\to\mathbf{0}} a_{\mathbf{n}} = a$$

if for any number $\varepsilon > 0$ one can find an integer N such that

$$|a_{\mathbf{n}} - a| < \varepsilon \qquad \text{for all} \quad n_1, n_2, n_3 > N\,.$$

One can also say that a triple sequence converging to a has only *finitely many* terms outside any interval $(a-\varepsilon, a+\varepsilon)$. The limit of a triple sequence is analogous to the limit of a function of three variables. It can be found by studying the corresponding limit of a function of three variables. Suppose

$$a_{\mathbf{n}} = f(1/n_1, 1/n_2, 1/n_3)$$

and

$$f(x, y, z) \to a \quad \text{as} \quad (x, y, z) \to (0, 0, 0)\,.$$

Then

$$a_{\mathbf{n}} \to a \quad \text{as} \quad \mathbf{n} \to \mathbf{0}\,.$$

Indeed, since f has the limit at the origin, for any $\varepsilon > 0$ one can find a ball of radius $\delta > 0$ in which the values of f deviate from a no more than by ε:

$$|f(\mathbf{r}) - a| < \varepsilon\,, \quad \|\mathbf{r}\| < \delta\,,$$

where $\mathbf{r} = \langle x, y, z\rangle$. In particular, for $\mathbf{r} = \langle 1/n_1, 1/n_2, 1/n_3\rangle$, the condition

$$\|\mathbf{r}\|^2 = \frac{1}{n_1^2} + \frac{1}{n_2^2} + \frac{1}{n_3^2} < \delta^2$$

is satisfied for all $n_1, n_2, n_3 > N \geq 3/\delta$ (e.g., the number N is the integer part of $3/\delta$). Hence,

$$|a_{\mathbf{n}} - a| < \varepsilon \quad \text{for all} \quad n_1, n_2, n_3 > N \geq \frac{3}{\delta}\,,$$

which means that $a_{\mathbf{n}} \to a$ as $\mathbf{n} \to \infty$ (only finitely many terms of $a_{\mathbf{n}}$ lie outside any interval $(a-\varepsilon, a+\varepsilon)$).

DEFINITION **34.1**. (Triple Integral).
If the limits of the upper and lower sums exist as $\mathbf{N} \to \infty$ (or $\Delta\mathbf{r} = \langle \Delta x, \Delta y, \Delta z\rangle \to \mathbf{0}$) and coincide, then f is said to be Riemann integrable on E, and the limit of the upper and lower sums

$$\iiint_E f(x, y, z)\, dV = \lim_{\mathbf{N}\to\infty} U(f, \mathbf{N}) = \lim_{\mathbf{N}\to\infty} L(f, \mathbf{N})$$

is called the triple integral *of f over the region E.*

Continuity and Integrability. Let the boundary of E be piecewise smooth (recall Definition **28.2**). In other words, the region E is bounded by finitely many level surfaces of functions that have continuous partial derivatives and whose gradients do not vanish; the level surfaces are adjacent along piecewise smooth curves. For example, E can be bounded by graphs of functions of two variables which have continuous partial derivatives. The relation between continuity and integrability is pretty much the same as in the case of double integrals.

THEOREM **34.1**. (Integrability of Continuous Functions).
Let E be a closed, bounded spatial region whose boundary is a piecewise smooth surface. If a function f is continuous on E, then it is integrable on E. Furthermore, if f is bounded and not continuous only on a finite number of smooth surfaces in E, then it is also integrable on E.

A set in space is said to have *zero volume* if it can be covered by open balls whose total volume is less than any preassigned positive number ε. For example, a straight line segment of length L can be covered by N balls of radius $R = L/N$ so that their total volume is $4\pi R^3 N/3 = 4\pi L^3/(3N^2)$. By taking N large enough, the total volume can be made less than any given $\varepsilon > 0$ and therefore the volume of the segment is zero. Similarly, a square piece of a plane with dimension L can be covered by N^2 balls of radius $R = L/N$. There total volume is $4\pi R^3 N^2/3 = 4\pi L^3/(3N)$ can be made arbitrary small with a large enough N. So, this piece of a plane has no volume. Using similar arguments, one can show that a smooth surface in space has zero volume. For this reason, the value of a triple integral does not change if the values of the integrand are changed on a smooth surface.

34.2. Properties of Triple Integrals. The properties of triple integrals are the same as those of double integrals discussed in Section **29**; that is, the linearity, additivity, positivity, integrability of the absolute value $|f|$, and upper and lower bounds holds for triple integrals.

A constant function is continuous and, hence, integrable by Theorem **34.1**. In particular, put $f(\mathbf{r}) = 1$. The corresponding triple integral is the volume of the region E

$$V(E) = \iiint_E dV\,. \tag{34.2}$$

If $m \le f(\mathbf{r}) \le M$ for all $\mathbf{r}$ in E, then

$$mV(E) \le \iiint_E f dV \le MV(E)\,.$$

The Integral Mean-Value Theorem. The integral mean value theorem (Theorem **29.1**) is extended to triple integrals. If f is continuous in E, then there is exists a point $\mathbf{r}_0$ in E such that

$$\iiint_E f(\mathbf{r})dV = V(E)f(\mathbf{r}_0)$$

Its proof follows the same lines as in the case of double integrals.

Riemann Sums. Let f be a continuous function on a closed bounded region E whose boundary is piecewise smooth. Let E be partitioned by piecewise smooth surfaces into partition elements E_p, $p = 1, 2, ..., N$, so that the union of E_p is D and $V(E) = \sum_{p=1}^{N} \Delta V_p$, where ΔV_p is the volume of E_p defined by Eq. (**34.2**). If R_p is the smallest radius of a ball that contains E_p, put $R_N^* = \max_p R_p$; that is, R_p characterizes the size of the partition element E_p and R_N^* is the size of the largest partition element. Suppose that $R_N^* \to 0$ as $N \to \infty$. Let $\mathbf{r}_p$ be a sample point in E_p. Under the aforementioned conditions the analog of Theorem **29.3** also holds for triple integrals

THEOREM **34.2**. (Independence of the Partition)
For any choice of sample points $\mathbf{r}_p$ and any choice of partition elements E_p,

$$\iiint_E f\, dV = \lim_{\substack{N\to\infty \\ (R_N^*\to 0)}} \sum_{p=1}^{N} f(\mathbf{r}_p)\, \Delta V_p\,. \tag{34.3}$$

A proof of this theorem goes along the same line of reasoning as the proof of Theorem **29.3**. Equation (**34.3**) can be used for approximations of triple integrals, when evaluating the latter numerically just like in the case of double integrals.

Symmetry. Let T be a transformation in space, that is, T is a rule that assigns a unique point $\mathbf{r}_s$ to a point $\mathbf{r}$ so that

$$T:\ E \to E_s = T(E)$$

for every set E in space. If a transformation preserves the volume of any region, $V(E) = V(E_s)$, then it is called *volume preserving*. Obviously, rotations, reflections, and translations in space are volume-preserving transformations as they preserve the distance between any two points in space. Suppose that, under a volume-preserving transformation, a region E is mapped onto itself; that is, E is *symmetric* relative to this transformation. If $\mathbf{r}_s$ in E is the image of $\mathbf{r}$ in E under this transformation and the integrand is skew-symmetric, $f(\mathbf{r}_s) = -f(\mathbf{r})$, then the triple integral of f over E vanishes:

$$\left.\begin{array}{l} T:\ \mathbf{r}\ \to\ \mathbf{r}_s \\ T(E) = E \\ f(\mathbf{r}_s) = -f(\mathbf{r}) \end{array}\right\} \quad \Rightarrow \quad \iiint_E f(\mathbf{r})\, dV = 0\,.$$

A proof of this assertion is postponed until the change of variables in triple integrals is introduced.

EXAMPLE **34.1**. *Evaluate the triple integral of* $f(x, y, z) = x^2 \sin(y^4 z) + 2$ *over a ball centered at the origin of radius* R.

SOLUTION: Put $g(x, y, z) = x^2 \sin(y^4 z)$ so that $f = g + h$, where $h = 2$ is a constant function. By the linearity property, the triple integral of f is the sum of triple integrals of g and h over the ball. The ball is symmetric relative to the reflection transformation

$$T: \quad (x, y, z) \to (x, y, -z)\,,$$

whereas the function g is skew-symmetric:

$$g(x, y, -z) = x^2 \sin(y^4(-z)) = -x^2 \sin(y^4 z) = -g(x, y, z)\,.$$

Therefore, its triple integral vanishes, and

$$\begin{aligned}\iiint_E f\,dV &= \iiint_E g\,dV + \iiint_E h\,dV \\ &= 0 + 2\iiint_E dV = 2V(E) = 8\pi R^3/3.\end{aligned}$$

□

One can think of the numerical value of a triple integral of f over E as the total amount of a quantity distributed in the region E with the density f (the amount of the quantity per unit volume). For example, f can be viewed as the density of electric charge distributed in a dielectric occupying a region E. The total electric charge stored in the region E is then given by the triple integral of the density over E. The electric charge can be positive and negative. So, if the total positive charge in E is exactly the same as the negative charge, the triple integral vanishes.

34.3. Iterated Triple Integrals. Similar to a double integral, a triple integral can be converted to a triple iterated integral, which can then be evaluated by means of ordinary single-variable integration.

DEFINITION **34.2**. (Simple Region).
A spatial region E *is said to be simple in the direction of a vector* $\mathbf{v}$ *if any straight line parallel to* $\mathbf{v}$ *intersects* E *along at most one straight line segment.*

A triple integral can be converted to an iterated integral if E is simple in a particular direction. If there is no such direction, then E should be split into a union of simple regions with the consequent use of the additivity property of triple integrals. Suppose that $\mathbf{v} = \hat{\mathbf{e}}_3$; that is, E is simple along the z axis. Consider all lines parallel to the z axis that intersect E. These lines also intersect the xy plane. The region D_{xy} in the xy plane is the set of all such points of intersection and is called the *projection* of E into the xy plane. One might think of D_{xy} as a shadow made by the solid E when it is illuminated by rays of light parallel to the z axis. Take any line through (x, y) in D_{xy} parallel to the z axis. By the simplicity of E, any such line

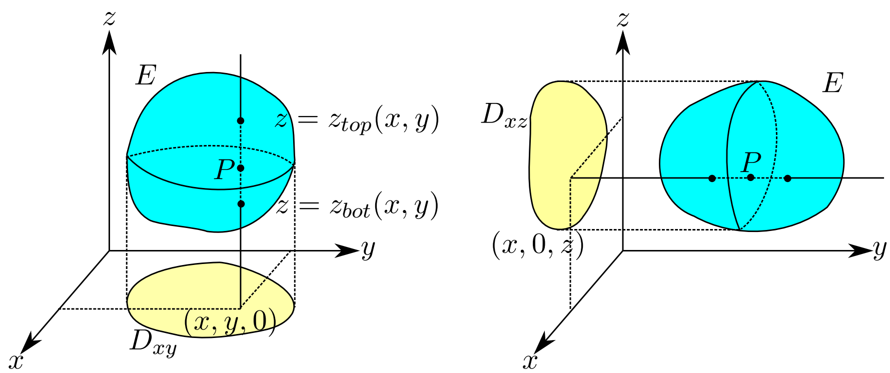


FIGURE **34.2**. **Left**: An algebraic description of a solid region simple in the direction of the z axis. The solid E is vertically projected into the xy plane: every point (x, y, z) of E goes into the point $(x, y, 0)$. The projection points form the region D_{xy}. Since E is simple in the z direction, for every $(x, y, 0)$ in D_{xy}, the z coordinate of the point $P = (x, y, z)$ in E ranges the interval $z_{\text{bot}}(x, y) \le z \le z_{\text{top}}(x, y)$. In other words, E lies between the graphs $z = z_{\text{bot}}(x, y)$ and $z = z_{\text{top}}(x, y)$. **Right**: An illustration to the algebraic description (**34.5**) of a solid E as a simple in the y direction. E is projected along the y axis to the xz plane, forming a region D_{xz}. For every $(x, 0, z)$ in D_{xz}, the y coordinate of the point $P = (x, y, z)$ in E ranges the interval $y_{\text{bot}}(x, z) \le y \le y_{\text{top}}(x, z)$. In other words, E lies between the graphs $y = y_{\text{bot}}(x, z)$ and $y = y_{\text{top}}(x, z)$.

intersects E along a single segment. If z_{bot} and z_{top} are the minimal and maximal values of the z coordinate along the intersection segment, then, for any (x, y, z) in E, $z_{\text{bot}} \le z \le z_{\text{top}}$ and (x, y) in D_{xy}. Naturally, the values z_{bot} and z_{top} may depend on (x, y) in D_{xy}. Thus, the region E is bounded from the *top* by the graph $z = z_{\text{top}}(x, y)$ and from the *bottom* by the graph $z = z_{\text{bot}}(x, y)$; it admits the following algebraic description:

$$E = \{(x, y, z) | z_{\text{bot}}(x, y) \le z \le z_{\text{top}}(x, y), \quad (x, y) \in D_{xy}\} . \tag{34.4}$$

If E is simple along the y or x axis, then E admits similar descriptions:

$$E = \{(x, y, z) | \, y_{\text{bot}}(x, z) \le y \le y_{\text{top}}(x, z), \quad (x, z) \in D_{xz}\} , \tag{34.5}$$

$$E = \{(x, y, z) | \, x_{\text{bot}}(y, z) \le x \le x_{\text{top}}(y, z), \quad (y, z) \in D_{yz}\} , \tag{34.6}$$

where D_{xz} and D_{yz} are projections of E into the xz and yz planes, respectively; they are defined analogously to D_{xy}. The "top" and "bottom" are now defined relative to the direction of the y or x axis.

Suppose f is integrable on E. Suppose E is simple in the direction of the z axis. According to (**34.3**), the limit of the Riemann sum is independent of partitioning E and choosing sample points. Since E is bounded, there are numbers s and q such that

$$s \le z_{\rm bot}(x,y) \le z_{\rm top}(x,y) \le q \quad \text{for all} \quad (x,y) \in D_{xy}\,;$$

that is, E always lies between two horizontal planes $z = s$ and $z = q$. The region E can therefore be embedded into a cylinder with the horizontal cross section having the shape of D_{xy}. The function f is extended outside E by zero values. The cylinder is partitioned as follows. Let D_p, $p = 1, 2, ..., N$, be a partition of the region D_{xy}. This partition induces partitioning of the cylinder by cylinders (or columns E_p) with the horizontal cross sections D_p. Each column E_p is partitioned by equispaced horizontal planes

$$z = s + k\,\Delta z\,, \quad k = 0, 1, ..., N_3\,, \quad \Delta z = (q - s)/N_3\,,$$

into small regions E_{pk}. The union of all E_{pk} forms a partition of the cylinder embedding E. This partition will be used in the Riemann sum (**34.3**). The volume of E_{pk} is

$$\Delta V_{pk} = \Delta z\,\Delta A_p\,,$$

where ΔA_p is the area of D_p. Next, let us chose sample points. Take a point $(x_p, y_p, 0)$ in each D_p and pick a number z_k^* from the interval $[z_{k-1}, z_k]$. Then a sample point in the partition element E_{pk} is (x_p, y_p, z_k^*).

The three-variable limit (**34.3**) exists and hence can be taken in any particular order. Let us take first the limit $N_3 \to \infty$ or $\Delta z \to 0$ and then the limit $N \to \infty$. The latter limit of the sum over the partition of D_{xy} is understood as before; that is, the radii R_p of smallest disks containing D_p go to 0 uniformly, $R_p \le \max_p R_p = R_N^* \to 0$ as $N \to \infty$. Therefore,

$$\begin{aligned}
\iiint_E f\,dV &= \lim_{\substack{N\to\infty \\ (R_N^*\to 0)}} \sum_{p=1}^{N} \left(\lim_{N_3\to\infty} \sum_{k=1}^{N_3} f(x_p, y_p, z_k^*)\,\Delta z \right) \Delta A_p \\
&= \lim_{\substack{N\to\infty \\ (R_N^*\to 0)}} \sum_{p=1}^{N} \left(\int_{z_{\rm bot}(x_p,y_p)}^{z_{\rm top}(x_p,y_p)} f(x_p, y_p, z)\,dz \right) \Delta A_p
\end{aligned}$$

because, for every (x_p, y_p) in D_{xy}, the function f vanishes outside the interval $z_{\rm bot}(x_p, y_p) \le z \le z_{\rm top}(x_p, y_p)$. The integral of f with respect to z over the interval $[z_{\rm bot}(x,y), z_{\rm top}(x,y)]$ defines a function

$$g(x,y) = \int_{z_{\rm bot}(x,y)}^{z_{\rm top}(x,y)} f(x,y,z)\,dz$$

on D_{xy}. Its values $g(x_p, y_p)$ at sample points in the partition elements D_p appear in the parentheses of the above Riemann sum. A comparison of the resulting expression with (**29.2**) leads to the conclusion that, after taking the second limit, one obtains the double integral of $g(x,y)$ over D_{xy}.

THEOREM **34.3**. (Iterated Triple Integral).
Let f be integrable on a solid region E bounded by a piecewise smooth surface. Suppose that E is simple in the z direction so that it is bounded by the graphs $z = z_{\rm bot}(x, y)$ and $z = z_{\rm top}(x, y)$ for all (x, y) in D_{xy}. Then

$$\begin{aligned}\iiint_E f(x,y,z)\,dV &= \iint_{D_{xy}} \int_{z_{\rm bot}(x,y)}^{z_{\rm top}(x,y)} f(x,y,z)\,dz\,dA \\ &= \iint_{D_{xy}} g(x,y)\,dA.\end{aligned}$$

34.4. Evaluation of Triple Integrals. By Theorem **34.3**, an evaluation of a triple integral over a region E can be carried out by the following steps:

Step 1. Determine the direction along which E is simple. If no such direction exists, split E into a union of simple regions and use the additivity property. For definitiveness, suppose that E happens to be simple in the direction of the z axis.
Step 2. Find the projection D_{xy} of E into the xy plane.
Step 3. Find the bottom and top boundaries of E as the graphs of some functions $z = z_{\rm bot}(x, y)$ and $z = z_{\rm top}(x, y)$.
Step 4. Evaluate the integral of f with respect to z to obtain $g(x, y)$.
Step 5. Evaluate the double integral of $g(x, y)$ over D_{xy} by converting it to a suitable iterated integral.

Similar iterated integrals can be written when E is simple in the y or x direction. According to (**34.5**) or (**34.6**), the first integration is carried out with respect to y or x, respectively, and the double integral is evaluated over D_{xz} or D_{yz}:

$$\iiint_E f(x,y,z)\,dV = \iint_{D_{xz}} \int_{y_{\rm bot}(x,z)}^{y_{\rm top}(x,z)} f(x,y,z)\,dy\,dA\,,$$

$$\iiint_E f(x,y,z)\,dV = \iint_{D_{yz}} \int_{x_{\rm bot}(y,z)}^{x_{\rm top}(y,z)} f(x,y,z)\,dx\,dA\,.$$

If E is simple in any direction, then any of the iterated integrals can be used. In particular, just like in the case of double integrals, the choice of an iterated integral for a simple region E should be motivated by the simplicity of an algebraic description of the top and bottom boundaries or by the simplicity of the integrations involved. Technical difficulties may strongly depend on the order in which the iterated integral is evaluated.

Let E be a rectangular region $[a, b] \times [c, d] \times [s, q]$. It is simple in any direction. If the integration with respect to z is to be carried out first, then $D_{xy} = [a, b] \times [c, d]$ and the top and bottom boundaries are the planes $z = q$ and $z = s$. The double integral over the rectangle D_{xy} can be evaluated by Fubini's theorem. Alternatively, one can take $D_{yz} = [c, d] \times [s, q]$, $x_{\rm bot}(y, z) = a$, and $x_{\rm top}(y, z) = b$ to obtain an iterated integral in a different order (where the x integration is carried out first). So, Fubini's theorem is extended to triple integrals.

THEOREM **34.4**. (Fubini's Theorem).
Let f be continuous on a rectangular region $E = [a, b] \times [c, d] \times [s, q]$. Then

$$\iiint_E f\, dV = \int_a^b \int_c^d \int_s^q f(x, y, z)\, dz\, dy\, dx$$

and the iterated integral can be evaluated in any order.

In particular, if $f(x, y, z) = g(x)h(y)w(z)$, then

$$\iiint_E f(x, y, z) dV = \int_a^b g(x) dx \int_c^d h(y) dy \int_q^s w(z) dz$$

which is an extension of the factorization property stated in Corollary **30.1** to triple integrals.

EXAMPLE **34.2**. *Evaluate the triple integral of $f(x, y, z) = xy^2z^2$ over the rectangular box $E = [0, 2] \times [1, 2] \times [0, 3]$.*

SOLUTION: By Fubini's theorem,

$$\iiint_E xy^2z^2\, dV = \int_0^2 x\, dx \int_1^2 y^2\, dy \int_0^3 z^2\, dz = \frac{4}{2} \cdot \frac{8-1}{3} \cdot \frac{27}{3} = 42.$$

□

EXAMPLE **34.3**. *Evaluate the triple integral of $f(x, y, z) = (x^2+y^2)z$ over the portion of the solid bounded by the cone $z = \sqrt{x^2 + y^2}$ and paraboloid $z = 2 - x^2 - y^2$ in the first octant.*

SOLUTION: Following the step-by-step procedure outlined above, the integration region is simple in the direction of the z axis. The top and bottom boundaries are the graphs of the functions:

$$z_{\rm top}(x, y) = 2 - x^2 - y^2\,, \quad z_{\rm bot}(x, y) = \sqrt{x^2 + y^2}\,.$$

To determine the region D_{xy}, note that it has to be bounded by the projection of the curve of the intersection of the cone and paraboloid onto the xy plane. The curve of intersection is defined by the condition

$$z_{\rm bot}(x, y) = z_{\rm top}(x, y) \quad \Rightarrow \quad r = 2 - r^2 \quad \Rightarrow \quad r = 1\,,$$

where $r = \sqrt{x^2 + y^2}$. The condition $r = 1$ defines the circle of unit radius centered at the origin. Since E is in the first octant, D_{xy} is the part of the

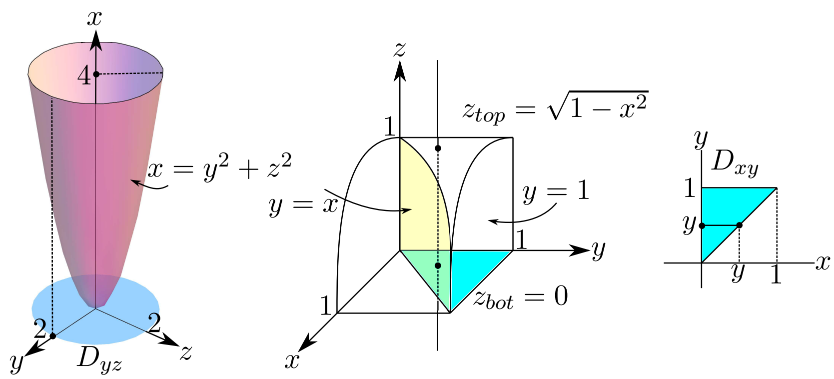


FIGURE **34.3**. **Left**: The integration region in Example **34.4**. The x axis is vertical. The region is bounded by the plane $x = 4$ (top) and the paraboloid $x = y^2 + z^2$ (bottom). Its projection into the yz plane is the disk of radius 2 as the plane and paraboloid intersect along the circle $4 = y^2 + z^2$. **Right**: An illustration to Study Problem **34.1**.

disk of unit radius in the first quadrant. One has

$$\begin{aligned}\iiint_E (x^2+y^2)z\,dV &= \iint_{D_{xy}} (x^2+y^2)\int_{\sqrt{x^2+y^2}}^{2-x^2-y^2} z\,dz\,dA \\ &= \frac{1}{2}\iint_{D_{xy}} (x^2+y^2)[(2-x^2-y^2)^2-(x^2+y^2)]\,dA \\ &= \frac{1}{2}\int_0^{\pi/2} d\theta \int_0^1 r^2[(2-r^2)^2-r^2]\,r\,dr \\ &= \frac{\pi}{8}\int_0^1 u[(2-u)^2-u]\,du = \frac{7\pi}{96},\end{aligned}$$

where the double integral has be transformed into polar coordinates; the region D_{xy} is the image of the rectangle $D'_{xy} = [0,1]\times[0,\pi/2]$ in the polar plane and $dA = rdrd\theta$. The integration with respect to r is carried out by the substitution $u = r^2$ so that $rdr = \frac{1}{2}du$. □

EXAMPLE **34.4**. *Evaluate the triple integral of* $f(x,y,z) = \sqrt{y^2+z^2}$ *over the region* E *bounded by the paraboloid* $x = y^2+z^2$ *and the plane* $x = 4$.

SOLUTION: It is convenient to choose an iterated integral for E described as an x simple region (see (**34.6**)). There are two reasons for doing so. First, the integrand f is independent of x, and hence the first integration with respect to x is trivial. Second, the boundaries of E are already given in the form required by (**34.6**):

$$x_{\text{bot}}(y,z) = y^2+z^2\,,\quad x_{\text{top}}(y,z) = 4\,.$$

The region D_{yz} is bounded by the curve of intersection of the boundaries of E:

$$x_{\rm top}(y,z) = x_{\rm bot}(y,z) \quad \Rightarrow \quad y^2+z^2 = 4\,.$$

Therefore, D_{yz} is the disk or radius 2 (see the left panel of Fig. **34.3**). One has

$$\begin{aligned}\iiint_E \sqrt{y^2+z^2}dV &= \iint_{D_{yz}} \sqrt{y^2+z^2}\int_{y^2+z^2}^4 dx\,dA\\ &= \iint_{D_{yz}} \sqrt{y^2+z^2}\;[4-(y^2+z^2)]\,dA\\ &= \int_0^{2\pi} d\theta \int_0^2 r[4-r^2]r\,dr = \frac{128\pi}{15},\end{aligned}$$

where the double integral over the disk D_{yz} has been converted to polar coordinates in the yz plane ($y = r\cos\theta$, $z = r\sin\theta$, and $dA = r dr d\theta$). □

34.5. Study Problems.

Problem **34.1**. *Evaluate the triple integral of $f(x,y,z) = z$ over the region E bounded by the cylinder $x^2+z^2 = 1$ and the planes $z = 0$, $y = 1$, and $y = x$ in the first octant.*

SOLUTION: The region is z simple and bounded by the xy plane from the bottom (i.e., $z_{\rm bot}(x,y) = 0$), and by the cylinder from the top (i.e., $z_{\rm top}(x,y) = \sqrt{1-x^2}$) (by taking the positive solution of $x^2+z^2 = 1$). The integration region is shown in the right panel of Fig. **34.3**. Since E is in the first octant, the region D_{xy} lies in the first quadrant and is bounded by the lines $y = x$ and $y = 1$. Thus, D_{xy} is the triangle bounded by the lines $x = 0$, $y = 1$, and $y = x$:

$$\begin{aligned}D_{xy} &= \{\ (x,y)\mid 0\le x\le y\,,\ y\in[0,1]\}\\ E &= \{\ (x,y,z)\mid 0\le z\le \sqrt{1-x^2}\,,\ \ (x,y)\in D_{xy}\}\,.\end{aligned}$$

One has

$$\begin{aligned}\iiint_E z\,dV &= \iint_{D_{xy}}\int_0^{\sqrt{1-x^2}} z\,dz\,dA\\ &= \frac{1}{2}\iint_{D_{xy}}(1-x^2)\,dA = \frac{1}{2}\int_0^1\int_0^y (1-x^2)dx\,dy\\ &= \frac{1}{2}\int_0^1 \Big(y-\frac{1}{3}y^3\Big)dy = \frac{5}{24},\end{aligned}$$

where the double integral has been evaluated by using the above description of D_{xy} as a horizontally simple region. □

Problem **34.2**. *Evaluate the triple integral of the function $f(x,y,z) = xy^2z^3$ over the region E that is a ball of radius 3 centered at the origin with a cubic cavity $[0,1]\times[0,1]\times[0,1]$.*

SOLUTION: The region E is not simple in any direction. The additivity property must be used. Let E_1 be the ball and let E_2 be the cavity. By the additivity property,

$$\begin{aligned}\iiint_E xy^2z^3\,dV &= \iiint_{E_1} xy^2z^3\,dV - \iiint_{E_2} xy^2z^3\,dV\\ &= 0 - \int_0^1 x\,dx \int_0^1 y^2\,dy \int_0^1 z^3\,dz = -\frac{1}{24}.\end{aligned}$$

The triple integral over E_1 vanishes by the symmetry argument (the ball is symmetric under the reflection $(x,y,z)\to(-x,y,z)$ whereas $f(-x,y,z) = -f(x,y,z)$). The second integral is evaluated by Fubini's theorem. □

34.6. Exercises.

1–11. Evaluate each of the following triple integrals over the specified solid region by converting it to an appropriate iterated integral.

1. $\iiint_E (xy-3z^2)dV$, where $E=[0,1]\times[1,2]\times[0,2]$;

2. $\iiint_E 6xzdV$, where E is defined by the inequalities $0\le y\le x+z$, $0\le x\le z$, and $0\le z\le 1$;

3. $\iiint_E ze^{y^2}dV$, E is defined by the inequalities $0\le x\le y$, $0\le y\le z$, and $0\le z\le 1$;

4. $\iiint_E 6xydV$, where E lies under the plane $x+y-z=-1$ and above the region in the xy plane bounded by the curves $x=\sqrt{y}$, $x=0$, and $y=1$;

5. $\iiint_E xydV$ where E is bounded by the parabolic cylinders $y=x^2$, $x=y^2$, and by the planes $x+y-z=0$, $x+y+z=0$;

6. $\iiint_E dV$ where E is bounded by the coordinate planes and the plane through the points $(a,0,0)$, $(0,b,0)$, and $(0,0,c)$ with a,b,c being positive numbers ;

7. $\iiint_E zx\,dV$ where E lies in the first octant between two planes $x=y$ and $x=0$ and is bounded by the cylinder $y^2+z^2=1$;

8. $\iiint_E (x^2z+y^2z)dV$ where E is enclosed by the paraboloid $z=1-x^2-y^2$ and the plane $z=0$;

9. $\iiint_E zdV$ where where E is enclosed by the elliptic paraboloid $z=1-x^2/a^2-y^2/b^2$ and the plane $z=0$. Hint: Use a suitable change of variable in the double integral ;

10. $\iiint_E xy^2z^3dV$ where E is bounded by the surfaces $z=xy$, $y=x$, $x=1$, $z=0$;

11. $\iiint_E (1+x+y+z)^{-3}dV$ where E is bounded by the plane $x+y+z=1$ and by the coordinate planes.

12–21. Use the triple integral to find the volume of the specified solid E.

12. E is bounded by the parabolic cylinder $x=y^2$ and the planes $z=0$ and $z+x=1$;

13. E lies in the first octant and is bounded by the parabolic sheet $z = 4 - y^2$ and by two planes $y = x$ and $y = 2x$;

14. E is bounded by the surfaces $z^2 = xy$, $x + y = a$, $x + y = b$, where $0 < a < b$;

15. E is bounded by the surfaces $z = x^2 + y^2$, $xy = a^2$, $xy = 2a^2$, $y = x/2$, $y = 2x$, $z = 0$. Hint: Use a suitable change of variable in the double integral;

16. E is bounded by the surfaces $z = x^{3/2} + y^{3/2}$, $z = 0$, $x + y = 1$, $x = 0$, and $y = 0$;

17. E is bounded by the surfaces $x^2/a^2 + y^2/b^2 + z/c = 1$, $(x/a)^{2/3} + (y/b)^{2/3} = 1$, and $z = 0$, where $c > 0$. Hint: use the generalized polar coordinates defined in Study Problem **33.1** to evaluate the double integral;

18. E is bounded by the surfaces $z = x + y$, $z = xy$, $x + y = 1$, $x = 0$, and $y = 0$;

19. E is bounded by the surfaces $x^2 + z^2 = a^2$, $x + y = \pm a$, and $x - y = \pm a$, where $a > 0$;

20. E is bounded by the surfaces $az = x^2 + y^2$, $z = a - x - y$, and by the coordinate planes, where $a > 0$;

21. E is bounded by the surfaces $z = 6 - x^2 - y^2$, and $z = \sqrt{x^2 + y^2}$.

22-24. Use symmetry and other properties of the triple integral to evaluate each of the following triple integrals.

22. $\iiint_E 24xy^2z^3\, dV$ where E is bounded by the elliptic cylinder $(x/a)^2 + (y/b)^2 = 1$ and by the paraboloids $z = \pm[c - (x/a)^2 - (y/b)^2]$ and has the rectangular cavity $[0, 1] \times [-1, 1] \times [0, 1]$. Assume that a, b, and c are larger than 2;

23. $\iiint_E (\sin^2(xz) - \sin^2(xy))dV$ where E lies between the spheres: $1 \le x^2 + y^2 + z^2 \le 4$.

24. $\iiint_E (\sin^2(xz) + \cos^2(xy))dV$ where E lies between the spheres: $1 \le x^2 + y^2 + z^2 \le 4$.

25–26. Express the integral $\iiint_E f dV$ as an iterated integral in six different ways, where E is the solid bounded by the specified surfaces.

25. $x^2 + y^2 = 4$, $z = -1$, and $z = 2$;

26. $z + y = 1$, $z = 0$, and $y = x^2$.

27–29. Reverse the order of integration in all possible ways.

27. $\int_0^1 \int_0^{1-x} \int_0^{x+y} f(x, y, z)\, dzdydx$;

28. $\int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \int_{\sqrt{x^2+y^2}}^1 f(x, y, z)\, dzdydx$;

29. $\int_0^1 \int_0^1 \int_0^{x^2+y^2} f(x, y, z)\, dzdydx$.

30–31. Reduce each of the following iterated integrals to a single integral by reversing the integration in a suitable order.

30. $\int_0^a \int_0^x \int_0^y f(z) dzdydx$;

31. $\int_0^1 \int_0^1 \int_0^{x+y} f(z)dzdydx$.

32. Use the interpretation of the triple integral f over a region E as the total amount of some quantity in E distributed with the density f to find E for which $\iiint_E (1 - x^2/a^2 - y^2/b^2 - z^2/c^2)dV$ is maximal.

33. Prove the following representation of the triple integral by iterated integrals.

$$\iiint_E f(x,y,z)dV = \int_a^b \iint_{D_z} f(x,y,z)dAdz$$

where D_z is the cross section of E by the plane $z = \text{const}$.

34. Prove that if $f(x,y,z)$ is continuous in E and for any subregion W of E, $\iiint_W f dV = 0$, then $f(x,y,z) = 0$ in E.

35. Use the lower bound for the triple integral in Exercise **11** to show that $\ln 2 \geq \frac{2}{3}$. Try to give an alternative prove of this inequality based on the definition of the log or exponential function.

35. Triple Integrals in Cylindrical and Spherical Coordinates

A change of variables has been proved to be quite useful in simplifying the technicalities involved in evaluating double integrals. An essential advantage is a simplification of the integration region. The concept of changing variables can be extended to triple integrals.

35.1. Cylindrical Coordinates. One of the simplest examples of curvilinear coordinates in space is cylindrical coordinates. They are defined by the transformation

$$(\mathbf{35}.\mathbf{1}) \qquad T: \ (r,\theta,z) \ \to \ (x,y,z)\,, \quad x = r\cos\theta\,,\ y = r\cos\theta\,,\ z = z\,.$$

In any plane parallel to the xy plane, the points are labeled by polar coordinates, while the z coordinate is not transformed. A set of triples (r,θ,z) can be viewed as a set of points E' in a Euclidean space in which the coordinate axes are spanned by r, θ, and z. Then, under the transformation (**35.1**), the region E' is mapped to an *image* region E. From the study of polar coordinates, the transformation (**35.1**) is one-to-one if (r,θ,z) in $(0,\infty)\times[0,2\pi)\times(-\infty,\infty)$. As noted before, the range of θ can be chosen to be any interval of length 2π. The inverse transformation is the same as in the case of polar coordinates. The transformation is not one-to-one on the boundaries $r=0$ or $\theta=2\pi$. Indeed, $T:\ (0,\theta,z)\to(0,0,z)$ (the plane $r=0$ is mapped to the z axis) and the points with θ different by 2π have the same image. This multiple-counting does not have any effect on the triple integral as it occurs on surfaces that have zero volume. So there is no harm to assume that $0\le r<\infty$ and θ ranges over a closed interval of length 2π, just like in polar coordinates. The z axis is called the axis of cylindrical coordinates. Note that the x or y axis may be chosen as the axis of cylindrical coordinates. In this case, the polar coordinates are introduced in the yz or xz planes.

Given a region E, to find the shape of E' as well as its algebraic description, the same strategy as in the two-variable case can be used:

$$T:\ \text{boundary of } E' \quad\to\quad \text{boundary of } E$$

under the transformation (**35.1**). A particularly important question is to investigate the shape of *coordinate surfaces* of cylindrical coordinates, that is, surfaces on which each of the cylindrical coordinates has a constant value. If E is bounded by coordinate surfaces only, then it is an image of a rectangular box E' which is the simplest, most desirable, shape when evaluating a multiple integral.

The coordinate surfaces of r are cylinders, $r=\sqrt{x^2+y^2}=r_0$ or $x^2+y^2=r_0^2$. In the xy plane, the equation $\theta=\theta_0$ defines a ray from the origin at the angle θ_0 to the positive x axis counted counterclockwise. Since θ depends only on x and y, the coordinate surface of θ is the half-plane bounded by the z axis that makes an angle θ_0 with the xz plane (it is swept by the ray

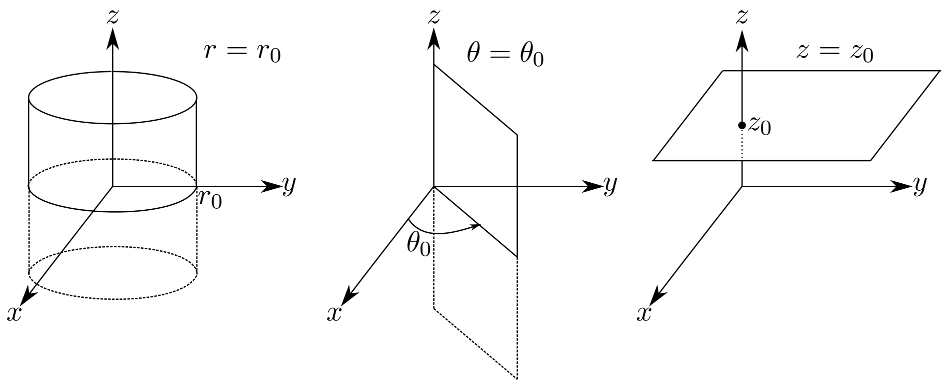


FIGURE **35.1**. Coordinate surfaces of cylindrical coordinates: Cylinders $r = r_0$, half-planes $\theta = \theta_0$ bounded by the z axis, and horizontal planes $z = z_0$. Any point in space can be viewed as the point of intersection of three coordinate surfaces.

when the latter is moved parallel up and down along the z axis). Since the z coordinate is not changed, neither changes its coordinate surfaces; they are planes parallel to the xy plane. So the coordinate surfaces of cylindrical coordinates are

$$\begin{array}{llll} T: & r = r_0 & \to\ x^2 + y^2 = r_0^2 & \text{(cylinder)}, \\ T: & \theta = \theta_0 & \to\ y\cos\theta_0 = x\sin\theta_0 & \text{(half-plane)}, \\ T: & z = z_0 & \to\ z = z_0 & \text{(plane)}. \end{array}$$

The coordinate surfaces of cylindrical coordinates are shown in Fig. **35.1**. A point in space corresponding to an ordered triple (r_0, θ_0, z_0) is the point of intersection of a cylinder, half-plane bounded by the cylinder axis, and a plane perpendicular to the cylinder axis.

EXAMPLE **35.1**. *Find the region E' whose image under the transformation* (**35.1**) *is the solid region E that is bounded by the paraboloid $z = x^2 + y^2$ and the planes $z = 4$, $y = x$, and $y = 0$ in the first (positive) octant.*

SOLUTION: In cylindrical coordinates, the equations of boundaries become

$$\begin{aligned} z = x^2 + y^2 \quad &\Rightarrow \quad z = r^2\,, \\ z = 4 \quad &\Rightarrow \quad z = 4\,, \\ y = x\,,\ x \ge 0 \quad &\Rightarrow \quad \theta = \pi/4\,, \\ y = 0\,,\ x \ge 0 \quad &\Rightarrow \quad \theta = 0\,. \end{aligned}$$

Since E lies below the plane $z = 4$ and above the paraboloid $z = r^2$, the range of r is determined by their intersection: $4 = r^2$ or $r = 2$ as $r \ge 0$. Thus,

$$T: \quad E' = \Big\{(r, \theta, z)\,|\, r^2 \le z \le 4\,,\quad (r, \theta) \in [0, 2] \times [0, \pi/4]\Big\} \ \to\ E\,.$$

□

35.2. Triple Integrals in Cylindrical Coordinates. To change variables in a triple integral to cylindrical coordinates, one has to consider a partition of the integration region E by *coordinate surfaces*, that is, by cylinders, half-planes, and horizontal planes, which corresponds to a rectangular partition of E' (the image of E under the transformation from rectangular to cylindrical coordinates). Then the limit of the corresponding Riemann sum (**34.3**) has to be evaluated. In the case of cylindrical coordinates, this task can be accomplished by simpler means.

Suppose E is simple in the direction of the axis of the cylindrical coordinates. Let the z be the axis of cylindrical coordinates. By Theorem **34.3**, the triple integral can be written as an iterated integral consisting of a double integral over D_{xy} and an ordinary integral with respect to z. The transformation (**35.1**) merely defines polar coordinates in the region D_{xy}. So, if D_{xy} is the image of D'_{xy} in the polar plane spanned by pairs (r, θ), then, by converting the double integral to polar coordinates, one infers that

$$\begin{aligned}\iiint_E f(x,y,z)\,dV &= \iint_{D'_{xy}} \int_{z_{\text{bot}}(r,\theta)}^{z_{\text{top}}(r,\theta)} f(r\cos\theta, r\sin\theta, z) r\,dz\,dA' \\ &= \iiint_{E'} f(r\cos\theta, r\sin\theta, z) r\,dV', \qquad (\mathbf{35.2})\end{aligned}$$

where the region E' is the image of E under the transformation from rectangular to cylindrical coordinates,

$$E' = \{(r,\theta,z) \,|\, z_{\text{bot}}(r,\theta) \le z \le z_{\text{top}}(r,\theta)\,, \quad (r,\theta) \in D'_{xy}\},$$

and $z = z_{\text{bot}}(r,\theta)$, $z = z_{\text{top}}(r,\theta)$ are equations of the bottom and top boundaries of E written in polar coordinates by substituting (**35.1**) into the equations for boundaries written in rectangular coordinates. Note that $dV' = dz\,dr\,d\theta = dz\,dA'$ is the volume of an infinitesimal rectangle in the space spanned by the triples (r,θ,z). Its image in the space spanned by (x,y,z) lies between two cylinders whose radii differ by dr, between two half-planes with the angle $d\theta$ between them, and between two horizontal planes separated by the distance dz as shown in the left panel of Fig. **35.2**. So its volume is the product of the area dA of the base and the height dz, $dV = dz\,dA = r\,dz\,dA'$ according to the area transformation law for polar coordinates, $dA = rdA'$. So the volume transformation law for cylindrical coordinates reads

$$dV = J\,dV'\,, \quad J = r,$$

where $J = r$ is the Jacobian of transformation to cylindrical coordinates.

Cylindrical coordinates are advantageous when the boundary of E contains cylinders, half-planes, horizontal planes, or any surfaces with *axial symmetry*. A set in space is said to be *axially symmetric* if there is an

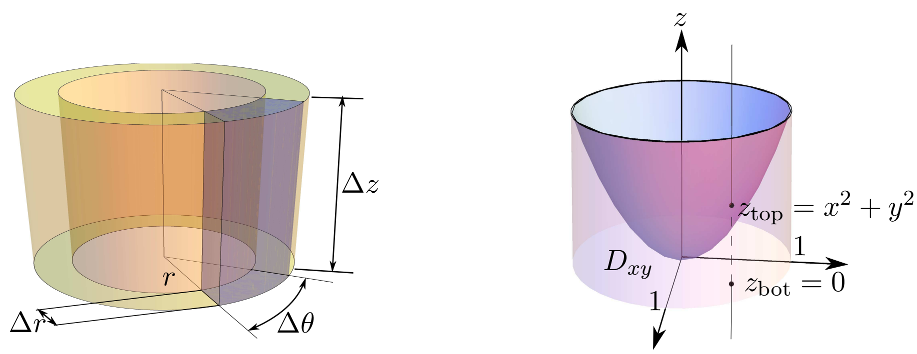


FIGURE **35.2**. **Left**: A partition element of the partition of E by cylinders, half-planes, and horizontal planes (coordinate surfaces of cylindrical coordinates). The partition is the image of a rectangular partition of E'. Keeping only terms linear in the differentials $dr = \Delta r$, $d\theta = \Delta\theta$, $dz = \Delta z$, the volume of the partition element is $dV = dAdz = rdrd\theta dz = rdV'$ where $dA = rdrd\theta$ is the area element in the polar coordinates. So, the Jacobian of cylindrical coordinates is $J = r$. **Right**: An illustration to Example **35.2**.

axis such that any rotation about it maps the set onto itself. For example, circular cones, circular paraboloids, and spheres are axially symmetric.

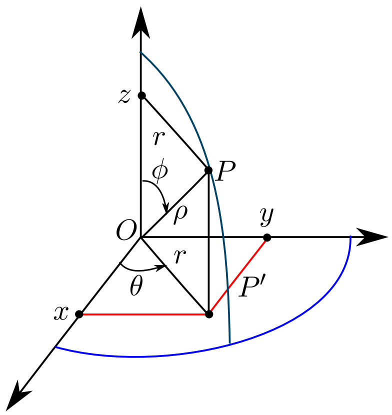
EXAMPLE **35.2**. *Evaluate the triple integral of $f(x, y, z) = x^2 z$ over the region E bounded by the cylinder $x^2 + y^2 = 1$, the paraboloid $z = x^2 + y^2$, and the plane $z = 0$.*

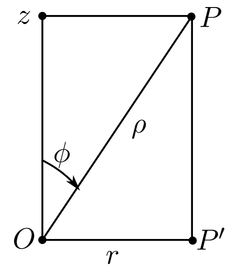
SOLUTION: The solid E is axially symmetric because it is bounded from below by the plane $z = 0$, by the circular paraboloid from above, and the side boundary is the cylinder. Hence, the projection D_{xy} of E onto the xy plane is a disk of unit radius. It is the image of the rectangle $D'_{xy} = [0, 1] \times [0, 2\pi]$ in the polar plane. The top and bottom boundaries are $z = z_{\text{top}}(r, \theta) = r^2$ and $z = z_{\text{bot}}(r, \theta) = 0$. Hence,

$$\begin{aligned}\iiint_E x^2 z\, dV &= \int_0^{2\pi}\int_0^1\int_0^{r^2} r^2\cos^2\theta\, z\, r\, dz\, dr\, d\theta \\ &= \frac{1}{2}\int_0^{2\pi}\cos^2\theta\, d\theta\int_0^1 r^7 dr = \frac{\pi}{16},\end{aligned}$$

where the double-angle formula, $\cos^2\theta = (1 + \cos(2\theta))/2$, has been used to evaluate the integral. □

35.3. Spherical Coordinates. Spherical coordinates are introduced by the following geometrical procedure. Let (x, y, z) be a point in space. Consider a





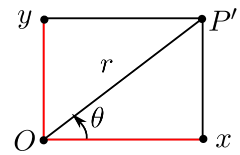


FIGURE **35.3**. Spherical coordinates and their relation to the rectangular coordinates. A point P in space is defined by its distance to the origin ρ, the angle ϕ between the positive z axis and the ray OP, and the polar angle θ.

ray from the origin through this point (see Figure **35.3**). Any such ray lies in the half-plane corresponding to a fixed value of the polar angle θ. Therefore, the ray is uniquely determined by the polar angle θ and the angle ϕ between the ray and the positive z axis. If ρ is the distance from the origin to the point (x, y, z), then the ordered triple of numbers (ρ, ϕ, θ) defines uniquely any point in space. The triples (ρ, ϕ, θ) are called *spherical coordinates* in space.

To find the transformation law from spherical to rectangular coordinates, consider the plane that contains the z axis and the ray from the origin through $P = (x, y, z)$ and the rectangle with vertices $(0, 0, 0)$, $(0, 0, z)$, $P' = (x, y, 0)$, and (x, y, z) in this plane as depicted in Figure **35.3**. The diagonal of this rectangle has length ρ (the distance between $(0, 0, 0)$ and (x, y, z)). Therefore, its vertical side has length $z = \rho\cos\phi$ because the angle between this side and the diagonal is ϕ. Its horizontal side has length $|OP'| = \rho\sin\phi$. On the other hand, $|OP'| = \sqrt{x^2 + y^2} = r$ is the distance between $(0, 0, 0)$ and $(x, y, 0)$. Therefore $r = \rho\sin\phi$. Since $x = r\cos\theta$ and $y = r\sin\theta$, it is concluded that

$$x = \rho\sin\phi\cos\theta\,,\quad y = \rho\sin\phi\sin\theta\,,\quad z = \rho\cos\phi\,. \tag{35.3}$$

These equations define a transformation T of an ordered triple (ρ, ϕ, θ) to an ordered triple (x, y, z). It follows from Figure **35.3** that the range of the *zenith* angle ϕ must be the interval $[0, \pi]$ because ϕ is the angle between the positive z axis and a ray from the origin. If $\phi = 0$, the ray coincides with the positive z axis. If $\phi = \pi$, the ray is the negative z axis. Any ray with $\phi = \pi/2$ lies in the xy plane. The variable ρ is non-negative as it is the distance from the origin to a point in space.

Under the transformation (**35.3**)

$$\begin{aligned} T:&\quad (0,\phi,\theta) \;\to\; (0,0,0)\,;\\ T:&\quad (\rho,0,\theta) \;\to\; (0,0,\rho)\,;\\ T:&\quad (\rho,\pi,\theta) \;\to\; (0,0,-\rho)\,. \end{aligned}$$

So the plane $\rho = 0$ collapses into a single point, while the planes $\phi = 0$ and $\phi = \pi$ are mapped to the positive and negative z axes, respectively. It follows from the geometrical interpretation of the spherical coordinates that the transformation (**35.3**) is one-to-one if the range of (ρ, ϕ, θ) is restricted to $(0, \infty) \times (0, \pi) \times [0, 2\pi)$. The inverse transformation is then defined by

$$(\mathbf{35}.4)\quad \rho = \sqrt{x^2 + y^2 + z^2}\,,\quad \cot\phi = \frac{z}{r} = \frac{z}{\sqrt{x^2+y^2}}\,,\quad x\sin\theta = y\cos\theta\,,$$

here the last equation is solved for θ as explained in Section **32** where the polar coordinates are discussed. On the other hand, the image of the set $[0,\infty)\times[0,\pi]\times[0,2\pi]$ covers the entire space spanned by (x, y, z). Therefore, any region E in space is the image of a region E' under the transformation (**35.3**) that lies in $[0,\infty)\times[0,\pi]\times[0,2\pi]$, and the transformation $T:\ E' \to E$ is one-to-one except possibly at the boundary of E'. However, the multiple-counting occurs on the set that has no volume and for this reason the full range $[0,\infty)\times[0,\pi]\times[0,2\pi]$ of spherical coordinates can be used for purposes of integration. As noted earlier, any closed interval of length 2π may be chosen as the full range of the polar angle instead of $[0, 2\pi]$.

Coordinate Surfaces of Spherical Coordinates. It follows from the inverse transformation (**35.4**) that all points (x, y, z) that have the same value of $\rho = \rho_0$ form a sphere of radius ρ_0 centered at the origin because they are at the same distance ρ_0 from the origin. Naturally, the coordinate surfaces of θ are the half-planes described earlier when discussing cylindrical coordinates. Consider a ray from the origin that has the angle $\phi = \phi_0$ (here $0 < \phi_0 < \pi$) with the positive z axis. By rotating this ray about the z axis, all rays with the fixed value of ϕ are obtained. Therefore, the coordinate surface $\phi = \phi_0$ is a circular cone whose axis is the z axis. For small values of ϕ, the cone is a narrow cone about the positive z axis. The cone becomes wider as ϕ increases so that it coincides with the xy plane when $\phi = \pi/2$. For $\phi > \pi/2$, the cone lies below the xy plane, and it eventually collapses into the negative z axis as soon as ϕ reaches the value π. Thus, the coordinate surfaces of the spherical coordinates are

$$\begin{aligned} T:&\ \rho = \rho_0 &&\to\quad x^2+y^2+z^2 = \rho_0^2 &&\text{(sphere)},\\ T:&\ \phi = \phi_0 &&\to\quad z = \cot(\phi_0)\sqrt{x^2+y^2} &&\text{(cone)},\\ T:&\ \theta = \theta_0 &&\to\quad y\cos\theta_0 = x\sin\theta_0 &&\text{(half-plane)}. \end{aligned}$$

The coordinate surfaces of spherical coordinates are depicted in Figure **35.4**. A point in space can be viewed as the point of intersection of three coordinate

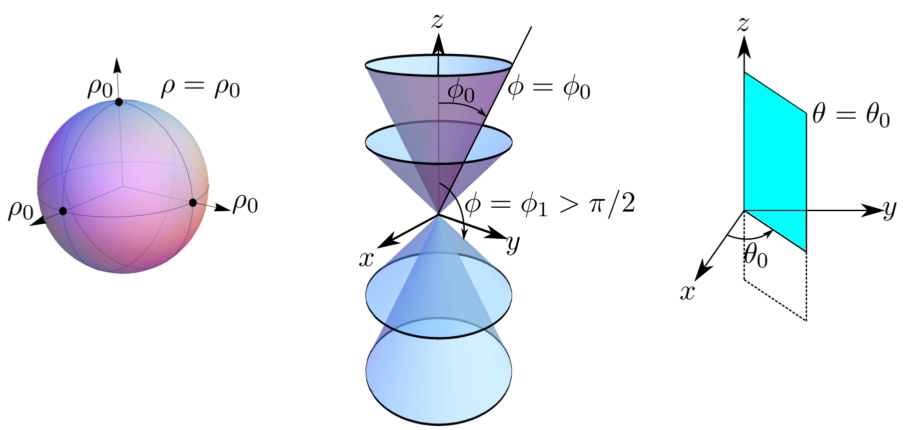


FIGURE **35.4**. Coordinate surfaces of spherical coordinates: Spheres $\rho = \rho_0$, circular cones $\phi = \phi_0$, and half-planes $\theta = \theta_0$ bounded by the z axis. In particular, $\phi = 0$ and $\phi = \pi$ describe the positive and negative z axes, respectively, and the cone with the angle $\phi = \pi/2$ becomes the xy plane.

surfaces: the sphere, cone, and half-plane (if the positive and negative z axes are viewed at the limit cases of the cone $\phi = \phi_0$ where $\phi_0 \to 0$ and $\phi_0 \to \pi$). If E is bounded by coordinate surfaces of spherical coordinates, then E is the image of a rectangular box $E' = [\rho_1, \rho_2] \times [\phi_1, \phi_2] \times [\theta_1, \theta_2]$.

EXAMPLE **35.3**. *Let E be the portion of the solid bounded by the sphere $x^2+y^2+z^2 = 4$ and the cone $z^2 = 3(x^2+y^2)$ that lies in the first octant. Find the region E' spanned by (ρ, ϕ, θ) in $[0, \infty) \times [0, \pi] \times [0, 2\pi]$ that is mapped into E by the transformation $(\rho, \phi, \theta) \to (x, y, z)$.*

SOLUTION: The region E has four boundaries: the sphere, the cone $z = \sqrt{3}\sqrt{x^2 + y^2}$, the xz plane ($x \geq 0$), and the yz plane ($y \geq 0$). Writing the equations of the boundary surfaces in spherical coordinates the corresponding boundary surfaces of E' are obtained:

$$\begin{aligned} x^2 + y^2 + z^2 = 4 \quad &\Rightarrow \quad \rho = 2\,, \\ z = \sqrt{3}\sqrt{x^2 + y^2} \quad &\Rightarrow \quad \cot\phi = \sqrt{3} \quad \text{or} \quad \phi = \pi/6\,, \\ y = 0,\ x > 0 \quad &\Rightarrow \quad \theta = 0\,, \\ x = 0,\ y > 0 \quad &\Rightarrow \quad \theta = \pi/2\,. \end{aligned}$$

The region E is intersected by all spheres with radii $0 < \rho \leq 2$, all cones with angles $0 < \phi \leq \pi/6$, and all half-planes with angles $0 \leq \theta \leq \pi/2$. Therefore

$$T: \ E' = [0, 2] \times [0, \pi/6] \times [0, \pi/2] \ \to \ E\,.$$

□

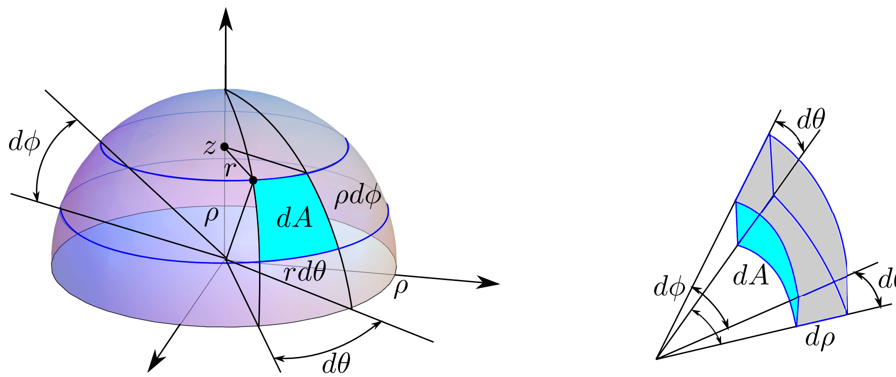


FIGURE **35.5**. **Left**: The base of a partition element in spherical coordinates is a portion of a sphere of radius ρ cut out by two cones with the angles ϕ and $\phi+d\phi$, where $d\phi = \Delta\phi$ and by two half-planes with the angles θ and $\theta + d\theta$, where $d\theta = \Delta\theta$. Its area is $dA = (\rho d\phi) \cdot (r d\theta) = \rho^2 \sin\phi \, d\phi \, d\theta$ if only terms linear in $d\phi$ and $d\theta$ are retained. **Right**: A partition element has the height $d\rho = \Delta\rho$ as it lies between two spheres whose radii differ by $d\rho$. So, its volume is $dV = dA d\rho = \rho^2 \sin\phi d\rho d\phi d\theta = J dV'$ and the Jacobian of spherical coordiantes is $J = \rho^2 \sin\phi$.

35.4. Triple integrals in spherical coordinates. Suppose that f is integrable on a bounded closed region E. A triple integral in spherical coordinates is obtained by partitioning the integration region E by coordinate surfaces of spherical coordinates, that is, spheres, cones, and half-planes, constructing the Riemann sum (**34.3**), and taking its limit under a refinement of the partition. Let E' be mapped onto a region E under the transformation (**35.3**). If E is bounded, then E' must be bounded, too. Indeed, E lies in a ball of sufficiently large radius and hence the range of ρ in E' must be bounded, whereas the range of (ϕ, θ) is always bounded. Therefore E' is contained in the rectangular box $[a, b] \times [c, d] \times [s, q]$ (E is contained in the image of this box). Consider a rectangular partition of E' by equispaced planes $\rho = \rho_i$, $\phi = \phi_j$, and $\theta = \theta_k$ such that

$$\begin{aligned}
&\rho_0 = a\,, \quad \rho_i = \rho_{i-1} + \Delta\rho\,, \quad \Delta\rho = (b-a)/N_1\,, \quad i = 1, 2, ..., N_1\,,\\
&\phi_0 = c\,, \quad \phi_j = \phi_{j-1} + \Delta\phi\,, \quad \Delta\phi = (d-c)/N_2\,, \quad j = 1, 2, ..., N_2\,,\\
&\theta_0 = s\,, \quad \theta_k = \theta_{k-1} + \Delta\theta\,, \quad \Delta\theta = (q-s)/N_3\,, \quad k = 1, 2, ..., N_3\,.
\end{aligned}$$

The volume of each partition element is $\Delta V' = \Delta\rho\Delta\phi\Delta\theta$. The rectangular partition of E' induces a partition of E by spheres, cones, and half-planes so that

$$T:\ E'_{ijk} = [\rho_i, \rho_i + \Delta\rho] \times [\phi_j, \phi_j + \Delta\phi] \times [\theta_k, \theta_k + \Delta\theta] \ \to\ E_{ijk}$$

Each partition element E_{ijk} is bounded by two spheres whose radii differ by $\Delta\rho$, by two cones whose angles differ by $\Delta\phi$, and by two half-planes the angle between which is $\Delta\theta$ as shown in Fig. **35.5**. In the limit $\mathbf{N} = \langle N_1, N_2, N_3\rangle \to \infty$, the numbers $\Delta\rho$, $\Delta\phi$, and $\Delta\theta$ tend to zero. When calculating the volume of partition elements E_{ijk} only terms linear in $d\rho = \Delta\rho$, $d\phi = \Delta\phi$, and $d\theta = \Delta\theta$ should be kept so that the volume of E_{ijk} can be written in the form

$$\Delta V_{ijk} = J_{ijk}\,\Delta V'\,.$$

To find J_{ijk}, note that E_{ijk} lies between two spheres of radii ρ_i and $\rho_i + \Delta\rho$, its volume can be written as

$$\Delta V_{ijk} = \Delta\rho\,\Delta A_{ijk}\,,$$

where ΔA_{ijk} is the area of the portion of the sphere of radius ρ_i that lies between two cones and two half-planes. Any half-plane $\theta = \theta_k$ intersects the sphere $\rho = \rho_i$ along a half-circle of radius ρ_i. The arc length of the portion of this circle that lies between the two cones $\phi = \phi_j$ and $\phi = \phi_j + \Delta\phi$ is therefore $\Delta l_{ij} = \rho_i\,\Delta\phi$. The cone $\phi = \phi_j$ intersects the sphere $\rho = \rho_i$ along a circle of radius $r_{ij} = \rho_i \sin\phi_j$ (see the text above (**35.3**)). Hence, the arc length of the portion of this circle of intersection that lies between the half-planes $\theta = \theta_k$ and $\theta = \theta_k + \Delta\theta$ is $\Delta m_{ijk} = r_{ij}\,\Delta\theta = \rho_i \sin\phi_j\,\Delta\theta$. The area ΔA_{ijk} can be approximated by the area of a rectangle with adjacent sides Δl_{ij} and Δm_{ijk} because the circles that contain the arcs of length Δl_{ij} and Δm_{ijk} are intersecting at the right angle. Since only terms linear in $\Delta\phi$ and $\Delta\theta$ are to be retained, one can write

$$\Delta A_{ijk} = \Delta l_{ij}\,\Delta m_{ijk} = \rho_i^2 \sin\phi_j\,\Delta\phi\,\Delta\theta \quad\Rightarrow\quad \Delta V_{ijk} = \rho_i^2 \sin\phi_j \Delta V'$$

Let us choose $(\rho_i, \phi_j, \theta_k)$ to be sample points in E'_{ijk} so that $T:\ (\rho_i, \phi_j, \theta_k) \to \mathbf{r}_{ijk}$, the corresponding sample points in E_{ijk}. Put $J = \rho^2 \sin\phi$. Then by Eq. (**34.3**),

$$\iiint_E f(x, y, z)dV = \lim_{\mathbf{N}\to\infty} \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{k=1}^{N_3} f(\mathbf{r}_{ijk}) J_{ijk} \Delta V'$$

$$= \iiint_{E'} f\Big(\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi\Big)\rho^2 \sin\phi\, dV' \qquad (\mathbf{35.5})$$

because the triple sum is the Riemann sum of the function fJ expressed in spherical coordinates over the region E' and, hence, should converge to the triple integral of fJ over E', provided fJ is integrable on E'. One can prove that if f is integrable on E, then fJ is integrable on E'.

Thus, Equation (**35.5**) defines the triple integral in spherical coordinates.

The coefficient J in the volume transformation

$$dV = J\,dV'\,,\quad J = \rho^2 \sin\phi\,,$$

is called the *Jacobian* of the transformation to spherical coordinates. It vanishes in the planes $\rho = 0$, $\phi = 0$, and $\phi = \pi$ of a Euclidean space spanned

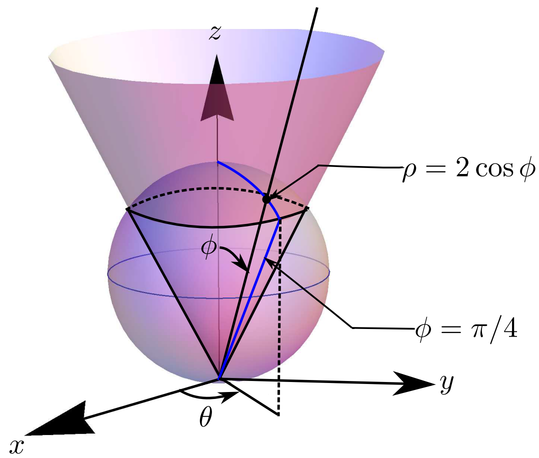


FIGURE **35.6**. An illustration to Example **35.5**. Any ray in space is defined as the intersection of a cone with an angle ϕ and a half-plane with an angle θ. To find E' whose image is the depicted solid E, note that any such ray intersects E along a *single* straight line segment if $0 \le \phi \le \pi/4$ where the cone $\phi = \phi/4$ is a part of the boundary of E. Due to the axial symmetry of E, there is no restriction on the range of θ, i.e., $0 \le \theta \le 2\pi$ in E'. The range of ρ is determined by the length of the segment of intersection of the ray at fixed ϕ and θ with E: $0 \le \rho \le 2\cos\phi$ where $\rho = 2\cos\phi$ is the equation of the top boundary of E in spherical coordinates.

by ordered triples (ρ, ϕ, θ). As noted before, the transformation (**35.3**) is not one-to-one on them. The Jacobian vanishes only in the boundary of $[0, \infty) \times [0, \pi] \times [0, 2\pi]$ whose image covers the whole space spanned by (x, y, z). It is worth noting that in any open region in $[0, \infty) \times [0, \pi] \times [0, 2\pi]$ the Jacobian is strictly positive (it does not vanish) and the transformation T defined in (**35.3**) has the inverse defined in (**35.4**).

The main advantage of converting a multiple integral to curvilinear coordinates is a simplification of the region of integration. Therefore *the conversion of a triple integral to spherical coordinates is advantageous if the region of integration is bounded by coordinate surfaces of spherical coordinates.*

EXAMPLE **35.4**. *Evaluate* $\iiint_E z\,dV$ *if* E *lies in the first octant and is bounded by the planes* $y = x$, $y = \sqrt{3}x$, $z = 0$, *and the sphere* $x^2+y^2+z^2 = 1$.

Solution: The equations of the boundaries of E in spherical coordinates are:

$$\begin{aligned} y = x\,,\ x > 0 \quad &\Rightarrow \quad \tan\theta = 1 \quad \Rightarrow \quad \theta = \pi/4\,, \\ y = \sqrt{3}x\,,\ x > 0 \quad &\Rightarrow \quad \tan\theta = \sqrt{3} \quad \Rightarrow \quad \theta = \pi/3\,, \\ z = 0 \quad &\Rightarrow \quad \phi = \pi/2\,, \\ x^2 + y^2 + z^2 = 1 \quad &\Rightarrow \quad \rho = 1\,. \end{aligned}$$

Since E is in the first octant, $0 \leq \phi \leq \pi/2$. Thus,

$$T: \ E' = [0,1] \times [0,\pi/2] \times [\pi/4, \pi/3] \ \to \ E\,.$$

Therefore, by Eq. (**35.5**) and Fubini's theorem

$$\begin{aligned} \iiint_E z\,dV &= \iiint_{E'} \rho\cos\phi J\,dV' = \iiint_{E'} \rho^3\cos\phi\sin\phi\,dV' \\ &= \int_{\pi/4}^{\pi/3} d\theta \int_0^{\pi/2} \cos\phi\sin\phi\,d\phi \int_0^1 \rho^3 d\rho \\ &= \frac{\pi}{12}\cdot\left(\frac{1}{2}\sin^2\phi\Big|_0^{\pi/2}\right)\cdot\frac{1}{4} = \frac{\pi}{96}\,. \end{aligned}$$

□

Example **35.5**. *Find the volume of the solid E bounded by the sphere $x^2+y^2+z^2 = 2z$ and the cone $z = \sqrt{x^2+y^2}$.*

Solution: By completing the squares, the equation $x^2+y^2+z^2 = 2z$ is written in the standard form $x^2+y^2+(z-1)^2 = 1$, which describes a sphere of unit radius centered at $(0,0,1)$. So E is bounded from the top by this sphere, while the bottom boundary of E is the cone, and E has no other boundaries (see Figure **35.6**). In spherical coordinates, the equations of the boundary surfaces are

$$\begin{aligned} x^2+y^2+z^2 = 2z \quad &\Rightarrow \quad \rho^2 = 2\rho\cos\phi \quad \Rightarrow \quad \rho = 2\cos\phi\,, \\ z = \sqrt{x^2+y^2} \quad &\Rightarrow \quad \cot\phi = 1 \quad \Rightarrow \quad \phi = \pi/4\,. \end{aligned}$$

The boundaries of E impose no restriction on θ, which can therefore be taken over its full range. Since E lies above the xy plane, $0 \leq \phi \leq \pi/4$. Hence,

$$T: \quad E' = \Big\{(\rho,\phi,\theta)\,|\,0 \leq \rho \leq 2\cos\phi\,, \quad (\phi,\theta) \in [0,\pi/4]\times[0,2\pi]\Big\} \ \to \ E\,.$$

Since the range of ρ depends on the other variables, the integration with respect to it must be carried out first when converting the triple integral over E' into an iterated integral (E' is simple in the direction of the ρ axis and the projection of E' onto the $\phi\theta$ plane is the rectangle $[0,\pi/4]\times[0,2\pi]$). The order in which the integration with respect to θ and ϕ is carried out is irrelevant because the angular variables range over a rectangle. One has

$$\begin{aligned} V(E) &= \iiint_E dV = \iiint_{E'} J\,dV' = \iiint_{E'} \rho^2 \sin\phi\, dV' \\ &= \int_0^{2\pi}\int_0^{\pi/4} \sin\phi \int_0^{2\cos\phi} \rho^2\, d\rho\, d\phi\, d\theta \\ &= \frac{8}{3}\int_0^{2\pi} d\theta \int_0^{\pi/4} \cos^3\phi \sin\phi\, d\phi = \frac{16\pi}{3}\int_{1/\sqrt{2}}^1 u^3\, du = \pi, \end{aligned}$$

where the change of variables $u = \cos\phi$ has been carried out in the last integral. □

35.5. Study Problems.

Problem **35.1**. *Convert the iterated integral in cylindrical coordinates*

$$\int_0^{\pi/2}\int_0^1\int_r^{\sqrt{2-r^2}} r^2 \cos\theta \sqrt{z^2+r^2}\, dz dr d\theta$$

to spherical coordinates and evaluate it.

SOLUTION: In order to convert the given iterated integral to spherical coordinates, one has first to reconstruct the triple integral in rectangular coordinates which has been converted to cylindrical coordinates. In particular, one has to find the integration region. The triple integral in rectangular coordinates is then converted to spherical coordinates. First, note that the volume transformation law in cylindrical coordinates is

$$dxdydz = Jdzdrd\theta\,, \quad J = r\,.$$

Therefore the integrand in the triple integral is

$$f(x,y,z) = r\cos\theta\sqrt{z^2+r^2} = x\sqrt{x^2+y^2+z^2}\,,$$

where the relations between the cylindrical and rectangular coordinates have been used. If E is the integration region of the triple integral in rectangular coordinates and D_{xy} is its projection onto the xy plane, then D_{xy} is the image of the rectangle $D'_{xy} = [0,1]\times[0,\pi/2]$ in the polar plane according to Eq. (**35.2**). Therefore D_{xy} is the part of the disk $x^2+y^2 \le 1$ in the first quadrant:

$$D_{xy} = \{\ (x,y)\mid x^2+y^2\le 1\,,\ x\ge 0\,,\ y\ge 0\}\,.$$

Converting the equation of the top and bottom boundaries of E back to the rectangular coordinates:

$$z=\sqrt{2-r^2} \quad\Rightarrow\quad x^2+y^2+z^2 = 2\,, \qquad z = r \quad\Rightarrow\quad z=\sqrt{x^2+y^2}$$

it is concluded that E is the part of a solid bounded by the sphere $x^2+y^2+z^2 = 2$ and the cone $z=\sqrt{x^2+y^2}$ that lies in the first octant (as $0\le\theta\le\pi/2$). Expressing the function and the volume element in spherical coordinates, the given iterated integral is proved to be equal to the triple integral in spherical coordinates

$$\iiint_E f dV = \iiint_{E'} \rho^2 \sin\phi \cos\theta \, J \, dV' , \quad J = \rho^2 \sin\phi .$$

Next, one has to find E'. Rewriting the equations of the boundary surfaces of E in spherical coordinates

$$x^2 + y^2 + z^2 = 2 \quad \Rightarrow \quad \rho^2 = 2 \quad \Rightarrow \quad \rho = \sqrt{2} ,$$
$$z = \sqrt{x^2 + y^2} \quad \Rightarrow \quad \cot\phi = 1 \quad \Rightarrow \quad \phi = \pi/4 ,$$

and taking into account the side boundaries $\theta = 0$ and $\theta = \pi/2$, it is concluded that

$$E' = [0, \sqrt{2}] \times [0, \pi/4] \times [0, \pi/2] .$$

Therefore by Fubini's theorem

$$\begin{aligned} \iiint_E f dV &= \int_0^{\pi/2} \cos\theta \, d\theta \int_0^{\pi/4} \sin^2\phi \, d\phi \int_0^{\sqrt{2}} \rho^4 d\rho \\ &= 1 \cdot \Big(\frac{1}{2}\phi - \frac{1}{4}\sin(2\phi)\Big)\Big|_0^{\pi/4} \cdot \frac{1}{5}\,(\sqrt{2})^5 \\ &= \Big(\frac{\pi}{8} - \frac{1}{4}\Big)\frac{4\sqrt{2}}{5} = \frac{\pi - 2}{5\sqrt{2}} , \end{aligned}$$

where the double angle formula $\sin^2\phi = (1 - \cos(2\phi))/2$ has been used to evaluate the integral with respect to ϕ. □

35.6. Exercises.

1–3. Sketch the solid E onto which the specified region E' is mapped by the transformation $(r, \theta, z) \to (x, y, z)$.

1. $E' = [0, 3] \times [-\pi/4, \pi/4] \times [0, 1]$;

2. $E' = \{(r, \theta, z) \,|\, r - 1 \le z \le 1 - r , \quad (r, \theta) \in [0, 1] \times [0, \pi/2]\}$;

3. $E' = \{(r, \theta, z) \,|\, 0 \le z \le 4 - r^2 , \quad (r, \theta) \in [0, 2] \times [0, \pi/2]\}$.

4–7. Given the solid E, find the region E' whose image is E under the transformation to cylindrical coordinates and the transformation is one-to-one except perhaps on the boundary of E'.

4. E is bounded by the cylinder $x^2 + y^2 = 1$, the paraboloid $z = x^2 + y^2$, and the plane $z = 0$;

5. E is bounded by the cone $(z - 1)^2 = x^2 + y^2$ and the cylinder $x^2 + y^2 = 1$;

6. E is bounded by the paraboloid $z = x^2 + y^2$, the cylinder $x^2 + y^2 = 2x$, and the plane $z = 0$;

7. E is the part of the ball $x^2 + y^2 + z^2 \le a^2$ in the first octant.

8–14. Evaluate the triple integral by converting it to cylindrical coordinates.

8. $\iiint_E |z| dV$, where E is bounded by the sphere $x^2 + y^2 + z^2 = 4$ and lies inside the cylinder $x^2 + y^2 = 1$;

9. $\iiint_E (x^2y + y^3)dV$, where E lies beneath the paraboloid $z = 1 - x^2 - y^2$ in the first octant;
10. $\iiint_E y dV$, where E is enclosed by the planes $z = 0$, $x + y - z = -5$, and by the cylinders $x^2 + y^2 = 1$, $x^2 + y^2 = 4$;
11. $\iiint_E dV$, where E is enclosed by the cylinder $x^2 + y^2 = 2x$, by the plane $z = 0$, and by the cone $z = \sqrt{x^2 + y^2}$;
12. $\int\int\int_E yz\, dV$, where E lies beneath the paraboloid $z = a^2 - x^2 - y^2$ in the first octant;
13. $\iiint_E (x^2 + y^2)dV$, where E is bounded by the surfaces $x^2 + y^2 = 2z$ and $z = 2$;
14. $\int_E xyzdV$, where E lies in the first octant and is bounded by the surfaces $x^2 + y^2 = az$, $x^2 + y^2 = bz$, $xy = c^2$, $xy = k^2$, $y = \alpha x$, $y = \beta x$, and $0 < a < b$, $0 < \alpha < \beta$, $0 < c < k$.

15–17. Sketch the solid E onto which the specified region E' is mapped by the transformation $(\rho, \phi, \theta) \to (x, y, z)$

15. $E' = [0, 1] \times [0, \pi/2] \times [0, \pi/4]$;
16. $E' = [1, 2] \times [0, \pi/4] \times [0, \pi/2]$;
17. $E' = \{(\rho, \phi, \theta) \,|\, \frac{1}{\cos\phi} \le \rho \le 2\,,\ (\phi, \theta) \text{ in } [0, \pi/6] \times [0, \pi]\}$.

18–20. Given the solid E, find the region E' whose image is E under the transformation to spherical coordinates and the transformation is one-to-one except perhaps on the boundary of E'.

18. E lies between two spheres $x^2 + y^2 + z^2 = 1$ and $x^2 + y^2 + z^2 = 4$ in the first octant;
19. E is defined by the inequalities $z^2 \le 3(x^2+y^2)$ and $x^2+y^2+z^2 \le a^2$;
20. E is bounded by the sphere $x^2+y^2+z^2 = a^2$ and by the half-planes $y = \sqrt{3}x$, $y = x/\sqrt{3}$ where $x \ge 0$;
21. E is bounded by the surface $x^2 + y^2 + z^2 = 4z$.

22–27. Evaluate each of the following triple integrals by converting it to spherical coordinates.

22. $\iiint_E (x^2 + y^2 + z^2)^3 dV$, where E is the ball of radius a centered at the origin;
23. $\iiint_E y^2 dV$, where E is bounded by yz plane and the hemispheres $x = \sqrt{1 - y^2 - z^2}$, $x = \sqrt{4 - y^2 - z^2}$. Hint: Use spherical coordinates in which the polar angle is defined in the yz plane;
24. $\iiint_E xyzdV$, where E is enclosed by the cone $z = \sqrt{3}\sqrt{x^2 + y^2}$ and the spheres $x^2 + y^2 + z^2 = a^2$, and $a = 1, 2$;
25. $\iiint_E zdV$, where E is the part of the ball $x^2 + y^2 + z^2 \le 1$ that lies below the cone $z = \sqrt{3x^2 + 3y^2}$;
26. $\iiint_E zdV$, where E lies in the first octant between the planes $y = 0$ and $x = \sqrt{3}\,y$, above the cone $z = \sqrt{x^2 + y^2}$, and inside the sphere $x^2 + y^2 + z^2 = 4$;

27. $\iiint_E \sqrt{x^2+y^2+z^2}dV$, where E is bounded by the sphere $x^2+y^2+z^2=z$.

28–29. Sketch the region of integration in the triple integral corresponding to the given iterated integral, write the triple integral in spherical coordinates, and then evaluate it.

28. $\int_0^1 \int_0^{\sqrt{1-x^2}} \int_{\sqrt{x^2+y^2}}^1 zdzdydx$;

29. $\int_0^1 \int_0^{\sqrt{1-x^2}} \int_{\sqrt{x^2+y^2}}^{\sqrt{2-x^2-y^2}} z^2dzdydx$.

30. Sketch the solid whose volume is given by the iterated integral in the spherical coordinates:
$\int_0^{\pi/2} \int_0^{\pi/4} \int_0^{2/\cos\phi} \rho^2 \sin\phi d\rho d\phi d\theta$
Write the integral in cylindrical coordinates and evaluate it.

31. Sketch the region of integration in the triple integral corresponding to the given iterated integral write the triple integral in cylindrical coordinates, and then evaluate it:
$\int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \int_0^{1-x^2-y^2} zdzdydx$

32–33. Convert the triple integral $\iiint_E f(x^2+y^2+z^2)dV$ to iterated integrals in cylindrical and spherical coordinates if E is bounded by the given surfaces.

32. $z=x^2+y^2$, $y=x$, $x=1$, $y=0$, $z=0$;

33. $z^2=x^2+y^2$, $x^2+y^2+z^2=2z$, $x=y/\sqrt{3}$, $x=y\sqrt{3}$ where $x \geq 0$ and $y \geq 0$.

34–36. Use spherical coordinates to find the volume of a solid bounded by the given surfaces.

34. $x^2+y^2+z^2=a^2$, $x^2+y^2+z^2=b^2$, $z=c\sqrt{x^2+y^2}$, $c>0$, and $0<a<b$;

35. $(x^2+y^2+z^2)^2=a^2(x^2+y^2-z^2)$, $a>0$;

36. $(x^2+y^2+z^2)^3=3xyz$.

37. Find the volume of a solid bounded by the surfaces $x^2+z^2=a^2$, $x^2+z^2=b^2$, $x^2+y^2=z^2$ where $x>0$ and $0<a<b$.

36. Change of Variables in Triple Integrals

36.1. Change of variables in space. Consider a transformation T of an open region E' in space into a region E defined by $x = x(u, v, w)$, $y = y(u, v, z)$, and $z = z(u, v, w)$; that is, for every point (u, v, w) in E', these functions define an image point (x, y, z) in E.

DEFINITION **36.1**. (Jacobian of a Transformation).
Suppose that a transformation of an open region E' into E has continuous partial derivatives. The quantity

$$\frac{\partial(x,y,z)}{\partial(u,v,w)} = \det\begin{pmatrix} x'_u & y'_u & z'_u \\ x'_v & y'_v & z'_v \\ x'_w & y'_w & z'_w \end{pmatrix}$$

is called the Jacobian of the transformation.

If the determinant is expanded over the first column, then it can also be written as the triple product:

$$\frac{\partial(x,y,z)}{\partial(u,v,w)} = \boldsymbol{\nabla} x \cdot (\boldsymbol{\nabla} y \times \boldsymbol{\nabla} z) \tag{36.1}$$

The technical details are left to the reader as an exercise.

EXAMPLE **36.1**. *Find the Jacobian of the transformation to spherical coordinates.*

SOLUTION: Let $(u, v, w) = (\rho, \phi, \theta)$ be the spherical coordinates. Then using Eq. (**35.3**) and Definition **36.1**,

$$\begin{aligned}
\frac{\partial(x,y,z)}{\partial(\rho,\phi,\theta)} &= \det\begin{pmatrix} x'_\rho & y'_\rho & z'_\rho \\ x'_\phi & y'_\phi & z'_\phi \\ x'_\theta & y'_\theta & z'_\theta \end{pmatrix} \\
&= \det\begin{pmatrix} \sin\phi\cos\theta & \sin\phi\sin\theta & \cos\phi \\ \rho\cos\phi\cos\theta & \rho\cos\phi\sin\theta & -\rho\sin\phi \\ -\rho\sin\phi\sin\theta & \rho\sin\phi\cos\theta & 0 \end{pmatrix} \\
&= \sin\phi\cos\theta(\rho^2\sin^2\phi\cos\theta - 0) - \sin\phi\sin\theta(0 - \rho^2\sin^2\phi\sin\theta) \\
&\quad + \cos\phi[\rho^2\cos\phi\sin\phi(\cos^2\theta + \sin^2\theta)] \\
&= \rho^2\sin^3\phi(\cos^2\theta + \sin^2\theta) + \rho^2\sin\phi\cos^2\phi \\
&= \rho^2\sin\phi(\sin^2\phi + \cos^2\phi) \\
&= \rho^2\sin\phi\,.
\end{aligned}$$

□

If no two points in E' have the same image point, the transformation is *one-to-one*, and there is a *one-to-one correspondence* between points of E and E'. The inverse transformation exists and is defined by the functions $u = u(x, y, z)$, $v = v(x, y, z)$, and $w = w(x, y, z)$. Suppose that these functions have continuous partial derivatives so that the gradients of these functions do not vanish. Then, as shown in Section **24.2**, the equations

$$u(x, y, z) = u_0\,, \quad v(x, y, z) = v_0\,, \quad w(x, y, z) = w_0\,,$$

define smooth surfaces. A point (x_0, y_0, z_0) in E is the point of intersection of three coordinate planes $x = x_0$, $y = y_0$, and $z = z_0$. Alternatively, it can be viewed as the point of intersection of three *coordinate surfaces*, $u(x, y, z) = u_0$, $v(x, y, z) = v_0$, and $w(x, y, z) = w_0$, where the point (u_0, v_0, w_0) in E' is mapped to (x_0, y_0, z_0) by the transformation T.

Definition **36.2**. (Change of Variables in Space). *A one-to-one transformation of an open region E' defined by $x = x(u, v, w)$, $y = y(u, v, w)$, and $z = z(u, v, w)$ is called a* change of variables *(or a change of coordinates) if the functions $x(u, v, w)$, $y(u, v, w)$, and $z(u, v, w)$ have continuous partial derivatives on E'.*

The ordered triples (u, v, w) are also called *curvilinear coordinates* in space. There is a three-dimensional analog of Theorem **33.1** that establishes a useful criterion for a transformation to be a change of variables.

Theorem **36.1**. (Inverse Function Theorem for Three Variables) *Let a transformation $(u, v, w) \to (x, y, z)$ be defined on an open set U' containing a point (u_0, v_0, w_0). Let the point (x_0, y_0, z_0) be the image of the point (u_0, v_0, w_0). Suppose that the functions $x(u, v, w)$, $y(u, v, w)$, and $z(u, v, w)$ have continuous partial derivatives in U' and the Jacobian of the transformation does not vanish at the point (u_0, v_0, w_0). Then there exists a open neighborhood U of (x_0, y_0, z_0) in which the inverse transformation exists, the functions $u = u(x, y, z)$, $v = v(x, y, z)$, and $w = w(x, y, z)$ have continuous partial derivatives, and the Jacobian of the inverse transformation is given by*

$$(\mathbf{36.2}) \qquad \frac{\partial(u, v, w)}{\partial(x, y, z)} = \det\begin{pmatrix} u'_x & v'_x & w'_x \\ u'_y & v'_y & w'_y \\ u'_z & v'_z & w'_z \end{pmatrix} = \cfrac{1}{\cfrac{\partial(x, y, z)}{\partial(u, v, w)}}$$

Equation (**36.2**) shows that the Jacobian of the direct and inverse transformations are reciprocals of one another. The left side of (**36.2**) defines the Jacobian as a function of (x, y, z), while the right side defines the Jacobian as a function of new variables (u, v, w). By expanding the determinant over the first column it is not difficult to verify that

$$(\mathbf{36.3}) \qquad \frac{\partial(u, v, w)}{\partial(x, y, z)} = \boldsymbol{\nabla} u \cdot (\boldsymbol{\nabla} v \times \boldsymbol{\nabla} w)$$

which the analog of (**36.1**).

The non-vanishing gradients $\boldsymbol{\nabla} u$, $\boldsymbol{\nabla} v$, and $\boldsymbol{\nabla} w$ are normal to the level surfaces of the functions $u(x, y, z)$, $v(x, y, z)$, and $w(x, y, z)$. If the Jacobian does not vanish at a particular point, then by its continuity it cannot vanish in a neighborhood of that point. Therefore, by Eq. (**36.3**), the gradients are not coplanar in a neighborhood of a point at which the Jacobian does not vanish and the level surfaces in this neighborhood are always intersecting at a *single* point. The latter means that the system of equations $u(x, y, z) = u$, $v(x, y, z) = v$, $w(x, y, z) = w$ has a *unique* solution $x = x(u, v, w)$, $y = y(u, v, w)$, $z = z(u, v, w)$ (or that the transformation is one-to-one) near any point at which the Jacobian does not vanish. Furthermore, the image of a rectangular box in a neighborhood of (u_0, v_0, w_0) looks like a "deformed" rectangular box in a neighborhood of (x_0, y_0, z_0) whose six faces are level surfaces of the functions $u(x, y, z)$, $v(x, y, z)$, and $w(x, y, z)$. In other words, level surfaces of the functions $u(x, y, z)$, $v(x, y, z)$, and $w(x, y, z)$ are *coordinate surfaces* in a region where the Jacobian is not zero, just like planes are coordinate surfaces of the rectangular coordinate system, or spheres, cones, and half-planes are coordinate surfaces of the spherical coordinate system. This observation is crucial to determine the volume transformation law under a change of variables.

36.2. The Volume Transformation Law. It is convenient to introduce the following notations: $\langle u, v, w\rangle = \mathbf{r}'$ and $\langle x, y, z\rangle = \mathbf{r}$ so that the change of variables is written as

$$(\mathbf{36}.4)\qquad \mathbf{r} = \Big\langle x(\mathbf{r}'),\ y(\mathbf{r}'),\ z(\mathbf{r}')\Big\rangle \quad\text{or}\quad \mathbf{r}' = \Big\langle u(\mathbf{r}),\ v(\mathbf{r}),\ w(\mathbf{r})\Big\rangle .$$

It is assumed that the above relations define a change variable in an open region E'. Let

$$E_0' = [u_0, u_0 + \Delta u] \times [v_0, v_0 + \Delta v] \times [w_0, w_0 + \Delta w]$$

be a rectangular box in E' and E_0 be the image of E_0' under the change of variables. As noted before, the image E_0 is bounded by smooth surfaces as depicted in Figure **36.1**. If the values of Δu, Δv, and Δw can be made arbitrary small, owing to the smoothness of coordinate surfaces, the boundary surfaces of E_0 can be approximated by tangent planes to them and the volume of E_0 is then approximated by the volume of the polyhedron bounded by these planes. This implies, in particular, that when calculating the volume, only terms linear in Δu, Δv, and Δw are to be retained, while their higher powers are neglected. Therefore the volumes of E_0 and E_0' must be proportional:

$$\Delta V = J\,\Delta V' , \qquad \Delta V' = \Delta u\,\Delta v\,\Delta w .$$

The objective is to calculate J. By the examples of cylindrical and spherical coordinates, J is a function of the point (u_0, v_0, w_0) at which the rectangle E_0' is taken. The derivation of J is fully analogous to the two-variable case.

An infinitesimal rectangular box in E_0' and its image under the coordinate transformation are shown in Fig. **36.1**. Let the position vectors of the

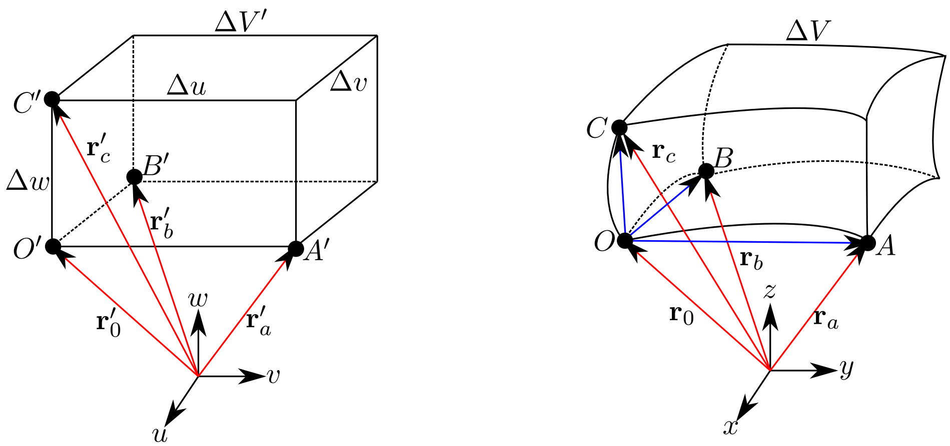


FIGURE **36.1**. **Left**: A rectangular box E'_0 with small sides $du = \Delta u$, $dv = \Delta v$, $dw = \Delta w$ so that its volume $\Delta V' = dudvdw$. **Right**: The image of the rectangular box under a change of variables. The position vectors $\mathbf{r}_p$, where $p = 0, a, b, a$, are images of the position vectors $\mathbf{r}'_p$. The volume ΔV of the image is approximated by the volume of the parallelepiped with adjacent sides OA, OB, and OC. It is computed by linearization of ΔV in du, dv, and dw so that $\Delta V = Jdudvdw = J\Delta V'$ where $J > 0$ is the Jacobian of the change of variables. The approximation is justified in the limit $d\mathbf{r}' \to \mathbf{0}$ owing to the smoothness of the boundary surfaces of E_0.

points O', A', B', and C' be, respectively,

$$\begin{aligned}
\mathbf{r}'_0 &= \langle u_0, v_0, w_0 \rangle\,,\\
\mathbf{r}'_a &= \langle u_0 + \Delta u, v_0, w_0 \rangle = \mathbf{r}'_0 + \hat{\mathbf{e}}_1\,\Delta u,\\
\mathbf{r}'_b &= \langle u_0, v_0 + \Delta v, w_0 \rangle = \mathbf{r}'_0 + \hat{\mathbf{e}}_2\,\Delta v,\\
\mathbf{r}'_c &= \langle u_0, v_0, w_0 + \Delta w \rangle = \mathbf{r}'_0 + \hat{\mathbf{e}}_3\,\Delta w,
\end{aligned}$$

where $\hat{\mathbf{e}}_1 = \langle 1, 0, 0\rangle$, $\hat{\mathbf{e}}_2 = \langle 0, 1, 0\rangle$, and $\hat{\mathbf{e}}_3 = \langle 0, 0, 1\rangle$ are unit vectors along the first, second, and third coordinate axes, respectively. In other words, the segments $O'A'$, $O'B'$, and $O'C'$ are the adjacent sides of the rectangular box E'_0. Let O, A, B, and C be the images of O', A', B', and C'. Owing to the smoothness of the boundary surfaces of E_0, the volume ΔV of E_0 can be approximated by the volume of the parallelepiped with adjacent sides

$\mathbf{a} = \overrightarrow{OA}$, $\mathbf{b} = \overrightarrow{OB}$, and $\mathbf{c} = \overrightarrow{OC}$. Then

$$\begin{aligned}
\mathbf{a} &= \Big(x(\mathbf{r}_a') - x(\mathbf{r}_0'),\ y(\mathbf{r}_a') - y(\mathbf{r}_0'),\ z(\mathbf{r}_a') - z(\mathbf{r}_0')\Big) = (x_u',\ y_u',\ z_u')\,\Delta u,\\
\mathbf{b} &= \Big(x(\mathbf{r}_b') - x(\mathbf{r}_0'),\ y(\mathbf{r}_b') - y(\mathbf{r}_0'),\ z(\mathbf{r}_b') - z(\mathbf{r}_0')\Big) = (x_v',\ y_v',\ z_v')\,\Delta v,\\
\mathbf{c} &= \Big(x(\mathbf{r}_c') - x(\mathbf{r}_0'),\ y(\mathbf{r}_c') - y(\mathbf{r}_0'),\ z(\mathbf{r}_c') - z(\mathbf{r}_0')\Big) = (x_w',\ y_w',\ z_w')\,\Delta w,
\end{aligned}$$

where all the differences have been linearized, for instance,

$$x(\mathbf{r}_a') - x(\mathbf{r}_0') = x(\mathbf{r}_0' + \hat{\mathbf{e}}_1\,\Delta u) - x(\mathbf{r}_0') = \frac{\partial x}{\partial u}(\mathbf{r}_0')\Delta u = x_u'(\mathbf{r}_0')\,\Delta u\,.$$

The function $x(u,v,w)$ has continuous partial derivatives and, hence, differentiable. Therefore the error of the above approximation tends to zero faster than Δu as $\Delta u \to 0$. Similarly, owing to differentiability of the functions $x(\mathbf{r}')$, $y(\mathbf{r}')$, and $z(\mathbf{r}')$, all the differences in the components of the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ can be linearized in Δu, Δv, and Δw. The error of this approximation decreases to zero faster than Δu, Δv, Δw as the latter approach zero values.

In the limit $(\Delta u, \Delta v, \Delta w) \to (0,0,0)$, the volume of the image of the rectangular box E_0' is well approximated by the volume of the parallelepiped with adjacent sides $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$:

$$(\mathbf{36.5})\qquad \Delta V = |\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})| = \left|\det\begin{pmatrix} x_u' & y_u' & z_u' \\ x_v' & y_v' & z_v' \\ x_w' & y_w' & z_w' \end{pmatrix}\right| \Delta u\,\Delta v\,\Delta w = J\,\Delta V',$$

where the derivatives are evaluated at (u_0, v_0, w_0). The function J in (**36**.**5**) is the *absolute value* of the Jacobian of the transformation. Since the considered transformation is a change of variables, the Jacobian does not vanish. Equation (**36**.**5**) defines J as a function of new variables (u,v,w). By Equation (**36**.**2**), J can also be determined as a function of old variables (x,y,z):

$$(\mathbf{36.6})\qquad J = \frac{1}{\left|\dfrac{\partial(u,v,w)}{\partial(x,y,z)}\right|} = \left|\det\begin{pmatrix} u_x' & u_y' & u_z' \\ v_x' & v_y' & v_z' \\ w_x' & w_y' & w_z' \end{pmatrix}\right|^{-1}$$

36.3. Triple Integral in Curvilinear Coordinates. As in the case of double integrals, a change of variables in space can be used to simplify the evaluation of triple integrals. For example, if there is a change of variables whose coordinate surfaces form the boundary of the integration region E, then the new integration region E' is a rectangular box, and the limits in the corresponding iterated integral are greatly simplified in accordance with Fubini's theorem.

The derivation of the triple integral in curvilinear coordinates follows the same conceptual steps as in the case of spherical coordinates. Suppose the integration region E is the image of a closed bounded region E' under a transformation that defines a change of variables in a rectangular box $R'_E = [a,b] \times [c,d] \times [s,q]$ except perhaps on the boundary of R'_E and the rectangular box R'_E contains E'. Then E is contained in the image R_E of R'_E. The integrand f is extended to R_E by setting its values to zero for all points that are not in E. Consider a rectangular partition of E' by the planes $u = u_i$, $v = v_j$, and $w = w_k$ such that

$$\begin{aligned}
&u_0 = a\,, \quad u_i = u_{i-1} + \Delta u\,, \quad \Delta u = (b-a)/N_1\,, \quad i = 1, 2, ..., N_1\,,\\
&v_0 = c\,, \quad v_j = u_{j-1} + \Delta v\,, \quad \Delta v = (d-c)/N_2\,, \quad j = 1, 2, ..., N_2\,,\\
&w_0 = s\,, \quad w_k = w_{k-1} + \Delta w\,, \quad \Delta w = (q-s)/N_3\,, \quad k = 1, 2, ..., N_3\,.
\end{aligned}$$

This rectangular partition of E' corresponds to a partition of E by the coordinate surfaces $u(\mathbf{r}) = u_i$, $v(\mathbf{r}) = v_j$, and $w(\mathbf{r}) = w_k$. Under the transformation considered

$$T:\ E'_{ijk} = [u_{i-1}, u_i] \times [v_{j-1}, v_j] \times [w_{k-1}, w_k] \ \to\ E_{ijk}\,,$$

where E_{ijk} is a partition element of E. Consider the Riemann sum of f for this partition of E. The triple integral of f over E is the three-variable limit of the Riemann sum (**34.3**) as $\mathbf{N} = \langle N_1, N_2, N_3\rangle \to \infty$ (or $(\Delta u, \Delta v, \Delta w) \to (0,0,0)$). The volume ΔV_{ijk} of E_{ijk} is related to the volume of the rectangular box E'_{ijk} by (**36.5**). By continuity of J, its value in (**36.5**) can be taken at any sample point in E'_{ijk} because variations of a sample point yields corrections that decreases to zero faster than $\Delta V'$. Therefore the limit of the Riemann sum is the triple integral of fJ over the region E'. The above qualitative consideration suggests that the following theorem holds (a full proof is considered in advanced calculus courses).

THEOREM **36.2**. (Change of Variables in Triple Integrals).
Let a transformation $E' \to E$ defined by functions $(u,v,w) \to (x,y,z)$ with continuous partial derivatves have a non-vanishing Jacobian, except perhaps on the boundary of E'. Suppose that f is continuous on E and E is bounded by piecewise smooth surfaces. Then

$$\iiint_E f(\mathbf{r})\,dV = \iiint_{E'} f(x(\mathbf{r}'), y(\mathbf{r}'), z(\mathbf{r}'))J(\mathbf{r}')\,dV'\,,$$
$$J(\mathbf{r}') = \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right|\,.$$

Evaluation of a triple integral in curvilinear coordinates follows the same steps as for a double integral in curvilinear coordinates.

EXAMPLE **36.2**. (Volume of an Ellipsoid).
Find the volume of a solid region E bounded by an ellipsoid $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$, where a, b, and c are positive numbers.

SOLUTION: The integration domain can be simplified by a scaling transformation $x = au$, $y = bv$, and $z = cw$ under which the ellipsoid is mapped onto a sphere of unit radius $u^2 + v^2 + w^2 = 1$ (see Figure **36.2**). The image

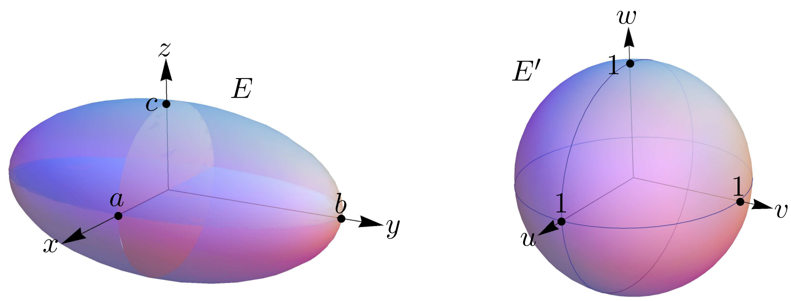


FIGURE **36.2**. An illustration to Example **36.2**. The ellipsoidal region $x^2/a^2 + y^2/b^2 + z^2/c^2 \le 1$ is mapped onto the ball $u^2 + v^2 + w^2 \le 1$ by the coordinate transformation $u = x/a$, $v = y/b$, $w = z/c$ with the Jacobian $J = abc$

E' of E is a ball of unit radius. The transformation defines a change of variables because its Jacobian vanishes nowhere:

$$J = \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| = \left|\det\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}\right| = abc \neq 0\,.$$

Therefore,

$$\begin{aligned} V(E) = \iiint_E dV &= \iiint_{E'} J\, dV' = abc \iiint_{E'} dV' \\ &= abcV(E') = \frac{4\pi}{3}\, abc. \end{aligned}$$

□

When $a = b = c = R$, the ellipsoid becomes a ball of radius R, and a familiar expression for the volume is recovered: $V = (4\pi/3)R^3$.

EXAMPLE **36.3**. *Let* $\mathbf{a}$, $\mathbf{b}$, *and* $\mathbf{c}$ *be non-coplanar vectors. Find the volume of a solid* E *bounded by the surface* $(\mathbf{a}\cdot\mathbf{r})^2 + (\mathbf{b}\cdot\mathbf{r})^2 + (\mathbf{c}\cdot\mathbf{r})^2 = R^2$ *where* $\mathbf{r} = \langle x, y, z\rangle$.

SOLUTION: Define new variables by the transformation $u = \mathbf{a}\cdot\mathbf{r}$, $v = \mathbf{b}\cdot\mathbf{r}$, $w = \mathbf{c}\cdot\mathbf{r}$. The Jacobian of this transformation is obtained by Eqs. (**36.6**) where it is convenient to use the representation (**36.3**)

$$\begin{aligned} \frac{\partial(x,y,z)}{\partial(u,v,w)} = \left(\frac{\partial(u,v,w)}{\partial(x,y,z)}\right)^{-1} &= \Big(\boldsymbol{\nabla} u\cdot(\boldsymbol{\nabla} v\times\boldsymbol{\nabla} w)\Big)^{-1} \\ &= \Big(\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})\Big)^{-1} \end{aligned}$$

The vector **a**, **b**, and **c** are non-coplanar and, hence, their triple product does not vanish. So, the transformation is a genuine change of variables. Under this transformation the boundary of E becomes a sphere $u^2+v^2+w^2=R^2$. So, the region E is mapped onto the ball E' of radius R. Therefore

$$V(E)=\iiint_E dV=\iiint_{E'} J dV'=\frac{1}{|\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})|}\iiint_{E'} dV'$$
$$=\frac{V(E')}{|\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})|}=\frac{4\pi R^3}{3|\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})|}$$

where $V(E')=4\pi R^3/3$ is the volume of a ball of radius R. □

36.4. Volume Preserving Transformations and Symmetry. Consider a transformation such that the absolute value of its Jacobian is one. Then the volume transformation law reads $dV=JdV'=dV'$ and therefore such a transformation preserves the volume:

$$V(E)=\iiint_E dV=\iiint_{E'} JdV'=\iiint_{E'} dV'=V(E')\,.$$

This allows us to prove the assertion stated in Section **34.2** about the use of symmetry in triple integrals which is the analog of Theorem **33.3** for triple integrals.

THEOREM **36.3**. (Symmetry of Triple Integrals)
Let a function f be integrable on a region E. Suppose that a volume-preserving transformation $x=x(u,v,w)$, $y=y(u,v,w)$, $z=z(u,v,w)$ maps a region E onto itself. Suppose that the function f is skew-symmetric under this transformation, that is,

$$f(x(u,v,w),y(u,v,w),z(u,v,w))=-f(u,v,w)\,.$$

Then the triple integral of f over E vanishes.

PROOF. Since $E'=E$ and $dV=dV'$, the change of variables yields

$$I=\iiint_E f(x,y,z)\,dV=\iiint_E f(x(u,v,w),y(u,v,w),z(u,v,w))\,dV'$$
$$=-\iiint_E f(u,v,z)\,dV'=-I\,;$$

that is, $I=-I$, or $I=0$. □

36.5. Study Problems.

Problem **36.1**. (Volume of a Tetrahedron).
A tetrahedron is a solid with four vertices and four triangular faces. Let the vectors **a**, **b**, *and* **c** *be three adjacent sides of the tetrahedron. Find its volume.*

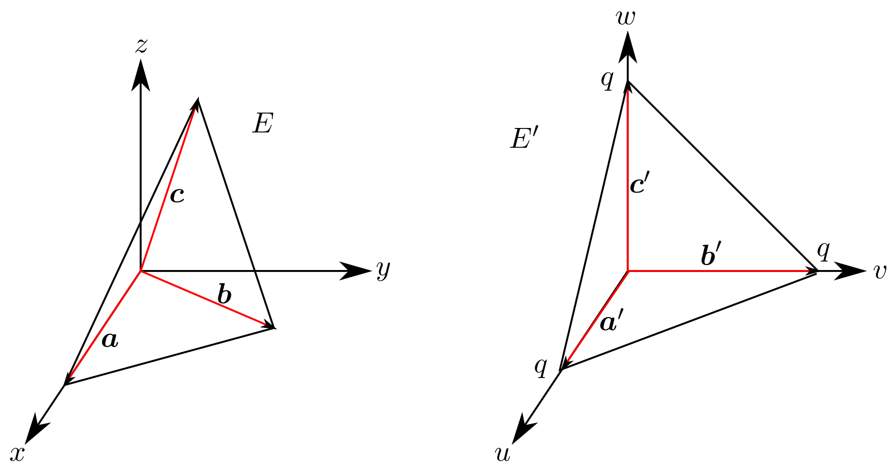


FIGURE **36.3**. An illustration to Study Problem **36.1**. A general tetrahedron is transformed to a tetrahedron whose faces lie in the coordinate planes by a change of variables.

SOLUTION: Consider first a tetrahedron whose adjacent sides are along the coordinate axes and have the same length q. From the geometry, it is clear that six such tetrahedrons form a cube of volume q^3. Therefore, the volume of each tetrahedron is $q^3/6$ (if so desired this can also be established by evaluating the corresponding triple integral; this is left to the reader). The idea is to make a change of variables such that a generic tetrahedron is mapped onto a tetrahedron whose adjacent faces lie in the three coordinate planes (see Figure **36.3**. The adjacent faces of a generic tetrahedron are portions of the planes through the origin. The face containing vectors $\mathbf{a}$ and $\mathbf{b}$ is perpendicular to vector $\mathbf{n} = \mathbf{a} \times \mathbf{b}$ so the equation of this boundary is $\mathbf{n} \cdot \mathbf{r} = 0$. The other adjacent faces are obtained similarly:

$$\begin{aligned} \mathbf{n} \cdot \mathbf{r} &= 0 \quad \text{or} \quad n_1x + n_2y + n_3z = 0\,, & \mathbf{n} &= \mathbf{a} \times \mathbf{b}, \\ \mathbf{l} \cdot \mathbf{r} &= 0 \quad \text{or} \quad l_1x + l_2y + l_3z = 0\,, & \mathbf{l} &= \mathbf{c} \times \mathbf{a}, \\ \mathbf{m} \cdot \mathbf{r} &= 0 \quad \text{or} \quad m_1x + m_2y + m_3z = 0\,, & \mathbf{m} &= \mathbf{b} \times \mathbf{c}, \end{aligned}$$

where $\mathbf{r} = \langle x, y, z\rangle$. So, by putting

$$u = \mathbf{m} \cdot \mathbf{r}\,, \quad v = \mathbf{l} \cdot \mathbf{r} \quad w = \mathbf{n} \cdot \mathbf{r}\,,$$

the images of these planes become the coordinate planes, $w = 0$, $v = 0$, and $u = 0$. The defined transformation is a genuine change of variables because its Jacobian does not vanish because the vectors $\mathbf{n}$, $\mathbf{l}$, and $\mathbf{m}$ are not coplanar (as they are normals to adjacent faces of the tetrahedron). To see this, it is convenient to use the representation (**36.6**) in combination with the relation

(**36.3**):

$$J = \frac{1}{\left|\dfrac{\partial(u,v,w)}{\partial(x,y,z)}\right|} = \frac{1}{|\boldsymbol{\nabla} u \cdot (\boldsymbol{\nabla} v \times \boldsymbol{\nabla} w)|} = \frac{1}{|\mathbf{m}\cdot(\mathbf{n}\times\mathbf{l})|}\,.$$

Furthermore a linear equation in the old variables becomes a linear equation in the new variables under a linear transformation. Therefore, an image of a plane is a plane. So the fourth boundary of E' is a plane through the points $\mathbf{a}'$, $\mathbf{b}'$, and $\mathbf{c}'$, which are the images of $\mathbf{r}=\mathbf{a}$, $\mathbf{r}=\mathbf{b}$, and $\mathbf{r}=\mathbf{c}$, respectively. One has

$$\mathbf{a}' = \langle u(\mathbf{a}),\ v(\mathbf{a}),\ w(\mathbf{a})\rangle = \langle q,0,0\rangle\,, \quad q = \mathbf{a}\cdot\mathbf{m} = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$$

because $\mathbf{a}\cdot\mathbf{n}=0$ and $\mathbf{a}\cdot\mathbf{l}=0$ by the geometrical properties of the cross product. Similarly,

$$\mathbf{b}' = \langle u(\mathbf{b}),\ v(\mathbf{b}),\ w(\mathbf{b})\rangle = \langle 0,q,0\rangle\,, \quad \mathbf{c}' = \langle u(\mathbf{c}),\ v(\mathbf{c}),\ w(\mathbf{c})\rangle = \langle 0,0,q\rangle\,.$$

Thus, the volume of the image region E' is $V(E') = |q|^3/6$ (the absolute value is needed because the triple product can be negative). Therefore,

$$V(E) = \iiint_E dV = \iiint_{E'} J\,dV' = J\iiint_{E'} dV' = JV(E') = \frac{|q|^3 J}{6}.$$

The volume $V(E)$ is independent of the orientation of the coordinate axes. It is convenient to direct the x axis along the vector $\mathbf{a}$. The y axis is directed so that $\mathbf{b}$ is in the xy plane. With this choice,

$$\mathbf{a} = \langle a_1,0,0\rangle\,, \quad \mathbf{b} = \langle b_1,b_2,0\rangle\,, \quad \mathbf{c} = \langle c_1,c_2,c_3\rangle\,.$$

A straight forward calculation shows that

$$q = a_1b_2c_3\,, \quad J = (a_1^2b_2^2c_3^2)^{-1} \quad \Rightarrow \quad V(E) = \frac{1}{6}|a_1b_2c_3|\,.$$

Finally, note that $|c_3| = h$ is the height of the tetrahedron, that is, the distance from the vertex $\mathbf{c}$ to the opposite face (to the xy plane). The area of that face is $A = \|\mathbf{a}\times\mathbf{b}\|/2 = |a_1b_2|/2$. Thus,

$$V(E) = \frac{1}{3}hA;$$

that is, the volume of a tetrahedron is one-third the distance from a vertex to the opposite face, times the area of that face. $\square$

36.6. Exercises.

1–4. Find the Jacobian of each of the following transformations.

1. $x = u/v$, $y = v/w$, $z = w/u$;

2. $x = v + w^2$, $y = w + u^2$, $z = u + v^2$;

3. $x = uv\cos w$, $y = uv\sin w$, $z = (u^2 - v^2)/2$ (these coordinates are called *parabolic* coordinates;

4. $x + y + z = u$, $y + z = uv$, $z = uvw$.

5. Find the region E' whose image E under the transformation defined in Exercise **4.** is bounded by the coordinate planes and by the plane $x+y+z=1$. In particular, investigate the image of those points in E' at which the Jacobian of the transformation vanishes.
6. Let E be the solid region in the first octant defined by the inequality $\sqrt{x}+\sqrt{y}+\sqrt{z}\leq a$ where $a>0$. Find its volume using the triple integral in the new variables $u=\sqrt{x}$, $v=\sqrt{y}$, $w=\sqrt{z}$.
7–11. Use a suitable change of variables in the triple integral to find the volume of a solid bounded by the given surfaces.

7. $(x/a)^{2/3}+(y/b)^{2/3}+(z/c)^{2/3}=1$, where a, b, and c are positive;
8. $(x/a)^{1/3}+(y/b)^{1/3}+(z/c)^{1/3}=1$ where $x\geq 0$, $y\geq 0$, $z\geq 0$ and a, b, and c are positive;
9. $(x/a)^{n}+(y/b)^{m}+(z/c)^{k}=1$ where $x\geq 0$, $y\geq 0$, $z\geq 0$, and the numbers a, b, c, n, m, and k are positive;
10. $(x+y+z)^2=ax+by$ where (x,y,z) lie in the first octant and a, b are positive;
11. $(x+y)^2+z^2=R^2$ where (x,y,z) lie in the first octant.

12. Evaluate the triple integral $\iiint_E zdV$ where E lies above the cone $z=c\sqrt{x^2/a^2+y^2/b^2}$ and bounded from above by the ellipsoid $x^2/a^2+y^2/b^2+z^2/c^2=1$.
13. Evaluate the triple integral $\iiint_E(4x^2-9y^2)dV$ where is enclosed by the paraboloid $z=x^2/9+y^2/4$ and the plane $z=10$.
14. Consider a linear transformation of the coordinates $x=\mathbf{a}\cdot\mathbf{r}'$, $y=\mathbf{b}\cdot\mathbf{r}'$, $z=\mathbf{c}\cdot\mathbf{r}'$ where $\mathbf{r}'=\langle u,v,w\rangle$ and the vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ have constant components. Show that this transformation is *volume preserving* if $|\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})|=1$.
15. If $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are constant vectors, $\mathbf{r}=\langle x,y,z\rangle$, and E is given by the inequalities $0\leq\mathbf{a}\cdot\mathbf{r}\leq\alpha$, $0\leq\mathbf{b}\cdot\mathbf{r}\leq\beta$, $0\leq\mathbf{c}\cdot\mathbf{r}\leq\gamma$, show that $\iiint_E(\mathbf{a}\cdot\mathbf{r})(\mathbf{b}\cdot\mathbf{r})(\mathbf{c}\cdot\mathbf{r})dV=\frac{1}{8}(\alpha\beta\gamma)^2/|\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})|$.
16. Consider parabolic coordinates $x=uv\cos w$, $y=uv\sin w$, $z=(u^2-v^2)$. Show that $2z=(x^2+y^2)/v^2-v^2$, $2z=-(x^2+y^2)/u^2+u^2$, and $\tan w=y/x$. Use these relations to sketch the coordinates surfaces $u(x,y,z)=u_0$, $v(x,y,z)=v_0$, and $w(x,y,z)=w_0$. Evaluate the triple integral of $f(x,y,z)=xyz$ over the region E that lies in the first octant beneath the paraboloid $2z-1=-(x^2+y^2)$ and above the paraboloid $2z+1=x^2+y^2$ by converting to parabolic coordinates.
17. Use a suitable change of variables to find the volume of a solid that is bounded by the surface

$$\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)^n+\frac{z^{2n}}{c^{2n}}=\frac{z}{h}\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)^{n-2},\quad n>1,\quad h>0\,.$$

18. (Generalized Spherical Coordinates) Generalized spherical coordinates (ρ, ϕ, θ) are defined by the equations

$$x = a\rho \sin^n \phi \cos^m \theta\,, \quad y = b\rho \sin^n \phi \sin^m \theta\,, \quad z = c\rho \cos^n \phi$$

where $0 \le \rho < \infty$, $0 \le \phi \le \pi$, $0 \le \theta < 2\pi$, and a, b, c, n, and m are positive parameters. Find the Jacobian of the generalized spherical coordinates.

19–22. Use generalized spherical coordinates with a suitable choice of parameters to find the volume of a solid bounded by the given surfaces.

19. $[(x/a)^2 + (y/b)^2 + (z/c)^2]^2 = (x/a)^2 + (y/b)^2\,;$
20. $[(x/a)^2 + (y/b)^2 + (z/c)^2]^2 = (x/a)^2 + (y/b)^2 - (z/c)^2\,;$
21. $(x/a)^2 + (y/b)^2 + (z/c)^4 = 1\,;$
22. $[(x/a)^2 + (y/b)^2]^2 + (z/c)^4 = 1\,.$

23. (Dirichlet's integral) Let n, m, p, and s be positive integers. Use the transformation defined by $x + y + z = u$, $y + z = uv$, $z = uvw$ to show that

$$\iiint_E x^n y^m z^p (1 - x - y - z)^s dV = \frac{n!\, m!\, p!\, s!}{(n + m + p + s + 3)!}$$

where E is the tetrahedron bounded by the coordinate planes and the plane $x + y + z = 1$.

24. (Orthogonal curvilinear coordinates)
Curvilinear coordinates (u, v, w) are called *orthogonal* if the normals to their coordinate surfaces are mutually orthogonal at any point of their intersection. In other words, the gradients $\boldsymbol{\nabla} u(x, y, x)$, $\boldsymbol{\nabla} v(x, y, z)$, and $\boldsymbol{\nabla} w(x, y, z)$ are mutually orthogonal. One can define unit vectors orthogonal to the coordinates surfaces:

$$(\mathbf{36.7}) \qquad \hat{\mathbf{e}}_u = \frac{\boldsymbol{\nabla} u}{\|\boldsymbol{\nabla} u\|}, \quad \hat{\mathbf{e}}_v = \frac{\boldsymbol{\nabla} v}{\|\boldsymbol{\nabla} v\|}, \quad \hat{\mathbf{e}}_w = \frac{\boldsymbol{\nabla} w}{\|\boldsymbol{\nabla} w\|}$$

Note that the Jacobian of a change of variables does not vanish and the relation (**36.6**) guarantees that these unit vectors are not coplanar and form a basis in space (any vector can be uniquely expanded into a linear combination of them).
(i) Show that

$$(\mathbf{36.8}) \qquad \|\boldsymbol{\nabla} r\| = 1, \quad \|\boldsymbol{\nabla} \theta\| = \frac{1}{r}\,, \quad \|\boldsymbol{\nabla} z\| = 1$$

$$(\mathbf{36.9}) \qquad \|\boldsymbol{\nabla} \rho\| = 1\,, \quad \|\boldsymbol{\nabla} \phi\| = \frac{1}{\rho}\,, \quad \|\boldsymbol{\nabla} \theta\| = \frac{1}{\rho \sin \phi}$$

for the cylindrical (r, θ, z) and spherical (ρ, ϕ, θ) coordinates.
(ii) Show that the spherical and cylindrical coordinates are orthogonal coordinates and, in particular,

$$(\mathbf{36.10}) \qquad \hat{\mathbf{e}}_r = \langle \cos\theta, \sin\theta, 0\rangle,\ \hat{\mathbf{e}}_\theta = \langle -\sin\theta, \cos\theta, 0\rangle,\ \hat{\mathbf{e}}_z = \langle 0, 0, 1\rangle$$

for the cylindrical coordinates, and

$$\begin{aligned} \hat{\mathbf{e}}_r &= \langle \sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi \rangle \\ \hat{\mathbf{e}}_\phi &= \langle \cos\phi\cos\theta, \cos\phi\sin\theta, -\sin\phi \rangle \\ \hat{\mathbf{e}}_\theta &= \langle -\sin\theta, \cos\theta, 0 \rangle \end{aligned} \tag{36.11}$$

for the spherical coordinates.

37. Improper Multiple Integrals

37.1. Preliminary Remarks. In the case of one-variable integration, improper integrals occur when the integrand is not bounded at a boundary point of the integration interval or the integration interval is not bounded. For example, for $\nu \neq 1$,

$$\int_0^1 \frac{dx}{x^\nu} = \lim_{a\to 0^+} \int_a^1 \frac{dx}{x^\nu} = \lim_{a\to 0^+} \frac{1-a^{1-\nu}}{1-\nu} \tag{37.1}$$

or

$$\int_R^\infty \frac{dx}{x^\nu} = \lim_{a\to\infty} \int_R^a \frac{dx}{x^\nu} = \lim_{a\to\infty} \frac{x^{1-\nu}}{1-\nu}\bigg|_R^a = -\frac{R^{1-\nu}}{1-\nu} + \lim_{a\to\infty} \frac{a^{1-\nu}}{1-\nu}\,. \tag{37.2}$$

The limit (**37**.**1**) exists if $\nu < 1$ and does not exist if $\nu > 1$. The limit (**37**.**2**) exists if $\nu > 1$ and does not exist if $\nu < 1$. If $\nu = 1$, the improper integrals diverge as $\ln a$ with either $a \to 0^+$ or $a \to \infty$, respectively.

By the very definition of a multiple integral, the construction of the lower and upper sums requires that the function is bounded and the region of integration is bounded as well. If the function is not bounded, then its supremum and infimum do not exist for some partition elements. If the region of integration is not bounded, then it cannot be partitioned into finitely many pieces of finite areas. In either case, the upper and lower sums cannot be defined. Just like in the one-variable case, multiple integrals of unbounded functions, or over unbounded regions, or both are called *improper multiple integrals.*

For definitiveness, the discussion will be given for triple integrals. The case of double integrals is treated along the same line of reasoning and all the equations hold for double integrals if the symbol $\iiint$ is replaced by $\iint$. It will always be assumed that *the boundary of a closed region is piecewise smooth.* Let E be a bounded closed region in space and f is a function on E (regions in a plane and in space will be denoted by E here). Suppose now that E contains a set S of zero volume and in any neighborhood of each point of S, the function f is not bounded. The function f is said to be *singular* on S. The objective is to give a definition of the integral of f over E. Note that the values of f on S are irrelevant (f can be given any values on S), only the fact that f is not bounded near any point of S requires a modification of the definition of the integral of f over E. For example, the function defined by the rule

$$f(\mathbf{r}) = \frac{1}{\|\mathbf{r}\|} \quad \text{if} \quad \mathbf{r} \neq \mathbf{0}\,, \qquad f(\mathbf{0}) = 1\,,$$

is not bounded in any neighborhood of the origin. So, the function is singular at the origin (despite that it has a value at the origin).

Suppose first that S consists of a single point $\mathbf{r}_0$; that is, in any small open ball B_ε of radius ε centered at $\mathbf{r}_0$ the values of $f(\mathbf{r})$ are not bounded, whereas f is bounded on the closed region E_ε obtained from E by removing

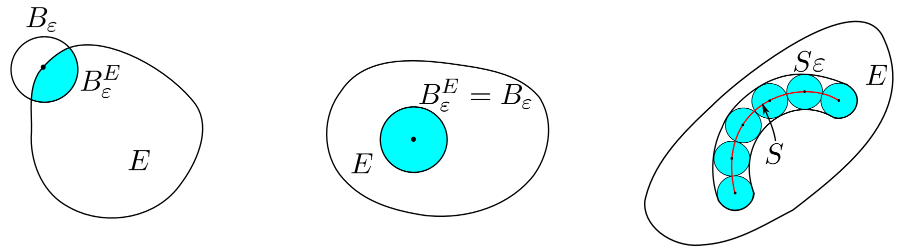


FIGURE **37.1**. A regularization of an improper integral. **Left**: B_ε is a ball centered at a singular point of the integrand. B_ε^E is the intersection of B_ε with E. The integration is carried out over the region E with B_ε removed. Then the limit $\varepsilon \to 0$ is taken. **Middle**: The same regularization procedure when the singular point is an interior point of E. **Right**: A regularization procedure when the set S on which the integrand is singular has more than one point. By removing the set S_ε from E, the region E_ε is obtained. The distance between any point of E_ε and the set S is no less than ε.

the ball B_ε. Suppose that f is integrable on E_ε for any $\varepsilon > 0$ (e.g., f is continuous on E_ε). Then, by analogy with the one-variable case, one can *define* the integral of f over E as the limit

$$\iiint_E f\, dV = \lim_{\varepsilon \to 0^+} \iiint_{E_\varepsilon} f\, dV\,, \tag{37.3}$$

provided, of course, the limit exists. Similarly, if S contains more than one point, one can construct a set S_ε that is the union of open balls of radius ε centered at each point of S. Then E_ε is obtained by removing S_ε from E. The regularization procedure is illustrated in Fig. **37.1**.

Suppose a region E is not bounded. Let E_R be the part of E that lies in the closed ball $\|\mathbf{r}\| \le R$ and a function f be integrable on E_R for all R for which E_R has non-zero volume. Then by analogy with the one variable case, the improper integral of f can be defined by the rule

$$\iiint_E f dV = \lim_{R \to \infty} \iiint_{E_R} f dV \tag{37.4}$$

provided the limit exists.

Although the rules (**37.3**) and (**37.4**) seem rather natural generalizations of one-variable improper integrals, there are subtleties that are specific to multivariable integrals. This is illustrated by the following example. Suppose that the function

$$f(x,y) = \frac{y^2 - x^2}{(x^2+y^2)^2} \tag{37.5}$$

is to be integrated over the region that is the part of the unit disk (centered at the origin) in the positive quadrant:

$$E = \{(x, y)\,|x^2 + y^2 \le 1\,,\ x \ge 0\,,\ y \ge 0\}.$$

The function is singular at the origin that lies on the boundary of the region of integration. The value of f at the origin is not relevant and one can set $f(0, 0)$ to any number because a regularization eliminates a neighborhood of the points $(0, 0)$ from the region of integration. If the rule (**37.3**) is applied, then one can choose

$$E_\varepsilon = \{(x, y)\,|\,R_\varepsilon^2 \le x^2 + y^2 \le 1\,,\ x \ge 0\,,\ y \ge 0\}\,, \quad R_\varepsilon = e^{-1/\varepsilon}$$

so that $R_\varepsilon \to 0^+$ as $\varepsilon \to 0^+$ (this choice of the dependence of R_ε on ε is a matter of convenience which will soon become clear). The integration region E_ε is symmetric under the reflection about the line $y = x$, while the function f is skew symmetric:

$$T: \quad (x, y) \to (y, x)\,, \quad T(E_\varepsilon) = E_\varepsilon\,, \quad f(y, x) = -f(x, y)\,.$$

So the integral of f over E_ε vanishes by the symmetry argument. The integrand is positive in the part of the domain where $x^2 < y^2$ and negative if $y^2 > x^2$, and there is a mutual cancellation of contributions from these regions.

Now consider the portion E'_ε of E_ε corresponding to the following interval of the polar angle θ:

$$0 \le \theta \le \theta_0 \le \frac{\pi}{2}$$

for some θ_0. Then, by evaluating the integral in polar coordinates, one finds that

$$\begin{aligned} \iint_{E'_\varepsilon} \frac{y^2 - x^2}{(x^2 + y^2)^2} dA &= \int_0^{\theta_0} (\sin^2\theta - \cos^2\theta) d\theta \int_{R_\varepsilon}^1 \frac{dr}{r} \\ &= \ln R_\varepsilon \int_0^{\theta_0} \cos(2\theta)\, d\theta \\ &= -\frac{\sin(2\theta_0)}{2\varepsilon}\,, \end{aligned} \tag{37.6}$$

where the double angle formula has been used to transform the integrand to $\cos(2\theta)$. Note also that the explicit form of R_ε was used (the choice of R_ε is justified by the simplicity (**37.6**)). Put

$$\theta_0 = \frac{\pi}{2} - \varphi\varepsilon$$

for some numerical parameter $\varphi \ge 0$ and ε small enough to make $\theta_0 > 0$ so that $\theta_0 \to \pi/2$ as $\varepsilon \to 0^+$. The integral (**37.6**) can be viewed as another regularization of the improper integral of f over E. However, this regularization can give *any value* of the improper integral! Indeed,

$$\lim_{\varepsilon\to 0^+} \iint_{E'_\varepsilon} f dA = -\lim_{\varepsilon\to 0^+} \frac{\sin(2\theta_0)}{2\varepsilon} = -\lim_{\varepsilon\to 0^+} \frac{\sin(2\varphi\varepsilon)}{2\varepsilon} = -\varphi\,.$$

This observation suggests that *the value of the improper integral may depend on the way a regularization is introduced.* Naturally, if the improper integral is to be given a value, then this value should not depend on a regularization used to obtain it. From this point of view, the rule (**37.3**) cannot be regarded as a proper definition of a multiple improper integral.

A similar observation can be made for an improper integral over an unbounded region. Suppose the function (**37.5**) is to be integrated over the unbounded region that is the part of the positive quadrant outside the unit disk centered at the origin:

$$E = \{(x, y) \,|\, 1 \le x^2 + y^2\,,\ x \ge 0\,,\ y \ge 0\}\,.$$

If the rule (**37.4**) is applied to evaluate the integral, one can choose

$$E_\varepsilon = \{(x, y) \,|1 \le x^2 + y^2 \le R_\varepsilon^2\,,\ x \ge 0\,,\ y \ge 0\}\,, \quad R_\varepsilon = e^{1/\varepsilon}$$

so that $R_\varepsilon \to \infty$ as $\varepsilon \to 0^+$. The integral over E_ε has zero value by the symmetry argument given above. Let E'_ε be the part of E_ε corresponding to the interval $0 \le \theta \le \theta_0 \le \pi/2$ of the polar angle θ and, as before, put $\theta_0 = \pi/2 - \varphi\varepsilon$. The integral over E'_ε can be viewed as another regularization of the improper integral over E. The integral over E'_ε is given by Eq. (**37.6**) if the limits of integration over r are swapped as in the present case $R_\varepsilon > 1$. Therefore

$$\lim_{\varepsilon \to 0^+} \iint_{E'_\varepsilon} f dA = -\frac{1}{2} \lim_{\varepsilon \to 0^+} \ln R_\varepsilon \sin(2\theta_0) = -\lim_{\varepsilon \to 0^+} \frac{\sin(2\varphi\varepsilon)}{2\varepsilon} = -\varphi\,.$$

37.2. Definition of an Improper Integral. Let E be a region in space (or in a plane) possibly unbounded. An *exhaustion* of E is a sequence of regions E_k, $k = 1, 2, ...$, such that

- each region E_k is bounded, closed, and contained in E,
- the region E_{k+1} contains E_k,
- the union of all E_k coincides with E excluding possibly a set of zero volume (or zero area) in E.

One-variable improper integrals are defined as the limit of integrals over *ever-expanding* intervals of integration, e.g., as in (**37.1**) or (**37.2**), that eventually cover the original interval of integration. An exhaustion is a multidimensional analog of ever-expanding intervals (E_{k+1} contains E_k).

For example, if E is the entire space, then the union of balls

$$E_k = \{(x, y, z) \,|\, x^2 + y^2 + z^2 \le k^2\}\,, \quad k = 1, 2, ...$$

So, this sequence of ever-expanding balls is an exhaustion of the entire space. Another exhaustion the entire space is given by the following sequence of ever-expanding cubes:

$$E_k = \{(x, y, z) \,|\, |x| \le k\,,\ |y| \le k\,,\ |z| \le k\} \quad k = 1, 2, ...$$

If E is the part of the disk of unit radius in the first quadrant discussed in the above example and θ is the polar angle, then for each fixed $0 \le \varphi < \pi/2$, the regions

$$E_k = \{(x,y)\,|\, e^{-2k} \le x^2 + y^2 \le 1\,,\ 0 \le \theta \le \pi/2 - \varphi/k\} \tag{37.7}$$

where $k = 1, 2, ...,$ are exhaustions of E (in the above consideration of E, put $\varepsilon = 1/k$).

If a bounded function f is integrable on a closed bounded region E that has the volume $V(E)$, then it can be proved that for any exhaustion of E

$$\lim_{k\to\infty} \iiint_{E_k} f dV = \iiint_E f dV\,,$$
$$\lim_{k\to\infty} V(E_k) = \lim_{k\to\infty} \iiint_{E_k} dV = \iiint_E dV = V(E)\,.$$

The above properties are called *continuity of the Riemann integral.* Its one-dimensional analog is a familiar property of the integral over an interval:

$$\lim_{c\to b^-} \int_a^c f(x)\,dx = \int_a^b f(x)\,dx$$

if f is integrable on $[a,b]$. In other words, the definition of the integral via an exhaustion of the region of integration leads to the *same* result as the definition via the upper and lower sums. This is just a consistency check.

However, an exhaustion of E can be used to regularize an improper integral of f over E. For example, if f is singular at a zero-volume set S in a closed bounded region E, then one can take an exhaustion of E such that none of E_k contains S. A simple possibility is take a sequence of the unions of open balls with centers at each point of S and of radii $\varepsilon = 1/k$ and, for each k, remove the union from E to obtain E_k as depicted in the right panel of Fig. **37.1**.

Definition **37.1**. (Improper Multiple Integral)
Let E_k be an exhaustion of a region E. Suppose that a function f on E is integrable on each E_k. Then the function f is said to be integrable on E if the limit

$$\lim_{k\to\infty} \iiint_{E_k} f dV = \iiint_E f dV$$

exists and is independent of the choice of E_k. The value of the limit is called an improper integral of f over E.

Note that the region of integration E is no longer required to be bounded and the function f may not be bounded in E, but it is bounded on each E_k in order to define the integral of f over E_k. An improper double integral is defined in the same way. The condition that the limit should not depend on the choice of an exhaustion means that *the value of the improper integral should not depend on its regularization.* According to this definition the function (**37.5**) is not integrable on any closed bounded region containing the origin because the limit depends on the way the regularization is imposed. For instance, let E be the part of the disk $x^2+y^2\le 1$ in the first quadrant. Take the exhaustion defined in Eq. (**37.7**). Then using the previous result (**37.6**) with $\varepsilon=1/k$ and $\theta_0=\pi/2-\varphi/k$,

$$\iint_{E_k} f(x,y)dA = \frac{1}{2}\sin\Big(\pi-\frac{2\varphi}{k}\Big)\ln e^{-k} = -\frac{1}{2}\,k\,\sin(2\varphi/k) \;\to\; -\varphi$$

as $k\to\infty$, where $\sin u = u+O(u^3)$ with $u=2\varphi/k$ has been used to find the limit. The limit value depends on an arbitrary parameter φ and therefore the improper integral does not exist.

37.3. Evaluation of an Improper Multiple Integral. Definition **37.1** eliminates the aforementioned potential ambiguity of the rule (**37.3**), but, unfortunately, it is rather difficult to use. It turns out that the difficult task of investigating the regularization-independence of an improper integral can be avoided for non-negative functions, thanks to the following theorem.

THEOREM **37.1**. (Improper Integrals of Non-negative Functions) *Let E_k and E'_k be two exhaustions of a region E. Let a function f be non-negative on E, $f(\mathbf{r})\ge 0$ for all $\mathbf{r}$ in E. Suppose that f is integrable on each E_k and each E'_k. Then*

$$\lim_{k\to\infty}\iiint_{E_k} f dV = \lim_{k\to\infty}\iiint_{E'_k} f dV\,,$$

where the limit may be $+\infty$. In particular, if the limit is a number, then f is integrable on E and

$$\iiint_E f dV = \lim_{k\to\infty}\iiint_{E_k} f dV\,.$$

The same statement holds for double integrals. The conclusion of this theorem can intuitively be understood in the following way. Take an exhaustion E_k (a regularization of the improper integral). Suppose the improper integral converges if the rule (**37.3**) or (**37.4**) is used. Since the integrand is non-negative, the value of the improper integral is positive (it is zero only if the integrand is zero). The sequence I_k of integrals over each E_k increases monotonically because the region of integration E_{k+1} contains E_k and the integrand is non-negative. Since each E_k is contained in E_R for a sufficiently large R in (**37.4**) or a sufficiently small ε in (**37.3**), the sequence I_k is also bounded by the value of the limit (**37.4**) or (**37.3**). Every monotonically increasing bounded sequence I_k converges (Calculus 2). Therefore

- *the improper integral of a non-negative function converges for any exhaustion and its value can be found for a particular exhaustion, e.g., defined in (**37.4**) or (**37.3**).*

EXAMPLE **37.1**. *Find the integral of* $f(x, y, z) = z(x^2+y^2+z^2)^{-7/4}$ *over the half-ball* $x^2 + y^2 + z^2 \leq 1$, $z \geq 0$, *if it exists.*

SOLUTION: The function is non-negative in the region of integration and singular at the origin. Therefore if the improper integral of f exists for a particular regularization, then by Theorem **37.1**, it exists for any regularization and has the same value. Let us use the rule (**37.3**) to regularize the improper integral in question. Put

$$E_\varepsilon = \{(x, y, z) \,|\, \varepsilon^2 \leq x^2 + y^2 + z^2 \leq 1\,,\ z \geq 0\}\,.$$

The region E_ε is the image of the rectangular box $E'_\varepsilon = [\varepsilon, 1]\times[0, \pi/2]\times[0, 2\pi]$ in spherical coordinates. By converting the integral of f over E_ε to spherical coordinates and using Fubini's theorem to evaluate it,

$$\begin{aligned}
\iiint_{E_\varepsilon} f dV &= \iiint_{E'_\varepsilon} \rho\cos\phi \cdot \rho^{-7/2} \cdot \rho^2 \sin\phi dV' \\
&= \int_0^{2\pi} d\theta \int_0^{\pi/2} \sin\phi\cos\phi\, d\phi \int_\varepsilon^1 \rho^{-1/2} d\rho \\
&= 2\pi \cdot \left(\frac{1}{2}\sin^2\phi\Big|_0^{\pi/2}\right) \cdot \left(2\rho^{1/2}\Big|_\varepsilon^1\right) \\
&= 2\pi(1-\sqrt{\varepsilon})\,.
\end{aligned}$$

Taking the limit $\varepsilon \to 0^+$,

$$\iiint_E f dV = \lim_{\varepsilon\to 0^+} \iiint_{E_\varepsilon} f dV = \lim_{\varepsilon\to 0^+} 2\pi(1-\sqrt{\varepsilon}) = 2\pi\,.$$

$\square$

Given a function f on a region E, define two functions on E

$$f_+(\mathbf{r}) = \begin{cases} f(\mathbf{r}), & \text{if } f(\mathbf{r}) \geq 0 \\ 0, & \text{otherwise} \end{cases}, \qquad f_-(\mathbf{r}) = \begin{cases} -f(\mathbf{r}), & \text{if } f(\mathbf{r}) \leq 0 \\ 0, & \text{otherwise} \end{cases}.$$

Then $f_\pm(\mathbf{r}) \geq 0$ and $f(\mathbf{r}) = f_+(\mathbf{r}) - f_-(\mathbf{r})$ in E. Suppose that the functions $f_\pm$ are integrable on E. Then f is also integrable on E. Indeed, if E_k is an exhaustion of E, then

$$\begin{aligned}
\lim_{k\to\infty} \iiint_{E_k} f dV &= \lim_{k\to\infty} \Big(\iiint_{E_k} f_+ dV - \iiint_{E_k} f_- dV\Big) \\
&= \lim_{k\to\infty} \iiint_{E_k} f_+ dV - \lim_{k\to\infty} \iiint_{E_k} f_- dV \\
&= \iiint_E f_+ dV - \iiint_E f_- dV
\end{aligned}$$

because the limit of the difference of two sequences exists and is equal to the difference of the limits of the sequences, provided the latter exist, and they do exist since the improper integrals of $f_\pm$ exist by the hypothesis.

THEOREM **37.2**. (Sufficient Condition for Integrability)
*Let f be a function on a region E possibly unbounded. Suppose that the functions $f_\pm$ are integrable on E in the sense of the rule (**37.3**) or (**37.4**). Then f is integrable on E and the improper integral of f over E can be evaluated by the rule (**37.3**) or (**37.4**), or by any convenient regularization of the improper integral.*

Next note that

$$0 \le f_\pm(\mathbf{r}) \le |f(\mathbf{r})| \quad \text{for all } \mathbf{r} \in E\,.$$

Suppose there exists a function g that is integrable on E such that

$$|f(\mathbf{r})| \le g(\mathbf{r}) \quad \text{for all } \mathbf{r} \in E\,.$$

Since $f_\pm$ and g are non-negative, by Theorem **37.1** and the positivity property of the integral

$$I_k^\pm = \iiint_{E_k} f_\pm dV \le \iiint_{E_{k+1}} f_\pm dV \le \iiint_{E_{k+1}} g dV \le \iiint_E g dV = I_g$$

for an exhaustion E_k of E, assuming that $f_\pm$ are integrable on each E_k. As E_{k+1} contains E_k, the value of the integral over E_{k+1} of a non-negative function cannot be less than the value of the integral over E_k. Therefore the numerical sequences $I_k^\pm$ are increasing and bounded by the integral I_g. Recall from Calculus 2 that any such sequence is convergent. Therefore $f_\pm$ are integrable on E and so is f. Moreover the rules (**37.3**) and (**37.4**) can be used to evaluate the improper integral of f over E by Theorem **37.2**.

THEOREM **37.3**. (Integrability Test).
*Suppose there exists an integrable function g on a region E such that $|f(\mathbf{r})| \le g(\mathbf{r})$ for all $\mathbf{r}$ in E. If f is integrable on any closed bounded subregion of E on which f is bounded, then f is integrable on E and its improper integral over E can be evaluated by the rules (**37.3**) and (**37.4**), or by any convenient regularization of the improper integral.*

Since g is non-negative, its integrability can be verified by the rule (**37.3**) or (**37.4**), or by any convenient regularization of the improper integral of g.

EXAMPLE **37.2**. *Evaluate the triple integral of*

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

over a ball of radius R centered at the origin if it exists.

SOLUTION: The function is singular only at the origin and continuous otherwise so f is integrable on any closed bounded subregion of the ball that does not include the origin. Put $\rho = (x^2 + y^2 + z^2)^{1/2}$ (the distance from the origin) so that $|x| \le \rho$ and $|z| \le \rho$. Using the inequality $|\sin u| \le |u|$ for $u = x^2 - z^2$,

$$|f(x, y, z)| \le \frac{|x^2 - z^2|}{\rho^4} \le \frac{x^2 + z^2}{\rho^4} \le \frac{2\rho^2}{\rho^4} = \frac{2}{\rho^2} = \frac{2}{x^2 + y^2 + z^2} = g(x, y, z)$$

If g is integrable on the ball, then f is integrable and the rule (**37.3**) can be applied to evaluate the improper integral. To verify the integrability of g, take the restricted region E_ε that lies between two spheres:

$$E_\varepsilon : \quad \varepsilon^2 \le x^2 + y^2 + z^2 \le R^2 \,.$$

It is the image of the rectangular box $E'_\varepsilon = [\varepsilon, R] \times [0, \pi] \times [0, 2\pi]$ in spherical coordinates. The improper integral of g becomes a proper integral in spherical coordinates because

$$g(x,y,z)dV = \frac{2}{\rho^2} \cdot JdV' = \frac{2}{\rho^2} \cdot \rho^2 \sin\phi dV' = 2\sin\phi dV'$$

and the singularity at $\rho = 0$ is cancelled by the Jacobian. Using the rule (**37.3**),

$$\iiint_E f dV = \lim_{\varepsilon\to 0^+} \iiint_{E_\varepsilon} \frac{\sin(x^2 - z^2)}{(x^2+y^2+z^2)^2} dV = 0$$

because for any $0 < \varepsilon < R$ the integral vanishes as E_ε is symmetric under the reflection about the plane $x = z$:

$$T : \quad (x,y,z) \to (z,y,x)\,, \quad T(E_\varepsilon) = E_\varepsilon\,,$$

whereas the function is skew-symmetric $f(z,y,x) = -f(x,y,z)$. □

Note that the use of symmetry in the above example is justified only after proving that the function is in fact integrable! For example, changing the denominator $(x^2 + y^2 + z^2)^2$ of the integrand to $(x^2 + y^2 + z^2)^3$ does not violate the symmetry of the integrand, but the improper integral does not exist. Integrands singular at a point will be discussed in the next subsection.

Although Theorems **37.1**, **37.2**, and **37.3** appear to be helpful when analyzing improper multiple integrals, they do not exhaust all the cases when the rules (**37.3**) and (**37.4**) are valid. It is important to understand that these theorems provide only *sufficient* conditions for the existence of improper integrals. In particular, if the integrals of $f_\pm$ defined by the rule (**37.3**) or (**37.4**) diverge, this does not generally imply that the improper integral of f does not exist. A further investigation is needed to verify the conditions of Definition **37.1**. The following improper integrals are known as the Fresnel integrals

$$\int_0^\infty \sin(x^2)dx = \lim_{a\to\infty} \int_0^a \sin(x^2)dx = \sqrt{\pi/8}\,, \quad \int_0^\infty \cos(x^2)dx = \sqrt{\pi/8}\,.$$

On the other hand, the integrals of $|\sin(x^2)|$ and $|\cos(x^2)|$ over $[0,\infty)$ are proved to diverge. For a positive integer n, the following inequality holds:

$$\int_{\sqrt{\pi(n-1)}}^{\sqrt{\pi n}} |\sin(x^2)|dx = \int_{\pi(n-1)}^{\pi n} |\sin u| \frac{du}{2\sqrt{u}} = \int_0^\pi \sin v \frac{dv}{2\sqrt{v + \pi(n-1)}}$$
$$\ge \frac{1}{2\sqrt{\pi n}} \int_0^\pi \sin v dv = \frac{1}{\sqrt{\pi n}}\,,$$

where $u = x^2$, $v = u - \pi(n-1)$, and $\sqrt{\pi n} \ge \sqrt{v + \pi(n-1)}$ for $v \le \pi$ by monotonicity of the power function. Using this inequality, the improper integral of $|\sin(x^2)|$ can be bounded from below:

$$\int_0^{\sqrt{\pi N}} |\sin(x^2)|dx = \sum_{n=1}^{N} \int_{\sqrt{\pi(n-1)}}^{\sqrt{\pi n}} |\sin(x^2)|dx \ge \frac{1}{\sqrt{\pi}} \sum_{n=1}^{N} \frac{1}{\sqrt{n}} .$$

The series $\sum n^p$ diverges for $p \ge -1$ (Calculus 2) and, by the comparison test, the limit $N \to \infty$ does not exist in the left side of this equation. By a similar argument, one can show that the functions $f_\pm$ for $f(x) = \sin(x^2)$ are not integrable, too (their improper integral diverges). The case of $\cos(x^2)$ can be studied in the same way and the same conclusion holds. There are a lot of cancellations in the improper integrals of $\sin(x^2)$ and $\cos(x^2)$ that ensure their convergence. The analogy can be made with alternating series (e.g., in Calculus 2 it has been shown that the alternating $p-$series, $\sum(-1)^{n+1}n^p$, converges for $p < 0$, but it converges absolutely only for $p < -1$). Multiple improper integrals may also behave similarly.

37.4. Functions singular at a point.

Example **37.3**. *Find the integral of $f(x,y) = x(x^2+y^2)^{-1}$ over the half-disk, $x^2+y^2 \le 1$, $x \ge 0$, if it exists.*

Solution: The function is singular at the origin. Since f is non-negative and continuous everywhere except the origin, it is sufficient to investigate the existence of the improper integral in a particular regularization. Take

$$E_\varepsilon = \{(x,y)\,|\,\varepsilon^2 \le x^2+y^2 \le 1\,,\ x \ge 0\}\,.$$

It is the image of the rectangle $E'_\varepsilon = [\varepsilon, 1] \times [-\pi/2, \pi/2]$ in the polar plane. Therefore

$$f(x,y)dA = \frac{r\cos\theta}{r^2}\, r dA' = \cos\theta\, dA'$$

So, the singularity at $r = 0$ is cancelled by the Jacobian of polar coordinates and the integral becomes proper in polar coordinates and, hence, the function is integrable

$$\iint_E f dA = \int_{-\pi/2}^{\pi/2} \cos\theta d\theta \int_0^1 dr = 2\cdot 1 = 2\,.$$

□

Examples **37.2** and **37.3** exhibit a common feature: An improper integral becomes a proper integral in curvilinear coordinates if the Jacobian vanishes at a point where the integrand is singular. The following theorem provides sufficient conditions under which a function singular at a point is integrable over a bounded closed region that includes this point.

Theorem **37.4**. *Let E be a closed bounded region in an $n-$dimensional Euclidean space ($n = 1, 2, 3$). Let a function f be singular at a point $\mathbf{r}_0$*

of E and integrable on any closed subregion of E that does not include $\mathbf{r}_0$. Suppose that

$$|f(\mathbf{r})| \leq \frac{M}{\|\mathbf{r}-\mathbf{r}_0\|^\nu}\,, \quad \text{if} \quad 0 < \|\mathbf{r}-\mathbf{r}_0\| \leq R \quad \text{and} \quad \nu < n\,,$$

for some positive R and M. Then the improper integral of f over E exists and can be evaluated in any convenient regularization.

Proof. One can always set the origin of the coordinate system at $\mathbf{r}_0$ by the shift transformation $\mathbf{r} \to \mathbf{r}-\mathbf{r}_0$. Evidently, its Jacobian is 1. So, without loss of generality, assume that f is singular at the origin. Let B_R be the ball $\|\mathbf{r}\| \leq R$ and B_R^E be the intersection of B_R and E (compare with Fig. **37.1** with $\varepsilon = R$). It is sufficient to show the existence of the improper integral of f over B_R^E as f is integrable over any closed subregion of E that does not include the origin. For $n = 1$, the integrability follows from (**37.1**). In the two-variable case, the use of the polar coordinates yields $dA = r\,dr\,d\theta$, $\|\mathbf{r}\| = r$, and

$$\iint_{B_R^E} |f|\,dA \leq M \iint_{B_R^E} \frac{dA}{\|\mathbf{r}\|^\nu} \leq M \iint_{B_R} \frac{dA}{\|\mathbf{r}\|^\nu} = 2\pi M \int_0^R \frac{dr}{r^{\nu-1}},$$

which is finite if $\nu < 2$; the second inequality follows from that the part B_R^E is contained in B_R and the integrand is positive. In the three-variable case, the use of spherical coordinates gives (with $\|\mathbf{r}\| = \rho$)

$$|f|dV \leq \frac{M}{\rho^\nu}\,J\,dV' = \frac{M}{\rho^{\nu-2}}\,\sin\phi\,d\rho\,d\phi\,d\theta$$

So a similar estimate of the improper triple integral of f over B_R^E yields an upper bound $4\pi M \int_0^R \rho^{2-\nu}\,d\rho$, which is finite if $\nu < 3$. □

37.5. Multiple Integrals Over Unbounded Regions. Equation (**37.2**) shows that the improper integral exists if the function decreases sufficiently fast at infinity, e.g., $|f(x)| \leq M/x^\nu$, $\nu > 1$, for all $x > R$ and some constants M and R. A similar sufficient criterion for the existence of a multiple improper integral can be established.

Theorem **37.5**. *Let E be an bounded region in an $n-$dimensional Euclidean space ($n = 1, 2, 3$). Let a function f be integrable on any closed bounded subregion of E. Suppose that*

$$|f(\mathbf{r})| \leq \frac{M}{\|\mathbf{r}\|^\nu} \quad \text{if } \|\mathbf{r}\| \geq R \quad \text{and} \quad \nu > n$$

for some positive R and M. Then the improper integral of f over E exists and can be evaluated in any convenient regularization.

Proof. Let $R > 0$. Let E'_R be the part of E that lies outside the ball B_R of radius R and let B'_E be the part of the space outside B_R (see Fig. **37.2** (left panel)). Note that B'_R includes E'_R. In the two-variable case, the use of polar coordinates gives

$$\iint_{E'_R} |f| dA \le \iint_{B'_R} |f| dA \le \iint_{B'_R} \frac{M dA}{\|\mathbf{r}\|^\nu} = M \int_0^{2\pi} d\theta \int_R^\infty \frac{r\,dr}{r^\nu}$$
$$= 2\pi M \int_R^\infty \frac{dr}{r^{\nu-1}},$$

which is finite, provided $\nu - 1 > -1$ or $\nu > 2$. The case of triple integrals is proved similarly by means of spherical coordinates. The volume element is $dV = \rho^2 \sin\phi\, d\rho\, d\phi\, d\theta$. The integration over the spherical angles yields the factor 4π as $0 \le \phi \le \pi$ and $0 \le \theta \le 2\pi$ for the region B'_R so that

$$\iiint_{E'_R} |f| dV \le \iiint_{B'_R} |f| dV \le \iiint_{B'_R} \frac{M\,dV}{\|\mathbf{r}\|^\nu} = 4\pi M \int_R^\infty \frac{\rho^2\,d\rho}{\rho^\nu}$$
$$= 4\pi M \int_R^\infty \frac{d\rho}{\rho^{\nu-2}},$$

which converges if $\nu > 3$. □

Example **37.4**. *Evaluate the double integral of $f(x,y) = \exp(-x^2 - y^2)$ over the entire plane. Use Fubini's theorem to find the numerical value of the integral of e^{-x^2} over $(-\infty, \infty)$.*

Solution: In polar coordinates $|f| = e^{-r^2}$. So, as $r \to \infty$, $|f|$ decreases faster than any inverse power r^{-n}, $n > 0$, and by virtue of the integrability test and Theorem **37.5**, the improper integral of f exists and can be evaluated in any suitable regularization. By making use of the polar coordinates,

$$\iint_E e^{-x^2-y^2} dA = \lim_{R\to\infty} \int_0^{2\pi} \int_0^R e^{-r^2} r\,dr\,d\theta = \pi \lim_{R\to\infty} \int_0^{R^2} e^{-u}\,du$$
$$= \pi \lim_{R\to\infty} (1 - e^{-R^2}) = \pi,$$

where the substitution $u = r^2$ has been made. On the other hand, choosing a rectangle $[-a,a] \times [-b,b]$ as the regularization, by virtue of Fubini's theorem one infers that

$$\pi = \iint_E e^{-x^2-y^2} dA = \lim_{a\to\infty} \int_{-a}^a e^{-x^2} dx \cdot \lim_{b\to\infty} \int_{-b}^b e^{-y^2} dy = I^2\,,$$
$$\Rightarrow \quad I = \int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}\,.$$

A direct evaluation of I by means of the fundamental theorem of calculus is problematic as an antiderivative of e^{-x^2} cannot be expressed in elementary functions. □

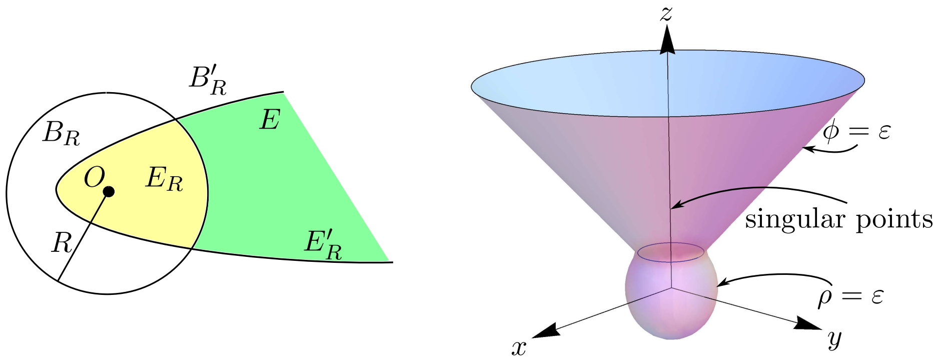


FIGURE **37.2**. **Left**: An unbounded region E is split into two parts E_R that lies inside the ball B_R of radius R and E'_R is the part of E that lies outside the ball B_R. The region B'_R is the entire space with the ball B_R removed. The region E'_R is contained in B'_R. **Right**: A regularization procedure for the integral in Study Problem **37.1**. The integration region E contains singular points along the z axis. The integral is regularized by removing the ball $\rho < \varepsilon$ and the solid cone $\phi < \varepsilon$ from E. After the evaluation of the integral, the limit $\varepsilon \to 0$ is taken.

Volume and area of unbounded regions. Let $f(\mathbf{r}) = 1$ in an unbounded region E. Since f is positive, its improper integral exists if it exists in any particular regularization.

DEFINITION **37.2**. (Area and volume of unbounded regions)
The double and triple integrals

$$A(D) = \iint_D dA\,, \quad V(E) = \iiint_E dV$$

over unbounded regions D and E are called the area of D and the volume of E, respectively, provided they converge.

37.6. Fubini's Theorem and Integrability. If a function is not integrable, its iterated integrals may still exist as improper integrals. However, the value of the iterated integral depends on the order of integration and Fubini's theorem does not hold. For example, consider the function (**37.5**) over the rectangle $[0, 1] \times [0, 1]$. As argued, the function is not integrable on the rectangle because it contains the origin. The improper integral can be regularized by reducing the domain to a rectangle $[a,1]\times[b,1]$, where $a\to 0^+$ and $b\to 0^+$. Consider the iterated integral

$$\begin{aligned}\lim_{a\to 0^+}\int_a^1 \lim_{b\to 0^+}\int_b^1 \frac{x^2-y^2}{(x^2+y^2)^2}dy\,dx &= \lim_{a\to 0^+}\int_a^1 \lim_{b\to 0^+}\int_b^1 \frac{\partial}{\partial y}\frac{y}{x^2+y^2}dy\,dx\\ &= \lim_{a\to 0^+}\int_a^1 \lim_{b\to 0^+}\Big(\frac{1}{1+x^2}-\frac{b}{x^2+b^2}\Big)dx\\ &= \lim_{a\to 0^+}\int_a^1 \frac{dx}{1+x^2} = \int_0^1\frac{dx}{1+x^2} = \frac{\pi}{4}.\end{aligned}$$

So it exists as an improper integral. The iterated integral in the reverse order also exists but has a different value:

$$\begin{aligned}\lim_{b\to 0^+}\int_b^1 \lim_{a\to 0^+}\int_a^1 \frac{x^2-y^2}{(x^2+y^2)^2}dx\,dy &= -\lim_{b\to 0^+}\int_b^1 \lim_{a\to 0^+}\int_a^1 \frac{\partial}{\partial x}\frac{x}{x^2+y^2}dx\,dy\\ &= -\lim_{b\to 0^+}\int_b^1 \lim_{a\to 0^+}\Big(\frac{1}{1+y^2}-\frac{a}{y^2+a^2}\Big)dy\\ &= -\lim_{b\to 0^+}\int_b^1 \frac{dy}{1+y^2} = -\int_0^1\frac{dy}{1+y^2} = -\frac{\pi}{4}.\end{aligned}$$

This example shows that Fubini's theorem *cannot be used unless the existence of the improper integral has been established*. The same observation holds for improper integrals over unbounded regions.

37.7. Study Problems.

Problem **37.1**. *Evaluate the triple integral of* $f(x,y,z) = (x^2+y^2)^{-1/2}(x^2+y^2+z^2)^{-1/2}$ *over* E*, which is bounded by the cone* $z=\sqrt{x^2+y^2}$ *and the sphere* $x^2+y^2+z^2=1$ *if it exists.*

SOLUTION: The function is singular at all points on the z axis. It is positive on the domain of integration. Therefore it is sufficient to investigate the convergence of the improper integral in a particular regularization. Under the transformation T from spherical to rectangular coordinates, E is the image of a rectangular box:

$$T:\quad E' = [0,1]\times[0,\pi/4]\times[0,2\pi]\ \to\ E\,.$$

Consider E_ε obtained from E by eliminating from the latter a solid cone $\phi<\varepsilon$ and a ball $\rho<\varepsilon$, where ρ and ϕ are spherical coordinates (as depicted in the right panel of Fig. **37.2**) so that

$$T:\quad E'_\varepsilon = [\varepsilon,1]\times[\varepsilon,\pi/4]\times[0,2\pi]\ \to\ E_\varepsilon\,.$$

Then transforming the integral over E_ε to spherical coordinates, in which $\sqrt{x^2+y^2}=\rho\sin\phi$, one infers that the improper integral becomes a proper integral over E' because the singularity is cancelled by the Jacobian:

$$f\,dV = (\rho^2\sin\phi)^{-1}\rho^2\sin\phi dV' = dV'$$

So the improper integral in question exists and is equal to

$$\lim_{\varepsilon\to 0}\iiint_{E_\varepsilon} f\,dV = \iiint_{E'} dV' = V(E') = 1\cdot\frac{\pi}{4}\cdot 2\pi = \frac{\pi^2}{2}\,.$$

$\square$

37.8. Exercises.
1–5. Let a function $g(x,y)$ be integrable on any bounded closed region. Assume that $0 < m \le g(x,y) \le M$ for all (x,y). Investigate the existence of each of the following improper double integrals.

1. $\iint_D g(x,y)(x^2+y^2)^{-1}dA$ where D is defined by the conditions $|y|\le x^2$, $x^2+y^2\le 1$;
2. $\iint_D g(x,y)(|x|^p+|y|^q)^{-1}dA$, $p>0$, $q>0$, where D is defined by the condition $|x|+|y|\le 1$;
3. $\iint_D g(x,y)(1-x^2-y^2)^{-p}dA$ where D is defined by the condition $x^2+y^2\le 1$;
4. $\iint_D g(x,y)|x-y|^{-p}dA$ where D is the square $[0,a]\times[0,a]$;
5. $\iint_D e^{-(x+y)}dA$ where D is defined by $0\le x\le y$.

6–9. Let a function $g(x,y,z)$ be integrable on any bounded closed region. Assume that $0 < m \le g(x,y,z) \le M$ for all (x,y,z). Investigate the existence of each of the following improper triple integrals.

6. $\iiint_E g(x,y,z)(x^2+y^2+z^2)^{-\nu}dV$ where E is defined by $x^2+y^2+z^2\ge 1$;
7. $\iiint_E g(x,y,z)(x^2+y^2+z^2)^{-\nu}dV$ where E is defined by $x^2+y^2+z^2\le 1$;
8. $\iiint_E g(x,y,z)(|x|^p+|y|^q+|z|^s)^{-1}dV$, where p, q, s are positive numbers, and E is defined by $|x|+|y|+|z|\ge 1$;
9. $\iiint_E g(x,y,z)|x+y-z|^{-\nu}dV$ where $E=[-1,1]\times[-1,1]\times[-1,1]$.

10. Let n be an integer. Use the Fresnel integral to show that

$$\lim_{n\to\infty}\iint_{D_n}\sin(x^2+y^2)dA = \pi\,,\quad D_n:\ |x|\le n\,,\ |y|\le n;$$
$$\lim_{n\to\infty}\iint_{D_n}\sin(x^2+y^2)dA = 0\,,\quad D_n:\ x^2+y^2\le 2\pi n$$

Note that in each case D_n covers the entire plane as $n\to\infty$. What can be said about the convergence of the integral over the entire plane?
11–30. Evaluate each of the following improper integrals if it exists. Use appropriate coordinates when needed.

11. $\iiint_E (x^2+y^2+z^2)^{-1/2}(x^2+y^2)^{-1/2}dV$ where E is the region in the first octant bounded from above by the sphere $x^2+y^2+z^2=2z$ and from below by the cone $z=\sqrt{3}\sqrt{x^2+y^2}$;

12. $\iiint_E z(x^2+y^2)^{-1/2}$ where E is in the first octant and bounded from above by the cone $z = 2 - \sqrt{x^2+y^2}$ and from below by the paraboloid $z = x^2 + y^2$;
13. $\iiint_E xy(x^2+y^2)^{-1}(x^2+y^2+z^2)^{-1}dV$ where E is the portion of the ball $x^2+y^2+z^2 \le a^2$ above the plane $z = 0$;
14. $\iiint_E e^{-x^2-y^2-z^2}(x^2+y^2+z^2)^{-1/2}dV$ where E is the entire space;
15. $\iint_D (x^2+y^2)^{-1/2}dA$ where D lies between the two circles $x^2+y^2 = 4$ and $(x-1)^2+y^2 = 1$ in the first quadrant, $x, y \ge 0$;
16. $\iint_D \ln(x^2+y^2)dA$ where D is the disk $x^2+y^2 \le a^2$;
17. $\iiint_E (x^2+y^2+z^2)^\nu \ln(x^2+y^2+z^2)dV$ where E is the ball $x^2+y^2+z^2 \le a^2$ and ν is real. Does integral exist for all ν?
18. $\iint_D (x^2+y^2)^\nu \ln(x^2+y^2)dA$ where D is defined by $x^2+y^2 \ge a^2 > 0$ and ν is real. Does integral exist for all ν?
19. $\iiint_E (x^2+y^2+z^2)^\nu \ln(x^2+y^2+z^2)dV$ where E is defined by $x^2+y^2+z^2 \ge a^2 > 0$ and ν is real. Does integral exist for all ν?
20. $\iint_D [(a-x)(x-y)]^{-1/2}dA$ where D is the triangle bounded by the lines $y = 0$, $y = x$, and $x = a$;
21. $\iint_D \ln\sin(x-y)dA$ where D is bounded by the lines $y = 0$, $y = x$, and $x = \pi$;
22. $\iint_D (x^2+y^2)^{-1}dA$ where D is defined by $x^2+y^2 \le x$;
23. $\iiint_E x^{-p}y^{-q}z^{-s}dV$ where $E = [0,1]\times[0,1]\times[0,1]$;
24. $\iiint_E (x^2+y^2+z^2)^{-3}dV$ where E is defined by $x^2+y^2+z^2 \ge 1$;
25. $\iiint_E (1-x^2-y^2-z^2)^{-\nu}dV$ where E is defined by $x^2+y^2+z^2 \le 1$;
26. $\iiint_E e^{-x^2-y^2-z^2}dV$ where E is the entire space;
27. $\iint_D e^{-x^2-y^2}\sin(x^2+y^2)dA$ where D is the entire plane;
28. $\iint_D e^{-(x/a)^2-(y/b)^2}dA$ where D is the entire plane;
29. $\iint_D e^{ax^2+2bxy+cy^2}dA$ where $a < 0$, $ac - b^2 > 0$, and D is the entire plane. Hint: Find a rotation that transforms x and y so that in the new variables the bilinear term "xy" is absent in the exponential;
30. $\iiint_E e^{-(x/a)^2-(y/b)^2-(z/c)^2+\alpha x+\beta y+\gamma z}dV$ where E is the entire space.

31–32. Show that each of the following improper integrals converges. Use the geometric series to show that their values are given by the specified convergent series.

31. $\lim_{a\to 1^-}\iint_{D_a}(1-xy)^{-1}dA = \sum_{n=1}^\infty \frac{1}{n^2}$ where $D_a = [0,a]\times[0,a]$;
32. $\lim_{a\to 1^-}\iiint_{E_a}(1-xyz)^{-1}dV = \sum_{n=1}^\infty \frac{1}{n^3}$ where $E_a = [0,a]\times[0,a]\times[0,a]$.

38. Line Integrals

Consider a wire made of a nonhomogeneous material. The inhomogeneity means that, if one takes a small piece of the wire of length Δs at a point $\mathbf{r}$, then its mass Δm depends on the point $\mathbf{r}$. It can therefore be characterized by a *linear* mass density (the mass per unit length at a point $\mathbf{r}$):

$$\sigma(\mathbf{r}) = \lim_{\Delta s \to 0} \frac{\Delta m(\mathbf{r})}{\Delta s}.$$

Suppose that the linear mass density is known as a function of $\mathbf{r}$. What is the total mass of the wire that occupies a space curve C? If the curve C has a length L, then it can be partitioned into N small segments of length $\Delta s = L/N$. If $\mathbf{r}_p^*$ is a sample point in the pth segment, then the total mass reads

$$M = \lim_{N \to \infty} \sum_{p=1}^{N} \sigma(\mathbf{r}_p^*)\, \Delta s,$$

where the mass of the p^{th} segment is approximated by $\Delta m_p \approx \sigma(\mathbf{r}_p^*)\, \Delta s$ and the limit is required because this approximation becomes exact only in the limit $\Delta s \to 0$. The expression for M resembles the limit of a Riemann sum and leads to the concept of a *line integral* of σ along a curve C.

38.1. Line Integral of a Function. Let f be a bounded function in a region E and let C be a smooth (or piecewise-smooth) curve in E. Suppose C has a finite arclength. Recall Section **13** where the arclength of a smooth curve is defined. Consider a partition of C by its N pieces C_p of length Δs_p, $p = 1, 2, ..., N$, which is the arclength of C_p. Put $m_p = \inf_{C_p} f$ and $M_p = \sup_{C_p} f$; that is, m_p is the largest lower bound of values of f for all $\mathbf{r}$ in C_p, and M_p is the smallest upper bound on the values of f for all $\mathbf{r}$ in C_p. The upper and lower sums are defined by, respectively,

$$U(f, N) = \sum_{p=1}^{N} M_p\, \Delta s_p\,, \quad L(f, N) = \sum_{p=1}^{N} m_p\, \Delta s_p\,.$$

Suppose that $\max_p \Delta s_p = \Delta s_N^* \to 0$ as $N \to \infty$. In other words, the partition element of the maximal arclength becomes smaller with increasing N. The upper and lower sums are the least upper and greatest lower bounds of the total mass of the wire for a given partition in the above mass problem. Naturally, with increasing N these bounds should become closer and coincide with the total mass in the limit $N \to \infty$. Therefore the following definition of the line integral can be adopted.

Definition **38.1**. (Line Integral of a Function).
The line integral of a function f along a piecewise-smooth curve C is

$$\int_C f(\mathbf{r})\, ds = \lim_{N \to \infty} U(f, N) = \lim_{N \to \infty} L(f, N),$$

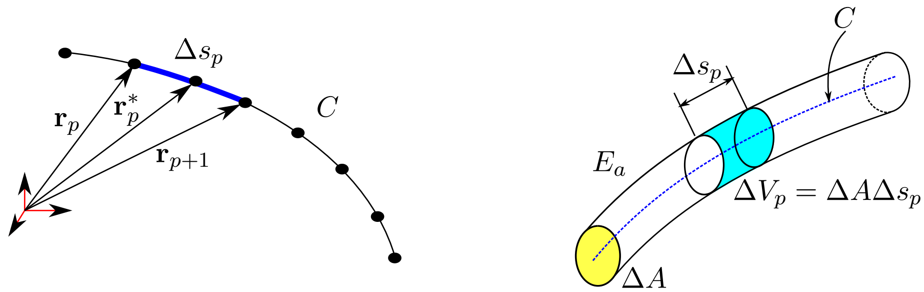


FIGURE **38.1**. **Left**: A partition of a smooth curve C by segments of arclength Δs_p used in the definition of the line integral and its Riemann sum. **Right**: The region E_a is a neighborhood of a smooth curve C. It consists of points whose distance to C cannot exceed $a > 0$ (recall Definition **5.2**). For a and Δs_p small enough, planes normal to C through the points $\mathbf{r}_p$ partition E_a into elements whose volume is $\Delta V_p = \Delta A \Delta s_p$ where $\Delta A = \pi a^2$ is the area of the cross section of E_a (for a small enough). This partition is used to establish the relation (**38.1**) between the triple and line integrals.

provided the limits of the upper and lower sums exist and coincide. The limit is understood in the sense that $\max_p \Delta s_p \to 0$ *as* $N \to \infty$*.*

It follows from this definition that the line integral can also be represented by the limit of a Riemann sum (see the left panel of Fig. **38.1**).

THEOREM **38.1**. (Riemann Sums for a Line Integral)
Suppose f is integrable along a smooth curve C. Let C_p be a partition of C and for each p, a point $\mathbf{r}_p^*$ lies in the curve segment C_p. Then

$$\int_C f(\mathbf{r})\, ds = \lim_{N\to\infty} \sum_{p=1}^{N} f(\mathbf{r}_p^*)\, \Delta s_p = \lim_{N\to\infty} R(f,N)\,.$$

PROOF. For any partition

$$m_p \le f(\mathbf{r}_p^*) \le M_p \quad \Rightarrow \quad L(f,N) \le R(f,N) \le U(f,N)\,.$$

with any choice of sample points $\mathbf{r}_p^*$. The conclusion of the theorem follows from the squeeze principle for limits. □

The following theorem provides sufficient conditions for the existence of the line integral.

THEOREM **38.2**. (Integrability on a Curve)
If f is bounded and possibly not continuous at finitely many points of a piecewise smooth curve C, then the line integral of f along C exists.

It is also interesting to establish a relation of the line integral with a triple integral. Suppose that f is continuous on a spatial region that contains a smooth curve C. Let E_a be a *neighborhood* of C that is defined as the set of points whose distance (in the sense of Definition **5.2**) to C cannot exceed $a > 0$ (think of the union of balls of radius a centered at each point of C). Consider a partition of C by curve segments of arclength Δs_p, $p = 1, 2, ..., N$. Then E_a is partitioned by elements E_{ap} obtained by cutting E_a by normal planes through the endpoints of curve segments C_p (see the right panel of Fig. **38.1**). Recall that a plane normal to a smooth curve is the plane whose normal is a vector tangent to the curve. In the limit $a \to 0^+$, the cross section of E_a by any such plane is a disk of radius a and therefore the volume ΔV_p of the partition element E_{ap} has the property

$$\lim_{a\to 0^+} \frac{\Delta V_p}{\pi a^2} = \Delta s_p\,.$$

By the integral mean value theorem for multiple integrals

$$\iiint_{E_a} f dV = \sum_{p=1}^{N} \iiint_{E_{ap}} f dV = \sum_{p=1}^{N} f(\mathbf{r}_p)\Delta V_p\,,$$

for some points $\mathbf{r}_p$ in E_{ap} and any integer N. Note that $\mathbf{r}_p$ depends on a. However, in the limit $a \to 0^+$, the point $\mathbf{r}_p$ should approach a point $\mathbf{r}_p^*$ on the curve segment C_p by continuity of f. Therefore

$$\lim_{a\to 0^+} \frac{1}{\pi a^2} \iiint_{E_a} f(\mathbf{r})\, dV = \lim_{a\to 0^+} \sum_{p=1}^{N} f(\mathbf{r}_p)\frac{\Delta V_p}{\pi a^2} = \sum_{p=1}^{N} f(\mathbf{r}_p^*)\Delta s_p\,.$$

The latter relation holds for any N and therefore one can take the limit $N \to \infty$, assuming that $\max_p \Delta s_p = \Delta s_N^* \to 0$ as $N \to \infty$. Since f is continuous, the Riemann sum of f over the curve C converges to the line integral:

$$\lim_{a\to 0^+} \frac{1}{\pi a^2} \iiint_{E_a} f(\mathbf{r})\, dV = \int_C f(\mathbf{r})\, ds. \tag{38.1}$$

If C is a planar curve, then a relation similar to (**38.1**) can be established for double integrals by considering the double integral over the planar region D_a defined similarly to E_a. The region D_a is partitioned by normal lines to the curve C. The area ΔA_p of each partition element has the property that $\Delta A_p/(2a) \to \Delta s_p$ as $a \to 0^+$ so that in Eq. (**38.1**) the factor $(\pi a^2)^{-1}$ is replaced by $(2a)^{-1}$ and the triple integral is replaced by the double integral:

$$\lim_{a\to 0^+} \frac{1}{2a} \iint_{D_a} f(\mathbf{r})\, dA = \int_C f(\mathbf{r})\, ds. \tag{38.2}$$

Thus, line integrals can be viewed as the limiting case of multiple integrals when the other dimensions of the integration region become small as compared to the arclength of the curve. In particular, *the line integral inherits all the properties of multiple integrals.*

THEOREM **38.3**. (Mean Value Theorem for Line Integrals)
Suppose f is continuous and C is a smooth curve of length L. Then there exists a point $\mathbf{r}^*$ on C such that

$$\int_C f(\mathbf{r})ds = f(\mathbf{r}^*)L\,.$$

38.2. Evaluation of Line Integrals. The evaluation of a line integral is based on the following theorem.

THEOREM **38.4**. (Evaluation of a Line Integral).
Suppose that f is continuous in a region that contains a smooth curve C. Let $\mathbf{r} = \mathbf{r}(t)$, $a \le t \le b$, be a smooth parameterization of C. Then

$$\int_C f(\mathbf{r})\,ds = \int_a^b f(\mathbf{r}(t))\|\mathbf{r}'(t)\|\,dt. \tag{38.3}$$

PROOF. Consider a partition of $[a, b]$,

$$t_p = a + p\,\Delta t\,, \quad \Delta t = (b-a)/N\,, \quad p = 0, 1, 2, ..., N\,.$$

It induces a partition of C by curve segments C_p so that $\mathbf{r}(t)$ traces out C_p if $t_{p-1} \le t \le t_p$, $p = 1, 2, ..., N$. The arclength of C_p is

$$\Delta s_p = \int_{t_{p-1}}^{t_p} \|\mathbf{r}'(t)\|\,dt = \|\mathbf{r}'(t_p^*)\|\Delta t\,,$$

for some $t_p^* \in [t_{p-1}, t_p]$. The latter equality follows from the integral mean value theorem. Indeed, since C is smooth, the tangent vector $\mathbf{r}'(t)$ is a continuous function and so is its length $\|\mathbf{r}'(t)\|$. By the integral mean value theorem, there exists t_p^* in $[t_{p-1}, t_p]$ such that the value of the integrand at t_p^* times the length of the integration interval gives the value of the integral. Note that $\|\mathbf{r}'(t)\|$ is bounded on $[a, b]$ as any continuous function on a closed bounded interval. This ensures that $\max_p \Delta s_p \to 0$ as $N \to \infty$ (or $\Delta t \to 0$). Since f is integrable along C, the limit of its Riemann sum is independent of the choice of sample points and a partition of C. Choose the sample points to be $\mathbf{r}_p^* = \mathbf{r}(t_p^*)$. Therefore,

$$\int_C f\,ds = \lim_{N\to\infty} \sum_{p=1}^{N} f(\mathbf{r}(t_p^*))\|\mathbf{r}'(t_p^*)\|\Delta t = \int_a^b f(\mathbf{r}(t))\|\mathbf{r}'(t)\|dt.$$

Note that the Riemann sum for the line integral becomes a Riemann sum of the function $g(t) = f(\mathbf{r}(t))\|\mathbf{r}'(t)\|$ over an interval t in $[a, b]$. Its limit exists by the continuity of g and equals the integral of g over $[a, b]$. $\square$

The conclusion of the theorem still holds if f is bounded and not continuous at finitely many points on C, and C is piecewise smooth. The latter implies that the function g is bounded and not continuous at finitely many points in the interval $[a, b]$ and, hence, g is integrable on $[a, b]$.

Thus, the evaluation of a line integral includes the following basic steps:

Step 1. Find a suitable parameterization of a smooth curve C, $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$, where $a \le t \le b$, so that $\mathbf{r}(t)$ traverses C only once as t increases from a to b;
Step 2. Calculate the derivative $\mathbf{r}'(t)$ and its norm $\|\mathbf{r}'(t)\|$;
Step 3. Substitute $x = x(t)$, $y = y(t)$, and $z = z(t)$ into $f(x, y, z)$ and evaluate the integral (**38.3**).

Remark. A curve C may be traced out by different vector functions. The value of the line integral is *independent* of the choice of parametric equations because Definition **38.1** is stated only in parameterization-invariant terms (the arclength and values of the function on the curve). The integrals (**38.3**) written for two different parameterizations of C are equal and can be transformed to one another by changing the integration variable. Recall from Section **13.2** that if $\mathbf{r}(t)$ and $\mathbf{R}(u)$ are two smooth parameterizations of C, then

$$ds = \|\mathbf{r}'(t)\|dt = \|\mathbf{R}'(u)\|du\,.$$

Since neither $\|\mathbf{r}'(t)\|$ nor $\|\mathbf{R}'(u)\|$ vanishes for a smooth parameterization, there is a change of variables $t = t(u)$ or $u = u(t)$ that transforms the integral (**38.3**) with respect to the parameter t to the integral (**38.3**) with respect to the parameter u.

EXAMPLE **38.1**. *Evaluate the line integral of $f(x, y) = x^2y$ over a circle of radius R centered at the point $(0, a)$.*

SOLUTION: Step 1. The equation of a circle of radius R centered at the origin is $x^2 + y^2 = R^2$. It has familiar parametric equations $x = R\cos t$ and $y = R\sin t$, where t is the angle between $\mathbf{r}(t)$ and the positive x axis counted counterclockwise. The equation of the circle in question is $x^2+(y-a)^2 = R^2$. So, by analogy, one can put $x = R\cos t$ and $y - a = R\sin t$ (by shifting the origin to the point $(0, a)$). Parametric equations of the circle can be taken in the form

$$\mathbf{r}(t) = \langle R\cos t, a + R\sin t\rangle\,, \quad 0 \le t \le 2\pi\,.$$

Step 2. The derivative of this vector function and its norm are

$$\mathbf{r}'(t) = \langle -R\sin t, R\cos t\rangle \quad \Rightarrow \quad \|\mathbf{r}'(t)\| = \sqrt{R^2\sin^2 t + R^2\cos^2 t} = R\,.$$

Step 3. Substituting $x = R\cos t$, $y = a - R\sin t$, and $ds = Rdt$ into the line integral,

$$\int_C x^2y\,ds = \int_0^{2\pi} (R\cos t)^2(a + R\sin t)R\,dt = R^3a\int_0^{2\pi} \cos^2 t\,dt = \pi R^3a\,,$$

where the integral

$$\int_0^{2\pi} \cos^2 t\sin t\,dt = -\frac{1}{3}\cos^3 t\Big|_0^{2\pi} = 0$$

vanishes by periodicity of the cosine function. The other integral is evaluated with the help of the double-angle formula $\cos^2 t = (1 + \cos(2t))/2$. □

Example **38.2**. *Evaluate the line integral of* $f(x, y, z) = x^2 + xy + zy$ *along the curve that consists of three straight line segments* $(0, 0, 0) \to (a, 0, 0) \to (a, b, 0) \to (a, b, c)$.

Solution: The curve consists of three smooth pieces (straight line segments): $C_1 : (0, 0, 0) \to (a, 0, 0)$, $C_2 : (a, 0, 0) \to (a, b, 0)$, and $C_3 : (a, b, 0) \to (a, b, c)$. By the additivity of the line integral

$$\int_C f ds = \int_{C_1} f ds + \int_{C_2} f ds + \int_{C_3} f ds$$

The segments can be parameterized as

$$\begin{aligned} C_1 : &\quad (x, y, z) = (x, 0, 0)\,, \quad 0 \le x \le a\,, \quad ds = dx\,; \\ C_2 : &\quad (x, y, z) = (a, y, 0)\,, \quad 0 \le y \le b\,, \quad ds = dy\,; \\ C_3 : &\quad (x, y, z) = (a, b, z)\,, \quad 0 \le z \le c\,, \quad ds = dz\,. \end{aligned}$$

Therefore

$$\begin{aligned} \int_C f ds &= \int_0^a f(x, 0, 0)dx + \int_0^b f(a, y, 0)dy + \int_0^c f(a, b, z)dz \\ &= \int_0^a x^2 dx + \int_0^b (a^2 + ay)dy + \int_0^c (a^2 + ab + bz)dz \\ &= \frac{1}{3}a^3 + a^2 b + \frac{1}{2}ab^2 + (a^2 + ab)c + \frac{1}{2}bc^2\,. \end{aligned}$$

□

Example **38.3**. *Evaluate the line integral of* $f(x, y, z) = \sqrt{3x^2 + 3y^2 - z^2}$ *over the curve of intersection of the cylinder* $x^2 + y^2 = 1$ *and the plane* $x + y + z = 0$.

Solution: Since the curve lies on the cylinder, one can always put

$$x = \cos t\,, \quad y = \sin t\,, \quad z = z(t)\,,$$

where $z(t)$ is to be found from the condition that the curve also lies in the plane:

$$x(t) + y(t) + z(t) = 0 \quad \Rightarrow \quad z(t) = -\cos t - \sin t\,.$$

So C is traversed by the vector function

$$\mathbf{r}(t) = \langle \cos t\,,\ \sin t\,,\ -\cos t - \sin t \rangle\,, \quad 0 \le t \le 2\pi\,.$$

Therefore, using the identity $2 \sin t \cos t = \sin(2t)$,

$$\begin{aligned} \mathbf{r}'(t) &= \langle -\sin t, \cos t, \sin t - \cos t \rangle\,, \\ \|\mathbf{r}'(t)\| &= \Big(\sin^2 t + \cos^2 t + (\sin t - \cos t)^2\Big)^{1/2} = \Big(2 - 2 \sin t \cos t\Big)^{1/2} \\ &= (2 - \sin(2t))^{1/2}\,, \\ f(\mathbf{r}(t)) &= \Big(3 - (\cos t + \sin t)^2\Big)^{1/2} = (2 - \sin(2t))^{1/2} \end{aligned}$$

Note that the function is defined only in the region $3(x^2+y^2) \geq z^2$ (outside the double cone). It happens that the curve C lies in the domain of f and its values along C are well defined as $2 > \sin(2t)$ for any t. Hence,

$$\int_C f\,ds = \int_0^{2\pi} f(\mathbf{r}(t))\|\mathbf{r}'(t)\|dt = \int_0^{2\pi} (2-\sin(2t))\,dt = \int_0^{2\pi} 2dt = 4\pi\,.$$

The integral of $\sin(2t)$ vanishes by periodicity. □

38.3. Exercises.

1–17. Evaluate each of the following line integrals.

1. $\int_C xy^2ds$, where C is the half-circle $x^2+y^2=4$, $x \geq 0$;
2. $\int_C x\sin y ds$, where C is the line segment from $(0,a)$ to $(b,0)$;
3. $\int_C xyzds$, where C is the helix $x = 2\cos t$, $y = t$, $z = -2\sin t$, $0 \leq t \leq \pi$;
4. $\int_C (2x+9z)ds$, where C is the curve $x = t$, $y = t^2$, $z = t^3$ from $(0,0,0)$ to $(1,1,1)$;
5. $\int_C zds$, where C is the intersection of the paraboloid $z = x^2+y^2$ and the plane $z = 4$;
6. $\int_C yds$, where C is the part of the graph $y = e^x$ for $0 \leq x \leq 1$;
7. $\int_C (x+y)ds$, where C is the triangle with vertices $(0,0)$, $(1,0)$, and $(0,1)$;
8. $\int_C y^2ds$, where C is an arc of the cycloid $x = R(t-\sin t)$, $y = R(1-\cos t)$ from $(0,0)$ to $(2\pi R,0)$;
9. $\int_C xyds$, where C is an arc of the hyperbola $x = a\sinh t$, $y = a\cosh t$ for $0 \leq t \leq T$,;
10. $\int_C (x^{4/3}+y^{4/3})ds$, where C is the astroid $x^{2/3}+y^{2/3} = a^{2/3}$;
11. $\int_C xds$, where C is the part of the spiral $r = ae^{\theta}$ that lies in the disk $r \leq a$; here (r,θ) are polar coordinates;
12. $\int_C \sqrt{x^2+y^2}ds$, where C is the circle $x^2+y^2 = ax$;
13. $\int_C y^{-2}ds$, where C is $y = a\cosh(x/a)$;
14. $\int_C (x^2+y^2+z^2)ds$, where C is one turn of the helix $x = R\cos t$, $y = R\sin t$, $z = ht$ $(0 \leq t \leq 2\pi)$;
15. $\int_C y^2ds$, where C is the circle $x^2+y^2+z^2 = 1$, $x+y+z = 0$;
16. $\int_C zds$, where C is the conic helix $x = t\cos t$, $y = t\sin t$, $z = t$ $(0 \leq t \leq T)$;
17. $\int_C zds$, where C is the curve of intersection of the surfaces $x^2+y^2 = z^2$ and $y^2 = ax$ from the origin to the point $(a,a,a\sqrt{2})$, $a > 0$.

18. Find the mass of an arc of the parabola $y^2 = 2ax$, $0 \leq x \leq a/2$, $a > 0$, if its linear mass density is $\sigma(x,y) = |y|$.

19. Find the mass of the curve $x = at$, $y = at^2/2$, $z = at^3/3$, $0 \leq t \leq 1$, $a > 0$, if its linear mass density is $\sigma(x,y,z) = \sqrt{2y/a}$.

39. Surface Integrals

39.1. Surface Area. Suppose a function g in space has continuous partial derivatives and its gradient does not vanish. Then level sets of g are smooth surfaces in space. What is the area of a smooth surface? The question can be answered by the standard trick of integral calculus. The equation $g(x,y,z)=k$ that describes a smooth surface can be solved for one of the variables (by the implicit function theorem), say, $z=f(x,y)$ where (x,y) is in some region D, and the function f has continuous partial derivatives. The equation $z=f(x,y)$ defines the graph of f over D. In general, the level surface of g can always be represented as the union of several graphs. So, it is sufficient to answer the question about the surface area for the graph of a function that has continuous partial derivatives.

Let D be a bounded closed region in the xy plane. It can be embedded into a rectangle $R_D=[a,b]\times[c,d]$. Consider a rectangular partition of D:

$$x_i = a + i\Delta x\,, \quad x_0 = a\,, \quad \Delta x = (b-a)/N_1\,, \quad i=1,2,...,N_1\,;$$
$$y_j = c + j\Delta y\,, \quad y_0 = c\,, \quad \Delta y = (d-c)/N_2\,, \quad j=1,2,...,N_2\,.$$

Let a partition rectangle

$$R_{ij} = [x_i, x_i+\Delta x]\times[y_j, y_j+\Delta y]$$

be contained in the interior of D. Let ΔS_{ij} be the area of the part of the graph of f that lies above R_{ij}. As f has continuous partial derivatives, it is differentiable, and its linearization at any point in R_{ij} defines a tangent plane to the graph. Then ΔS_{ij} can be approximated by the area of the parallelogram that lies above R_{ij} in the tangent plane to the graph through a point (x_i^*, y_j^*, z_{ij}^*), where $z_{ij}^* = f(x_i^*, y_j^*)$ and (x_i^*, y_j^*) is any sample point in R_{ij}. Recall that the differentiability of f means that the deviation of f from its linearization tends to zero faster than $\sqrt{(\Delta x)^2+(\Delta y)^2}$ as $(\Delta x, \Delta y)\to(0,0)$. Therefore, in this limit, only terms linear in Δx and Δy must be retained, when calculating ΔS_{ij}, and hence the surface area ΔS_{ij} and the area of the partition rectangle $\Delta A = \Delta x\,\Delta y$ have to be proportional in this limit:

$$\Delta S_{ij} = J_{ij}\,\Delta A.$$

The coefficient J_{ij} is found by comparing the area of the parallelogram in the tangent plane above R_{ij} with the area ΔA of R_{ij}. Think of the roof of a building of shape $z=f(x,y)$ covered by shingles of area ΔS_{ij}. The equation of the tangent plane is

$$z = z_{ij}^* + f_x'(x_i^*, y_j^*)(x-x_i^*) + f_y'(x_i^*, y_j^*)(y-y_j^*) = L(x,y).$$

Let O', A', and B' be, respectively, the vertices $(x_i, y_j, 0)$, $(x_i+\Delta x, y_j, 0)$, and $(x_i, y_j+\Delta y, 0)$ of the rectangle R_{ij}; that is, the segments $O'A'$ and $O'B'$ are the adjacent sides of R_{ij} (see the left panel of Fig. **39.1**). If O, A, and B are the points in the tangent plane above O', A', and B', respectively, then

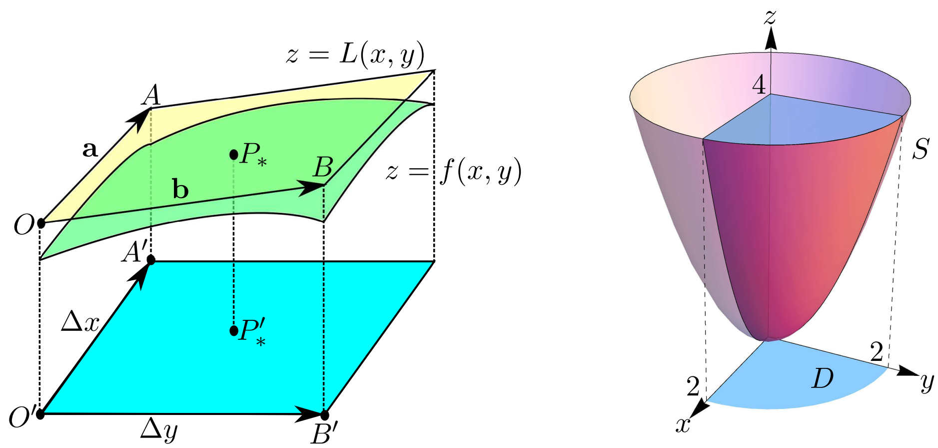


FIGURE **39.1**. **Left**: The rectangle with adjacent sides $O'A'$ and $O'B'$ is an element of a rectangular partition of D and P'_* is a sample point. The point P_* is the point on the graph $z = f(x, y)$ for $(x, y) = P'_*$. The linearization of f at P_* defines the tangent plane $z = L(x, y)$ to the graph through P_*. The surface area of the portion of the graph above the partition rectangle is approximated by the area of the portion of the tangent plane above the partition rectangle which is the area of the parallelogram with adjacent sides OA and OB. It equals $\|\mathbf{a} \times \mathbf{b}\|$. **Right**: An illustration to Example **39.1**. The part of the paraboloid whose area is to be evaluated is obtained by restricting (x, y) to the part D of the disk of radius 2 that lies in the first quadrant.

the adjacent sides of the parallelogram in question are $\mathbf{a} = \overrightarrow{OA}$ and $\mathbf{b} = \overrightarrow{OB}$ and

$$\Delta S_{ij} = \|\mathbf{a} \times \mathbf{b}\|$$

if only the leading term, proportional to $\Delta x \Delta y$, is retained in the limit $(\Delta x, \Delta y) \to (0, 0)$.

By substituting the coordinates of O', A', and B' into the equation of the tangent plane, the coordinates of the points O, A, and B are found:

$$\begin{aligned} O &= (x_i, y_j, L(x_i, y_j)) , \\ A &= (x_i + \Delta x, y_j, L(x_i + \Delta x, y_j)) , \\ B &= (x_i, y_j + \Delta y, L(x_i, y_j + \Delta y) . \end{aligned}$$

By the linearity of the function L,

$$\begin{aligned} L(x_i+\Delta x, y_j) - L(x_i, y_j) &= f'_x(x_i^*, y_j^*)\,\Delta x\,, \\ L(x_i, y_j+\Delta y) - L(x_i, y_j) &= f'_y(x_i^*, y_j^*)\Delta y \end{aligned}$$

and therefore

$$\begin{aligned} \mathbf{a} &= \langle \Delta x, 0, f'_x(x_i^*, y_j^*)\,\Delta x\rangle = \Delta x\,\langle 1,\ 0,\ f'_x(x_i^*, y_j^*)\rangle\,, \\ \mathbf{b} &= \langle 0, \Delta y, f'_y(x_i^*, y_j^*)\,\Delta y\rangle = \Delta y\,\langle 0,\ 1,\ f'_y(x_i^*, y_j^*)\rangle\,, \\ \mathbf{a}\times\mathbf{b} &= \Delta x\,\Delta y\ \langle -f'_x(x_i^*, y_j^*),\ -f'_y(x_i^*, y_j^*),\ 1\rangle\,, \\ \Delta S_{ij} &= \|\mathbf{a}\times\mathbf{b}\| = J(x_i^*, y_j^*)\,\Delta x\Delta y = J(x_i^*, y_j^*)\,\Delta A = J_{ij}\Delta A\,, \\ J(x,y) &= \sqrt{1+(f'_x)^2+(f'_y)^2}\,. \end{aligned}$$

If the intersection of a partition rectangle R_{ij} with D contains at most only points of the boundary of D, then it is natural to set $\Delta S_{ij} = 0$. If R_{ij} is not contained in D, but intersects the interior of D, then $\Delta S_{ij} = J_{ij}\Delta A$ where the sample point (x_i^*, y_j^*) can be chosen in the interior of D. With this agreement, the sum of ΔS_{ij} over the partition is a Riemann sum of a continuous function $J(x,y)$ over D. Assuming that the boundary of D is piecewise smooth, the Riemann sum should converge to the double integral of J over D, and by the geometrical construction of the Riemann sum, this limit is the surface area:

$$A(S) = \lim_{N_1,N_2\to\infty}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} J(x_i^*, y_j^*)\ \Delta A = \iint_D J(x,y)dA\,.$$

If $f(x,y) = \text{const}$, then $f'_x = f'_y = 0$ and $A(S) = A(D)$ as required. Note that the continuity of partial derivatives and the linearization of f can be established only for interior points of D if f is not defined outside of a *closed* region D. However the above double integral still exists by Corollary **28.1**, provided f'_x and f'_y are bounded on the interior of D and the boundary of D is piecewise smooth. Thus, the following definition of the surface area can be adopted.

Definition **39.1**. (Surface Area).
Suppose that $f(x,y)$ has continuous first-order partial derivatives on a closed bounded region D bounded by a piece-wise smooth curve. Then the surface area of the graph $z = f(x,y)$ is given by

$$A(S) = \iint_D \sqrt{1+(f'_x)^2+(f'_y)^2}\ dA.$$

Example **39.1**. *Find the area of the part of the paraboloid $z = x^2+y^2$ in the first octant and below the plane $z = 4$.*

Solution: The surface in question is the graph $z = f(x,y) = x^2+y^2$. Next, the region D must be specified (it determines the part of the graph whose area is to be found). One can view D as the vertical projection of the surface onto the xy plane. The plane $z = 4$ intersects the paraboloid above the circle $4 = x^2 + y^2$ of radius 2. Since the surface also lies in the first octant, D is the part of the disk $x^2+y^2 \le 4$ in the first quadrant. Then $f'_x = 2x$, $f'_y = 2y$, and $J = (1 + 4x^2 + 4y^2)^{1/2}$. The surface area is

$$\begin{aligned} A(S) &= \iint_D \sqrt{1+4x^2+4y^2}\, dA = \int_0^{\pi/2} d\theta \int_0^2 \sqrt{1+4r^2}\, r\, dr \\ &= \frac{\pi}{2}\int_0^2 \sqrt{1+4r^2}\, r\, dr = \frac{\pi}{16}\int_1^{17} \sqrt{u}\, du = \frac{\pi}{24}(17^{3/2}-1), \end{aligned}$$

where the double integral has been converted to polar coordinates and the substitution $u = 1 + 4r^2$ has been used to evaluate the last integral. □

Remark. Suppose that partial derivatives of f are continuous but not bounded in the interior of a region D or they do not exist at the boundary of D. The surface area may still exist in the sense of Definition **39.1** if the double integral is treated as an improper integral. Since $J > 0$ in the interior of D, it can be regularized in any convenient way according to Theorems **37.1** and **37.2**. Similarly *the surface area of an unbounded surface* is defined as the corresponding improper integral, provided it converges. Since J is positive, the improper integral can be evaluated in any convenient regularization. If in either of the two cases the improper integral diverges, the surface area is said to be *infinite*.

EXAMPLE **39.2**. *Show that the surface area of a sphere of radius R is $4\pi R^2$.*

SOLUTION: The hemisphere is the graph $z = f(x, y) = \sqrt{R^2 - x^2 - y^2}$ on the disk $x^2 + y^2 \le R^2$ of radius R. The area of the sphere is twice the area of this graph. One has

$$\begin{aligned} f'_x &= -\frac{x}{\sqrt{R^2-x^2-y^2}} = -\frac{x}{f}\,, \quad f'_y = -\frac{y}{\sqrt{R^2-x^2-y^2}} = -\frac{y}{f}\,, \\ J &= \left(1+\frac{x^2}{f^2}+\frac{y^2}{f^2}\right)^{1/2} = \frac{(f^2+x^2+y^2)^{1/2}}{f} = \frac{R}{\sqrt{R^2-x^2-y^2}}\,. \end{aligned}$$

The partial derivatives do not exist at the boundary of D, the circle $x^2+y^2 = R^2$. The surface area integral is not proper. One can regularize it by reducing D to the disk $D_a:\ x^2 + y^2 = a^2 < R^2$ and after evaluation of the integral take the limit $a \to R^-$. Hence, by converting the double integral to polar coordinates

$$\begin{aligned} A(S) &= 2R \lim_{a\to R^-} \iint_{D_a} \frac{dA}{\sqrt{R^2-x^2-y^2}} = 2R \lim_{a\to R^-} \int_0^{2\pi} d\theta \int_0^a \frac{r\,dr}{\sqrt{R^2-r^2}} \\ &= 4\pi R \lim_{a\to R^-} \int_0^a \frac{r\,dr}{\sqrt{R^2-r^2}} = 4\pi R \lim_{a\to R^-} \left(-\sqrt{R^2-r^2}\right)\Big|_0^a = 4\pi R^2, \end{aligned}$$

where the substitution $u = R^2 - r^2$ has been used to evaluate the last integral. □

39.2. Surface Integral of a Function. An intuitive idea of the concept of the surface integral of a function can be understood from the following example. Suppose one wants to find the total human population on the globe. The data about the population is usually supplied as the population *density* (i.e., the number of people per unit area). The population density is not a constant function on the globe. It is high in cities and low in deserts and jungles. Let $\sigma(\mathbf{r})$ be the population density as a function of position $\mathbf{r}$ on the globe ($\mathbf{r}$ is taken relative to some coordinate system in space). Consider a partition of the surface of the globe by surface elements of area ΔS_p. Then the population on each partition element is approximately $\sigma(\mathbf{r}_p^*)\,\Delta S_p$, where $\mathbf{r}_p^*$ is a sample point in the partition element. The approximation neglects variations of σ within each partition element. The total population is approximately the Riemann sum $\sum_p \sigma(\mathbf{r}_p^*)\,\Delta S_p$. To get an exact value, the partition has to be refined so that the size of each partition element becomes smaller. The limit is the surface integral of σ over the surface of the globe, which is the total population. In general, one can think of some quantity distributed over a surface with some density (the amount of this quantity per unit area as a function of position on the surface). The total amount is the surface integral of the density over the surface.

Let f be a bounded function in an open region E and let S be a surface in E that has a finite surface area. Consider a partition of S by N pieces S_p, $p = 1, 2, ..., N$, which have surface area ΔS_p. Suppose that S is defined as a level surface $g(x, y, z) = k$ of a function g that has continuous partial derivatives on E and whose gradient does not vanish. Then for any point P on S there is a function of two variables whose graph coincides with S in a neighborhood of P and the function has continuous partial derivatives. So the surface area ΔS_p of a partition element S_p can be evaluated by Definition **39.1**. Put $m_p = \inf_{S_p} f$ and $M_p = \sup_{S_p} f$; that is, m_p is the greatest lower bound of values of f for all $\mathbf{r}$ in S_p and M_p is the least upper bound on the values of f for all $\mathbf{r}$ in S_p. The upper and lower sums are defined by

$$U(f, N) = \sum_{p=1}^{N} M_p\,\Delta S_p\,, \quad L(f, N) = \sum_{p=1}^{N} m_p\,\Delta S_p\,.$$

Let R_p be the radius of the smallest ball that contains S_p and $\max_p R_p = R_N^*$. A partition of S is said to be refined if R_N^* is decreasing with increasing N so that $R_N^* \to 0$ as $N \to \infty$. In other words, under a refinement, the sizes R_p of partition elements become *uniformly* smaller with increasing the number N of partition elements.

Definition **39.2**. (Surface Integral of a Function).
The surface integral of a bounded function f over a smooth bounded surface

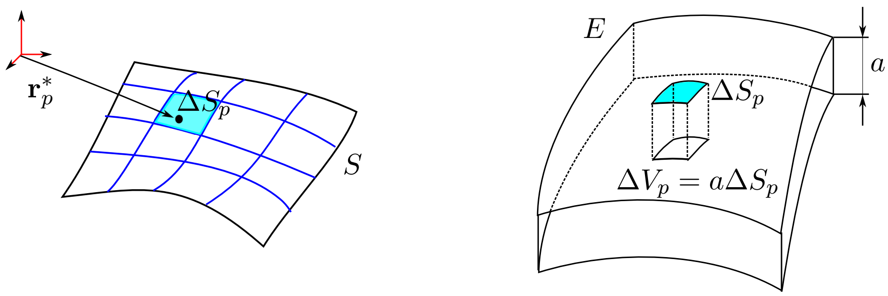


FIGURE **39.2**. **Left**: A partition of a surface S by elements with surface area ΔS_p. It is used in the definition of the surface integral and also to construct its Riemann sums. **Right**: A neighborhood E_a of a smooth surface S defined as the set of points whose distance to S cannot exceed $a > 0$. For sufficiently fine partition of S and small a, the region E_a is partitioned by elements of volume $\Delta V_p = a\Delta S_p$.

S is

$$\iint_S f(\mathbf{r})\, dS = \lim_{N\to\infty} U(f,N) = \lim_{N\to\infty} L(f,N),$$

provided the limits of the upper and lower sums exist and coincide. The limit is understood in the sense $R_N^* \to 0$ as $N \to \infty$.

If the surface integral of f exists, then it can also be represented by the limit of a Riemann sum:

$$\iint_S f(\mathbf{r})\, dS = \lim_{N\to\infty} \sum_{p=1}^{N} f(\mathbf{r}_p^*)\, \Delta S_p = \lim_{N\to\infty} R(f,N)\,, \tag{39.1}$$

for any choice of sample points $\mathbf{r}_p^*$ in partition elements S_p. Indeed, it follows from the definition of supremum and infimum that

$$m_p \le f(\mathbf{r}) \le M_p \quad \Rightarrow \quad L(f,N) \le R(f,N) \le U(f,N)\,.$$

The Riemann sum converges to the surface integral by the squeeze principle and its limit is *independent* of the choice of sample points $\mathbf{r}_p^*$. Riemann sums can be used in numerical approximations of the surface integral.

The following theorem provides sufficient conditions for the existence of the surface integral.

THEOREM **39.1**. (Integrability on a Surface)
If f is bounded and possibly not continuous at finitely many smooth curves in a piecewise smooth bounded surface S, then the surface integral of f over the surface S exists.

Similar to line integrals, surface integrals are related to triple integrals. Consider a neighborhood E_a of a smooth surface S which is defined as the set of points whose distance to S cannot exceed $a/2 > 0$ (in the sense of Definition **5.2**). The region E_a looks like a shell with thickness a (see the right panel of Fig. **39.2**). Suppose that f is continuous on E_a. Since S is smooth, there is a normal line through each point of S. For a small enough, the segments of normal lines of length a through any two neighboring points of S do not intersect. The region E_a can be partitioned by solid regions E_{ap} of volume ΔV_p such that the intersection of E_{ap} with the surface S is a part S_p of S and all S_p form a partition of S. If ΔS_p is the surface area of S_p, then

$$\lim_{a\to 0^+}\frac{\Delta V_p}{a} = \Delta S_p\,.$$

Using the integral mean value theorem for the triple integral over E_{ap}, it is concluded that

$$\iiint_{E_a} f dV = \sum_{p=1}^{N}\iiint_{E_{ap}} f dV = \sum_{p=1}^{N} f(\mathbf{r}_p)\Delta V_p\,,$$

for some points $\mathbf{r}_p$ in E_{ap} and any integer N. Note that $\mathbf{r}_p$ depends on a. However, in the limit $a \to 0^+$, the point $\mathbf{r}_p$ should approach a point $\mathbf{r}_p^*$ on the surface S_p by continuity of f. Therefore

$$\lim_{a\to 0^+}\frac{1}{a}\iiint_{E_a} f(\mathbf{r})\,dV = \lim_{a\to 0^+}\sum_{p=1}^{N} f(\mathbf{r}_p)\frac{\Delta V_p}{a} = \sum_{p=1}^{N} f(\mathbf{r}_p^*)\Delta S_p\,.$$

The latter relation holds for any N and therefore one can take the limit $N \to \infty$ in the sense that $R_N^* \to 0$ (as in Definition **39.2**). Since f is continuous, the Riemann sum of f over the surface S converges to the surface integral:

$$\lim_{a\to 0^+}\frac{1}{a}\iiint_{E_a} f(\mathbf{r})\,dV = \iint_S f(\mathbf{r})\,dS. \tag{39.2}$$

This shows that *the surface integral inherits all the properties of multiple integrals*. In particular, if f is *continuous* on S, then there exists a point $\mathbf{r}^*$ in S such that

$$\iint_S f(\mathbf{r})dS = f(\mathbf{r}^*)A(S)\,,$$

which is nothing but *the integral mean value theorem for surface integrals*.

39.3. Evaluation of a Surface Integral.

THEOREM **39.2**. (Evaluation of a Surface Integral).
Suppose that f is continuous in a region that contains a surface S defined by the graph $z = g(x, y)$ on D. Suppose that g has continuous partial derivatives on an open region that contains D. Then

$$\iint_S f(x,y,z)\,dS = \iint_D f(x,y,g(x,y))\sqrt{1+(g_x')^2+(g_y')^2}\;dA. \tag{39.3}$$

Proof. By Theorem **39.1**, the surface integral exists. So, it can be evaluated as the limit of any convenient Riemann sum as the limit is independent of the choice of partition or sample points. Consider a partition of D by elements D_p of area ΔA_p, $p = 1, 2, ..., N$. Let $J(x, y) = \sqrt{1 + (g'_x)^2 + (g'_y)^2}$. By the continuity of g'_x and g'_y, J is continuous on D. By the integral mean value theorem, the area of the part of the graph $z = g(x, y)$ over D_p is given by

$$\Delta S_p = \iint_{D_p} J(x, y)\, dA = J(x_p^*, y_p^*)\, \Delta A_p$$

for some (x_p^*, y_p^*) in D_p. In the Riemann sum for the surface integral (**39.1**), take the sample points to be $\mathbf{r}_p^* = \langle x_p^*, y_p^*, g(x_p^*, y_p^*)\rangle$ in S_p. The Riemann sum becomes the Riemann sum (**29.2**) of the function

$$F(x, y) = f(x, y, g(x, y))J(x, y)$$

on D. By the continuity of F (because f, g and J are continuous functions), it converges to the double integral of F over D. $\square$

The evaluation of the surface integral involves the following steps:

- Step 1. Represent S as a graph $z = g(x, y)$; that is, find the function g using a geometrical description of S. If S cannot be represented as a graph of a single function, cut the surface into several pieces each of which is a graph, use the additivity of the surface integral;
- Step 2. Find the region D that defines the part of the graph that coincides with S (if S is not the entire graph);
- Step 3. Calculate the derivatives g'_x and g'_y and the area transformation function J, $dS = J\, dA$;
- Step 4. Evaluate the double integral (**39.3**).

Example **39.3**. *Evaluate the integral of $f(x, y, z) = z$ over the part of the saddle surface $z = xy$ that lies inside the cylinder $x^2 + y^2 = 1$ in the first octant.*

Solution: Step 1. The surface is a part of the graph $z = g(x, y) = xy$. Step 2. Since the surface lies within the cylinder, its projection onto the xy plane is bounded by the circle of unit radius, $x^2 + y^2 = 1$. The first octant is projected onto the first (positive) quadrant in the xy plane. Thus, D is the part of the disk $x^2 + y^2 \le 1$ in the first quadrant.
Step 3. One has $g'_x = y$, $g'_y = x$, and $J(x, y) = (1 + x^2 + y^2)^{1/2}$.

Step 4. The surface integral is

$$\begin{aligned}\iint_S z\,dS &= \iint_D xy\sqrt{1+x^2+y^2}\,dA \\ &= \int_0^{\pi/2}\cos\theta\sin\theta\,d\theta\int_0^1 r^2\sqrt{1+r^2}\;r\,dr \\ &= \frac{\sin^2\theta}{2}\Big|_0^{\pi/2}\cdot\frac{1}{2}\int_1^2(u-1)\sqrt{u}\,du \\ &= \frac{1}{2}\Big(\frac{u^{5/2}}{5}-\frac{u^{3/2}}{3}\Big)\Big|_1^2 = \frac{\sqrt{2}+1}{15},\end{aligned}$$

where the double integral has been converted to polar coordinates and the last integral is evaluated by the substitution $u = 1 + r^2$ so that $du = 2rdr$. □

39.4. Parametric Equations of a Surface. The graph $z = g(x, y)$, where (x, y) in D, defines a surface S in space. Consider the vectors

$$\mathbf{r}(u,v) = \langle x(u,v),\, y(u,v),\, z(u,v)\rangle = \langle u,\ v,\ g(u,v)\rangle\,,$$

where the ordered pair of parameters (u, v) spans the region D. For every pair (u, v), the rule $\mathbf{r} = \mathbf{r}(u, v)$ defines a vector in space which is the position vector of a point on the surface. Consider a change of variables defined by a transformation

$$T:\ D' \ \to D\,,\quad u = u(u',v')\,,\ v = v(u',v')\,.$$

Then the components of position vectors of points of S become general functions of the new variables (u', v'):

$$\mathbf{r} = \mathbf{r}(u',v') = \langle x(u',v'),\ y(u',v')\ \ z(u',v')\rangle\,.$$

This observation suggests that a surface in space can be defined by specifying three functions of two variables that span a planar region; these functions are viewed as components of the position vector in space.

Definition **39.3**. (Parametric Surface)
A mapping of a planar region D into space defined by the rule

$$\mathbf{r} = \mathbf{r}(u,v) = \langle x(u,v),\ y(u,v),\ z(u,v)\rangle\,,$$

where $x(u, v)$, $y(u, v)$, and $z(u, v)$ are continuous functions on D, is called a parametric surface in space, *and the equations $x = x(u, v)$, $y = y(u, v)$, and $z = z(u, v)$ are called* parametric equations *of the surface.*

For example, the equations

$$x = R\cos\theta\sin\phi\,,\quad y = R\sin\theta\sin\phi\,,\quad z = R\cos\phi \qquad (\mathbf{39.4})$$

are parametric equations of a sphere of radius R. Indeed, by comparing these equations with the spherical coordinates, one finds that $(\rho, \phi, \theta) = (R, u, v)$; that is, when $(u, v) = (\phi, \theta)$ range over the rectangle $D = [0, \pi] \times [0, 2\pi]$, the vector $\langle x, y, z\rangle = \mathbf{r}(\phi, \theta)$ traces out the sphere $\rho = R$. An apparent advantage of using parametric equations of a surface is that the surface no longer needs be represented as the union of graphs. For example, the whole sphere is described by the single vector-valued function (**39**.**4**) of two variables instead of the union of two graphs $z = \pm\sqrt{R^2 - x^2 - y^2}$.

DEFINITION **39**.**4**. (Smooth Parametric Surface)
Let a vector function $\mathbf{r}(u, v)$ be defined on a closed planar region D. If the vector function is one-to-one, has continuous partial derivatives $\mathbf{r}'_u$ and $\mathbf{r}'_v$ in the interior of D such that the vector $\mathbf{n} = \mathbf{r}'_u \times \mathbf{r}'_v$ does not vanish, and the components of the unit vector $\hat{\mathbf{n}}$ parallel to $\mathbf{n}$ are continuously extendable to the boundary of D, then the range S of the vector function on D is called a smooth parametric surface*.*

Let (u_0, v_0) be a point on the boundary of D. Since $\hat{\mathbf{n}} = \hat{\mathbf{n}}(u, v)$ is well defined in the interior of D ($\mathbf{n}(u, v)$ does not vanish), one can investigate the limit $\lim_{(u,v)\to(u_0,v_0)} \hat{\mathbf{n}}(u, v)$ which is understood as the limit of each component of $\hat{\mathbf{n}}$ (just like the limit of a vector function). Following the discussion of Section **17.4**, the components of $\hat{\mathbf{n}}$ are continuously extendable to the limit point (u_0, v_0) if the above limit exists. An analogy can be made with parametric equations of a curve in space. A curve in space is a mapping of an *interval* $[a, b]$ into space defined by a vector function of *one variable* $\mathbf{r}(t)$. If $\mathbf{r}'(t)$ is continuous and $\mathbf{r}'(t) \neq \mathbf{0}$, then the curve has a continuous tangent vector and the curve is smooth. Similarly, the condition $\mathbf{r}'_u \times \mathbf{r}'_v \neq \mathbf{0}$ ensures that the surface has a continuous normal vector just like a graph of a continuously differentiable function of two variables. If $\mathbf{r}'(t_0) = \mathbf{0}$ or $\mathbf{r}'(t_0)$ does not exists for a particular t_0, but the limit of the unit tangent vector $\hat{\mathbf{v}}(t) = \hat{\mathbf{r}}'(t)/\|\mathbf{r}'(t)\|$ as $t \to t_0$ exists, the curve is smooth at $\mathbf{r}(t_0)$ (recall the discussion in Section **11.3**).

EXAMPLE **39**.**4**. *Find the parametric equations of the double cone $z^2 = x^2 + y^2$.*

SOLUTION: Suppose $z \neq 0$. Then $(x/z)^2 + (y/z)^2 = 1$. The solution of this equation is $x/z = \cos u$ and $y/z = \sin u$, where u in $[0, 2\pi)$. Therefore, the parametric equations are

$$x = v\cos u\,, \quad y = v\sin u\,, \quad z = v,$$

where (u, v) in $[0, 2\pi) \times (-\infty, \infty)$ for the whole double cone. Of course, there are many different parameterizations of the same surface. They are related by a change of variables $u = u(s, t)$, $v = v(s, t)$, where (s, t) are new parameters of the same surface S. □

EXAMPLE **39**.**5**. *A torus is a surface obtained by rotating a circle about an axis outside the circle and parallel to its diameter. Find the parametric equations of a torus.*

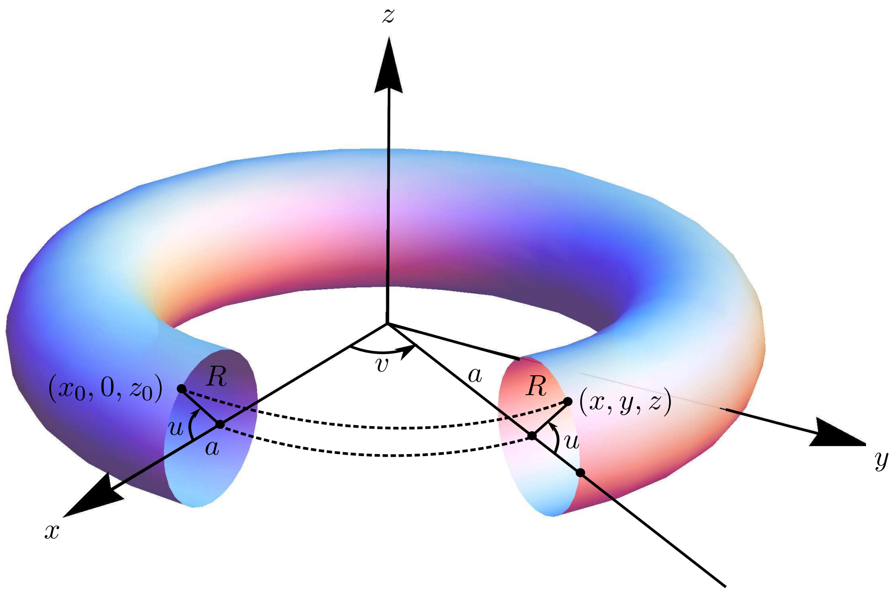


FIGURE **39.3**. A torus. Consider a circle of radius R in the zx plane whose center is positioned on the positive x axis at a distance $a > R$. Any point $(x_0, 0, z_0)$ on the circle is obtained from the point $(a + R, 0, 0)$ by rotation about the center of the circle through an angle $0 \leq u \leq 2\pi$ so that $x_0 = a + R\cos u$ and $z_0 = R\sin u$. A torus is a surface swept by the circle when the xz plane is rotated about the z axis. A generic point (x, y, z) on the torus is obtained from $(x_0, 0, z_0)$ by rotating the latter about the z axis through an angle $0 \leq v \leq 2\pi$. Under this rotation z_0 does not change and $z = z_0$, while the pair $(x_0, 0)$ in the xy plane changes to $(x, y) = (x_0 \cos v, x_0 \sin v)$. Parametric equations of a torus are $x = (a+R\cos u)\cos v$, $y = (a+R\cos u)\sin v$, $z = R\sin u$, where (u, v) ranges over the rectangle $[0, 2\pi] \times [0, 2\pi]$.

SOLUTION: Let the rotation axis be the z axis. Let a be the distance from the z axis to the center of the rotated circle and let R be the radius of the latter, $a \geq R$. In the xz plane, the rotated circle is $z^2 + (x - a)^2 = R^2$. Let $(x_0, 0, z_0)$ be a solution to this equation. The point $(x_0, 0, z_0)$ traces out the circle of radius x_0 upon the rotation about the z axis. All such points are $(x_0 \cos v, x_0 \sin v, z_0)$, where v in $[0, 2\pi]$. Since all points $(x_0, 0, z_0)$ are on the circle $z^2 + (x-a)^2 = R^2$, they can be parameterized as $x_0 - a = R\cos u$, $z_0 = R\sin u$ where u in $[0, 2\pi]$. Thus, the parametric equations of a torus are

$$x = (a + R\cos u)\cos v\,, \quad y = (a + R\cos u)\sin v\,, \quad z = R\sin u\,, \tag{39.5}$$

where (u, v) in $[0, 2\pi] \times [0, 2\pi]$. An alternative (geometrical) derivation of these parametric equations is given in the caption of Fig. **39.3**. □

A tangent plane to a parametric surface. The line $v = v_0$ in D is mapped onto the curve $\mathbf{r} = \mathbf{r}(u, v_0)$ in S (see Fig. **39.4**). The derivative $\mathbf{r}'_u(u, v_0)$ is tangent to the curve. Similarly, the line $u = u_0$ in D is mapped to the curve $\mathbf{r} = \mathbf{r}(u_0, v)$ in S and the derivatives $\mathbf{r}'_v(u_0, v)$ is tangent to it. If the cross product $\mathbf{r}'_u \times \mathbf{r}'_v$ does not vanish in D, then one can define a plane normal to the cross product at any point of S. Furthermore, if $\mathbf{r}'_u \times \mathbf{r}'_v \neq \mathbf{0}$ in a neighborhood of (u_0, v_0), then without loss of generality, one can assume that, say, the z component of the cross product is not zero: $x'_u y'_v - x'_v y'_u = \partial(x, y)/\partial(u, v) \neq 0$. This shows that the *transformation* $x = x(u, v)$, $y = y(u, v)$ *with continuous partial derivatives has a non-vanishing Jacobian.* By the inverse function theorem (Theorem **33.1**), there exists an inverse transformation $u = u(x, y)$, $v = v(x, y)$ which also has continuous partial derivatives. So the vector function $\mathbf{r}(u, v)$ can be written in the new variables (x, y) as

$$\mathbf{R}(x, y) = \mathbf{r}(u(x, y), v(x, y)) = (x, y, z(u(x, y), v(x, y)) = (x, y, g(x, y))$$

which is a vector function that traces out the graph $z = g(x, y)$. Thus, *a smooth parametric surface near any of its points can always be represented as the graph of a function of two variables.* By the chain rule, the function g has continuous partial derivatives. Therefore its linearization near $(x_0, y_0) = (x(u_0, v_0), y(u_0, v_0))$ defines the *tangent plane* to the graph and, hence, to the parametric surface at the point $\mathbf{r}_0 = \mathbf{r}(u_0, v_0)$. In particular, the vectors $\mathbf{r}'_v$ and $\mathbf{r}'_u$ must lie in this plane as they are tangent to two curves in the graph. Thus, the vector $\mathbf{r}'_u \times \mathbf{r}'_v$ is normal to the tangent plane. So Definition **39.4** of a smooth *parametric* surface agrees with the notion of a smooth surface as a level set of a function with continuous partial derivatives and a non-vanishing gradient and the following theorem holds.

THEOREM **39.3**. (Normal to a Smooth Parametric Surface).
Let $\mathbf{r} = \mathbf{r}(u, v)$ be a smooth parametric surface. Then the vector $\mathbf{n} = \mathbf{r}'_u \times \mathbf{r}'_v$ is normal to the surface .

Area of a smooth parametric surface. Owing to the definition of the surface area element of the graph and the established relation between graphs and smooth parametric surfaces, the area a smooth surface can be found using the tangent planes to it (see Fig. **39.1** (left panel)). Let a region D spanned by the parameters (u, v) be partitioned by rectangles of area $\Delta A = \Delta u\, \Delta v$, then the vector function $\mathbf{r}(u, v)$ defines a partition of the surface (a partition element of the surface is the image of a partition rectangle in D). Consider a rectangle $[u_0, u_0 + \Delta u] \times [v_0, v_0 + \Delta v] = R_0$. Let its vertices O', A', and B' have the coordinates (u_0, v_0), $(u_0 + \Delta u, v_0)$, and $(u_0, v_0 + \Delta v)$, respectively. The segments $O'A'$ and $O'B'$ are the adjacent sides of the rectangle R_0. Let

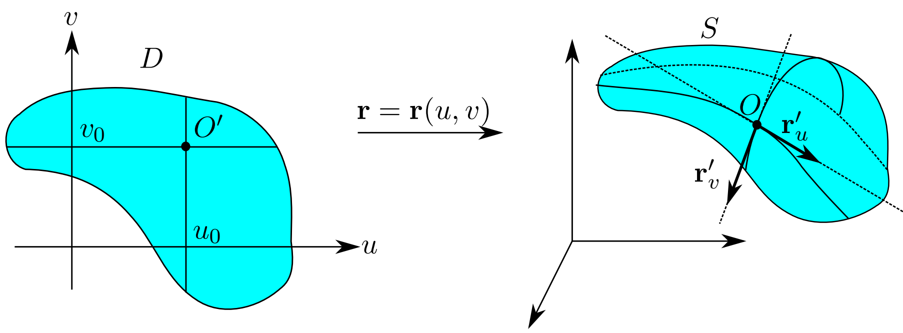


FIGURE **39.4**. The lines $u = u_0$ and $v = v_0$ in D are mapped onto the curves in S that are traced out by the vector functions $\mathbf{r} = \mathbf{r}(u_0, v)$ and $\mathbf{r} = \mathbf{r}(u, v_0)$, respectively. The curves intersect at the point O with the position vector $\mathbf{r}(u_0, v_0)$. The derivatives $\mathbf{r}'_v(u_0, v_0)$ and $\mathbf{r}'_u(u_0, v_0)$ are tangential to the curves. If they do not vanish and are not parallel, then their cross product is normal to the plane through O that contains $\mathbf{r}'_u$ and $\mathbf{r}'_v$. If the parametric surface is smooth, then $\mathbf{r}'_u \times \mathbf{r}'_v \neq \mathbf{0}$ is a normal vector to the plane tangent to the surface.

O, A, and B be the images of these points in the surface. Their position vectors are $\mathbf{r}_0 = \mathbf{r}(u_0, v_0)$, $\mathbf{r}_a = \mathbf{r}(u_0 + \Delta u, v_0)$, and $\mathbf{r}_b = \mathbf{r}(u_0, v_0 + \Delta v)$, respectively. The area ΔS of the image of the rectangle R_0 can be approximated by the area of the parallelogram $\|\mathbf{a} \times \mathbf{b}\|$ with adjacent sides:

$$\mathbf{a} = \overrightarrow{OA} = \mathbf{r}_a - \mathbf{r}_0 = \mathbf{r}(u_0 + \Delta u, v_0) - \mathbf{r}(u_0, v_0) = \mathbf{r}'_u(u_0, v_0)\,\Delta u,$$
$$\mathbf{b} = \overrightarrow{OB} = \mathbf{r}_b - \mathbf{r}_0 = \mathbf{r}(u_0, v_0 + \Delta v) - \mathbf{r}(u_0, v_0) = \mathbf{r}'_v(u_0, v_0)\,\Delta v.$$

The last equalities are obtained by the linearization of the components of $\mathbf{r}(u, v)$ near (u_0, v_0), which is justified because the surface has a tangent plane at any point. The area transformation law is now easy to find:

$$\Delta S = \|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{r}'_u \times \mathbf{r}'_v\| \Delta u \Delta v = J \Delta A\,.$$

Having found the surface area of a partition element of a parametric smooth surface, the total surface area can be found in the same way as it was done for a smooth surface defined by the graph of a function. The following theorem can be proved.

THEOREM **39.4**. (Surface Integral over Parametric Surface)
Let D be a closed bounded planar region with a piecewise smooth boundary. Let $\mathbf{r} = \mathbf{r}(u, v)$ be parametric equations of a surface S. Suppose that a vector function $\mathbf{r}(u, v)$ is one-to-one and has continuous and bounded partial derivatives, $\mathbf{r}'_u$ and $\mathbf{r}'_v$, in the interior of D such that the vector $\mathbf{r}'_u \times \mathbf{r}'_v$ does

not vanish. Then the surface area of S is given by the double integral

$$A(S) = \iint_D \|\mathbf{r}'_u \times \mathbf{r}'_v\| dA\,.$$

If a function f is bounded and possibly not continuous at finitely many smooth curves in S, then the surface integral of f over S exists and is given by

$$\iint_S f(\mathbf{r})\, dS = \iint_D f(\mathbf{r}(u,v))\|\mathbf{r}'_u \times \mathbf{r}'_v\|\, dA\,.$$

Note that the partial derivatives $\mathbf{r}'_u$ and $\mathbf{r}'_v$ are not required to be defined on the boundary of D and, if they are defined, the cross product $\mathbf{r}_u \times \mathbf{r}_v$ is allowed to vanish on the boundary of D. Also, the function $\mathbf{r}(u,v)$ is required to be one-to-one on D except possibly on the boundary of D. The behavior of the function $\mathbf{r}(u,v)$ on the boundary of D would have no effect on the value of the double integral over D by Corollary **28.1** as long as the partial derivatives $\mathbf{r}'_u$ and $\mathbf{r}'_v$ remain continuous and bounded in the interior of D because the boundary of D is a piecewise smooth curve and has zero area. The situation is similar to the change of variables in the double integral stated in Theorem **33.2** because, as has been argued before, parametric equations of a surface may be obtained from the graph of a function of two variables by a change of variables so that the surface integral over the graph (or the union of several graphs) is related to the surface integral over the same surface described by parametric equations by a change of variables.

Example **39.6**. *Find the surface area of the torus* (**39.5**).

Solution: To shorten the notation, put $w = a + R\cos u$. One has

$$\begin{aligned}
\mathbf{r}'_u &= \langle -R\sin u\cos v,\ -R\sin u\sin v,\ R\cos u\rangle\,,\\
&= -R\,\langle \sin u\cos v,\ \sin u\sin v,\ -\cos u\rangle\,,\\
\mathbf{r}'_v &= \langle -(a+R\cos u)\sin v,\ (a+R\cos u)\cos v,\ 0\rangle\\
&= w\,\langle -\sin v,\ \cos v,\ 0\rangle\,,\\
\mathbf{n} &= \mathbf{r}'_u \times \mathbf{r}'_v = -Rw\,\langle \cos v\cos u,\ \sin v\cos u,\ \sin u\rangle\,,\\
J &= \|\mathbf{r}'_u \times \mathbf{r}'_v\| = Rw\Big(\cos^2 u(\cos^2 v + \sin^2 v) + \sin^2 u\Big)^{1/2}\\
&= Rw = R(a + R\cos u).
\end{aligned}$$

The surface area is

$$A(S) = \iint_D J(u,v)\, dA = \int_0^{2\pi}\int_0^{2\pi} R(a+R\cos u)\, dv\, du = 4\pi^2 Ra\,.$$

$\square$

Example **39.7**. *Evaluate the surface integral of $f(x,y,z) = z^2(x^2 + y^2)$ over a sphere of radius R centered at the origin.*

SOLUTION: Using the parametric equations (**39**.**4**), one finds

$$\begin{aligned}
\mathbf{r}'_\phi &= \langle R\cos\theta\cos\phi,\ R\sin\theta\cos\phi,\ -R\sin\phi\rangle\,,\\
\mathbf{r}'_\theta &= \langle -R\sin\theta\sin\phi,\ R\cos\theta\sin\phi,\ 0\rangle\\
&= R\sin\phi\langle -\sin\theta,\ \cos\theta,\ 0\rangle\,,\\
\mathbf{n} &= \mathbf{r}'_\phi\times\mathbf{r}'_\theta = R\sin\phi\langle R\sin\phi\cos\theta,\ R\sin\phi\sin\theta,\ R\cos\phi\rangle\\
&= R\sin\phi\ \mathbf{r}(\phi,\theta),\\
J &= \|\mathbf{r}'_\phi\times\mathbf{r}'_\theta\| = R\sin\phi\|\mathbf{r}(\phi,\theta)\| = R^2\sin\phi\,,\\
f(\mathbf{r}(\phi,\theta)) &= (R\cos\phi)^2R^2\sin^2\phi = R^4\cos^2\phi(1-\cos^2\phi)\,.
\end{aligned}$$

Note that $\sin\phi\geq 0$ as $0\leq\phi\leq\pi$. Therefore, the normal vector $\mathbf{n}$ is outward (parallel to the position vector; the inward normal would be opposite to the position vector.) The surface integral is

$$\begin{aligned}
\iint_S f\,dS &= \iint_D f(\mathbf{r}(\phi,\theta))J(\phi,\theta)\,dA\\
&= R^6\int_0^{2\pi}d\theta\int_0^{\pi}\cos^2\phi(1-\cos^2\phi)\sin\phi\,d\phi\\
&= 2\pi R^6\int_{-1}^{1}w^2(1-w^2)\,dw = \frac{8\pi}{15}\,R^6,
\end{aligned}$$

where the substitution $w=\cos\phi$ has been made to evaluate the last integral. □

Note that the vector function defined by the parametric equations (**39**.**4**) is not one-to-one on the boundary of the rectangle $[0,\pi]\times[0,2\pi]$. The points $(\phi,0)$ and $(\phi,2\pi)$ have the same image. Furthermore all points $(0,\theta)$ on the boundary are mapped to the point $(0,0,R)$, while the points (π,θ) are mapped to $(0,0,-R)$ of the sphere, and the function J vanishes at these boundary points. But all the double-counting has no effect on the surface integral because it occurs on the set that has no area in full accord with Theorem **39**.**4**. For example, the surface area of the sphere is

$$A(S)=\iint_S dS=\iint_D J(\phi,\theta)dA' = R^2\int_0^{2\pi}d\theta\int_0^{\pi}\sin\phi d\phi = 4\pi R^2\,.$$

The parametric equations (**39**.**4**) define a smooth surface (the sphere is a smooth surface!) despite the fact that $\mathbf{n}(0,\theta)=\mathbf{n}(\pi,\theta)=\mathbf{0}$. Indeed the unit vector parallel to $\mathbf{n}$ is well defined and continuous for $0<\phi<\pi$:

$$\hat{\mathbf{n}} = \frac{1}{\|\mathbf{n}\|}\,\mathbf{n} = \langle\sin\phi\cos\theta,\ \sin\phi\sin\theta,\ \cos\phi\rangle$$

and has the continuous extension to $\phi=0$ and $\phi=\pi$:

$$\lim_{\phi\to 0^+}\hat{\mathbf{n}}(\phi,\theta)=\langle 0,0,1\rangle\,,\quad \lim_{\phi\to\pi^-}\hat{\mathbf{n}}(\phi,\theta)=\langle 0,0,-1\rangle$$

for any θ by continuity of the trigonometric functions. The limits are the unit normal vectors of the sphere at the points $(0,0,R)$ and $(0,0,-R)$, respectively.

39.5. Exercises.

1–5. Find the surface area of each of the following surfaces.

1. The part of the plane in the first octant that intersects the coordinate axes at $(a,0,0)$, $(0,b,0)$ and $(0,0,c)$ where a, b, and c are positive numbers;

2. The part of the plane $3x+2y+z=1$ that lies inside the cylinder $x^2+y^2=4$;

3. The part of the hyperbolic paraboloid $z=y^2-x^2$ that lies between the cylinders $x^2+y^2=1$ and $x^2+y^2=4$;

4. The part of the paraboloid $z=x^2+y^2$ that lies between two planes $z=1$ and $z=9$;

5. The part of the surface $y=4x+z^2$ that lies between the planes $x=0$, $z=1$, and $z=x$.

6–10. Evaluate the integral over the specified surface.

6. $\iint_S yz\,dS$ where S is the part of the plane $x+y+z=1$ that lies in the first octant;

7. $\iint_S x^2z^2\,dS$ where S is the part of the cone $z^2=x^2+y^2$ that lies in between the planes $z=1$ and $z=2$;

8. $\iint_S xz\,dS$ where S is the boundary of the solid region enclosed by the cylinder $y^2+z^2=1$ and the planes $x=0$ and $x+y=3$. Hint: use the additivity of the surface integral;

9. $\iint_S z\,dS$ where S is the part of the sphere $x^2+y^2+z^2=2$ that lies above the plane $z=1$;

10. $\iint_S z(\sin(x^2)-\sin(y^2))\,dS$ where S is the part of the paraboloid $z=1-x^2-y^2$ that lies in the first octant. Hint: Use the symmetry.

11. Suppose that $f(\mathbf{r})=g(\|\mathbf{r}\|)$ where $\mathbf{r}=(x,y,z)$. If $g(a)=2$, use the geometrical interpretation of the surface integral to find $\iint_S f\,dS$ where S is the sphere of radius a centered at the origin.

12–13. Identify and sketch the given parametric surface.

12. $\mathbf{r}(u,v)=\langle u+v,\ 2-v,\ 2-2v+3u\rangle$;

13. $\mathbf{r}(u,v)=\langle a\cos u,\ b\sin u,\ v\rangle$, where a and b are positive constants.

14–15. For the given parametric surface sketch the curves $\mathbf{r}(u,v_0)$ for several fixed values $v=v_0$ and the curves $\mathbf{r}(u_0,v)$ for several fixed values $u=u_0$. Use them to visualize the parametric surface.

14. $\mathbf{r}(u,v)=\langle \sin v,\ u\sin v,\ \sin u\sin(2v)\rangle$;

15. $\mathbf{r}(u,v)=(u\cos v\sin\theta, u\sin u\sin\theta, u\cos\theta)$ where $0\le\theta\le\pi/2$ is a parameter.

16–19. Find parametric equations of the specified surface.

16. The plane through $\mathbf{r}_0$ that contains two non-zero and non-parallel vectors $\mathbf{a}$ and $\mathbf{b}$;
17. The elliptic cylinder $y^2/a^2 + z^2/b^2 = 1$;
18. The part of the sphere $x^2 + y^2 + z^2 = a^2$ that lies below the cone $z = \sqrt{x^2 + y^2}$;
19. The ellipsoid $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$.

20–21. Find an equation of the tangent plane to the given parametric surface at the specified point P.

20. $\mathbf{r}(u, v) = \langle u^2,\ u - v, u + v\rangle$ at $P = (1, -1, 3)$;
21. $\mathbf{r}(u, v) = \langle \sin v,\ u \sin v,\ \sin u \sin(2v)\rangle$ at $P = (1, \pi/2, 0)$.

22–25. Evaluate each of the following surface integrals over the specified parametric surface.

22. $\iint_S z^2 dS$, where S is the torus (**39.5**) with $R = 1$ and $a = 2$;
23. $\iint_S (1 + x^2 + y^2)^{1/2} dS$, where S is the *helicoid* with parametric equations $\mathbf{r}(u, v) = \langle u \cos v,\ u \sin v,\ v\rangle$ and (u, v) in $[0, 1] \times [0, \pi]$;
24. $\iint_S z dS$, where S is the part of the helicoid $\mathbf{r}(u, v) = \langle u \cos v,\ u \sin v,\ v\rangle$, (u, v) in $[0, a] \times [0, 2\pi]$;
25. $\iint_S z^2 dS$, where S is the part of the cone $x = u \cos v \sin\theta$, $y = u \sin v \sin\theta$, $z = u \cos\theta$, (u, v) in $[0, a] \times [0, 2\pi]$, and $0 < \theta < \pi/2$ is a parameter.

26–32. Evaluate each of the following surface integrals. If necessary, use suitable parametric equations of the surface.

26. $\iint_S (x^2 + y^2 + z^2) dS$, where S is the sphere $x^2 + y^2 + z^2 = R^2$;
27. $\iint_S (x^2 + y^2 + z^2) dS$, where S is the surface $|x| + |y| + |z| = R$; compare the result with the previous exercise;
28. $\iint_S (x^2 + y^2) dS$ where S is the boundary of the solid $\sqrt{x^2 + y^2} \le z \le 1$;
29. $\iint_S (1 + x + y)^{-2} dS$ where S is the boundary of the tetrahedron bounded by the coordinate planes and by the plane $x + y + z = 1$;
30. $\iint_S |xyz| dS$ where S is the part of the paraboloid $z = x^2 + y^2$ below the plane $z = 1$;
31. $\iint_S (1/h) dS$ where S is an ellipsoid $(x/a)^2 + (y/b)^2 + (z/c)^2 = 1$ and h is the distance from the origin to the plane tangent to the ellipsoid at the point where the surface area element dS is taken;
32. $\iint_S (xy + yz + zx) dS$ where S is the part of the cone $z = \sqrt{x^2 + y^2}$ cut out by the cylinder $x^2 + y^2 = 2ax$.

33. Prove the *Poisson formula*

$$\iint_S f(ax + by + cz) dS = 2\pi \int_{-1}^{1} f(u\sqrt{a^2 + b^2 + c^2}) du$$

where S is the sphere $x^2 + y^2 + z^2 = 1$.

34. Evaluate $F(a, b, c, t) = \iint_S f(x, y, z) dS$ where S is the sphere $(x - a)^2 + (y - b)^2 + (z - c)^2 = t^2$, $f(x, y, z) = 1$ if $x^2 + y^2 + z^2 < R^2$ and $f(x, y, z) = 0$ elsewhere. Assume that $\sqrt{a^2 + b^2 + c^2} > R > 0$.

40. Moments of Inertia and Center of Mass

An important application of multiple integrals is finding the *center of mass* and *moments of inertia* of an extended object. The laws of mechanics say that the center of mass of an extended object on which no external force acts moves along a straight line with a constant speed. In other words, the center of mass is a particular point of an extended object that defines the trajectory of the object as a whole. The motion of an extended object can be viewed as a combination of the motion of its center of mass and rotation about its center of mass. The kinetic energy of the object is

$$K = \frac{mv^2}{2} + K_{\text{rot}},$$

where m is the total mass of the object, v is the speed of its center of mass, and K_{rot} is the kinetic energy of rotation of the object about its center of mass; K_{rot} is determined by *moments of inertia* discussed later. For example, when docking a spacecraft to a space station, one needs to know exactly how long the engine should be fired to achieve the required position of its center of mass and the orientation of the craft relative to it, that is, how exactly its kinetic energy has to be changed by firing the engines. So its center of mass and moments of inertia must be known to accomplish the task.

40.1. Center of Mass. Consider a point mass m fixed at an endpoint of a rod that can rotate about its other end. If the rod has length L and the gravitational force is normal to the rod, then the quantity gmL is called the *rotational moment* of the gravitational force mg, where g is the free-fall acceleration. If the rotation is clockwise (the mass is at the right endpoint), the moment is assumed to be positive, and it is negative, $-gmL$, for a counterclockwise rotation (the mass is at the left endpoint). More generally, if the mass has a position x on the x axis, then its rotation moment about a point x_c is $M = (x - x_c)m$ (omitting the constant g). It is negative if $x < x_c$ and positive when $x > x_c$. The center of mass is understood through the concept of rotational moments.

The simplest extended object consists of two point masses m_1 and m_2 connected by a massless rod. It is shown in the left panel of Fig. **40.1**. Suppose that one point of the rod is fixed so that it can only rotate about that point. The center of mass is the point on the rod such that the object would not rotate about it under a uniform gravitational force applied along the direction perpendicular to the rod. Evidently, the position of the center of mass is determined by the condition that the total rotational moment about it vanishes. Suppose that the rod lies on the x axis so that the masses have the coordinates x_1 and x_2. The total rotational moment of the object about the point x_c is $M = M_1 + M_2 = (x_1 - x_c)m_1 + (x_2 - x_c)m_2$. If x_c is

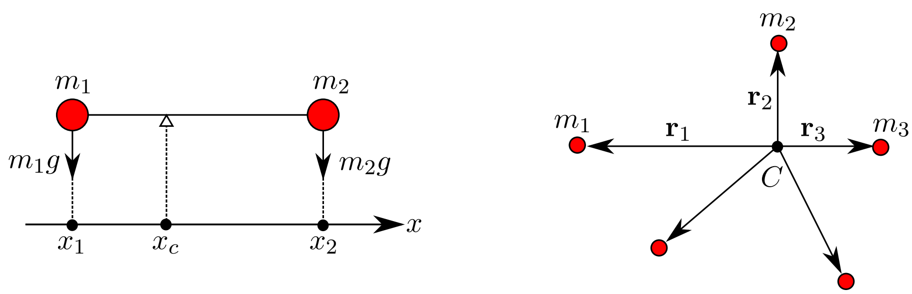


FIGURE **40.1**. **Left**: Two masses connected by a rigid massless rod (or its mass is much smaller than the masses m_1 and m_2) are positioned at x_1 and x_2. The gravitational force is perpendicular to the rod. The center of mass x_c is determined by the condition that the system does not rotate about x_c under the gravitational forces. **Right**: An extended object consisting of point masses with fixed distances between them. If the position vectors of the masses relative to the center of mass C are $\mathbf{r}_i$, then $m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \cdots + m_N\mathbf{r}_N = \mathbf{0}$.

such that $M = 0$, then

$$m_1(x_1 - x_c) + m_2(x_2 - x_c) = 0 \quad \Longrightarrow \quad x_c = \frac{m_1x_1 + m_2x_2}{m_1 + m_2}.$$

The center of mass (x_c, y_c) of point masses m_i, $i = 1, 2, ..., N$, positioned on a plane at (x_i, y_i) can be understood as follows. Think of the plane as a plate on which the masses are positioned. The gravitational force is normal to the plane. If a rod (a line) is put underneath the plate, parallel to the plate, then due to an uneven distribution of masses, the plate can rotate about the rod. When the rod is aligned along either the line $x = x_c$ or the line $y = y_c$, the plane with distributed masses on it does not rotate under the gravitational pull. In other words, the rotational moments about the lines $x = x_c$ and $y = y_c$ vanish. The rotational moment about the line $x = x_c$ or $y = y_c$ is determined by the distances of the masses from this line:

$$\sum_{i=1}^{N}(x_i - x_c)m_i = 0 \quad \Longrightarrow \quad x_c = \frac{1}{m}\sum_{i=1}^{N} m_i x_i = \frac{M_y}{m}\,, \qquad m = \sum_{i=1}^{N} m_i,$$

$$\sum_{i=1}^{N}(y_i - y_c)m_i = 0 \quad \Longrightarrow \quad y_c = \frac{1}{m}\sum_{i=1}^{N} m_i y_i = \frac{M_x}{m}\,,$$

where m is the total mass. The quantity M_y is the moment about the y axis (the line $x = 0$), whereas M_x is the moment about the x axis (the line $y = 0$).

Consider an extended object that is a collection of point masses shown in the right panel of Fig. **40.1**. Its center of mass is defined similarly by demanding that the total moments about either of the planes $x = x_c$, or

$y = y_c$, or $z = z_c$ vanish. Thus, if $\mathbf{r}_c$ is the position vector of the center of mass, it satisfies the condition:

$$\sum_i m_i(\mathbf{r}_i - \mathbf{r}_c) = \mathbf{0},$$

where the vectors $\mathbf{r}_i - \mathbf{r}_c$ are position vectors of masses *relative to the center of mass.*

Definition **40.1**. (Center of Mass).
Suppose that an extended object consists of N point masses m_i, $i = 1, 2, ..., N$, whose position vectors are $\mathbf{r}_i$. Then its center of mass is a point with the position vector

$$(\mathbf{40.1}) \qquad \mathbf{r}_c = \frac{1}{m}\sum_{i=1}^{N} m_i \mathbf{r}_i\,, \qquad m = \sum_{i=1}^{N} m_i,$$

where m is the total mass of the object. The quantities

$$M_{yz} = \sum_{i=1}^{N} m_i x_i, \quad M_{xz} = \sum_{i=1}^{N} m_i y_i, \quad M_{xy} = \sum_{i=1}^{N} m_i z_i,$$

are called the moments about the coordinate planes.

If an extended object contains continuously distributed masses, then the object can be partitioned into N small pieces. Let B_i be the smallest ball of radius R_i within which the i^{th} partition piece lies. Although all the partition pieces are small, they still have finite sizes R_i, and the definition (**40.1**) cannot be used because the point $\mathbf{r}_i$ could be any point in B_i. By making the usual trick of integral calculus, this uncertainty can be eliminated by taking the limit $N \to \infty$ in the sense that all the partition sizes tend to 0 uniformly, $R_i \leq \max_i R_i = R_N^* \to 0$ as $N \to \infty$. In this limit, the position of each partition piece can be described by any sample point $\mathbf{r}_i^*$ in B_i. The limit of the Riemann sum is given by the integral over the region E in space occupied by the object. If $\sigma(\mathbf{r})$ is the mass density of the object, then $\Delta m_i = \sigma(\mathbf{r}_i^*)\,\Delta V_i$, where ΔV_i is the volume of the i^{th} partition element and

$$(\mathbf{40.2}) \qquad \mathbf{r}_c = \frac{1}{m}\lim_{N\to\infty}\sum_{i=1}^{N} \mathbf{r}_i^*\,\Delta m_i = \frac{1}{m}\iiint_E \mathbf{r}\,\sigma(\mathbf{r})\,dV\,,$$

$$m = \iiint_E \sigma(\mathbf{r})\,dV\,.$$

Relation (**40.2**) is adopted as the definition of the position vector of the center of mass of an extended object (with continuously distributed mass). In practical applications, one often encounters extended objects whose one or two dimensions are small relative to the other (e.g., shell-like objects or wire-like objects). In this case, the triple integral is simplified to either a surface (or double) integral for shell-like E, according to (**39.2**), or to a line integral, according to (**38.1**). For two- and one-dimensional extended objects, the center of mass can be written as, respectively,

$$\mathbf{r}_c = \frac{1}{m}\iint_S \mathbf{r}\,\sigma(\mathbf{r})\,dS\,, \qquad m = \iint_S \sigma(\mathbf{r})\,dS,$$
$$\mathbf{r}_c = \frac{1}{m}\int_C \mathbf{r}\,\sigma(\mathbf{r})\,ds\,, \qquad m = \int_C \sigma(\mathbf{r})\,ds,$$

where, accordingly, σ is the surface mass density or the line mass density for two- or one-dimensional objects. In particular, when S is a part of the plane, the surface integral turns into a double integral.

The concept of rotational moments is also useful for finding the center of mass using the symmetries of the mass distribution of an extended object. For example, the center of mass of a disk with a uniform mass distribution apparently coincides with the disk center (the disk would not rotate about its diameter under the gravitational pull).

EXAMPLE **40.1**. *Find the center of mass of the half-disk $x^2 + y^2 \le R^2$, $y \ge 0$, if the mass density at any point is proportional to the distance of that point from the x axis.*

SOLUTION: The mass is distributed evenly to the left and right from the y axis because the mass density is independent of x, $\sigma(x, y) = ky$ (k is a constant). So, the rotational moment about the y axis vanishes; $M_y = 0$ by symmetry and hence $x_c = M_y/m = 0$. The total mass is

$$m = \iint_D \sigma\,dA = k\iint_D y\,dA = k\int_0^{\pi}\int_0^R r\sin\theta\ r\,dr\,d\theta$$
$$= 2k\int_0^R r^2\,dr = \frac{2kR^3}{3},$$

where the integral has been converted to polar coordinates. The moment about the x axis (about the line $y = 0$) is

$$M_x = \iint_D y\sigma\,dA = \int_0^{\pi}\int_0^R k(r\sin\theta)^2 r\,dr\,d\theta = \frac{\pi k}{2}\int_0^R r^3\,dr = \frac{\pi k R^4}{8}.$$

So $y_c = M_x/m = 3\pi R/16$. □

EXAMPLE **40.2**. *Find the center of mass of the solid that lies between spheres of radii $a < b$ centered at the origin and is bounded by the cone $z = \sqrt{x^2 + y^2}/\sqrt{3}$ if the mass density is constant.*

SOLUTION: The mass is evenly distributed about the xz and yz planes. So the moments M_{xz} and M_{yz} about them vanish, and hence $y_c = M_{xz}/m = 0$ and $x_c = M_{yz}/m = 0$. The center of mass lies on the z axis. Put $\sigma = k =$ const. The total mass is

$$m = \iiint_E \sigma\,dV = k\int_0^{2\pi}\int_0^{\pi/3}\int_a^b \rho^2\sin\phi\,d\rho\,d\phi\,d\theta = \frac{\pi k}{3}\,(b^3 - a^3),$$

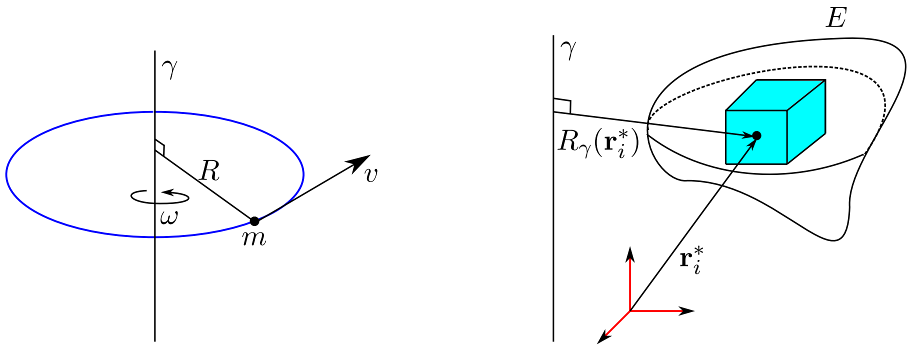


FIGURE **40.2**. **Left**: The moment of inertia of a point mass about an axis γ. A point mass rotates about an axis γ with the rate ω called the angular velocity. Its linear velocity is $v = \omega R$ where R is the distance from γ. So the kinetic energy of the rotational motion is $mv^2/2 = mR^2\omega^2/2 = I_\gamma\omega^2/2$ where $I_\gamma = mR^2$ is the moment of inertia. **Right**: The moment of inertia of an extended object E about an axis γ is defined as the sum of moments of inertia of partition elements of E: $\Delta I_i = \Delta m(\mathbf{r}_i^*)R_\gamma^2(\mathbf{r}_i^*)$ where $R_\gamma(\mathbf{r}_i^*)$ is the distance to the axis γ from a sample point $\mathbf{r}_i^*$ in the i^{th} partition element and $\Delta m(\mathbf{r}_i^*)$ is its mass.

where the triple integral has been converted to spherical coordinates. The boundaries of E are the spheres $\rho = a$ and $\rho = b$ and the cone defined by the condition $\cot\phi = 1/\sqrt{3}$ or $\phi = \pi/3$. Therefore, the region E is the image of the rectangular box $E' = [a, b] \times [0, \pi/3] \times [0, 2\pi]$ under the transformation to spherical coordinates. The full range is taken for the polar angle θ as the equations of the boundaries impose no condition on it. The moment about the xy plane is

$$M_{xy} = \iiint_E z\sigma\, dV = k\int_0^{2\pi}\int_0^{\pi/3}\int_a^b \rho\cos\phi\,\rho^2\sin\phi\, d\rho\, d\phi\, d\theta$$
$$= \frac{3\pi k}{16}\,(b^4 - a^4).$$

So $z_c = M_{xy}/m = (9/16)(a + b)(a^2 + b^2)/(a^2 + ab + b^2)$. □

Centroid. The center of mass of an extended object with a *constant* mass density is called the *centroid*. The centroid of a region depends only on the shape of the region. In this sense, the centroid is an intrinsic (geometrical) characteristic of the region.

40.2. Moments of Inertia. Consider a point mass m rotating about an axis γ at a constant rate of ω rad/s (called the *angular velocity*). The system is shown in Fig. **40.2** (left panel). If the radius of the circular trajectory is R, then the linear velocity of the object is $v = \omega R$. The object has the kinetic energy

$$K_{\rm rot} = \frac{mv^2}{2} = \frac{mR^2\omega^2}{2} = \frac{I_\gamma \omega^2}{2}.$$

The constant I_γ is called the *moment of inertia* of the point mass m about the axis γ. Similarly, consider an extended solid object consisting of N point masses. The distances between the masses do not change when the object moves (the object is *solid*). So, if the object rotates about an axis γ at a constant rate ω, then each point mass rotates at the same rate and hence has kinetic energy $m_i R_i^2 \omega^2/2$, where R_i is the distance from the mass m_i to the axis γ. The total kinetic energy is $K_{\rm rot} = I_\gamma \omega^2/2$, where the constant

$$I_\gamma = \sum_{i=1}^{N} m_i R_i^2$$

is called the *moment of inertia of the object about the axis* γ. It is independent of the motion itself and determined solely by the mass distribution and distances of the masses from the rotation axis.

Suppose that the mass is continuously distributed in a region E with the mass density $\sigma(\mathbf{r})$ (see the right panel of Fig. **40.2**). Let $R_\gamma(\mathbf{r})$ be the distance from a point $\mathbf{r}$ in E to an axis (line) γ. Consider a partition of E by small elements E_i of volume ΔV_i. The mass of each partition element is $\Delta m_i = \sigma(\mathbf{r}_i^*)\,\Delta V_i$ for some sample point $\mathbf{r}_i^*$ in E_i in the limit when all the sizes of partition elements tend to 0 uniformly. The moment of inertia about the axis γ is

$$I_\gamma = \lim_{N\to\infty} \sum_{i=1}^{N} R_\gamma^2(\mathbf{r}_i^*)\sigma(\mathbf{r}_i^*)\,\Delta V_i = \iiint_E R_\gamma^2(\mathbf{r})\sigma(\mathbf{r})\,dV$$

in accordance with the Riemann sum for triple integrals (**34.3**). In particular, the distance of a point (x, y, z) from the x-, y-, and z axes is, respectively, $R_x = \sqrt{y^2+z^2}$, $R_y = \sqrt{x^2+z^2}$, and $R_z = \sqrt{x^2+y^2}$. So the moments of inertia about the coordinate axes are

$$I_x = \iiint_E (y^2+z^2)\sigma\,dV\,, \quad I_y = \iiint_E (x^2+z^2)\sigma\,dV\,,$$
$$I_z = \iiint_E (x^2+y^2)\sigma\,dV.$$

In general, if the axis γ goes through the origin parallel to a unit vector $\hat{\mathbf{u}}$, then by the distance formula between a point $\mathbf{r}$ and the line,

$$\begin{aligned} R_\gamma^2(\mathbf{r}) &= \|\hat{\mathbf{u}}\times\mathbf{r}\|^2 = (\hat{\mathbf{u}}\times\mathbf{r})\cdot(\hat{\mathbf{u}}\times\mathbf{r}) = \hat{\mathbf{u}}\cdot(\mathbf{r}\times(\hat{\mathbf{u}}\times\mathbf{r})) \\ &= \mathbf{r}^2 - (\hat{\mathbf{u}}\cdot\mathbf{r})^2, \end{aligned} \tag{40.3}$$

where the $bac - cab$ rule (**4.2**) has been used to transform the double cross product.

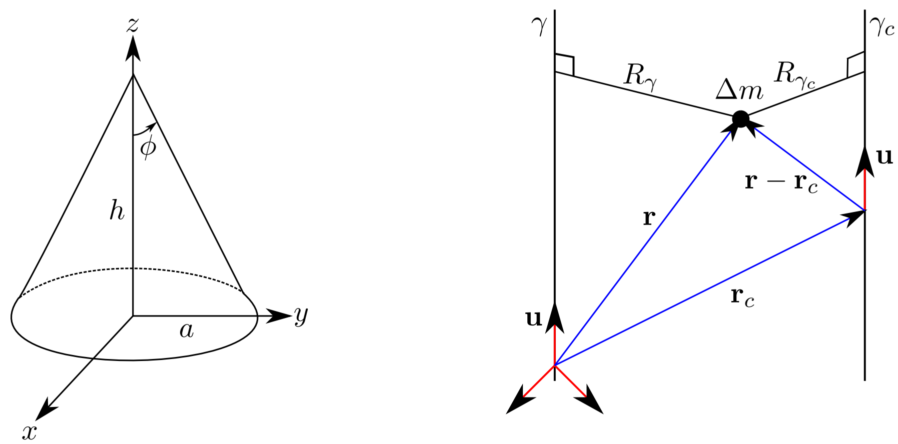


FIGURE **40.3**. **Left**: An illustration to Example **40.3**. **Right**: An illustration to the proof of the parallel axis theorem for moments of inertia (Study Problem **40.2**). The axis γ_c is parallel to γ and goes through the center of mass with the position vector $\mathbf{r}_c$. The vectors $\mathbf{r}$ and $\mathbf{r} - \mathbf{r}_c$ are position vectors of a partition element of mass Δm relative to the origin and the center of mass, respectively.

If one or two dimensions of the object are small relative to the other, the triple integral is reduced to either a surface integral or a line integral, respectively, in accordance with (**39.2**) or (**38.1**); that is, for two- or one-dimensional objects, the moment of inertia becomes, respectively,

$$I_\gamma = \iint_S R_\gamma^2(\mathbf{r})\sigma(\mathbf{r})\,dS\,, \qquad I_\gamma = \int_C R_\gamma^2(\mathbf{r})\sigma(\mathbf{r})\,ds,$$

where σ is either the surface or linear mass density.

EXAMPLE **40.3**. *A rocket tip is made of thin plates with a constant surface mass density $\sigma = k$. It has a circular conic shape with base diameter $2a$ and distance h from the tip to the base. Find the moment of inertia of the tip about its axis of symmetry.*

SOLUTION: Set up the coordinate system so that the tip is on the z axis at the point $(0, 0, h)$, and the base is the disk of radius a in the xy plane (see the left panel of Fig. **40.3**). A cone with its tip at the origin obtained by rotating a ray extended from the origin about the z axis is described by the equation $z = \cot\phi\sqrt{x^2 + y^2}$, where ϕ is the angle between the ray and the z axis. Putting $\cot\phi = -h/a$ (as $\phi > \pi/2$) and shifting the surface up along the z axis ($z \to z - h$), the equation of the surface in question is obtained:

$$z = g(x, y) = h - \frac{h}{a}\sqrt{x^2 + y^2}\,,$$

where (x, y) range over the disk

$$D: \quad x^2 + y^2 \leq a^2 .$$

To evaluate the needed surface integral, the area transformation law $dS = J\,dA$ should be established. One has

$$g'_x = -\frac{hx}{a}(x^2 + y^2)^{-1/2}, \quad g'_y = -\frac{hy}{a}(x^2 + y^2)^{-1/2},$$
$$J = \sqrt{1 + (g'_x)^2 + (g'_y)^2} = \sqrt{1 + (h/a)^2} = \frac{\sqrt{h^2 + a^2}}{a} .$$

The moment of inertia about the z axis is

$$I_z = \iint_S (x^2 + y^2)\sigma\,dS = k \iint_D (x^2 + y^2) J\,dA$$
$$= kJ \int_0^{2\pi} d\theta \int_0^a r^3\,dr = \frac{\pi k}{2} a^3 \sqrt{h^2 + a^2} ,$$

where the double integral has been converted to polar coordinates. □

Example **40.4**. *Find the moment of inertia of a homogeneous ball of radius a and mass m about its diameter.*

Solution: Set up the coordinate system so that the origin is at the center of the ball. Then the moment of inertia about the z axis has to be evaluated. Since the ball is homogeneous, its mass density is constant, $\sigma = m/V$, where $V = 4\pi a^3/3$ is the volume of the ball. By converting the triple integral to spherical coordinates,

$$I_z = \iiint_E (x^2 + y^2)\sigma\,dV = \frac{3m}{4\pi a^3} \int_0^{2\pi} \int_0^{\pi} \int_0^a (\rho \sin\phi)^2 \rho^2 \sin\phi\,d\rho\,d\phi\,d\theta$$
$$= \frac{3}{10} ma^2 \int_0^{\pi} \sin^3\phi\,d\phi = \frac{3}{10} ma^2 \int_{-1}^{1} (1 - u^2)\,du = \frac{2}{5} ma^2 ,$$

where the substitution $u = \cos\phi$ has been made to evaluate the integral. It is noteworthy that the problem admits a smarter solution by noting that $I_z = I_x = I_y$ owing to the rotational symmetry of the mass distribution. By the identity $I_z = (I_x + I_y + I_z)/3$, the triple integral can be simplified:

$$I_z = \frac{1}{3}\sigma \iiint_E 2(x^2 + y^2 + z^2)\,dV = \frac{1}{3}\sigma 8\pi \int_0^a \rho^4\,d\rho = \frac{2}{5} ma^2 .$$

□

Example **40.5**. *Find the center of mass and the moment of inertia of a homogeneous rod of mass m bent into a half-circle of radius R about the line through the endpoints of the rod.*

Solution: Set up the coordinate system so the half-circle lies above the x axis: $x^2 + y^2 = R^2$, $y \geq 0$. The linear mass density is constant $\sigma = m/(\pi R)$

where πR is the length of the rod. By the symmetry of the mass distribution, the center of mass lies on the y axis,

$$x_c = 0\,,\quad y_c = \frac{1}{m}\int_C y\sigma ds\,.$$

To evaluate the line integral, choose the following parametric equation of the half-circle $\mathbf{r}(t) = \langle R\cos t, R\sin t\rangle$, $0 \le t \le \pi$. Then $\mathbf{r}'(t) = \langle -R\sin t, R\cos t\rangle$ and $ds = \|\mathbf{r}'(t)\|dt = Rdt$. Therefore

$$y_c = \frac{1}{m}\int_C y\sigma ds = \frac{1}{\pi R}\int_0^\pi R\sin tRdt = \frac{2R}{\pi}$$

If $R_\gamma(x, y)$ is the distance from the line connecting the end point of the rod to its point (x, y), then in the chosen coordinate system $R_\gamma(x, y) = y$. Therefore the moment of inertia in question is

$$I_\gamma = \int_C R_\gamma^2\sigma ds = \frac{m}{\pi R}\int_C y^2 ds = \frac{mR^2}{\pi}\int_0^\pi \sin^2 t dt = \frac{mR^2}{2}$$

□

40.3. Study Problems.

Problem **40.1**. *Find the center of mass of the shell described in Example* **40.3**.

Solution: By the symmetry of the mass distribution about the axis of the conic shell, the center of mass must be on that axis:

$$x_c = y_c = 0\,,\quad z_c = \frac{M_{xy}}{m}\,.$$

Using the algebraic description of a shell given in Example **40.3**, the total mass of the shell is

$$m = \iint_S \sigma\, dS = k\iint_S dS = kJ\iint_D dA = kJA(D) = \pi ka\sqrt{h^2 + a^2}.$$

Using polar coordinates, the moment about the xy plane is

$$\begin{aligned}
M_{xy} &= \iint_S z\sigma\, dS = k\iint_D \left(h - \frac{h}{a}\sqrt{x^2 + y^2}\right) J\, dA \\
&= \frac{k\sqrt{h^2 + a^2}}{a}\left(hA(D) - \frac{h}{a}\iint_D \sqrt{x^2 + y^2} dA\right) \\
&= \frac{k\sqrt{h^2 + a^2}}{a}\left(\pi ha^2 - \int_0^{2\pi}\int_0^a r^2\, dr\, d\theta\right) \\
&= \frac{k\sqrt{h^2 + a^2}}{a}\left(\pi ha^2 - \frac{2}{3}\pi ha^2\right) = \frac{\pi kha}{3}\sqrt{h^2 + a^2}.
\end{aligned}$$

Thus, the center of mass is at the distance $z_c = M_{xy}/m = h/3$ from the base of the cone. □

Problem **40.2**. (Parallel Axis Theorem).
Let I_γ be the moment of inertia of an extended object about an axis γ and let γ_c be a parallel axis through the center of mass of the object. Prove that

$$I_\gamma = I_{\gamma_c} + mR_c^2,$$

where R_c is the distance between the axis γ and the center of mass, and m is the total mass.

SOLUTION: Choose the coordinate system so that the axis γ goes through the origin (see the right panel of Fig. **40.3**). Let it be parallel to a unit vector $\hat{\mathbf{u}}$. The difference $I_\gamma - I_{\gamma_c}$ is to be investigated. If $\mathbf{r}_c$ is the position vector of the center of mass, then the axis γ_c is obtained from γ by parallel transport of the latter along the vector $\mathbf{r}_c$. Therefore, the distance $R^2_{\gamma_c}(\mathbf{r})$ is obtained from $R^2_\gamma(\mathbf{r})$ (see (**40.3**)) by changing the position vector $\mathbf{r}$ in the latter to the position vector relative to the center of mass, $\mathbf{r} - \mathbf{r}_c$. In particular, $R^2_\gamma(\mathbf{r}_c) = R_c^2$ by the definition of the function $R_\gamma(\mathbf{r})$. Hence, by Eq. (**40.3**)

$$\begin{aligned} R^2_\gamma(\mathbf{r}) - R^2_{\gamma_c}(\mathbf{r}) &= R^2_\gamma(\mathbf{r}) - R^2_\gamma(\mathbf{r} - \mathbf{r}_c) \\ &= \mathbf{r}^2 - (\mathbf{r}\cdot\hat{\mathbf{u}})^2 - (\mathbf{r} - \mathbf{r}_c)^2 + (\mathbf{r}\cdot\hat{\mathbf{u}} - \mathbf{r}_c\cdot\hat{\mathbf{u}})^2 \\ &= 2\mathbf{r_c}\cdot\mathbf{r} - \mathbf{r}_c^2 - (\hat{\mathbf{u}}\cdot\mathbf{r}_c)(2\hat{\mathbf{u}}\cdot\mathbf{r} - \hat{\mathbf{u}}\cdot\mathbf{r}_c) \\ &= \mathbf{r}_c^2 - (\hat{\mathbf{u}}\cdot\mathbf{r}_c)^2 + 2\mathbf{r}_c\cdot(\mathbf{r} - \mathbf{r}_c) - 2(\hat{\mathbf{u}}\cdot\mathbf{r}_c)\hat{\mathbf{u}}\cdot(\mathbf{r} - \mathbf{r}_c) \\ &= R_c^2 + 2\mathbf{a}\cdot(\mathbf{r} - \mathbf{r}_c), \end{aligned}$$

where $\mathbf{a} = \mathbf{r}_c - (\hat{\mathbf{u}}\cdot\mathbf{r}_c)\hat{\mathbf{u}}$. Therefore,

$$\begin{aligned} I_\gamma - I_{\gamma_c} &= \iiint_E \Big(R^2_\gamma(\mathbf{r}) - R^2_{\gamma_c}(\mathbf{r})\Big)\sigma(\mathbf{r})\,dV \\ &= R_c^2 \iiint_E \sigma(\mathbf{r})\,dV + 2\mathbf{a}\ \cdot \iiint_E (\mathbf{r} - \mathbf{r}_c)\sigma(\mathbf{r})\,dV = R_c^2 m, \end{aligned}$$

where the second integral vanishes by the definition of the center of mass. □

Problem **40.3**. *Find the moment of inertia of a homogeneous ball of radius a and mass m about an axis that is at a distance R from the ball center.*

SOLUTION: The center of mass of the ball coincides with its center because the mass distribution is invariant under rotations about the center. The moment of inertia of the ball about its diameter is $I_{\gamma_c} = (2/5)ma^2$ by Example **40.4**. By the parallel axis theorem, for any axis γ at a distance R from the center of mass, $I_\gamma = I_{\gamma_c} + mR^2 = m(R^2 + 2a^2/5)$. □

40.4. Exercises.

1–19. Find the center of mass of each the following extended objects.

1. A homogeneous thin rod of length L;

2. A homogeneous thin wire that occupies the part a circle of radius R that lies in the first quadrant;

3. A homogeneous thin wire bent into one turn of the helix of radius R that rises by the distance h per each turn;

4. A homogeneous thin shell that occupies a hemisphere of radius R;

5. A homogeneous thin disk of radius R that has a circular hole of radius $a < R/2$ and its center is at the distance $R/2$ from the disk center;

6. A homogeneous solid enclosed by the ellipsoid $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$ that has a square box cavity $[0, h] \times [0, h] \times [0, h]$;

7. The part of the ball $x^2 + y^2 + z^2 \le 4$ that lies above the cone $z\sqrt{3} = \sqrt{x^2 + y^2}$ and the mass density at any point is proportional to its distance from the origin;

8. The part of the spherical shell $a^2 \le x^2 + y^2 + z^2 \le b^2$ that lies above the xy plane and whose mass density at any point is proportional to its distance from the z axis;

9. The part of the disk $x^2 + y^2 \le a^2$ in the first quadrant bounded by the lines $y = x$ and $y = \sqrt{3}x$ if the mass density at any point is proportional to its distance from the origin;

10. The part of the solid enclosed by the paraboloid $z = 2 - x^2 - y^2$ and the cone $z = \sqrt{x^2 + y^2}$ that lies in the first octant and whose mass density at any point is proportional to its distance from the z axis;

11. A homogeneous surface cut from the cone $z = \sqrt{x^2 + y^2}$ by the cylinder $x^2 + y^2 = ax$;

12. The part of a homogeneous sphere defined by $z = \sqrt{a^2 - x^2 - y^2}$, $x \ge 0$, $y \ge 0$, $x + y \le a$, $a > 0$;

13. The arc of the homogeneous cycloid $x = a(t - \sin t)$, $y = a(1 - \cos t)$, $0 \le t \le \pi$;

14. The arc of the homogeneous curve $y = a \cosh(x/a)$ from the point $(0, a)$ to the point (b, h);

15. The arc of the homogeneous astroid $x^{2/3} + y^{2/3} = a^{2/3}$ in the first quadrant;

16. The homogeneous lamina bounded by the curves $\sqrt{x} + \sqrt{y} = \sqrt{a}$, $x = 0$, $y = 0$;

17. The part of the homogeneous lamina bounded by the curve $x^{2/3} + y^{2/3} = a^{2/3}$ in the first quadrant;

18. The homogeneous solid bounded by the surfaces $x^2 + y^2 = 2z$, $x + y = z$;

19. The homogeneous solid bounded by the surfaces $z = x^2 + y^2$, $2z = x^2 + y^2$, $x + y = \pm 1$, $x - y = \pm 1$.

20. Show that the centroid of a triangle is the point of intersection of its medians (the lines joining each vertex with the midpoint of the opposite side).

21. Show that the centroid of a pyramid is located on the line segment that connects the apex to the centroid of the base and is 1/4 the distance from the base to the apex.

22–34. Find the specified moment of inertia of each of the following extended objects.

22. The smaller wedge cut out from a ball of a radius R by two planes that intersect along the diameter of the ball at an angle $0 < \theta_0 \leq \pi$. The wedge is homogeneous and has the mass m. Find the moment of inertia about the diameter;

23. The moment of inertia about the z axis of the solid that is enclosed by the cylinder $x^2 + y^2 \leq 1$ and the planes $z = 0$, $y + z = 5$ and has the mass density $\sigma(x, y, z) = 10 - 2z$;

24. A thin homogeneous shell in the shape of the torus with radii R and $a > R$ that has mass m. The moment of inertia about the symmetry axis of the torus;

25. The moments of inertia I_x and I_y of the part of the disk of radius a that lies in the first quadrant and whose mass density at any point is proportional to its distance from the y axis;

26. The moments of inertia of a solid circular homogeneous cone with height h and the radius of the base a about its symmetry axis, the axis through its vertex and perpendicular to the symmetry axis, and an axis that contains a diameter of the base;

27. The moments of inertia of the part of the homogeneous plane $x + y + z = a$, $a > 0$, in the first octant about the coordinate axes;

28. The *polar* moment of inertia $I_0 = I_x + I_y$ of the homogeneous triangle of mass m whose vertices in polar coordinates are $(r, \theta) = (a, 0)$, $(a, 2\pi/3)$, $(a, 4\pi/3)$;

29. The moment of inertia of the homogeneous solid cylinder $x^2 + y^2 \leq a^2$, $-h \leq z \leq h$, of mass m about the line parallel to the z axis through the point $(a, 0, 0)$;

30. The sum of moments of inertia $I_x + I_y + I_z$ of the homogeneous solid of mass density σ_0 bounded by the surface $(x^2 + y^2 + z^2)^2 = a^2(x^2 + y^2)$;

31. The moments of inertia of the lamina with a constant mass density σ_0 bounded by the circle $(x - a)^2 + (y - a)^2 = a^2$ and by the segments $0 \leq y \leq a$, $0 \leq x \leq a$ about the coordinate axes;

32. The moments of inertia of the lamina with a constant mass density σ_0, bounded by the curves $xy = a^2$, $xy = 2a^2$, $x = 2y$, and $2x = y$, about the coordinate axes;

33. The moments of inertia of the solid that has a constant mass density σ_0 and is bounded by the ellipsoid $(x/a)^2 + (y/b)^2 + (z/c)^2 = 1$ about the coordinate axes;

34. The moment of inertial of a thin spherical homogeneous shell of mass m and radius R about its diameter.

Selected Answers and Hints to Exercises

Section 28.6. **6.** Hints: $1+2+3+\cdots+n=\frac{1}{2}n(n+1)$ and $1+2^2+3^2+\cdots+n^2=\frac{1}{6}(2n+1)n(n+1)$ so that $L(N)=\frac{40}{3}-\frac{32}{3N}+\frac{2}{N^2}\to\frac{40}{3}$ as $N\to\infty$. **10.** πk. **11**. $\frac{2}{3}\pi$ **12.** $\frac{1}{6}$. **14.** $\frac{1}{2}k^2a$. **15.** $\frac{1}{3}\pi$.

Section 29.1. **1.** $k(16-\pi)$. **2.** 3π. **3.** 6π. **4.** $-2\pi/3$. **5.** $\frac{5}{3}$. **8.** $1\le I\le 16$. **9.** $1\le I\le\sqrt{2}$. **10.** $0\le I\le\pi^2/8$. **11.** $100/51\le I\le 2$, where I denotes the double integral in question. **13.** $\pi(e-1)$.

Section 30.3. **1.** 3. **2.** $\frac{1}{3}$. **3.**$2(2^5-2^{5/2}-3^{5/2}+1]/15$. **4.** 4. **5.** $e-2$. **6.** -1. **7.** $\frac{\pi}{2}$. **8.** $2\tan^{-1}(1/2)+\frac{1}{2}\ln(5/2)-\frac{\pi}{4}$. **9.**$(1+(-1)^n)/[(n+1)(n+2)]$. **10.** $4[(2+e)^{5/2}-3^{5/2}-e^{5/2}+1]/15$. **11.** $\pi^2/4$. **12.** $F(4)+F(2)-2F(3)-1$ where $F(u)=\frac{1}{2}u^2(\ln u-\frac{1}{2})$. **13.** $\frac{1}{2}(6\ln 2-3\ln 3)$. **14.** $(2\ln 2)^{-1}-[2(\ln 2)^3]^{-1}$. **15.** $\frac{77}{288}$. **16.** 40. **17.** 16. **18.** 3. **22.** $\frac{1}{2}R^2+a^2+b^2$.

Section 31.6. **1.** $\int_0^1\int_{-2y}^{2y}f\,dxdy$. **2.** $\int_0^1\int_0^{x+1}f\,dydx$. **3.** $\int_{-1}^1\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}f\,dydx$. **4.** $\int_0^1\int_{-\sqrt{y-y^2}}^{\sqrt{y-y^2}}f\,dxdy$. **5.** $\int_{-2}^{-1}\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}}f\,dydx+\int_{-1}^1\int_{\sqrt{1-x^2}}^{\sqrt{4-x^2}}f\,dydx+\int_{-1}^1\int_{-\sqrt{4-x^2}}^{-\sqrt{1-x^2}}f\,dydx+\int_1^2\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}}f\,dydx$. **6.** $\frac{1}{24}$. **7.** $\frac{8}{3}$. **8.** $\frac{5}{18}$. **9.** 18. **10.** $\frac{8}{5}$. **11.** $\frac{1}{3}$. **12.** $-\frac{5}{8}$. **13.** $\frac{1}{12}$. **14.** $[2(\sqrt{2}-1)-\frac{2}{3}]a^{3/2}$. **15.** $\frac{1}{2}a^4$. **16.** $14a^4$. **17.** $\frac{35\pi}{12}a^4$. **18.** A solid bounded above by the paraboloid $z=x^2+y^2$ and below by the triangle with vertices $(0,0)$, $(0,1)$, and $(1,0)$. **19.** A solid bounded above by the plane $z=x+y$ and below by the triangle with vertices $(0,0)$, $(0,1)$, and $(1,0)$. **20.** A solid bounded above by the cone $z=\sqrt{x^2+y^2}$ and below by the disk $(x-\frac{1}{2})^2+y^2\le\frac{1}{4}$. **21.** A solid bounded above by the paraboloid $z=x^2+y^2$ and below by the square with vertices $(0,1)$, $(1,0)$, $(0,-1)$, and $(-1,0)$. **22.** A part of the solid ellipsoid $(x/2)^2+(y/3)^2+z^2\le 1$ that lies above the xy plane. **23.** $\frac{1}{6}$. **24.** $\frac{208}{105}$. **25.** $\frac{1}{3}$. **26.** $\frac{16}{3}a^3$. **27.** 32. **28.** 0. **29.** $\int_0^1\int_{y^2}^{y^{1/3}}f\,dxdy$. **30.** $\int_0^1\int_x^{\sqrt{x}}f\,dydx$. **31.** $\int_1^e\int_{\ln y}^1 f\,dxdy$. **32.** $\int_1^2\int_1^{x^2}f\,dydx$. **33.** $\int_0^3\int_x^{6-x}f\,dydx$. **34.** $\frac{2}{3}\ln 3$. **35.** $\int_{-1}^0\int_{-2\sqrt{y+1}}^{2\sqrt{y+1}}f\,dxdy+\int_0^8\int_{-2\sqrt{y+1}}^{2-y}f\,dxdy$. **36.** $\int_{-1}^0\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}f\,dxdy+\int_0^1\int_{-\sqrt{1-y}}^{\sqrt{1-y}}f\,dxdy$. **37.** $\int_0^a(\int_{y^2/(2a)}^{a-\sqrt{a^2-y^2}}+\int_{a+\sqrt{a^2-y^2}}^{2a})f\,dxdy+\int_a^{2a}\int_{y^2/(2a)}^{2a}f\,dxdy$. **38.** $\int_0^1\int_{\sin^{-1}(y)}^{\pi-\sin^{-1}(y)}f\,dxdy$. **39.** 0. **40.** 0. **41.** $-\frac{1}{3}ba^2$. **42.** 0.

Section 32.5. **1.** $D=\{(x,y)|1\le x^2+y^2\le 4,\ y\ge 0\}$, the integral is the area of D which is $A(D)=\frac{3}{2}\pi$. **2.** $D=\{(x,y)|(x-a)^2+y^2\le a^2\}$, the integral is the area of the disk D, which is $A(D)=\pi a^2$. **3.** D is the right-angled triangle with the vertices $(0,0)$, $(1,1)$, and $(1,-1)$; its area is 1.

4. D is bounded by the cardioid $r = 1 + \cos\theta$ (symmetric about the x axis); its area is $3\pi/2$. **9.** $\frac{1}{8}(b^4 - a^4)$. **10.** $\pi(1 - \cos(a^2))$. **11.** $\frac{\pi^2}{48}(b^2 - a^2)$. **12.** $\frac{\pi}{4}[b^2 \ln(\frac{b^2}{e}) - a^2 \ln(\frac{a^2}{e})]$. **13.** $-6\pi^2$. **17.** $D' = [0,1] \times [-\frac{\pi}{2}; \frac{\pi}{2}]$; $I = \frac{\pi}{2}(e-1)$. **18.** $D' = [0,1] \times [\frac{\pi}{2}, \frac{3\pi}{2}]$; $I = -\frac{2}{3}$. **19.** $D' = \{(r,\theta) | 0 \le r \le 2\sin\theta, 0 \le \theta \le \frac{\pi}{2}\}$; $I = \frac{16}{9}$. **20.** $D' = [1,2] \times [0, \frac{\pi}{4}]$; $I = \frac{15}{16}$. **23.** $2\pi \int_0^1 f(r) r dr$. **27.** $\frac{3}{2}\pi$. **28.** $\frac{1}{4}\pi^3$. **29.** $\frac{9}{8}\sqrt{3} - \frac{\pi}{4}$. **34.** $\frac{4}{3}\pi$. **35.** $\frac{14}{3}\pi$. **36.** 8π. **37.** $\frac{4}{3}\pi$. **38.** $\frac{3}{2}\pi$. **39.** $f(0,0)$.

Section 33.5. **7.** D lies in the first quadrant and is bounded by the coordinate lines and by the parabola $y = 1 - x^2$. **8.** D lies in the first quadrant and is bounded by the curves $x = 0$, $y = 1$, and $y = \sqrt{x}$. **9.** $D = [-1,1] \times [-1,1]$. **11.** 192. **12.** $\frac{\pi}{\sqrt{3}}$. **13.** $\ln 3 - \frac{1}{2}\ln 5$. **14.** $\frac{7}{3}(e - \sqrt{e})$. **19.** 1; $u = xy$, $v = yx^2$. **20.** $e - \frac{1}{e}$; $u = x + y$, $v = y - x$. **21.** 1; $u = x + y$, $v = y - x^3$. **22.** 2; $u = xy$, $v = y - x^2$. **23.** $\frac{3}{2}$; $u = \frac{y}{x}$, $v = x + y$.

Section 34.6. **1.** $-\frac{13}{2}$. **2.** 1. **3.** $\frac{1}{4}(e-2)$. **4.** $\frac{65}{28}$. **7.** $\frac{1}{30}$. **8.** $\frac{\pi}{24}$. **12.** $\frac{8}{15}$. **13.** 2. **18.** $\frac{7}{24}$. **21.** $\frac{32}{3}\pi$. **22.** -2. **23.** 0 by symmetry $(x,y,z) \to (x,z,y)$. **24.** $28\pi/3$ by symmetry $(x,y,z) \to (x,z,y)$. **32.** The maximum is achieved on the largest E in which the integrand is positive, which the solid ellipsoid $x^2/a^2 + y^2/b^2 + z^2/c^2 \le 1$.

Section 35.6. **1.** A wedge of the solid cylinder $x^2 + y^2 \le 3$ between the planes $z = 0$ and $z = 1$ that lies between the half-planes $y = \pm x$, $x \ge 0$. **2.** The solid bounded by the cones $z = 1 - \sqrt{x^2+y^2}$ and $z = \sqrt{x^2+y^2} - 1$. **3.** The part of the solid between the paraboloid $z = 4 - x^2 - y^2$ and the xy plane that lies in the first octant. **4.** $0 \le z \le r^2$, (r,θ) in $[0,1] \times [0,2\pi]$. **5.** $1 - r \le z \le 1 + r$, (r,θ) in $[0,1] \times [0,2\pi]$. **6.** $0 \le z \le r^2$, $0 \le r \le 2\cos\theta$, θ in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. **7.** $0 \le z \le \sqrt{a^2 - r^2}$, (r,θ) in $[0,a] \times [0, \frac{\pi}{2}]$. **8.** $\frac{7}{2}\pi$. **9.** $\frac{2}{35}$. **10.** $\frac{15}{4}\pi$. **11.** $\frac{32}{9}$. **12.** $\frac{4}{105}a^7$. **13.** $\frac{16}{3}\pi$. **18.** (ρ,ϕ,θ) in $[1,2] \times [0,\frac{\pi}{2}] \times [0,\frac{\pi}{2}]$. **19.** (ρ,ϕ,θ) in $[0,a] \times [\frac{\pi}{6}, \frac{5\pi}{6}] \times [0,2\pi]$. **20.** (ρ,ϕ,θ) in $[0,a] \times [0,\pi] \times [\frac{\pi}{6}, \frac{\pi}{3}]$. **21.** $0 \le \rho \le 4\cos\phi$, $(\phi,\theta) \in [0,\pi/2] \times [0,2\pi]$. **22.** $\frac{4}{9}\pi a^9$. **23.** $\frac{62}{15}\pi$. **24.** 0. **25.** $-\frac{\pi}{16}$. **26.** $\frac{\pi}{6}$. **27.** $\frac{\pi}{10}$. **28.** $\frac{\pi}{4}$. **29.** $\frac{\pi}{15}(2\sqrt{2} - 1)$. **30.** $\frac{2}{3}\pi$; the solid is the part of the cone $z = \sqrt{x^2+y^2}$ below the plane $z = 2$ that lies in the first octant. **31.** $\frac{\pi}{6}$; the region of integration is bounded by the paraboloid $z = 1 - x^2 - y^2$ and the xy plane.

Section 36.3. **1.** 0. **2.** $8uvw + 1$. **3.** $uv(u^2 + v^2)$. **6.** $\frac{1}{90}a^6$. **7.** $\frac{4\pi}{35}abc$. **8.** $\frac{abc}{1680}$. **12.** $\frac{\pi}{8}abc^2$. **13.** 0; put $u = x/3$, $v = y/2$, $w = z$ and use the symmetry $(u,v,w) \to (v,u,w)$.

Section 38.3. **1.** $\frac{32}{3}$. **2.** $\frac{b}{a}\sqrt{a^2+b^2}(1 - \frac{\sin a}{a})$. **3.** $\pi\sqrt{5}$. **4.** $\frac{1}{6}(14^{3/2} - 1)$. **5.** 16π. **16.** $\frac{1}{3}[(T^2+2)^{3/2} - 2^{3/2}]$. **18.** $\frac{2}{3}a^2(2\sqrt{2} - 1)$.

Section 39.5. **1.** $\frac{1}{2}\sqrt{a^2b^2+b^2c^2+a^2c^2}$. **2.** $4\pi\sqrt{14}$. **3.** $\frac{\pi}{6}(17^{3/2}-5^{3/2})$. **4.** $\frac{\pi}{6}(37^{3/2}-5^{3/2})$. **5.** $\frac{1}{12}(21^{3/2}-17^{3/2})$. **6.** $\frac{\sqrt{3}}{24}$. **7.** $\frac{21\pi}{\sqrt{2}}$. **8.** 0. **9.** $\pi\sqrt{2}$. **10.** 0.

Section 40.4. **3.** $(0,0,\frac{h}{2})$ if the axis of the helix is the z axis. **4.** $(0,0,\frac{R}{2})$ if the hemisphere lies above the xy plane and is centered at the origin. **5.** Let the center of the hole be at $(\frac{R}{2},0)$. Then $x_c=-\frac{Ra^2}{2(R^2-a^2)}$ and $y_c=0$. **8.** $x_c=y_c=0$, $z_c=\frac{16}{15\pi}\frac{b^5-a^5}{b^4-a^4}$. **9.** $x_c=\frac{9a}{2\pi}(\sqrt{3}-\sqrt{2})$, $y_c=\frac{9a}{2\pi}(\sqrt{2}-1)$. **22.** $\frac{2}{5}mR^2$. **23.** $\frac{38}{3}\pi$. **29.** $\frac{3}{2}ma^2$. (Hint: use the Parallel Axis Theorem, Study problem **40.2**). **34.** $\frac{2}{3}mR^2$.