Chapter 16
Applications of Integration

One of the main goals of mathematical analysis, besides applications in physics, is to compute the measure of sets (arc length, area, surface area, and volume). We have already spent time computing arc lengths, but only for graphs of functions. We saw examples of computing the area of certain shapes (mostly regions under graphs), and at the same time, we got a taste of computing volumes when we determined the volume of a sphere (see item 2 in Example 13.23). We also noted, however, that in computing area, some theoretical problems need to be addressed (as mentioned in point 5 of Remark 14.10). In this chapter, we turn to a systematic discussion of these questions.

When computing area, we obviously deal with sets in the plane, while when computing volume, we deal with sets in the space. However, when we deal with arc length we need to concern ourselves with both sets in the plane and in space, since some curves lie in the plane while others do not. Therefore, it is best to tackle questions concerning the plane and space simultaneously whenever possible. In mathematical analysis, points of the plane are associated with ordered pairs of real numbers, and the plane itself is associated with the set $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ (see the appendix of Chapter 9). We will proceed analogously for representing three-dimensional space as well. We consider three lines in space intersecting at a point that are mutually perpendicular, which we call the x-, y-, and z-**axes.** We call the plane spanned by the x- and y-axes the xy-plane, and we have similar definitions for the xz- and yz-planes. We assign an ordered triple (a,b,c) to every point P in space, in which a, b, and c denote the distance (with positive or negative sign) of the point from the yz-, xz-, and xy-planes respectively. We call the numbers a, b, and c the **coordinates** of P. The geometric properties of space imply that the map $P \mapsto (a,b,c)$ that we obtain in this way is a bijection. This justifies our representation of three-dimensional space by ordered triples of real numbers.

Thus if we want to deal with questions both in the plane and in space, we need to deal with sets that consist of ordered d-tuples of real numbers, where $d = 2$ or $d = 3$. We will see that the specific value of d does not usually play a role in the definitions and proofs coming up. Therefore, for every positive integer d, we can define **d-dimensional Euclidean space,** by which we simply mean the set of all

M. Laczkovich, V.T. Sós, *Real Analysis*, Undergraduate Texts
in Mathematics, DOI 10.1007/978-1-4939-2766-1_16

sequences of real numbers of length d, with the appropriately defined addition, multiplication by a constant, absolute value, and distance. If $d = 1$, then the Euclidean space is exactly the real line; if $d = 2$, then it is the plane; and if $d = 3$, then it is 3-dimensional space. For $d > 3$, a d-dimensional space does not have an observable meaning, but it is very important for both theory and applications.

Definition 16.1. $\mathbb{R}^d$ denotes the set of ordered d-tuples of real numbers, that is, the set

$$\mathbb{R}^d = \{(x_1, \dots, x_d) \colon x_1, \dots, x_d \in \mathbb{R}\}.$$

The points of the set $\mathbb{R}^d$ are sometimes called *d-dimensional vectors*. The *sum of the vectors $x = (x_1, \dots, x_d)$* and $y = (y_1, \dots, y_d)$ is the vector

$$x + y = (x_1 + y_1, \dots, x_d + y_d),$$

and the *product of the vector x and a real number c* is the vector

$$c \cdot x = (cx_1, \dots, cx_d).$$

The *absolute value* of the vector x is the nonnegative real number

$$|x| = \sqrt{x_1^2 + \cdots + x_d^2}.$$

It is clear that for all $x \in \mathbb{R}^d$ and $c \in \mathbb{R}$, $|cx| = |c| \cdot |x|$. It is also easy to see that if $x = (x_1, \dots, x_d)$, then

$$|x| \le |x_1| + \cdots + |x_d|. \tag{16.1}$$

The **triangle inequality** also holds:

$$|x + y| \le |x| + |y| \qquad (x, y \in \mathbb{R}^d). \tag{16.2}$$

To prove this, it suffices to show that $|x + y|^2 \le (|x| + |y|)^2$, since both sides are nonnegative. By the definition of the absolute value, this is exactly

$$(x_1 + y_1)^2 + \cdots + (x_d + y_d)^2 \le$$
$$\le (x_1^2 + \cdots + x_n^2) + 2 \cdot \sqrt{x_1^2 + \cdots + x_d^2} \cdot \sqrt{y_1^2 + \cdots + y_d^2} + y_1^2 + \cdots + y_d^2,$$

that is,

$$x_1 y_1 + \cdots + x_d y_d \le \sqrt{x_1^2 + \cdots + x_d^2} \cdot \sqrt{y_1^2 + \cdots + y_d^2},$$

which is the Cauchy–Schwarz–Bunyakovsky inequality (Theorem 11.19).

The **distance** between the vectors x and y is the number $|x - y|$. By (16.2), it is clear that

$$\big||x| - |y|\big| \le |x - y| \qquad \text{and} \qquad |x - y| \le |x - z| + |z - y|$$

for all $x, y, z \in \mathbb{R}^d$. We can consider these to be variants of the triangle inequality.

If we apply (16.1) to the difference of the vectors $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$, then we get that

$$||x| - |y|| \leq |x - y| \leq |x_1 - y_1| + \cdots + |x_d - y_d|. \tag{16.3}$$

The **scalar product** of the vectors $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$ is the real number $\sum_{i=1}^{d} x_i y_i$, which we denote by $\langle x, y \rangle$. With the help of the arguments in Remark 14.57, it is easy to see that $\langle x, y \rangle = |x| \cdot |y| \cdot \cos\alpha$, where α denotes the angle enclosed by the two vectors.

16.1 The General Concept of Area and Volume

We deal with the concepts of area and volume at once; we will use the word *measure* instead. We will actually define measure in every space $\mathbb{R}^d$, and area and volume will be the special cases $d = 2$ and $d = 3$.

Most of the concepts we define in the plane and in space can be generalized—purely through analogy—for the space $\mathbb{R}^d$, independent of the value of d. That includes, first of all, the concepts of axis-parallel rectangles or rectangular boxes. Since these are sets of the form $[a_1, b_1] \times [a_2, b_2]$ and $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$ in the plane and in space respectively, by an **axis-parallel rectangle** in $\mathbb{R}^d$, or just a **rectangle** for short, we will mean the set

$$[a_1, b_1] \times \cdots \times [a_d, b_d],$$

where $a_i < b_i$ for all $i = 1, \dots, d$. (Here we use the Cartesian product with a finite number of terms. This means that $A_1 \times \cdots \times A_d$ denotes the set of sequences $(x_1, \dots, x_d)$ that satisfy $x_1 \in A_1, \dots, x_d \in A_d$.) For the case $d = 1$, the definition of a rectangle agrees with the definition of a nondegenerate closed and bounded interval.

We get (open) balls in $\mathbb{R}^d$ in the same way through analogy. The **open ball** $B(a, r)$ with center $a \in \mathbb{R}^d$ and radius $r > 0$ is the set of points that are less than distance r away from a, that is,

$$B(a, r) = \{x \in \mathbb{R}^d : |x - a| < r\}.$$

For the case $d = 1$, $B(a, r)$ is exactly the open interval $(a - r, a + r)$, while when $d = 2$, it is an open disk of radius r centered at a (where "open" means that the boundary of the disk does not belong to the set).

We call the set $A \subset \mathbb{R}^d$ **bounded** if there exists a rectangle $[a_1, b_1] \times \ldots \times [a_d, b_d]$ that contains it. It is easy to see that a set is bounded if and only if it is contained in a ball (see exercise 16.1).

We say that x is an **interior point** of the set $H \subset \mathbb{R}^d$ if H contains a ball centered at x; that is, if there exists an $r > 0$ such that $B(x, r) \subset H$. Since every ball contains a rectangle and every rectangle contains a ball, a set A has an interior point if and only if A contains a rectangle.

We call the sets A and B **nonoverlapping** if they do not share any interior points.

If we want to convert the intuitive meaning of measure into a precise notion, then we should first list our expectations for the concept. Measure has numerous properties that we consider natural. We choose three of these (see Remark 14.10.5):

(a) The measure of the rectangle $R = [a_1, b_1] \times \cdots \times [a_d, b_d]$ equals the product of its side lengths, that is, $(b_1 - a_1) \cdots (b_d - a_d)$.
(b) If we decompose a set into the union of finitely many nonoverlapping sets, then the measure of the set is the sum of the measures of the parts.
(c) If $A \subset B$, then the measure of A is not greater than the measure of B.

We will see that these requirements naturally determine to which sets we can assign a measure, and what that measure should be.

Definition 16.2. If $R = [a_1, b_1] \times \cdots \times [a_d, b_d]$, then we let $m(R)$ denote the product $(b_1 - a_1) \cdots (b_d - a_d)$.

Let A be an arbitrary bounded set in $\mathbb{R}^d$. Cover A in every possible way by finitely many rectangles $R_1, \ldots, R_K$, and form the sum $\sum_{i=1}^{K} m(R_i)$ for each cover. The *outer measure* of the set A is defined as the infimum of the set of all the sums we obtain in this way (Figure 16.1). We denote the outer measure of the set A by $\overline{m}(A)$.

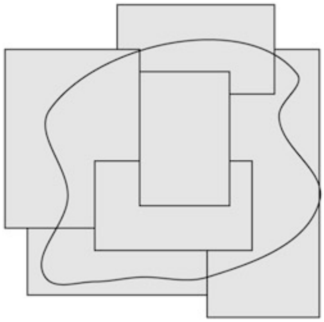

Fig. 16.1

If A does not have an interior point, then we define the *inner measure* to be zero. If A does have an interior point, then choose every combination of finitely many rectangles $R_1, \ldots, R_K$ each in A such that they are mutually nonoverlapping, and form the sum $\sum_{i=1}^{K} m(R_i)$ each time. The inner measure of A is defined as the supremum of the set of all such sums. The inner measure of the set A will be denoted by $\underline{m}(A)$.

It is intuitively clear that for every bounded set A, the values $\underline{m}(A)$ and $\overline{m}(A)$ are finite. Moreover, $0 \le \underline{m}(A) \le \overline{m}(A)$. Now by restrictions (a) and (c) above, it is clear that the measure of the set A should fall between $\underline{m}(A)$ and $\overline{m}(A)$. If $\underline{m}(A) < \overline{m}(A)$, then without further inspection, it is not clear which number (between $\underline{m}(A)$ and $\overline{m}(A)$) we should consider the measure of A to be. We will do what we did when we considered integrals, and restrict ourselves to sets for which $\underline{m}(A) = \overline{m}(A)$, and this shared value will be called the measure of A.

Definition 16.3. We call the bounded set $A \subset \mathbb{R}^d$ *Jordan*[1] *measurable* if $\underline{m}(A) = \overline{m}(A)$. The *Jordan measure* of the set A (the measure of A, for short) is the shared value $\underline{m}(A) = \overline{m}(A)$, which we denote by $m(A)$.

If $d \ge 3$, then instead of Jordan measure, we can say **volume**; if $d = 2$, **area**; and if $d = 1$, **length** as well.

[1] Camille Jordan (1838–1922), French mathematician.

If we want to emphasize that we are talking about the inner, outer, or Jordan measure of a d-dimensional set, then instead of $\underline{m}(A)$, $\overline{m}(A)$, or $m(A)$, we may write $\underline{m}_d(A)$, $\overline{m}_d(A)$, or $m_d(A)$.

Exercises

16.1. Prove that for every set $A \subset \mathbb{R}^d$, the following statements are equivalent.

(a) The set A is bounded.
(b) There exists an $r > 0$ such that $A \subset B(0,r)$.
(c) For all $i = 1,\ldots,d$, the ith coordinates of the points of A form a bounded set in $\mathbb{R}$.

16.2. Prove that the set

$$A = \left\{(x,y)\colon 0 \le x \le 1,\ 0 \le y \le 1,\ y \ne 1/n\ (n \in \mathbb{N}^+)\right\}$$

is Jordan measurable.

16.3. Prove that the set

$$A = \{(x,y)\colon 0 \le x \le 1,\ 0 \le y \le 1,\ x,y \in \mathbb{Q}\}$$

is not Jordan measurable.

16.4. Prove that if $A \subset \mathbb{R}^p$ and $B \subset \mathbb{R}^q$ are measurable sets, then $A \times B \subset \mathbb{R}^{p+q}$ is also measurable, and that $m_{p+q}(A \times B) = m_p(A) \cdot m_q(B)$. (S)

16.2 Computing Area

With the help of Definition 16.3, we can now prove that the area under the graph of a function agrees with the integral of the function (see Remark 14.10.5). We will actually determine the areas of slightly more general regions, and then the area under the graph of a function, as well as the reflection of that region in the x-axis, will be special cases.

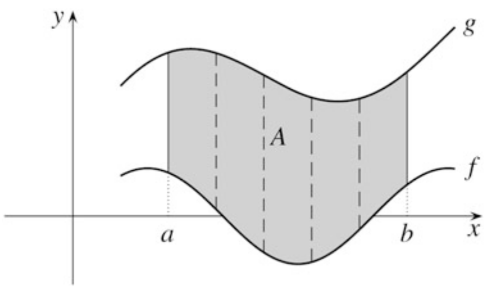

Fig. 16.2

Definition 16.4. We call the set $A \subset \mathbb{R}^2$ a *normal domain* if

$$A = \{(x,y) : x \in [a,b],\ f(x) \le y \le g(x)\}, \tag{16.4}$$

where f and g are integrable on $[a,b]$ and $f(x) \le g(x)$ for all $x \in [a,b]$ (Figure 16.2).

Theorem 16.5. *If f and g are integrable on $[a,b]$ and $f(x) \le g(x)$ for all $x \in [a,b]$, then the normal domain in (16.4) is measurable, and its area is*

$$m_2(A) = \int_a^b (g-f)\,dx.$$

Proof. For a given $\varepsilon > 0$, choose partitions F_1 and F_2 such that $\Omega_{F_1}(f) < \varepsilon$ and $\Omega_{F_2}(g) < \varepsilon$. If $F = (x_0, \ldots, x_n)$ is the union of the partitions F_1 and F_2, then $\Omega_F(f) < \varepsilon$ and $\Omega_F(g) < \varepsilon$. Let $m_i(f), m_i(g), M_i(f)$, and $M_i(g)$ be the infimum and supremum of the functions f and g respectively on the interval $[x_{i-1}, x_i]$. Then the rectangles $[x_{i-1}, x_i] \times [m_i(f), M_i(g)]\ (i = 1, \ldots, n)$ cover the set A, so

$$\begin{aligned}\overline{m}_2(A) &\le \sum_{i=1}^{n} (M_i(g) - m_i(f)) \cdot (x_i - x_{i-1}) = \\ &= S_F(g) - s_F(f) < \\ &< \int_a^b g\,dx + \varepsilon - \left(\int_a^b f\,dx - \varepsilon \right) = \\ &= \int_a^b (g-f)\,dx + 2\varepsilon. \end{aligned} \tag{16.5}$$

Let I denote the set of indices i that satisfy $M_i(f) \le m_i(g)$. Then the rectangles $[x_{i-1}, x_i] \times [M_i(f), m_i(g)]\ (i \in I)$ are contained in A and are nonoverlapping, so

$$\begin{aligned}\underline{m}_2(A) &\ge \sum_{i \in I} (m_i(g) - M_i(f)) \cdot (x_i - x_{i-1}) \ge \\ &\ge \sum_{i=1}^{n} (m_i(g) - M_i(f)) \cdot (x_i - x_{i-1}) = \\ &= s_F(g) - S_F(f) > \\ &> \int_a^b g\,dx - \varepsilon - \int_a^b f\,dx - \varepsilon = \\ &= \int_a^b (g-f)\,dx - 2\varepsilon. \end{aligned} \tag{16.6}$$

Since ε was arbitrary, by (16.5) and (16.6) it follows that A is measurable and has area $\int_a^b (g-f)\,dx$. □

Example 16.6. With the help of the theorem above, we can conclude that the domain bounded by the ellipse with equation $x^2/a^2 + y^2/b^2 = 1$ is a measurable set, whose area is $ab\pi$. Clearly, the set A in question is a normal domain given by the continuous functions $f(x) = -b \cdot \sqrt{1-\left(\frac{x}{a}\right)^2}$ and $g(x) = b \cdot \sqrt{1-\left(\frac{x}{a}\right)^2}$ over the interval $[-a,a]$. Then by Theorem 16.5, A is measurable and $m_2(A) = \int_a^b (g-f)\,dx = 2 \cdot \int_a^b g\,dx$, so by the computation done in Remark 15.20.2, we obtain $m_2(A) = ab\pi$.

We now turn to a generalization of Theorem 16.5 that can (theoretically) be used to compute the measure of any measurable plane set.

Definition 16.7. The *sections* of the set $A \subset \mathbb{R}^2$ are the sets

$$A_x = \{y \in \mathbb{R} \colon (x,y) \in A\}$$

and

$$A^y = \{x \in \mathbb{R} \colon (x,y) \in A\}$$

for every $x, y \in \mathbb{R}$ (Figure 16.3).

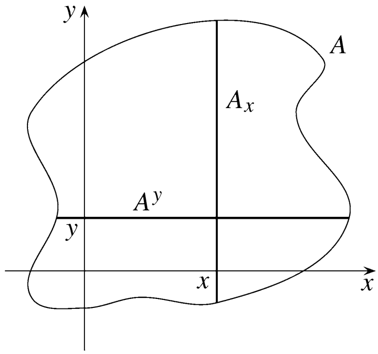

Fig. 16.3

Theorem 16.8. *Let $A \subset \mathbb{R}^2$ be a measurable set such that $A \subset [a,b] \times [c,d]$. Then the functions $x \mapsto \overline{m}_1(A_x)$ and $x \mapsto \underline{m}_1(A_x)$ are integrable in $[a,b]$, and*

$$m_2(A) = \int_a^b \overline{m}_1(A_x)\,dx = \int_a^b \underline{m}_1(A_x)\,dx. \tag{16.7}$$

Similarly, the functions $y \mapsto \overline{m}_1(A^y)$ and $y \mapsto \underline{m}_1(A^y)$ are integrable in $[c,d]$, and

$$m_2(A) = \int_c^d \overline{m}_1(A^y)\,dy = \int_c^d \underline{m}_1(A^y)\,dy.$$

Proof. It suffices to prove (16.7). Since $A \subset [a,b] \times [c,d]$, we have $A_x \subset [c,d]$ for all $x \in [a,b]$. It follows that if $x \in [a,b]$, then $\underline{m}_1(A_x) \le \overline{m}_1(A_x) \le d-c$, so the functions $\underline{m}_1(A_x)$ and $\overline{m}_1(A_x)$ are bounded in $[a,b]$.

Let $\varepsilon > 0$ be given, and choose rectangles $T_i = [a_i,b_i] \times [c_i,d_i]$ $(i = 1,\ldots,n)$ such that $A \subset \bigcup_{i=1}^n T_i$ and $\sum_{i=1}^n m_2(T_i) < m_2(A) + \varepsilon$. We can assume that $[a_i,b_i] \subset [a,b]$ for all $i = 1,\ldots,n$. Let

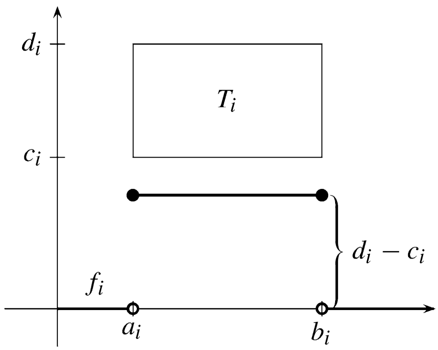

Fig. 16.4

$$f_i(x) = \begin{cases} 0, & \text{if } x \notin [a_i, b_i], \\ d_i - c_i, & \text{if } x \in [a_i, b_i] \end{cases} \qquad (i = 1, \ldots, n).$$

(See Figure 16.4.) Then f_i is integrable in $[a,b]$, and $\int_a^b f_i\,dx = m_2(T_i)$. For arbitrary $x \in [a,b]$, the sections A_x are covered by the intervals $[c_i, d_i]$ that correspond to indices i for which $x \in [a_i, b_i]$. Thus by the definition of the outer measure,

$$\overline{m}_1(A_x) \le \sum\nolimits_{x \in [a_i, b_i]} (d_i - c_i) = \sum_{i=1}^{n} f_i(x).$$

It follows that

$$\overline{\int_a^b} \overline{m}_1(A_x)\,dx \le \overline{\int_a^b} \sum\nolimits_{i=1}^n f_i\,dx = \int_a^b \sum\nolimits_{i=1}^n f_i\,dx = \sum\nolimits_{i=1}^n m_2(T_i) < m_2(A) + \varepsilon. \tag{16.8}$$

Now let $R_i = [p_i, q_i] \times [r_i, s_i]$ $(i = 1, \ldots, m)$ be nonoverlapping rectangles such that $A \supset \bigcup_{i=1}^m R_i$ and $\sum_{i=1}^m m_2(R_i) > m_2(A) - \varepsilon$. Then $[p_i, q_i] \subset [a,b]$ for all $i = 1, \ldots, m$. Let

$$g_i(x) = \begin{cases} 0, & \text{if } x \notin [p_i, q_i], \\ s_i - r_i, & \text{if } x \in [p_i, q_i] \end{cases} \qquad (i = 1, \ldots, m).$$

Then g_i is integrable in $[a,b]$, and $\int_a^b g_i\,dx = m_2(R_i)$. If $x \in [a,b]$, then the section A_x contains all the intervals $[r_i, s_i]$ whose indices i satisfy $x \in [a_i, b_i]$. We can also easily see that if x is distinct from all points p_i, q_i, then these intervals are nonoverlapping. Then by the definition of the inner measure,

$$\underline{m}_1(A_x) \ge \sum\nolimits_{x \in [p_i, q_i]} (s_i - r_i) = \sum\nolimits_{i=1}^m g_i(x).$$

It follows that

$$\underline{\int_a^b} \underline{m}_1(A_x)\,dx \ge \underline{\int_a^b} \sum\nolimits_{i=1}^m g_i\,dx = \int_a^b \sum\nolimits_{i=1}^m g_i\,dx = \sum\nolimits_{i=1}^m m_2(R_i) > m_2(A) - \varepsilon. \tag{16.9}$$

Now $\underline{m}_1(A_x) \le \overline{m}_1(A_x)$ for all x, so by (16.8) and (16.9), we get that

$$m_2(A) - \varepsilon < \underline{\int_a^b} \underline{m}_1(A_x)\,dx \le \overline{\int_a^b} \underline{m}_1(A_x)\,dx \le \overline{\int_a^b} \overline{m}_1(A_x)\,dx < m_2(A) + \varepsilon.$$

Since this holds for all ε, we have $\underline{\int_a^b} \underline{m}_1(A_x)\,dx = \overline{\int_a^b} \underline{m}_1(A_x)\,dx = m_2(A)$, which means that the function $x \mapsto \underline{m}_1(A_x)$ is integrable on $[a,b]$ with integral $m_2(A)$. We get that $\int_a^b \overline{m}_1(A_x)\,dx = m_2(A)$ the same way. □

Remark 16.9. Observe that we did not assume the measurability of the sections A_x and A^y in Theorem 16.8 (that is, that $\underline{m}_1(A_x) = \overline{m}_1(A_x)$ and $\underline{m}_1(A^y) = \overline{m}_1(A^y)$).

Corollary 16.10. *Let f be a nonnegative and bounded function on the interval $[a,b]$. The set $B_f = \{(x,y)\colon x \in [a,b],\ 0 \le y \le f(x)\}$ is measurable if and only if f is integrable, and then*

$$m_2(B_f) = \int_a^b f\,dx.$$

Proof. With the help of Theorem 16.5. we need to prove only that if B_f is measurable, then f is integrable. This, however, is clear by Theorem 16.8, since $(B_f)_x = [0, f(x)]$, and so $\underline{m}_1\big((B_f)_x\big) = \overline{m}_1\big((B_f)_x\big) = f(x)$ for all $x \in [a,b]$. □

Exercises

16.5. Determine the area of the set $\{(x,y)\colon 2 - x \le y \le 2x - x^2\}$.

16.6. Determine the area of the set $\{(x,y)\colon 2^x \le y \le x + 1\}$.

16.7. For a given $a > 0$, determine the area of the set $\{(x,y)\colon y^2 \le x^2(a^2 - x^2)\}$.

16.8. Let $0 < \delta < \pi/2$, $r > 0$, and $x_0 = r\cos\delta$. Prove, by computing the area under the graph of the function

$$f(x) = \begin{cases} (\operatorname{tg}\delta)\cdot x, & \text{if } 0 \le x \le x_0, \\ \sqrt{r^2 - x^2}, & \text{if } x_0 \le x \le r, \end{cases}$$

that a circular sector with central angle δ and radius r is measurable and has area $r^2\delta/2$. (S)

16.9. Let $u > 1$ and $v = \sqrt{u^2 - 1}$. The two segments connecting the origin to the points (u,v) and $(u,-v)$ and the hyperbola $x^2 - y^2 = 1$ between the points (u,v) and $(u,-v)$ define a region A_u. Determine the area of A_u. (S)

To determine the **center of mass** of a region, we will borrow the fact from physics that if we break up a region into smaller parts, then the "weighed average" of the centers of mass of the parts, where the weights are equal to the area of each part, gives us the center of mass of the whole region. More precisely, this means that if we break the region A into regions $A_1, \ldots, A_n$, and p_i is the center of mass of A_i, then the center of mass of A is

$$\frac{\sum_{i=1}^n m(A_i)p_i}{m(A)}.$$

Let $f\colon [a,b] \to \mathbb{R}$ be a nonnegative continuous function, and consider the region under the graph of f, $B_f = \{(x,y)\colon x \in [a,b],\ 0 \le y \le f(x)\}$. Let $F\colon a = x_0 < x_1 < \cdots < x_n = b$ be a fine partition, and break up B_f into the regions $A_i = \{(x,y)\colon x \in [x_{i-1}, x_i],\ 0 \le y \le f(x)\}$ $(i = 1, \ldots, n)$. If $x_i - x_{i-1}$ is small, then A_i can be well approximated by the rectangle $T_i = [x_{i-1}, x_i] \times [0, f(c_i)]$, where $c_i = (x_{i-1} + x_i)/2$. Its center of mass is the point $p_i = (c_i, f(c_i)/2)$, and its area is $f(c_i)\cdot(x_i - x_{i-1})$. Thus the center of mass of the region $\bigcup_{i=1}^{n} T_i$ approximating B_f is the weighed average of the points p_i with weights $f(c_i) \cdot (x_i - x_{i-1})$, that is, the point

$$\frac{1}{\sigma_F} \cdot \sum_{i=1}^{n} f(c_i) \cdot (x_i - x_{i-1}) \cdot (c_i, f(c_i)/2), \tag{16.10}$$

where $\sigma_F = \sum_{i=1}^{n} f(c_i) \cdot (x_i - x_{i-1})$. The first coordinate of the point (16.10) is

$$x_F = {\sigma_F}^{-1} \cdot \sum_{i=1}^{n} f(c_i) \cdot c_i \cdot (x_i - x_{i-1}),$$

and its second coordinate is $y_F = (1/2) \cdot \sigma_F^{-1} \cdot \sum_{i=1}^{n} f^2(c_i) \cdot (x_i - x_{i-1})$. Let $I = \int_a^b f\,dx$. If the partition F is fine enough (has small enough mesh), then σ_F is close to I, x_F is close to the value of $x_s = I^{-1} \int_a^b f(x) \cdot x\,dx$, and y_F is close to the value of $y_s = (1/2) \cdot I^{-1} \int_a^b f^2(x)\,dx$.

This motivates the following definition. Let $f \colon [a,b] \to \mathbb{R}$ be nonnegative and integrable, and suppose that $I = \int_a^b f\,dx > 0$. Then the **center of mass** of the region $B_f = \{(x,y) \colon x \in [a,b],\ 0 \le y \le f(x)\}$ is the point (x_s, y_s), where

$$x_s = \frac{1}{I} \int_a^b f(x) \cdot x\,dx \quad \text{and} \quad y_s = \frac{1}{2I} \int_a^b f(x)^2\,dx.$$

Exercise

16.10. Determine the centers of mass of the following regions.

(a) $\{(x,y) \colon x \in [0,m],\ 0 \le y \le c \cdot x\}$ $(c, m > 0)$;
(b) $\{(x,y) \colon x \in [-r,r],\ 0 \le y \le \sqrt{r^2 - x^2}\}$ $(r > 0)$;
(c) $\{(x,y) \colon x \in [0,a],\ 0 \le y \le x^n\}$ $(a, n > 0)$.

16.3 Computing Volume

Theorem 16.8 can be generalized to higher dimensions without trouble, and the proof of the generalization is the same. Consider, for example, the three-dimensional version. If $A \subset \mathbb{R}^3$, then let A_x denote the set $\{(y,z) \colon (x,y,z) \in A\}$.

Theorem 16.11. *Let $A \subset \mathbb{R}^3$ be a measurable set such that $A \subset [a,b] \times [c,d] \times [e,f]$. Then the functions $x \mapsto \overline{m}_2(A_x)$ and $x \mapsto \underline{m}_2(A_x)$ are integrable in $[a,b]$, and*

$$m_3(A) = \int_a^b \overline{m}_2(A_x)\,dx = \int_a^b \underline{m}_2(A_x)\,dx. \tag{16.11}$$

In the theorem above—just as in Theorem 16.8—the variable x can be replaced by the variable y or z.

With the help of equation (16.11), we can easily compute the volume of measurable sets whose sections are simple geometric shapes, for example rectangles or disks. One such family of sets are called solids of revolution, which we obtain by rotating the region under the graph of a function around the x-axis. More precisely, if the function f is nonnegative on the interval $[a,b]$, then the set

$$B_f = \left\{(x,y,z) \colon a \le x \le b,\ y^2 + z^2 \le f^2(x)\right\}$$

is called the **solid of revolution** determined by the function f.

Theorem 16.12. *If f is nonnegative and integrable on the interval $[a,b]$, then the solid of revolution determined by f is measurable, and its volume is*

$$m_3(B_f) = \pi \cdot \int_a^b f^2(x)\,dx. \tag{16.12}$$

Proof. By (16.11), it is clear that (16.12) gives us the volume (assuming that B_f is measurable). The measurability of B_f is, however, not guaranteed by Theorem 16.11, so we give a direct proof of this fact that uses the idea of squeezing the solid of revolution between solids whose volumes we know. We can obtain such solids if we rotate the inner and outer rectangles corresponding to the curve $\{(x,y) \colon a \le x \le b,\ 0 \le y \le f(x)\}$, which give us so-called inner and outer cylinders (Figure 16.5).

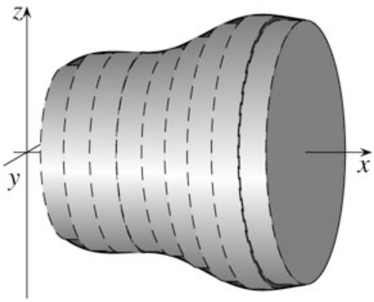

Fig. 16.5

Consider a partition $F \colon a = x_0 < \cdots < x_n = b$, and let

$$m_i = \inf\{f(x) \colon x \in [x_{i-1}, x_i]\} \quad \text{and} \quad M_i = \sup\{f(x) \colon x \in [x_{i-1}, x_i]\}\ (i = 1, \ldots, n).$$

The cylinders

$$\underline{C}_i = \left\{(x,y,z) \colon x_{i-1} \le x \le x_i,\ y^2 + z^2 \le m_i\right\},$$

and

$$\overline{C}_i = \left\{(x,y,z) : x_{i-1} \le x \le x_i,\ y^2+z^2 \le M_i\right\}$$

clearly satisfy

$$\bigcup_{i=1}^{n} \underline{C}_i \subset B_f \subset \bigcup_{i=1}^{n} \overline{C}_i. \tag{16.13}$$

Now we use the fact that the cylinder $\left\{(x,y,z)\colon c \le x \le d,\ y^2+z^2 \le r\right\}$ is measurable and has area $r^2\pi \cdot (d-c)$. This is a simple corollary of the fact that if $A \subset \mathbb{R}^p$ and $B \subset \mathbb{R}^q$ are measurable sets, then $A \times B$ is also measurable, and $m_{p+q}(A \times B) = m_p(A)m_q(B)$ (see Exercise 16.4). Then by (16.13),

$$\overline{m}_3(B_f) \le m_3\left(\bigcup_{i=1}^{n} \overline{C}_i\right) = \pi \sum_{i=1}^{n} M_i^2(x_i - x_{i-1}) = \pi \cdot S_F(f^2),$$

and

$$\underline{m}_3(B_f) \ge m_3\left(\bigcup_{i=1}^{n} \underline{C}_i\right) = \pi \sum_{i=1}^{n} m_i^2(x_i - x_{i-1}) = \pi \cdot s_F(f^2).$$

Since

$$\inf_F S(f^2) = \sup_F s(f^2) = \int_a^b f^2\,dx,$$

we have that

$$\pi \cdot \int_a^b f^2\,dx \le \underline{m}_3(B_f) \le \overline{m}_3(B_f) \le \pi \cdot \int_a^b f^2\,dx.$$

Thus B_f is measurable, and (16.12) holds. □

Exercises

16.11. Compute the area of the solids of revolution corresponding to the following functions:

(a) $\arcsin x$ $(x \in [0,1])$;
(b) $f(x) = e^{-x} \cdot \sqrt{\sin x}$ $(x \in [0,\pi])$;
(c) $f(x) = \operatorname{ch} x$ $(x \in [-a,a])$.

16.12. Prove that the ellipsoid that is a result of rotating the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ around the x-axis has volume $(4/3)ab^2\pi$.

16.13. Consider two right circular cylinders of radius R whose axes intersect and are perpendicular. Compute the volume of their intersection. (H)

16.14. Consider a right circular cylinder of radius R. Compute the volume of the part of the cylinder bounded by the side, the base circle, and a plane passing through a diameter of the base circle and forming an angle of $\frac{\pi}{4}$ with it.

16.15. Compute the volumes of the following solids (taking for granted that they are measurable):

(a) $\{(x,y,z)\colon 0 \le x \le y \le 1,\ 0 \le z \le 2x+3y+4\}$;
(b) $\{(x,y,z)\colon x^2 \le y \le 1,\ 0 \le z \le x^2+y^2\}$;
(c) $\{(x,y,z)\colon x^4 \le y \le 1,\ 0 \le z \le 2\}$;
(d) $\{(x,y,z)\colon (x^2/a^2)+(y^2/b^2)+(z^2/c^2) \le 1\}$ (ellipsoid);
(e) $\{(x,y,z)\colon |x|+|y|+|z| \le 1\}$;
(f) $\{(x,y,z)\colon x,y,z \ge 0, (x+y)^2+z^2 \le 1\}$.

16.16. Check, with the help of Theorem 16.11, that both of the integrals

$$\int_0^1 \left(\int_{y^2}^1 y \cdot e^{-x^2}\,dx \right) dy \quad \text{and} \quad \int_0^1 \left(\int_0^{\sqrt{x}} y \cdot e^{-x^2}\,dy \right) dx$$

give us the volume of the set

$$\{(x,y,z) : y \ge 0,\ y^2 \le x \le 1,\ 0 \le z \le y \cdot e^{-x^2}\}$$

(taking for granted that it is measurable); therefore, their values agree. Compute this common value.

16.17. Using the idea behind the previous question, compute the following integrals:

(a) $\int_0^1 \left(\int_x^1 \frac{x \sin y}{y}\,dy \right) dx$;
(b) $\int_0^1 \left(\int_{\sqrt{y}}^1 \sqrt{1+x^3}\,dx \right) dy$;
(c) $\int_0^1 \left(\int_{y^{2/3}}^1 y \cos x^2\,dx \right) dy$.

16.18. Let f be nonnegative and integrable on $[a,b]$. Prove that the volume of the solid of revolution

$$\{(x,y,z)\colon a \le x \le b,\ y^2+z^2 \le f^2(x)\}$$

is equal to the area of the set

$$A = \{(x,y)\colon a \le x \le b,\ 0 \le y \le f(x)\}$$

times the circumference of the circle described by rotating the center of mass of A. (This is sometimes called **Guldin's**[2] **second Theorem.**)

[2] Paul Guldin (1577–1643), Swiss mathematician.

16.4 Computing Arc Length

We defined the arc length and rectifiability of graphs of functions in Definition 10.78. Let the function $f:[a,b]\to\mathbb{R}$ be continuously differentiable. Then by Theorem 13.41, the graph of f is rectifiable. Moreover, if $s(x)$ denotes the arc length of the graph of the function over the interval $[a,x]$, then s is differentiable, and $s'(x) = \sqrt{1+(f'(x))^2}$ for all $x\in[a,b]$. Since the arc length of the graph of f is $s(b) = s(b)-s(a)$, the fundamental theorem of calculus gives us the following theorem.

Theorem 16.13. *If the function $f:[a,b]\to\mathbb{R}$ is continuously differentiable, then the arc length of the graph is $\int_a^b \sqrt{1+(f'(x))^2}\,dx$.*

Example 16.14 (The Arc Length of a Parabola). Let s denote the arc length of the function x^2 over the interval $[0,a]$. By the previous theorem, $s=\int_0^a \sqrt{1+4x^2}\,dx$. Since by (15.23), $\int\sqrt{x^2+1}\,dx = \frac{1}{2}\cdot x\cdot\sqrt{x^2+1}-\frac{1}{2}\log(\sqrt{x^2+1}-x)+c$, with the help of a linear substitution and the fundamental theorem of calculus, we obtain

$$s=\left[\frac{1}{2}\cdot x\cdot\sqrt{4x^2+1}-\frac{1}{4}\log(\sqrt{4x^2+1}-2x)\right]_0^a = \\ =\frac{a}{2}\cdot\sqrt{4a^2+1}-\frac{1}{4}\log(\sqrt{4a^2+1}-2a).$$

If we want to compute the arc length of curves more general than graphs of functions, we first need to clarify the notion of a curve. We can think of a curve as the path of a moving particle, and we can determine that path by defining the particle's position vector at every time t. Thus the movement of a particle is defined by a function that assigns a vector in whatever space the particle is moving to each point of the time interval $[a,b]$. If the particle is moving in d-dimensional space, then this means that we assign a d-dimensional vector to each $t\in[a,b]$.

We will accept this idea as the definition of a curve, that is, a **curve** is a map of the form $g\colon [a,b]\to\mathbb{R}^d$. If $d=2$, then we are talking about **planar curves**, and if $d=3$, then we are talking about **space curves**.

Consider a curve $g\colon [a,b]\to\mathbb{R}^d$, and let the coordinates of the vector $g(t)$ be denoted by $g_1(t),\dots,g_d(t)$ for all $t\in[a,b]$. This defines a function $g_i\colon [a,b]\to\mathbb{R}$ for each $i=1,\dots,d$, which is called the ith **coordinate function** of the curve g. Thus the curve is the map

$$t\mapsto (g_1(t),\dots,g_d(t)) \qquad (t\in[a,b]).$$

We say that a curve g is **continuous, differentiable, continuously differentiable, Lipschitz,** etc. if each coordinate function of g has the corresponding property.

We emphasize that when we talk about curves, we are talking about the mapping itself, and not the path the curve traces (that is, its image). More simply: a curve is a map, not a set in $\mathbb{R}^d$. If the set H agrees with the image of the curve $g\colon [a,b]\to\mathbb{R}^d$, that is, $H=g([a,b])$, then we say that g is a **parameterization** of H. A set can have several parameterizations. Let us see some examples.

The **segment** determined by the points $u, v \in \mathbb{R}^d$ is the set

$$[u,v] = \{u + t \cdot (v - u) : t \in [0,1]\}.$$

The segment $[u,v]$ is *not* a curve. On the other hand, the map $g\colon [0,1] \to \mathbb{R}^d$, which is defined by $g(t) = u + t \cdot (v - u)$ for all $t \in [0,1]$, is a curve that traces out the points of the segment $[u,v]$, that is, g is a parameterization of the segment $[u,v]$. Another curve is the map $h\colon [0,1] \to \mathbb{R}^d$, which is defined as $h(t) = u + t^2 \cdot (v - u)$ $(t \in [0,1])$. The curve h is also a parameterization of the segment $[u,v]$. Nevertheless, the curves g and h are different, since $g(1/2) = (u+v)/2$, while $h(1/2) = (3u+v)/4$.

Consider now the map $g\colon [0,2\pi] \to \mathbb{R}^2$, for which

$$g(t) = (\cos t, \sin t) \qquad (t \in [0,2\pi]).$$

The planar curve g defines the path of a particle that traces out the unit circle C centered at the origin, that is, g is a parameterization of the circle C. The same holds for the curve $g_1\colon [0,2\pi] \to \mathbb{R}^2$, where

$$g_1(t) = (\cos 2t, \sin 2t) \qquad (t \in [0,2\pi]).$$

The curve g_1 also traces out C, but "twice over." Clearly, $g \neq g_1$. Since the length of g is 2π, while the length of g_1 is 4π, this example shows that arc length should be assigned to the curve (that is, the map), and not to the image of the curve.

Note that the graph of any function $f\colon [a,b] \to \mathbb{R}$ can be parameterized with the planar curve $t \mapsto (t, f(t)) \in \mathbb{R}^2$ $(t \in [a,b])$.

We define the arc length of curves similarly to how we defined the arc length of graphs of functions. A broken or polygonal line is a set that is the union of connected segments. If $a_0, \ldots, a_n$ are arbitrary points of the space $\mathbb{R}^d$, then the polygonal line connecting the points a_i (in this order) consists of the segments $[a_0,a_1]$, $[a_1,a_2]$, $\ldots$, $[a_{n-1},a_n]$. The length of a polygonal line is the sum of the lengths of the segments that constitute it, that is, $|a_1 - a_0| + |a_2 - a_1| + \cdots + |a_n - a_{n-1}|$ (Figure 16.6).

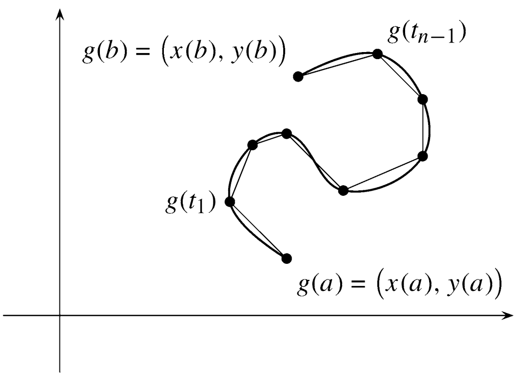

Fig. 16.6

Definition 16.15. An *inscribed polygonal path* of the curve $g\colon [a,b] \to \mathbb{R}^d$ is a polygonal line connecting the points $g(t_0), g(t_1), \ldots, g(t_n)$, where $a = t_0 < t_1 < \ldots < t_n = b$ is a partition of the interval $[a,b]$. The *arc length* of the curve g is the least upper bound of the set of lengths of inscribed polygonal paths of g (which can be infinite). We denote the arc length of the curve g by $s(g)$. Thus

$$s(g) = \sup\left\{\sum_{i=1}^{n} |g(t_i) - g(t_{i-1})| : a = t_0 < t_1 < \cdots < t_n = b,\ n = 1,2,\ldots\right\}.$$

We say that a curve g is *rectifiable* if $s(g) < \infty$.

Not every curve is rectifiable. It is clear that if the image of the curve $g\colon [a,b] \to \mathbb{R}^d$ is unbounded, then there exist arbitrarily long inscribed polygonal paths of g, and so $s(g) = \infty$. The following example shows that it is not enough for a curve to be continuous or even differentiable in order for it to be rectifiable.

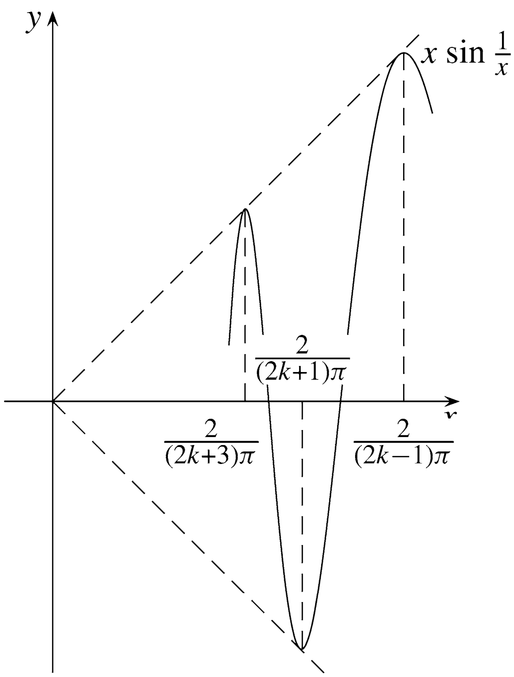

Fig. 16.7

Example 16.16. Consider the planar curve $g(t) = (t, f(t))$ $(t \in [0,1])$, where

$$f(t) = \begin{cases} t \cdot \sin(1/t) & \text{if } t \neq 0, \\ 0 & \text{if } t = 0. \end{cases}$$

(The curve g parameterizes the graph of f on the interval $[0,1]$) (Figure 16.7). We show that the curve g is not rectifiable. Let us compute the length of the inscribed polygonal path of g corresponding to the partition F_n, where F_n consists of the points 0, 1, and

$$x_i = \frac{2}{(2i-1)\pi} \qquad (i = 1, \ldots, n).$$

(We happen to have listed the inner points x_i in decreasing order.) Since

$$f(x_i) = (-1)^{i+1}\frac{2}{(2i-1)\pi} \qquad (i = 1, \ldots, n)$$

if $1 \le i \le n-1$, the length of the segment $[g(x_i), g(x_{i+1})]$ is

$$|g(x_{i+1}) - g(x_i)| \ge |f(x_{i+1}) - f(x_i)| = \frac{2}{\pi}\left(\frac{1}{2i-1} + \frac{1}{2i+1}\right) > \frac{2}{\pi} \cdot \frac{1}{i+1}.$$

Thus the length of the inscribed polygonal path is at least

$$\sum_{i=1}^{n} |g(x_i) - g(x_{i-1})| > \frac{2}{\pi} \sum_{i=1}^{n} \frac{1}{i}.$$

Since $\sum_{i=1}^{n}(1/i) \to \infty$ if $n \to \infty$ (see Theorem 7.8), the set of lengths of inscribed polygons is indeed unbounded. Thus the curve g is not rectifiable, even though the functions t and $f(t)$ defining g are continuous.

Now consider the planar curve $h(t) = (t^2, f(t^2))$ $(t \in [0,1])$, where f is the function above. The curve h also parameterizes the graph of f and has the same inscribed polygonal paths as g. Thus h is not rectifiable either, even though the functions t^2 and $f(t^2)$ are both differentiable on $[0,1]$ (see Example 13.46).

Remark 16.17. As we have seen, the arc length of the curve $g\colon [a,b] \to \mathbb{R}^d$ depends on the map g, since the inscribed polygonal paths already depend on g. This means that we cannot generally speak of the arc length of a set H—even when H is the image of a curve, or in other words, even if it is parameterizable. This is because H can have multiple parameterizations, and the arc lengths of these different parameterizations could be different. So for example, the curves $f(t) = (t^2, 0)$ $(t \in [0,1])$ and $g(t) = (t^2, 0)$ $(t \in [-1,1])$ parameterize the same set (namely the interval $[0,1]$ of the x-axis), but their arc lengths are different.

In some important cases, however, we do give sets H a unique arc length. We call sets that have a bijective and continuous parameterization **simple curves**. Some examples of simple curves are the segments, arcs of a circle, and the graph of any continuous function.

Theorem 16.18. *If $H \subset \mathbb{R}^d$ is a simple curve, then every bijective and continuous parameterization of H defines the same arc length.*

Proof. We outline a sketch of the proof. Let $\beta\colon [a,b] \to H$ and $\gamma\colon [c,d] \to H$ be bijective parameterizations. Then the function $h = \gamma^{-1} \circ \beta$ maps the interval $[a,b]$ onto the interval $[c,d]$ bijectively, and one can show that h is also continuous. (This step—which belongs to multivariable calculus—is not detailed here.) It follows that in this case, h is strictly monotone (see Exercise 10.54). Thus $\beta = \gamma \circ h$, where h is a strictly monotone bijection of $[a,b]$ onto $[c,d]$.

This property ensures that the curves $\beta\colon [a,b] \to H$ and $\gamma\colon [c,d] \to H$ have the same inscribed polygonal paths. If $F : a = t_0 < t_1 < \ldots < t_n = b$ is a partition of the interval $[a,b]$, then either $c = h(a) < h(t_1) < \cdots < h(t_n) = d$ or $c = h(t_n) < h(t_{n-1}) < \cdots < h(t_0) = d$, depending on whether h is increasing or decreasing. One of the two will give a partition of the interval $[c,d]$ that will give the same inscribed polygonal path as given by F under the map $\beta = \gamma \circ h$. That is, every inscribed polygonal path of $\beta\colon [a,b] \to H$ is an inscribed polygonal path of $\gamma\colon [c,d] \to H$. In the same way, every inscribed polygonal path of $\gamma\colon [c,d] \to H$ is also an inscribed polygonal path of $\beta\colon [a,b] \to H$. It then follows that the arc lengths of the curves $\beta\colon [a,b] \to H$ and $\gamma\colon [c,d] \to H$ are the suprema of the same set, so the arc lengths agree. □

According to what we said above, we can talk about arc lengths of simple curves: by this, we mean the arc length of a parameterization that is bijective and continuous.

The following theorem gives us simple sufficient conditions for a curve to be rectifiable.

Theorem 16.19. *Consider a curve $g\colon [a,b] \to \mathbb{R}^d$.*

(i) *If the curve is Lipschitz, then it is rectifiable.*
(ii) *If the curve g is differentiable, and the derivatives of the coordinate functions of g are bounded on $[a,b]$, then g is rectifiable.*
(iii) *If the curve is continuously differentiable, then it is rectifiable.*

Proof. (i) Let the curve g be Lipschitz, and suppose that $|g_i(x) - g_i(y)| \le K \cdot |x-y|$ for all $x \in [a,b]$ and $i = 1,\ldots,d$. Then using (16.3), we get that $|g(x) - g(y)| \le Kd \cdot |x-y|$ for all $x,y \in [a,b]$. Then it immediately follows that every inscribed polygonal path of g has length at most $Kd \cdot (b-a)$, so g is rectifiable.
(ii) By the mean value theorem, if the function $g_i\colon [a,b] \to \mathbb{R}$ is differentiable on the interval $[a,b]$ and its derivative is bounded, then g_i is Lipschitz. Thus the rectifiability of g follows from (i).

Statement (iii) is clear from (ii), since continuous functions on the interval $[a,b]$ are necessarily bounded (Theorem 10.52). □

Theorem 16.20. *Suppose that the curve $g\colon [a,b] \to \mathbb{R}^d$ is differentiable, and the derivatives of the coordinate functions of g are integrable on $[a,b]$. Then g is rectifiable, and*

$$s(g) = \int_a^b \sqrt{\left(g_1'(t)\right)^2 + \cdots + \left(g_d'(t)\right)^2}\, dt. \tag{16.14}$$

We give a proof of this theorem in the appendix of the chapter.

Remark 16.21. Let $f\colon [a,b] \to \mathbb{R}$, and apply the above theorem to the curve given by $g(t) = (t, f(t))$ $(t \in [a,b])$. We get that in Theorem 16.13, instead of having to assume the continuous differentiability of the function f, it is enough to assume that f is differentiable and that f' is integrable on $[a,b]$.

Remark 16.22. Suppose that the curve $g\colon [a,b] \to \mathbb{R}^d$ is differentiable. Let the coordinate functions of g be $g_1,\ldots,g_d$. If t_0 and t are distinct points of the interval $[a,b]$, then

$$\frac{g(t) - g(t_0)}{t - t_0} = \left(\frac{g_1(t) - g_1(t_0)}{t - t_0}, \ldots, \frac{g_d(t) - g_d(t_0)}{t - t_0} \right).$$

Here if we let t approach t_0, the jth coordinate of the right-hand side tends to $g_j'(t_0)$. Thus it is reasonable to call the vector $(g_1'(t_0),\ldots,g_d'(t_0))$ the **derivative** of the curve g at the point t_0. We denote it by $g'(t_0)$. With this notation, (16.14) takes on the form

$$s(g) = \int_a^b |g'(t)|\,dt. \tag{16.15}$$

The physical meaning of the derivative g' is the velocity vector of a particle moving along the curve g. Clearly, the displacement of the particle between the times t_0 and t is $g(t) - g(t_0)$. The vector $(g(t) - g(t_0))/(t - t_0)$ describes the average displacement of the particle during the time interval $[t_0, t]$. As $t \to t_0$, this average tends to the velocity vector of the particle. Since $(g(t) - g(t_0))/(t - t_0)$ tends to $g'(t_0)$ in each coordinate, $g'(t_0)$ is exactly the velocity vector.

On the other hand, the value $|(g(t) - g(t_0))/(t - t_0)|$ denotes the average magnitude of the displacement of the moving particle during the time interval $[t_0, t]$. The limit of this as $t \to t_0$ is the instantaneous velocity of the particle. Thus the absolute value of the velocity vector, $|g'(t_0)|$, is the instantaneous velocity. Thus the physical interpretation of (16.15) is that during movement (along a curve) of a particle, the distance traversed by the point is equal to the integral of its instantaneous velocity. We already saw this for motions along a straight path: this was the physical statement of the fundamental theorem of calculus (Remark 15.2). Thus we can consider (16.14), that is, (16.15), to be an analogue of the fundamental theorem of calculus for curves.

Example 16.23. Consider a circle of radius a that is rolling along the x-axis. The path traced out by a point P on the rolling circle is called a **cycloid**. Suppose that the point P was at the origin at the start of the movement. The circle rolls along the x-axis (without slipping), meaning that at each moment, the length of the circular arc between the point A of the circle touching the axis and P is equal to the length of the segment OA.

Let t denote the angle between the rays CA and CP, where C denotes the center of the circle. Then the length of the line segment AP is at, that is, $\overline{OA} = at$. In the triangle CPR seen in the figure, $\overline{PR} = a \sin t$ and $\overline{CR} = -a \cos t$, so the coordinates of the point P are $(at - a\sin t, a - a\cos t)$. After a full revolution of the circle, the point P is touching the axis again. Thus the parameterization of the cycloid is

$$g(t) = (at - a\sin t, a - a\cos t) \qquad (t \in [0, 2\pi]).$$

Since $g'(t) = (a - a\cos t, a\sin t)$ and

$$|g'(t)| = \sqrt{(a - a\cos t)^2 + (a\sin t)^2} = a \cdot \sqrt{2 - 2\cos t} =$$
$$= a \cdot \sqrt{4\sin^2 \frac{t}{2}} = 2a \sin \frac{t}{2},$$

by (16.15), the arc length of a cycloid is

$$\int_0^{2\pi} 2a \sin \frac{t}{2}\,dt = \left[-4a\cos\frac{t}{2}\right]_0^{2\pi} = 8a.$$

Thus the arc length of a cycloid is eight times the radius of the rolling circle (Figure 16.8).

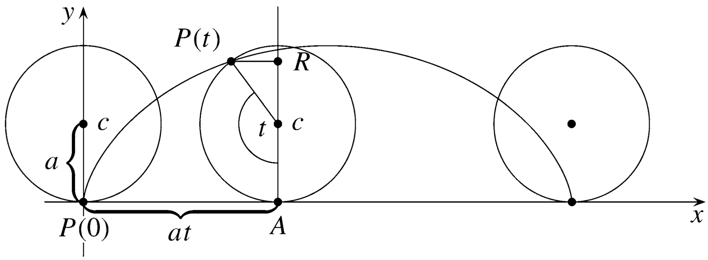

Fig. 16.8

Exercises

16.19. Construct (a) a segment; (b) the boundary of a square as the image of both a differentiable and a nondifferentiable curve.

16.20. Compute the arc lengths of the graphs of the following functions:

(a) $f(x) = x^{3/2} \quad (0 \le x \le 4)$;
(b) $f(x) = \log(1-x^2)$ $(0 \le x \le a < 1)$;
(c) $f(x) = \log\cos x$ $(0 \le x \le a)$;
(d) $f(x) = \log \frac{e^x+1}{e^x-1}$ $(a \le x \le b)$.

16.21. Let the arc length of the graph $f : [a,b] \to \mathbb{R}$ be denoted by L, and the arc length of $g(t) = (t, f(t))$ $(t \in [a,b])$ by S. Prove that $L \le S \le L + (b-a)$. Show that the graph of f is rectifiable if and only if the curve $g(t)$ $(t \in [a,b])$ is rectifiable.

16.22. In the following exercises, by the planar curve with parameterization $x = x(t)$, $y = y(t)$ $(t \in [a,b])$ we mean the curve $g(t) = (x(t), y(t))$ $(t \in [a,b])$. Compute the arc lengths of the following planar curves:

(a) $x = a \cdot \cos^3 t,\ y = a \cdot \sin^3 t$ $(0 \le t \le 2\pi)$ (**astroid**);
(b) $x = a \cdot \cos^4 t,\ y = a \cdot \sin^4 t$ $(0 \le t \le \pi/2)$;
(c) $x = e^t(\cos t + \sin t),\ y = e^t(\cos t - \sin t)$ $(0 \le t \le a)$;
(d) $x = t - \operatorname{th} t,\ y = 1/\operatorname{ch} t$ $(t \in [0,1])$;
(e) $x = \operatorname{ctg} t, y = 1/(2\sin^2 t)$ $(\pi/4 \le t \le \pi/2)$.

16.23. Let $n > 0$, and consider the curve $g = (\cos(t^n),\ \sin(t^n))$ $\left(t \in [0, \sqrt[n]{2\pi}]\right)$ (which is a parameterization of the unit circle). Check that the arc length of this curve is 2π (independent of the value of n).

16.24. For a given $b, d > 0$, compute the arc length of the catenary, that is, the graph of the function $f(x) = b^{-1} \cdot \operatorname{ch}(bx) \quad (0 \le x \le d)$.

16.25. Let a and b be fixed positive numbers. For which c will the arc length of the ellipse with semiaxes a and b be equal to the arc length of the function $c \cdot \sin x$ over the interval $[0, \pi]$?

16.26. How large can the arc length of the graph of a (a) monotone; (b) monotone and continuous; (c) strictly monotone; (d) strictly monotone and continuous function $f\colon [0,1] \to [0,1]$ be?

16.27. Let f be the Riemann function in $[0,1]$. For which $c > 0$ will the graph of f^c be rectifiable? (∗ H S)

16.28. Prove that if $f\colon [a,b] \to \mathbb{R}$ is differentiable and f' is bounded on $[a,b]$, then

(a) the graph of f is rectifiable, and
(b) the arc length of the graph of f lies between

$$\underline{\int_a^b} \sqrt{1 + \left(f'(x)\right)^2}\, dx \quad \text{and} \quad \overline{\int_a^b} \sqrt{1 + \left(f'(x)\right)^2}\, dx.$$

16.29. Prove that if the curve $g\colon [a,b] \to \mathbb{R}^2$ is continuous and rectifiable, then for every $\varepsilon > 0$, the image $g([a,b])$ can be covered by finitely many disks whose total area is less than ε. (∗ H S)

The center of mass of a curve. Imagine a curve $g\colon [a,b] \to \mathbb{R}^d$ made up of some homogeneous material. Then the weight of every arc of g is ρ times the length of that arc, where ρ is some constant (density). Consider a partition $F : a = t_0 < t_1 < \cdots < t_n = b$, and let $c_i \in [t_{i-1}, t_i]$ be arbitrary inner points. If the curve is continuously differentiable and the partition is fine enough, then the length of the arc $g([t_{i-1}, t_i])$ is close to the length of the segment $[g(t_{i-1}), g(t_i)]$, so the weight of the arc is close to $\rho \cdot |g(t_i) - g(t_{i-1})|$. Thus if for every i, we concentrate a weight $\rho \cdot |g(t_i) - g(t_{i-1})|$ at the point $g(c_i)$, then the weight distribution of the points of weights we get in this way will be close to the weight distribution of the curve itself. We can expect the center of mass of the collection of these points to be close to the center of mass of the curve.

The center of mass of the system of points above is the point $\frac{1}{L_F} \cdot \sum_{i=1}^{n} |g(t_i) - g(t_{i-1})| \cdot g(c_i)$, where $L_F = \sum_{i=1}^{n} |g(t_i) - g(t_{i-1})|$. If the partition is fine enough, then L_F is close to the arc length L.

Let the coordinate functions of g be $g_1, \ldots, g_d$. Then the length $|g(t_i) - g(t_{i-1})|$ is well approximated by the value $\sqrt{g_1'(c_i)^2 + \cdots + g_d'(c_i)^2} \cdot (t_i - t_{i-1})$, so the jth coordinate of the point $\sum_{i=1}^{n} |g(t_i) - g(t_{i-1})| \cdot g(c_i)$ will be close to the sum

$$\sum_{i=1}^{n} \sqrt{g_1'(c_i)^2 + \cdots + g_d'(c_i)^2} \cdot g_j(c_i) \cdot (t_i - t_{i-1}).$$

If the partition is fine enough, then this sum is close to the integral

$$s_j = \int_a^b \sqrt{g_1'(t)^2 + \cdots + g_d'(t)^2} \cdot g_j(t)\, dt. \tag{16.16}$$

This motivates the following definition.

Definition 16.24. If the curve $g\colon [a,b] \to \mathbb{R}^p$ is differentiable and the derivatives of the coordinate functions of g are integrable on $[a,b]$, then the **center of mass** of g is the point $(s_1/L, \dots, s_d/L)$, where L is the arc length of the curve, and the s_j are defined by (16.16) for all $j = 1, \dots, d$.

Exercise

16.30. Compute the center of mass of the following curves:

(a) $g(t) = (t, t^2)$ $(t \in [0,1])$;
(b) $g(t) = (t, \sin t)$ $(y \in [0,\pi])$;
(c) $g(t) = (a \cdot (1+\cos t)\cos t, a \cdot (1+\cos t)\sin t)$ $(0 \le t \le 2\pi)$, where $a > 0$ is constant (cardioid).

16.5 Polar Coordinates

The **polar coordinates** of a point P distinct from the origin are given by the ordered pair (r,φ), where r denotes the distance of P from the origin, and φ denotes the angle between $\overrightarrow{OP}$ and the positive direction of the x-axis (Figure 16.9). From the figure, it is clear that if the polar coordinates of P are (r,φ), then the usual (Cartesian) coordinates are $(r\cos\varphi, r\sin\varphi)$. The polar coordinates of the origin are given by $(0,\varphi)$, where φ can be arbitrary.

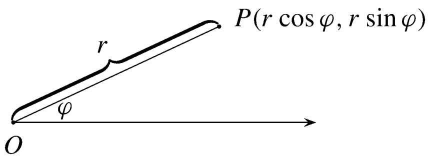

Fig. 16.9

If $[\alpha,\beta] \subset [0,2\pi)$, then every function $r\colon [\alpha,\beta] \to [0,\infty)$ describes a curve, the collection of points $(r(\varphi),\varphi)$ for $\varphi \in [\alpha,\beta]$. Since the Cartesian coordinates of the point $(r(\varphi),\varphi)$ are $(r(\varphi)\cos\varphi, r(\varphi)\sin\varphi)$, using our old notation we are actually talking about the curve

$$g(t) = (r(t)\cos t, r(t)\sin t) \qquad (t \in [\alpha,\beta]). \tag{16.17}$$

Definition 16.25. The function $r\colon [\alpha,\beta] \to [0,\infty)$ is called the *polar coordinate form* of the curve (16.17).

In this definition, we do not assume $[\alpha,\beta]$ to be part of the interval $[0,2\pi]$. This is justified in that for arbitrary $t \in \mathbb{R}$, if $r > 0$, then the polar coordinate form of the point $P = (r\cos t, r\sin t)$ is $(r, t-2k\pi)$, where k is an integer such that $0 \le t - 2k\pi < 2\pi$. In other words, t is equal to *one* of the angles between $\overrightarrow{OP}$ and the positive half of the x-axis, so in this more general sense, we can say that the points $(r(\varphi),\varphi)$ given in polar coordinate form give us the curve (16.17).

Examples 16.26. **1.** If $[\alpha,\beta] \subset [0,2\pi)$, then the function $r \equiv a$ ($\varphi \in [\alpha,\beta]$), where $a > 0$ is constant, is the polar coordinate form of a subarc of the circle of radius a centered at the origin.

2. The function

$$r(\varphi) = a\cdot\varphi \qquad (\varphi \in [0,\beta]) \tag{16.18}$$

describes what is called the **Archimedean spiral**. The Archimedean spiral is the path of a particle that moves uniformly along a ray starting from the origin while the ray rotates uniformly about the origin (Figure 16.10).

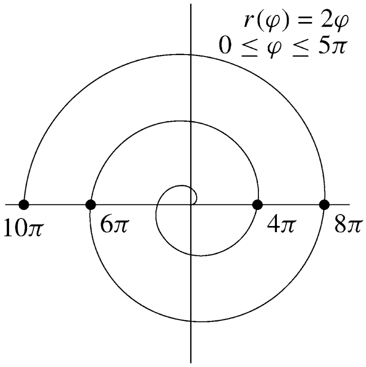

Fig. 16.10

Theorem 16.27. *Suppose that the function $r\colon [\alpha,\beta] \to [0,\infty)$ is differentiable, and its derivative is integrable on $[\alpha,\beta]$. Then the curve given by r in its polar coordinate form is rectifiable, and its arc length is*

$$\int_\alpha^\beta \sqrt{(r')^2 + r^2}\, d\varphi. \tag{16.19}$$

Proof. The curve g given by (16.17) is differentiable, and the derivatives of the coordinate functions, $r'\cos t - r\sin t$ and $r'\sin t + r\cos t$, are integrable on $[\alpha,\beta]$. Thus by Theorem 16.20, the curve is rectifiable. Since

$$|g'(t)| = \sqrt{(r'\cos t - r\sin t)^2 + (r'\sin t + r\cos t)^2} = \sqrt{(r')^2 + r^2},$$

by (16.15), we get (16.19). □

Example 16.28. The arc length of the Archimedean spiral given by (16.18) is

$$\int_0^\beta \sqrt{a^2t^2 + a^2}\, dt = a\cdot\left[\frac{1}{2}\cdot\beta\cdot\sqrt{\beta^2+1} - \frac{1}{2}\log(\sqrt{\beta^2+1} - \beta)\right].$$

Consider a curve given in its polar coordinate form $r\colon [\alpha,\beta] \to [0,\infty)$. The union of the segments connecting every point of the curve to the origin is called a **sectorlike region**. By the definition of polar coordinates, the region in question is exactly the set

$$A = \{(r\cos\varphi,\ r\sin\varphi) : 0 \le r \le r(\varphi),\ \alpha \le \varphi \le \beta\}. \tag{16.20}$$

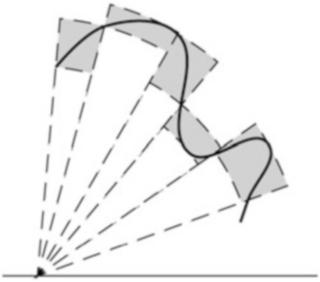

Fig. 16.11

Theorem 16.29. *Let* $0 \le \alpha < \beta \le 2\pi$. *If the function* f *is nonnegative and integrable on* $[\alpha,\beta]$, *then the sectorlike region given in (16.20) is measurable, and its area is* $\frac{1}{2}\int_\alpha^\beta r^2(\varphi)\,d\varphi$.

Proof. To prove this theorem, we use the fact that the circular sector with radius r and central angle δ is measurable, and has area $r^2\delta/2$ (see Exercise 16.8). Moreover, we use that if $A \subset \bigcup_{i=1}^n A_i$, then $\overline{m}(A) \le \sum_{i=1}^n \overline{m}(A_i)$, and if $A \supset \bigcup_{i=1}^n B_i$, where the sets B_i are nonoverlapping, then $\underline{m}(A) \ge \sum_{i=1}^n \underline{m}(B_i)$. These follow easily from the definitions of the inner and outer measure.

Consider a partition $F : \alpha = t_0 < t_1 < \cdots < t_n = \beta$ of the interval $[\alpha,\beta]$. If $m_i = \inf\{r(t) : t \in [t_{i-1},t_i]\}$ and $M_i = \sup\{r(t) : t \in [t_{i-1},t_i]\}$, then the set of points with polar coordinates (r,φ) ($\varphi \in [t_{i-1},t_i]$, $0 \le r \le m_i$) is a circular sector B_i that is contained in A (Figure 16.11). Since the circular sectors $B_1,\ldots,B_n$ are nonoverlapping, the inner area of A is at least as large as the sum of the areas of these sectors, which is $(1/2)\cdot\sum_{i=1}^n m_i^2(t_i - t_{i-1}) = (1/2)\cdot s_F(r^2)$.

Similarly, the set of points with polar coordinates (r,φ) ($\varphi \in [t_{i-1},t_i]$, $0 \le r \le M_i$) is a circular sector A_i, and the sectors $A_1,\ldots,A_n$ together cover A. Thus the outer area of A must be less than or equal to the sum of the areas of these sectors, which is $\frac{1}{2}\cdot\sum_{i=1}^n M_i^2(t_i - t_{i-1}) = \frac{1}{2}\cdot S_F(r^2)$. Thus

$$\tfrac{1}{2}s_F(r^2) \le \underline{m}(A) \le \overline{m}(A) \le \tfrac{1}{2}S_F(r^2)$$

for every partition F. Since r^2 is integrable,

$$\sup_F s_F(r^2) = \inf_F S_F(r^2) = \int_\alpha^\beta r^2(\varphi)\,d\varphi,$$

and

$$\frac{1}{2}\int_\alpha^\beta r^2(\varphi)\,d\varphi \le \underline{m}(A) \le \overline{m}(A) \le \frac{1}{2}\int_\alpha^\beta r^2(\varphi)\,d\varphi.$$

This shows that A is measurable, and its area is equal to half of the integral. □

Exercises

16.31. Compute the arc lengths of the following curves given in polar coordinate form:

(a) $r = a \cdot (1 + \cos\varphi)$ $(0 \le \varphi \le 2\pi)$, where $a > 0$ is constant (cardioid);
(b) $r = a/\varphi \quad (\pi/2 \le \varphi \le 2\pi)$, where $a > 0$ is constant;
(c) $r = a \cdot e^{c\cdot\varphi}$ $(0 \le \varphi \le \alpha)$, where $a > 0$, $c \in \mathbb{R}$, and $\alpha > 0$ are constants;
(d) $r = \frac{p}{1+\cos\varphi}$ $(0 \le \varphi \le \pi/2)$, where $p > 0$ is constant; what is this curve? (H)
(e) $r = \frac{p}{1-\cos\varphi}$ $(\pi/2 \le \varphi \le \pi)$, where $p > 0$ is constant; what is this curve? (H)

16.32. Let $a > 0$ be constant. The set of planar points whose distance from $(-a,0)$ times its distance from $(a,0)$ is equal to a^2 is called a *lemniscate* (Figure 16.12). Show that $r^2 = 2a^2 \cdot \cos 2\varphi$ $(-\pi/4 \le \varphi \le \pi/4$ or $3\pi/4 \le \varphi \le 5\pi/4)$ is a parameterization of the lemniscate in polar coordinate form. Compute the area of the region bounded by the lemniscate.

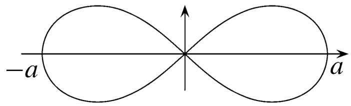

Fig. 16.12

16.33. Compute the area of the set of points satisfying $r^2 + \varphi^2 \le 1$.

16.34. Compute the area of the region bounded by the curve $r = \sin\varphi + e^{\varphi}$ $(0 \le \varphi \le \pi)$ given in polar coordinate form, and the segment $[-e^{\pi}, 1]$ of the x-axis.

16.35. The curve $r = a \cdot \varphi$ $(0 \le \varphi \le \pi/4)$ given in polar coordinate form is the graph of a function f.

(a) Compute the area of the region under the graph of f.
(b) Revolve the region under this graph about the x-axis. Compute the volume of the solid of revolution we obtain in this way. (H)

16.36. The cycloid with parameter a over $[0, 2a\pi]$ is the graph of a function g.

(a) Compute the area of the region under the graph of g.
(b) Revolve the region under this graph about the x-axis. Compute the volume of the solid of revolution we obtain in this way. (H)

16.37. Express the curve satisfying the conditions

(a) $x^4 + y^4 = a^2(x^2 + y^2)$; and
(b) $x^4 + y^4 = a \cdot x^2 y$

in polar coordinate form, and compute the area of the enclosed region.

16.6 The Surface Area of a Surface of Revolution

Determining the surface area of surfaces is a much harder task than finding the area of planar regions or the volume of solids; the definition of surface area itself already causes difficulties. To define surface area, the method used to define area—bounding the value from above and below—does not work. We could try to copy the method of defining arc length and use the known surface area of inscribed polygonal surfaces, but this already fails in the simplest cases: one can show that the inscribed polygonal surfaces of a right circular cylinder can have arbitrarily large surface area. To precisely define surface area, we need the help of differential geometry, or at least multivariable differentiation and integration, which we do not yet have access to.

Determining the surface area of a surface of revolution is a simpler task. Let $f\colon [a,b] \to \mathbb{R}$ be a nonnegative function, and let A^f denote the set we get by rotating graph f about the x-axis. It is an intuitive assumption that the surface area of A^f is well approximated by the surface area of the rotation of an inscribed polygonal path about the x-axis. Before we inspect this assumption in more detail, let us compute the surface area of the rotated inscribed polygonal paths.

Let $F : a = x_0 < x_1 < \cdots < x_n = b$ be an arbitrary partition. Rotating the inscribed polygonal paths corresponding to F about x gives us a set P^F, which consists of n parts: the ith part, which we will denote by P_i^F, is the rotated segment over the interval $[x_{i-1}, x_i]$ (Figure 16.13). We can see in the figure that the set P_i^F is the side of a right conical frustum with height $x_i - x_{i-1}$, and radii of bases $f(x_i)$ and $f(x_{i-1})$.

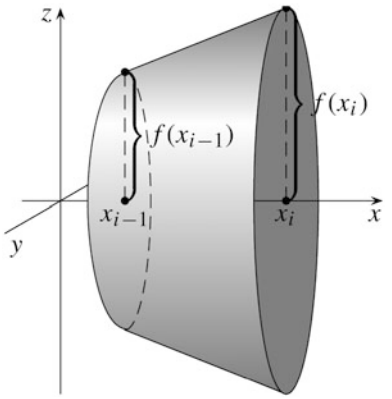

Fig. 16.13

It is intuitively clear that if we unroll the side of a right conical frustum, then the area of the region we get is equal to the surface area of the side of the frustum. Since the unrolled side is the difference of two circular sectors, the area of this can be computed easily, using that the area of the sector is half the radius times the arc length.

In the end, we get that if the height of the frustum is m, and the bases have radii r and R, then the lateral surface area is $\pi(R + r)\sqrt{(R-r)^2 + m^2}$ (see Exercise 16.38). Thus the surface area of the side P_i^F is

$$\pi \cdot (f(x_i) + f(x_{i-1})) \cdot h_i,$$

and the surface area of the set P^F is

$$\Phi_F = \pi \cdot \sum_{i=1}^{n} (f(x_i) + f(x_{i-1})) \cdot h_i, \tag{16.21}$$

where $h_i = \sqrt{(f(x_i) - f(x_{i-1}))^2 + (x_i - x_{i-1})^2}$. Here we should note that the sum $\sum_{i=1}^n h_i$ is equal to the length of the inscribed polygonal path (corresponding to F). Therefore, $\sum_{i=1}^n h_i \le L$, where L denotes the arc length of graph f.

Now let us return to figuring out in what sense the value Φ_F approximates the surface area of the set A^f. Since the length of the graph of f is equal to the supremum of the arc lengths of the inscribed polygonal paths, our first thought might be that the surface area of A^f needs to be equal to the supremum of the values Φ_F. However, this is already not the case with simple functions. Consider, for example, the function $|x|$ on the interval $[-1,1]$. In this case, the set A^f is the union of two sides of cones, and its surface area is $2 \cdot (2\pi \cdot \sqrt{2}/2) = 2\sqrt{2}\pi$ (Figure 16.14). But if the partition F consists only of the points -1 and 1, then P^F is a cylinder whose surface area is $2\pi \cdot 2 = 4\pi$, which is larger than the surface area of A^f.

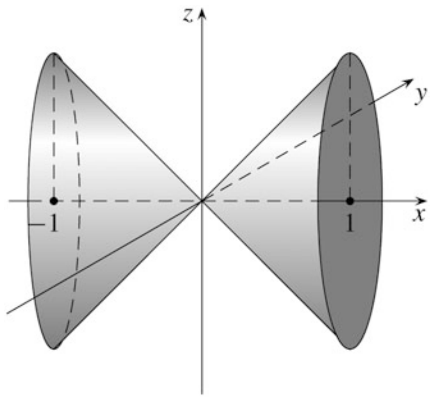

Fig. 16.14

By the example above, we can rule out being able to define the surface area of A^f as the supremum of the set of values Φ_F. However, the example hints at the correct definition, since an arbitrary partition F of $[-1,1]$ makes P^F equal either to A^f (if 0 is a base point of F) or to the union of the sides of three frustums. If the mesh of the partition is small, then the surface area of the middle frustum will be small, and the two other frustums will be close to the two cones making up A^f. This means that the surface area of P^F will be arbitrarily close to the surface area of A^f if the partition becomes fine enough. This observation motivates the following definition.

Definition 16.30. Let f be nonnegative on $[a,b]$. We say that the *surface area* of

$$A^f = \{(x,y,z) : a \le x \le b,\ \sqrt{y^2+z^2} = f(x)\}$$

exists and equals Φ if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for every partition F of $[a,b]$ with mesh smaller than δ, we have $|\Phi_F - \Phi| < \varepsilon$, where Φ_F is the surface area of the set we get by rotating the inscribed polygonal path corresponding to F about the x-axis, defined by (16.21).

Theorem 16.31. *Let f be a nonnegative and continuous function on the interval $[a,b]$ whose graph is rectifiable. Suppose that f is differentiable on (a,b), and $f \cdot \sqrt{1+(f')^2}$ is integrable on $[a,b]$. Then the surface area of A^f exists, and its value is*

$$2\pi \int_a^b f(x)\sqrt{1+\left(f'(x)\right)^2}\,dx.$$

Proof. We did not assume the function f to be differentiable at the points a and b, so the function $g = f \cdot \sqrt{1+(f')^2}$ might not be defined at these points. To prevent any ambiguity, let us define g to be zero at the points a and b; by Theorem 14.46 and Remark 14.47, the integrability of g and the value of the integral $I = \int_a^b g\,dx$ are unchanged when we do this.

Let $\varepsilon > 0$ be given. Since f is uniformly continuous on $[a,b]$, there exists a number $\eta > 0$ such that if $x, y \in [a,b]$ and $|y-x| < \eta$, then $|f(y) - f(x)| < \varepsilon$. By Theorem 14.23, we can choose a number $\delta > 0$ such that for every partition F of $[a,b]$ with mesh smaller than δ and every approximating sum σ_F corresponding to F, we have $|\sigma_F - I| < \varepsilon$.

Let $F : a = x_0 < x_1 < \cdots < x_n = b$ be a partition with mesh smaller than $\min(\eta, \delta)$. We show that the value Φ_F defined by (16.21) is close to $2\pi I$.

By the mean value theorem, for each i, there exists a point $c_i \in (x_{i-1}, x_i)$ such that $f(x_i) - f(x_{i-1}) = f'(c_i) \cdot (x_i - x_{i-1})$. Then

$$h_i \stackrel{\text{def}}{=} \sqrt{(f(x_i) - f(x_{i-1}))^2 + (x_i - x_{i-1})^2} = \sqrt{(f'(c_i))^2 + 1} \cdot (x_i - x_{i-1})$$

for all i. Thus the approximating sum of g with inner points c_i is

$$\sigma_F(g;(c_i)) = \sum_{i=1}^{n} f(c_i) \cdot \sqrt{(f'(c_i))^2 + 1} \cdot (x_i - x_{i-1}) = \sum_{i=1}^{n} f(c_i) \cdot h_i.$$

Since the partition F has smaller mesh than δ, we have $|\sigma_F(g;(c_i)) - I| < \varepsilon$. The partition F has smaller mesh than η, too, so by the choice of η, we have $|f(x_i) - f(c_i)| < \varepsilon$ and $|f(x_{i-1}) - f(c_i)| < \varepsilon$, so $|f(x_i) + f(x_{i-1}) - 2f(c_i)| < 2\varepsilon$ for all i. Thus

$$\left| \frac{1}{2\pi} \cdot \Phi_F - \sigma_F(g;(c_i)) \right| = \left| \sum_{i=1}^{n} \frac{f(x_i) + f(x_{i-1}) - 2f(c_i)}{2} \cdot h_i \right| \leq$$
$$\leq \varepsilon \cdot \sum_{i=1}^{n} h_i \leq \varepsilon \cdot L,$$

where L denotes the arc length of graph f. This gives

$$\left| \frac{1}{2\pi} \cdot \Phi_F - I \right| \leq \left| \frac{1}{2\pi} \cdot \Phi_F - \sigma_F(g;(c_i)) \right| + |\sigma_F(g;(c_i)) - I| < (L+1) \cdot \varepsilon. \quad (16.22)$$

Since $\varepsilon > 0$ was arbitrary and (16.22) holds for every partition with small enough mesh, we have shown that the surface area of A^f is $2\pi I$. □

Example 16.32. We compute the surface area of a **spherical segment**, that is, part of a sphere centered at the origin with radius r that falls between the planes $x = a$ and $x = b$, where $-r \leq a < b \leq r$.

The sphere centered at the origin with radius r is given by the rotated graph of the function $f(x) = \sqrt{r^2 - x^2}$ $(x \in [-r, r])$ about the x-axis.

Since f is monotone on the intervals $[-r,0]$ and $[0,r]$, its graph is rectifiable on $[-r,r]$ and thus on $[a,b]$, too. On the other hand, f is continuous on $[-r,r]$ and differentiable on $(-r,r)$, where we have

$$f(x)\sqrt{1+(f'(x))^2} = \sqrt{r^2-x^2}\cdot\sqrt{1+\frac{x^2}{r^2-x^2}} = r.$$

Thus we can apply Theorem 16.31. We get that the surface area we are looking for is

$$2\pi\cdot\int_a^b r\,dx = 2\pi r\cdot(b-a),$$

so *the area of a spherical segment agrees with the distance between the planes that define it times the circumference of a great circle of the sphere.* As a special case, *the surface area of a sphere of radius r is* $4r^2\pi$*.*

Remark 16.33. Suppose that f is a nonnegative continuous function on the interval $[a,b]$ whose graph is rectifiable. Denote the arc length of graph f over the interval $[a,x]$ by $s(x)$. Then s is strictly monotone increasing and continuous on $[a,b]$, and so it has an inverse function s^{-1}.

By a variant of the proof of Theorem 16.31, one can show that *with these conditions, A^f has a surface area, and it is equal to* $2\pi\cdot\int_0^L (f\circ s^{-1})\,dx$*, where L denotes the arc length of* graph f.

Exercises

16.38. Prove that if the height of a frustum is m, and the radii of the lower and upper circles are r and R, then flattening the side of the frustum gives us a region whose area is $\pi(R+r)\sqrt{(R-r)^2+m^2}$.

16.39. Compute the surface area of the surfaces that we get by revolving the graphs of the following functions about the x-axis:

(a) e^x over $[a,b]$;
(b) $\sqrt{x}$ over $[a,b]$, where $0<a<b$;
(c) $\sin x$ over $[0,\pi]$;
(d) $\operatorname{ch} x$ over $[-a,a]$.

16.40. Call a region falling between two parallel lines a strip. By the width of the strip, we mean the distance between the two lines. Prove that if we cover a circle with finitely many strips, then the sum of the widths of the strips we used is at least as large as the diameter of the circle. (H S)

16.41. Let f be nonnegative and continuously differentiable on $[a,b]$. Prove that the surface area of the surface of revolution $A^f=\{(x,y,z): a\le x\le b,\ y^2+z^2=f^2(x)\}$

equals the length of graph f times the circumference of the circle traced by the center of mass of graph f during its revolution (this is sometimes called **Guldin's first theorem**).

16.7 Appendix: Proof of Theorem 16.20

Proof. Since every integrable function is already bounded, statement (ii) of Theorem 16.19 implies that g is rectifiable. Let $f = \sqrt{(g_1')^2 + \cdots + (g_d')}$. We will show that for every $\varepsilon > 0$, there exists a partition F such that the inscribed polygonal path of g corresponding to F has length ℓ_F, which differs from $s(g)$ by less than ε, and also that every Riemann sum $\sigma_F(f)$ of f differs from ℓ_F by less than ε. It will follow from this that $|\sigma_F(f) - s(g)| < 2\varepsilon$ for every Riemann sum corresponding to the partition F, and so by Theorem 14.19, f is integrable with integral $s(g)$.

Let $F : a = t_0 < \cdots < t_n = b$ be a partition of the interval $[a,b]$, and let ℓ_F denote the length of the corresponding inscribed polygonal path, that is, let $\ell_F = \sum_{i=1}^{n} |g(t_i) - g(t_{i-1})|$. For all $i = 1, \ldots, n$,

$$|g(t_i) - g(t_{i-1})| = \sqrt{(g_1(t_i) - g_1(t_{i-1}))^2 + \cdots + (g_d(t_i) - g_d(t_{i-1}))^2}.$$

By the mean value theorem, there exist points $c_{i,1}, \ldots, c_{i,d} \in (t_{i-1}, t_i)$ such that

$$g_j(t_i) - g_j(t_{i-1}) = g_j'(c_{i,j})(t_i - t_{i-1}) \qquad (j = 1, \ldots, d).$$

Then

$$\begin{aligned}\ell_F &= \sum_{i=1}^{n} |g(t_i) - g(t_{i-1})| = \\ &= \sum_{i=1}^{n} \sqrt{\left(g_1'(c_{i,1})\right)^2 + \cdots + \left(g_d'(c_{i,d})\right)^2} \cdot (t_i - t_{i-1}).\end{aligned} \tag{16.23}$$

Now let $e_i \in [t_{i-1}, t_i]$ $(i = 1, \ldots, n)$ be arbitrary inner points, and consider the corresponding Riemann sum of f:

$$\begin{aligned}\sigma_F(f; (e_i)) &= \sum_{i=1}^{n} f(e_i)(t_i - t_{i-1}) = \\ &= \sum_{i=1}^{n} \sqrt{\left(g_1'(e_i)\right)^2 + \cdots + \left(g_d'(e_i)\right)^2} \cdot (t_i - t_{i-1}).\end{aligned} \tag{16.24}$$

By the similarities of the right-hand sides of the equalities (16.23) and (16.24), we can expect that for a suitable partition F, ℓ_F and $\sigma_F(f; (e_i))$ will be close to each other. By inequality (16.3),

$$\left|\sqrt{\left(g_1'(c_{i,1})\right)^2+\cdots+\left(g_d'(c_{i,d})\right)^2}-\sqrt{\left(g_1'(e_i)\right)^2+\cdots+\left(g_d'(e_i)\right)^2}\right|\le$$

$$\le \sum_{j=1}^{d}|g_j'(c_{i,d})-g_j'(e_i)|\le \sum_{j=1}^{d}\omega(g_j';[t_{i-1},t_i]).$$

Thus subtracting (16.23) and (16.24), we get that

$$|\ell_F-\sigma_F(f;(e_i))|\le \sum_{j=1}^{d}\sum_{i=1}^{n}\omega(g_j';[t_{i-1},t_i])(t_i-t_{i-1})\le \sum_{j=1}^{d}\Omega_F(g_j') \tag{16.25}$$

for every partition F and every e_i. Let $\varepsilon>0$ be fixed. Since $s(g)$ is the supremum of the numbers ℓ_F, there exists a partition F_0 such that $s(g)-\varepsilon<\ell_{F_0}\le s(g)$. It is easy to check that if we add new base points to a partition, then the value of ℓ_F cannot decrease. Clearly, if we add another base point, then we replace a term $|g(t_{k-1})-g(t_k)|$ in the sum by $|g(t_{k-1})-g(t_k')|+|g(t_k')-g(t_k)|$. The triangle inequality ensures that the value of ℓ_F does not decrease with this. Thus if F is a refinement of F_0, then

$$s(g)-\varepsilon<\ell_{F_0}\le\ell_F\le s(g). \tag{16.26}$$

Since the functions g_j' are integrable, there exist partitions F_j such that $\Omega_{F_j}(g_j')<\varepsilon$ $(j=1,\ldots,d)$. Let F be the union of partitions $F_0,F_1,\ldots,F_d$. Then

$$\Omega_F(g_j')\le\Omega_{F_j}(g_j')<\varepsilon$$

for all $j=1,\ldots,d$. If we now combine (16.25) and (16.26), we get that

$$|\sigma_F(f;(e_i))-s(g)|\le|\sigma_F(f;(e_i))-\ell_F|+|\ell_F-s(g)|<(d+1)\varepsilon. \tag{16.27}$$

In the end, for every $\varepsilon>0$, we have constructed a partition F that satisfies (16.27) with an arbitrary choice of inner points e_i. Then by Theorem 14.19, f is integrable, and its integral is $s(g)$. □