Solutions Manual to Walter Rudin's *Principles of Mathematical Analysis*

Roger Cooke, University of Vermont

Contents

Chapter 1

The Real and Complex Number Systems

Exercise 1.1 If r is rational ($r \neq 0$) and x is irrational, prove that $r + x$ and rx are irrational.

Solution. If r and $r + x$ were both rational, then $x = r + x - r$ would also be rational. Similarly if rx were rational, then $x = \frac{rx}{r}$ would also be rational.

Exercise 1.2 Prove that there is no rational number whose square is 12.

First Solution. Since $\sqrt{12} = 2\sqrt{3}$, we can invoke the previous problem and prove that $\sqrt{3}$ is irrational. If m and n are integers having no common factor and such that $m^2 = 3n^2$, then m is divisible by 3 (since if m^2 is divisible by 3, so is m). Let $m = 3k$. Then $m^2 = 9k^2$, and we have $3k^2 = n^2$. It then follows that n is also divisible by 3 contradicting the assumption that m and n have no common factor.

Second Solution. Suppose $m^2 = 12n^2$, where m and n have no common factor. It follows that m must be even, and therefore n must be odd. Let $m = 2r$. Then we have $r^2 = 3n^2$, so that r is also odd. Let $r = 2s + 1$ and $n = 2t + 1$. Then

$$4s^2 + 4s + 1 = 3(4t^2 + 4t + 1) = 12t^2 + 12t + 3,$$

so that

$$4(s^2 + s - 3t^2 - 3t) = 2.$$

But this is absurd, since 2 cannot be a multiple of 4.

Exercise 1.3 Prove Proposition 1.15, i.e., prove the following statements:

(*a*) *If* $x \neq 0$ *and* $xy = xz$, *then* $y = z$.
(*b*) *If* $x \neq 0$ *and* $xy = x$, *then* $y = 1$.
(*c*) *If* $x \neq 0$ *and* $xy = 1$, *then* $y = 1/x$.
(*d*) *If* $x \neq 0$, *then* $1/(1/x) = x$.

Solution. (*a*) Suppose $x \neq 0$ and $xy = xz$. By Axiom (M5) there exists an element $1/x$ such that $1/x = 1$. By (M3) and (M4) we have $(1/x)(xy) = ((1/x)x)y = 1y = y$, and similarly $(1/x)(xz) = z$. Hence $y = z$.

(*b*) Apply (*a*) with $z = 1$.

(*c*) Apply (*a*) with $z = 1/x$.

(*d*) Apply (*a*) with x replaced by $1/x$, $y = 1/(1/x)$, and $z = x$.

Exercise 1.4 Let E be a nonempty subset of an ordered set; suppose α is a lower bound of E, and β is an upper bound of E. Prove that $\alpha \leq \beta$.

Solution. Since E is nonempty, there exists $x \in E$. Then by definition of lower and upper bounds we have $\alpha \leq x \leq \beta$, and hence by property *ii* in the definition of an ordering, we have $\alpha < \beta$ unless $\alpha = x = \beta$.

Exercise 1.5 Let A be a nonempty set of real numbers which is bounded below. Let $-A$ be the set of all numbers $-x$, where $x \in A$. Prove that

$$\inf A = -\sup(-A).$$

Solution: We need to prove that $-\sup(-A)$ is the greatest lower bound of A. For brevity, let $\alpha = -\sup(-A)$. We need to show that $\alpha \leq x$ for all $x \in A$ and $\alpha \geq \beta$ if β is any lower bound of A.

Suppose $x \in A$. Then, $-x \in -A$, and, hence $-x \leq \sup(-A)$. It follows that $x \geq -\sup(-A)$, i.e., $\alpha \leq x$. Thus α is a lower bound of A.

Now let β be any lower bound of A. This means $\beta \leq x$ for all x in A. Hence $-x \leq -\beta$ for all $x \in A$, which says $y \leq -\beta$ for all $y \in -A$. This means $-\beta$ is an upper bound of $-A$. Hence $-\beta \geq \sup(-A)$ by definition of sup, i.e., $\beta \leq -\sup(-A)$, and so $-\sup(-A)$ is the greatest lower bound of A.

Exercise 1.6 Fix $b > 1$.

(*a*) If m, n, p, q are integers, $n > 0$, $q > 0$, and $r = m/n = p/q$, prove that

$$(b^m)^{1/n} = (b^p)^{1/q}.$$

Hence it makes sense to define $b^r = (b^m)^{1/n}$.

(*b*) Prove that $b^{r+s} = b^r b^s$ if r and s are rational.

(c) If x is real, define $B(x)$ to be the set of all numbers b^t, where t is rational and $t \le x$. Prove that

$$b^r = \sup B(r)$$

when r is rational. Hence it makes sense to define

$$b^x = \sup B(x)$$

for every real x.

(d) Prove that $b^{x+y} = b^x b^y$ for all real x and y.

Solution. (a) Let $k = mq = np$. Since there is only one positive real number c such that $c^{nq} = b^k$ (Theorem 1.21), if we prove that both $(b^m)^{1/n}$ and $(b^p)^{1/q}$ have this property, it will follow that they are equal. The proof is then a routine computation: $\left((b^m)^{1/n}\right)^{nq} = (b^m)^q = b^{mq} = b^k$, and similarly for $(b^p)^{1/q}$.

(b) Let $r = \frac{m}{n}$ and $s = \frac{v}{w}$. Then $r + s = \frac{mw+vn}{nw}$, and

$$b^{r+s} = (b^{mw+vn})^{1/nw} = \left((b^{mw}b^{vn})\right)^{1/nw},$$

by the laws of exponents for integer exponents. By the corollary to Theorem 1.21 we then have

$$b^{r+s} = (b^{mw})^{1/nw}(b^{nv})^{1/nw} = b^r b^s,$$

where the last equality follows from part (a).

(c) It will simplify things later on if we amend the definition of $B(x)$ slightly, by defining it as $\{b^t : t \text{ rational}, t < x\}$. It is then slightly more difficult to prove that $b^r = \sup B(r)$ if r is rational, but the technique of Problem 7 comes to our rescue. Here is how: It is obvious that b^r is an upper bound of $B(r)$. We need to show that it is the least upper bound. The inequality $b^{1/n} < t$ if $n > (b-1)/(t-1)$ is proved just as in Problem 7 below. It follows that if $0 < x < b^r$, there exists an integer n with $b^{1/n} < b^r/x$, i.e., $x < b^{r-1/n} \in B(r)$. Hence x is not an upper bound of $B(r)$, and so b^r is the least upper bound.

(d) By definition $b^{x+y} = \sup B(x+y)$, where $B(x+y)$ is the set of all numbers b^t with t rational and $t < x+y$. Now any rational number t that is less than $x+y$ can be written as $r+s$, where r and s are rational, $r < x$, and $s < y$. To do this, let r be any rational number satisfying $t - y < r < x$, and let $s = t - r$. Conversely any pair of rational numbers r, s with $r < x$, $s < y$ gives a rational sum $t = r + s < x + y$. Hence $B(x+y)$ can be described as the set of all numbers $b^r b^s$ with $r < x$, $s < y$, and r and s rational, i.e., $B(x+y)$ is the set of all products uv, where $u \in B(x)$ and $v \in B(y)$.

Since any such product is less than $\sup B(x) \sup B(y)$, we see that the number $M = \sup B(x) \sup B(y)$ is an upper bound for $B(x+y)$. On the other hand, suppose $0 < c < \sup B(x) \sup B(y)$. Then $c/(\sup B(x)) < \sup B(y)$. Let $m = (1/2)\big(c/(\sup B(x)) + \sup B(y)\big)$. Then $c/\sup B(x) < m < \sup B(y)$, and there exist $u \in B(x)$, $v \in B(y)$ such that $c/m < u$ and $m < v$. Hence we have

$c = (c/m)m < uv \in B(x+y)$, and so c is not an upper bound for $B(x+y)$. It follows that $\sup B(x) \sup B(y)$ is the least upper bound of $B(x+y)$, i.e.,

$$b^{x+y} = b^x b^y,$$

as required.

Exercise 1.7 Fix $b > 1$, $y > 0$, and prove that there is a unique real x such that $b^x = y$, by completing the following outline. (This x is called the *logarithm of y to the base b*.)

(a) For any positive integer n, $b^n - 1 \geq n(b-1)$.

(b) Hence $b - 1 \geq n(b^{1/n} - 1)$.

(c) If $t > 1$ and $n > (b-1)/(t-1)$, then $b^{1/n} < t$.

(d) If w is such that $b^w < y$, then $b^{w+(1/n)} < y$ for sufficiently large n; to see this apply part (c) with $t = y \cdot b^{-w}$.

(e) If $b^w > y$, then $b^{w-(1/n)} > y$ for sufficiently large n.

(f) Let A be the set of all w such that $b^w < y$, and show that $x = \sup A$ satisfies $b^w = y$.

(g) Prove that this x is unique.

Solution. (a) The inequality $b^n - 1 \geq n(b-1)$ is equality if $n = 1$. Then, by induction $b^{n+1} - 1 = b^{n+1} - b + (b-1) = b(b^n - 1) + (b-1) \geq bn(b-1) + (b-1) = (bn+1)(b-1) \geq (n+1)(b-1)$.

(b) Replace b by $b^{1/n}$ in part (a).

(c) The inequality $n > (b-1)/(t-1)$ can be rewritten as $n(t-1) > (b-1)$, and since $b - 1 \geq n(b^{1/n} - 1)$, we have $n(t-1) > n(b^{1/n} - 1)$, which implies $t > b^{1/n}$.

(d) The application of part (c) with $t = y \cdot b^{-w} > 1$ is immediate.

(e) The application of part (c) with $t = b^w \cdot (1/y)$ yields the result, as in part (d) above.

(f) There are only three possibilities for the number $x = \sup A$: 1) $b^x < y$; 2) $b^x > y$; 3) $b^x = y$. The first assumption, by part (d), implies that $x + (1/n) \in A$ for large n, contradicting the assumption that x is an upper bound for A. The second, by part (e), implies that $x - (1/n)$ is an upper bound for A if n is large, contradicting the assumption that x is the smallest upper bound. Hence the only remaining possibility is that $b^x = y$.

(g) Suppose $z \neq x$, say $z > x$. Then $b^z = b^{x+(z-x)} = b^x b^{z-x} > b^x = y$. Hence x is unique. (It is easy to see that $b^w > 1$ if $w > 0$, since there is a positive rational number $r = \frac{m}{n}$ with $0 < r < w$, and $b^r = (b^m)^{1/n}$. Then $b^m > 1$ since $b > 1$, and $(b^m)^{1/n} > 1$ since $1^n = 1 < b^m$.)

Exercise 1.8 Prove that no order can be defined in the complex field that turns it into an ordered field. *Hint:* -1 is a square.

Solution. By Part (a) of Proposition 1.18, either i or $-i$ must be positive. Hence $-1 = i^2 = (-i)^2$ must be positive. But then $1 = (-1)^2$, must also be positive, and this contradicts Part (a) of Proposition 1.18, since 1 and -1 cannot both be positive.

Exercise 1.9 Suppose $z = a + bi$, $w = c + di$. Define $z < w$ if $a < c$, and also if $a = c$ but $b < d$. Prove that this turns the set of all complex numbers into an ordered set. (This type of order relation is called a *dictionary order*, or *lexicographic order*, for obvious reasons.) Does this ordered set have the least upper bound property?

Solution. We need to show that either $z < w$ or $z = w$, or $w < z$. Now since the *real* numbers are ordered, we have $a < c$ or $a = c$, or $c < a$. In the first case $z < w$; in the third case $w < z$. Now consider the second case. We must have $b < d$ or $b = d$ or $d < b$. In the first of these cases $z < w$, in the third case $w < z$, and in the second case $z = w$.

We also need to show that if $z < w$ and $w < u$, then $z < u$. Let $u = e + fi$. Since $z < w$, we have either $a < c$ or $a = c$ and $b < d$. Since $w < u$ we have either $c < f$ or $c = f$ and $d < g$. Hence there are four possible cases:

Case 1: $a < c$ and $c < f$. Then $a < f$ and so $z < u$, as required.
Case 2: $a < c$ and $c = f$ and $d < g$. Again $a < f$, and $z < u$.
Case 3: $a = c$ and $b < d$ and $c < f$. Once again $a < f$ and so $z < u$.
Case 4: $a = c$ and $b < d$ and $c = f$, and $d < g$. Then $a = f$ and $b < g$, and so $z < u$.

Exercise 1.10 Suppose $z = a + bi$, $w = u + iv$, and

$$a = \left(\frac{|w| + u}{2}\right)^{1/2}, \quad b = \left(\frac{|w| - u}{2}\right)^{1/2}.$$

Prove that $z^2 = w$ if $v \geq 0$ and that $(\bar{z})^2 = w$ if $v \leq 0$. Conclude that every complex number (with one exception) has two complex square roots.

Solution.

$$z^2 = (a + bi)^2 = (a^2 - b^2) + 2abi.$$

Now

$$a^2 - b^2 = \frac{|w| + u}{2} - \frac{|w| - u}{2} = u,$$

and, since $(xy)^{1/2} = x^{1/2}y^{1/2}$,

$$2ab = 2\left(\frac{|w| + u}{2}\frac{|w| - u}{2}\right)^{1/2} = 2\left(\frac{|w|^2 - u^2}{4}\right)^{1/2}.$$

Hence

$$2ab = 2\left(\left(\frac{v}{2}\right)^2\right)^{1/2}$$

Now $(x^2)^{1/2} = x$ if $x \geq 0$ and $(x^2)^{1/2} = -x$ if $x \leq 0$. We conclude that $2ab = v$ if $v \geq 0$ and $2ab = -v$ if $v \leq 0$. Hence $z^2 = w$ if $v \geq 0$. Replacing b by $-b$, we find that $(\bar{z})^2 = w$ if $v \leq 0$.

Hence every non-zero complex number has (at least) two complex square roots.

Exercise 1.11 If z is a complex number, prove that there exists an $r \geq 0$ and a complex number w with $|w| = 1$ such that $z = rw$. Are w and r always uniquely determined by z?

Solution. If $z = 0$, we take $r = 0$, $w = 1$. (In this case w is not unique.) Otherwise we take $r = |z|$ and $w = z/|z|$, and these choices are unique, since if $z = rw$, we must have $r = r|w| = |rw| = |z|$, z/r.

Exercise 1.12 If $z_1, \ldots, z_n$ are complex, prove that

$$|z_1 + z_2 + \cdots + z_n| \leq |z_1| + |z_2| + \cdots + |z_n|.$$

Solution. The case $n = 2$ is Part (e) of Theorem 1.33. We can then apply this result and induction on n to get

$$\begin{aligned} |z_1 + z_2 + \cdots z_n| &= |(z_1 + z_2 + \cdots + z_{n-1}) + z_n| \\ &\leq |z_1 + z_2 + \cdots + z_{n-1}| + |z_n| \\ &\leq |z_1| + |z_2| + \cdots + |z_{n-1}| + |z_n|. \end{aligned}$$

Exercise 1.13 If x, y are complex, prove that

$$\big||x| - |y|\big| \leq |x - y|.$$

Solution. Since $x = x - y + y$, the triangle inequality gives

$$|x| \leq |x - y| + |y|,$$

so that $|x| - |y| \leq |x - y|$. Similarly $|y| - |x| \leq |x - y|$. Since $|x| - |y|$ is a real number we have either $\big||x| - |y|\big| = |x| - |y|$ or $\big||x| - |y|\big| = |y| - |x|$. In either case, we have shown that $\big||x| - |y|\big| \leq |x - y|$.

Exercise 1.14 If z is a complex number such that $|z| = 1$, that is, such that $z\bar{z} = 1$, compute

$$|1+z|^2 + |1-z|^2.$$

Solution. $|1+z|^2 = (1+z)(1+\bar{z}) = 1+\bar{z}+z+z\bar{z} = 2+z+\bar{z}$. Similarly $|1-z|^2 = (1-z)(1-\bar{z}) = 1-z-\bar{z}+z\bar{z} = 2-z-\bar{z}$. Hence

$$|1+z|^2 + |1-z|^2 = 4.$$

Exercise 1.15 Under what conditions does equality hold in the Schwarz inequality?

Solution. The proof of Theorem 1.35 shows that equality can hold if $B = 0$ or if $Ba_j - Cb_j = 0$ for all j, i.e., the numbers a_j are proportional to the numbers b_j. (In terms of linear algebra this means the vectors $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in complex n-dimensional space are linearly dependent. Conversely, if these vectors are linearly independent, then strict inequality holds.)

Exercise 1.16 Suppose $k \geq 3$, $\mathbf{x}, \mathbf{y} \in R^k$, $|\mathbf{x}-\mathbf{y}| = d > 0$, and $r > 0$. Prove:

(*a*) If $2r > d$, there are infinitely many $\mathbf{z} \in R^k$ such that

$$|\mathbf{z}-\mathbf{x}| = |\mathbf{z}-\mathbf{y}| = r.$$

(*b*) If $2r = d$, there is exactly one such $\mathbf{z}$.

(*c*) If $2r < d$, there is no such $\mathbf{z}$.

How must these statements be modified if k is 2 or 1?

Solution. (*a*) Let $\mathbf{w}$ be any vector satisfying the following two equations:

$$\begin{aligned} \mathbf{w}\cdot(\mathbf{x}-\mathbf{y}) &= 0, \\ |\mathbf{w}|^2 &= r^2 - \frac{d^2}{4}. \end{aligned}$$

From linear algebra it is known that all but one of the components of a solution $\mathbf{w}$ of the first equation can be arbitrary. The remaining component is then uniquely determined. Also, if $\mathbf{w}$ is any non-zero solution of the first equation, there is a unique positive number t such that $t\mathbf{w}$ satisfies both equations. (For example, if $x_1 \neq y_1$, the first equation is satisfied whenever

$$z_1 = \frac{z_2(x_2-y_2) + \cdots + z_k(x_k-y_k)}{y_1 - x_1}.$$

If $(z_1, z_2, \ldots, z_k)$ satisfies this equation, so does $(tz_1, tz_2, \ldots, tz_k)$ for any real number t.) Since at least two of these components can vary independently, we can find a solution with these components having any prescribed ratio. This ratio does not change when we multiply by the positive number t to obtain a solution of both equations. Since there are infinitely many ratios, there are infinitely many distinct solutions. For each such solution $\mathbf{w}$ the vector $\mathbf{z} = \frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y} + \mathbf{w}$ is a solution of the required equation. For

$$\begin{aligned} |\mathbf{z}-\mathbf{x}|^2 &= \left|\frac{\mathbf{y}-\mathbf{x}}{2} + \mathbf{w}\right|^2 \\ &= \left|\frac{\mathbf{y}-\mathbf{x}}{2}\right|^2 + 2\mathbf{w}\cdot\frac{\mathbf{x}-\mathbf{y}}{2} + |\mathbf{w}|^2 \\ &= \frac{d^2}{4} + 0 + r^2 - \frac{d^2}{4} \\ &= r^2, \end{aligned}$$

and a similar relation holds for $|\mathbf{z}-\mathbf{y}|^2$.

(*b*) The proof of the triangle inequality shows that equality can hold in this inequality only if it holds in the Schwarz inequality, i.e., one of the two vectors is a scalar multiple of the other. Further examination of the proof shows that the scalar must be nonnegative. Now the conditions of this part of the problem show that

$$|\mathbf{x}-\mathbf{y}| = d = |\mathbf{x}-\mathbf{z}| + |\mathbf{z}-\mathbf{y}|.$$

Hence it follows that there is a nonnegative scalar t such that

$$\mathbf{x}-\mathbf{z} = t(\mathbf{z}-\mathbf{y}).$$

However, the hypothesis also shows immediately that $t = 1$, and so $\mathbf{z}$ is uniquely determined as

$$\mathbf{z} = \frac{\mathbf{x}+\mathbf{y}}{2}.$$

(*c*) If $\mathbf{z}$ were to satisfy this condition, the triangle inequality would be violated, i.e., we would have

$$|\mathbf{x}-\mathbf{y}| = d > 2r = |\mathbf{x}-\mathbf{z}| + |\mathbf{z}-\mathbf{y}|.$$

When $k = 2$, there are precisely 2 solutions in case (*a*). When $k = 1$, there are no solutions in case (*a*). The conclusions in cases (*b*) and (*c*) do not require modification.

Exercise 1.17 Prove that

$$|\mathbf{x}+\mathbf{y}|^2 + |\mathbf{x}-\mathbf{y}|^2 = 2|\mathbf{x}|^2 + 2|\mathbf{y}|^2$$

if $\mathbf{x} \in R^k$ and $\mathbf{y} \in R^k$. Interpret this geometrically as a statement about parallelograms.

Solution. The proof is a routine computation, using the relation

$$|\mathbf{x} \pm \mathbf{y}|^2 = (\mathbf{x} \pm \mathbf{y}) \cdot (\mathbf{x} \pm \mathbf{y}) = |\mathbf{x}|^2 \pm 2\mathbf{x}\cdot\mathbf{y} + |\mathbf{y}|^2.$$

If $\mathbf{x}$ and $\mathbf{y}$ are the sides of a parallelogram, then $\mathbf{x}+\mathbf{y}$ and $\mathbf{x}-\mathbf{y}$ are its diagonals. Hence this result says that the sum of the squares on the diagonals of a parallelogram equals the sum of the squares on the sides.

Exercise 1.18 If $k \geq 2$ and $\mathbf{x} \in R^k$, prove that there exists $\mathbf{y} \in R^k$ such that $\mathbf{y} \neq \mathbf{0}$ but $\mathbf{x} \cdot \mathbf{y} = 0$. Is this also true if $k = 1$?

Solution. If $\mathbf{x}$ has any components equal to 0, then $\mathbf{y}$ can be taken to have the corresponding components equal to 1 and all others equal to 0. If all the components of $\mathbf{x}$ are nonzero, $\mathbf{y}$ can be taken as $(-x_2, x_1, 0, \ldots, 0)$. This is, of course, not true when $k = 1$, since the product of two nonzero real numbers is nonzero.

Exercise 1.19 Suppose $\mathbf{a} \in R^k$, $\mathbf{b} \in R^k$. Find $\mathbf{c} \in R^k$ and $r > 0$ such that

$$|\mathbf{x}-\mathbf{a}| = 2|\mathbf{x}-\mathbf{b}|$$

if and only if $|\mathbf{x}-\mathbf{c}| = r$. (*Solution:* $3\mathbf{c} = 4\mathbf{b}-\mathbf{a}$, $3r = 2|\mathbf{b}-\mathbf{a}|$.)

Solution. Since the solution is given to us, all we have to do is verify it, i.e., we need to show that the equation

$$|\mathbf{x}-\mathbf{a}| = 2|\mathbf{x}-\mathbf{b}|$$

is equivalent to $|\mathbf{x}-\mathbf{c}| = r$, which says

$$\left|\mathbf{x} - \frac{4}{3}\mathbf{b} + \frac{1}{3}\mathbf{a}\right| = \frac{2}{3}|\mathbf{b}-\mathbf{a}|.$$

If we square both sides of both equations, we an equivalent pair of equations, the first of which reduces to

$$3|\mathbf{x}|^2 + 2\mathbf{a}\cdot\mathbf{x} - 8\mathbf{b}\cdot\mathbf{x} - |\mathbf{a}|^2 + 4|\mathbf{b}|^2 = 0,$$

and the second of which reduces to this equation divided by 3. Hence these equations are indeed equivalent.

Exercise 1.20 With reference to the Appendix, suppose that property (III) were omitted from the definition of a cut. Keep the same definitions of order and addition. Show that the resulting ordered set has the least-upper-bound property, that addition satisfies axioms (A1) to (A4) (with a slightly different zero element!) but that (A5) fails.

Solution. We are now defining a cut to be a proper subset of the rational numbers that contains, along with each of its elements, all smaller rational numbers. Order is defined by containment. Now given a set A of cuts having an upper bound β, let α be the union of all the cuts in A. Obviously α is properly contained in β, and so is a proper subset of the rationals. It also obviously satisfies the property that if $p \in \alpha$ and $q < p$, then $q \in \alpha$; hence α is a cut. It is further obvious that α contains each elements of A, and so is an upper bound for A. It remains to prove that there is no smaller upper bound.

To that end, suppose, $\gamma < \alpha$, then α contains an element x not in γ. By definition of α, x must belong to some cut δ in A. But then $\gamma < \delta$, and so γ is not an upper bound for A. Thus α is the least upper bound.

The proof given in the text goes over without any change to show that (A1), (A2), and (A3) hold. As for (A4) let $O = \{r : r \le 0\}$. We claim $O + \alpha = \alpha$. The proof is easy. First, we obviously have $O + \alpha \subseteq \alpha$. For $r + s \le s$ if $r \le 0$. Hence $r + s \in \alpha$ if $s \in \alpha$. Conversely $\alpha \subseteq O + \alpha$, since each s in α can be written as $0 + s$.

Unfortunately, if $O' = \{r : r < 0\}$, there is no element α such that $\alpha + O' = O$. For $\alpha + O'$ has no largest element. If $x = r + s \in \alpha + O'$, where $r \in \alpha$ and $s \in O'$, there is an element $t \in O'$ with $t > s$, and so $r + t \in \alpha + O'$ and $r + t > s$. Since O has a largest element (namely 0), these two sets cannot be equal.

Chapter 2

Basic Topology

Exercise 2.1 Prove that the empty set is a subset of every set.

Solution. Let $\varnothing$ denote the empty set, and let E be any set. The statement $\varnothing \subset E$ is equivalent to the statement, "If $x \in \varnothing$, then $x \in E$." Since the hypothesis of this if-then statement is false, the implication is true, and we are done.

Exercise 2.2 A complex number z is said to be *algebraic* if there are integers $a_0, \ldots, a_n$, not all zero, such that

$$a_0 z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n = 0.$$

Prove that the set of all algebraic numbers is countable. *Hint:* For every positive integer N there are only finitely many equations with

$$n + |a_0| + |a_1| + \cdots + |a_n| = N.$$

Solution. Following the hint, we let A_N be the set of numbers satisfying one of the equations just listed with $n+|a_0|+|a_1|+\cdots+|a_n| = N$. The set A_N is finite, since each equation has only a finite set of solutions and there are only finitely many equations satisfying this condition. By the corollary to Theorem 2.12 the set of algebraic numbers, which is the union $\bigcup\limits_{N=2}^{\infty} A_N$, is at most countable. Since all rational numbers are algebraic, it follows that the set of algebraic numbers is exactly countable.

Exercise 2.3 Prove that there exist real numbers which are not algebraic.

Solution. By the previous exercise, the set of real algebraic numbers is countable. If every real number were algebraic, the entire set of real numbers would be countable, contradicting the remark after Theorem 2.14.

Exercise 2.4 Is the set of irrational real numbers countable?

Answer. No. If it were, the set of all real numbers, being the union of the rational and irrational numbers, would be countable.

Exercise 2.5 Construct a bounded set of real numbers with exactly three limit points.

Solution. Let E be the set of numbers of the form $a + \frac{1}{n}$, where $a \in \{1, 2, 3\}$ and $n \in \{2, 3, 4, 5, \ldots, \}$. It is clear that $\{1, 2, 3\} \subseteq E'$, since every deleted neighborhood of 1, 2, or 3, contains a point in E. Conversely, if $x \notin \{1, 2, 3\}$, let $\delta = \min\{|x-1|, |x-2|, |x-3|\}$. Then the set U of y such that $|x - y| < \delta/2$ contains at most a finite number of points of E, since the set $V = (1, 1 + \frac{\delta}{2}) \cup (2, 2 + \frac{\delta}{2}) \cup (3, 3 + \frac{\delta}{2})$ is disjoint from U, and V contains all the points of the set E except possibly the finite set of points $a + \frac{1}{n}$ for which $n \le \frac{2}{\delta}$. If $p_1, \ldots, p_r$ are the points of E in U, let η be the minimum of $\frac{\delta}{2}$ and the $|x - p_j|$ for which $x \ne p_j$. Then the set W of points y such that $|y - x| < \eta$ contains no points of E except possibly x. Hence $x \notin E'$. Thus $E' = \{1, 2, 3\}$.

Exercise 2.6 Let E' be the set of all limit points of a set E. Prove that E' is closed. Prove that E and $\overline{E}$ have the same limit points. (Recall that $\overline{E} = E \cup E'$.) Do E and E' always have the same limit points?

Solution. To show that E' is closed, we shall show that $(E')' \subseteq E'$. In fact, we shall show the even stronger statement that $\left(\overline{E}\right)' \subseteq E'$. To do this let $x \in \left(\overline{E}\right)'$, and let $r > 0$. We need to show that $x \in E'$; that is, since $r > 0$ is arbitrary, we need to find a point $z \in E$ with $0 < d(z, x) < r$. There certainly is a point y of $\overline{E}$ such that $0 < d(y, x) < r$. If $y \in E$, we can take $z = y$, and we are done. If $y \notin E$, then $y \in E'$. Let $s = \min\big(d(x, y), r - d(x, y)\big)$, so that $s > 0$. Since $y \in E'$, there exists $z \in E$ with $0 < d(x, z) < s$. But it then follows that $d(z, x) \ge d(x, y) - d(x, z) > 0$ and $d(z, x) \le d(x, y) + d(y, z) < d(x, y) + r - d(x, y) = r$, and we are done in any case.

To show that E and $\overline{E}$ have the same limit points, we need only show the converse of the preceding containment. But this is easy. Suppose $x \in E'$. Since every deleted neighborhood of x contains a point of E, *a fortiori* every deleted neighborhood of x contains a point of $\overline{E}$. Hence $E' \subseteq (\overline{E})'$.

Certainly E and E' may have different sets of limit points. For example if $E = \{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots\}$, then $E' = \{0\}$, while $(E')' = \varnothing$.

Exercise 2.7 Let $A_1, A_2, A_3, \ldots$ be subsets of a metric space.

(*a*) If $B_n = \cup_{i=1}^{n} A_i$, prove that $\overline{B}_n = \cup_{i=1}^{n} \overline{A}_i$, for $n = 1, 2, 3, \ldots$.

(*b*) If $B = \cup_{i=1}^{\infty} A_i$, prove that $\overline{B} \supset \cup_{i=1}^{\infty} \overline{A}_i$.

Show, by an example, that this inclusion can be proper.

Solution. We first show that $\overline{E \cup F} = \overline{E} \cup \overline{F}$, which follows from the stronger fact that $(E \cup F)' = E' \cup F'$. To show this, in turn, we note that if $x \in E'$, then certainly $x \in (E \cup F)'$, and similarly if $x \in F'$. Hence $E' \cup F' \subseteq (E \cup F)'$. To show the converse, suppose $x \notin E' \cup F'$. Then there is a positive number r such that there is no element y of E with $0 < d(x,y) < r$, and a positive number s such that there is no element y of F with $0 < d(x,y) < s$. Hence if $t = \min(r,s)$, then $t > 0$, and there is no element y of $E \cup F$ with $0 < d(x,y) < t$. Therefore $x \notin (E \cup F)'$.

The general result of (a) now follows easily by induction on n, since

$$\begin{aligned} \overline{B}_n &= \overline{\cup_{i=1}^n A_i} \\ &= \overline{A_1 \cup \cup_{i=2}^n A_i} \\ &= \overline{A}_1 \cup \overline{\cup_{i=2}^n A_i} \\ &= \overline{A}_1 \cup \cup_{i=2}^n \overline{A}_i \\ &= \cup_{i=1}^n \overline{A}_i. \end{aligned}$$

Part (b) amounts to the trivial observation that, since $B \supseteq A_i$ for all i, then $\overline{B} \supseteq \overline{A}_i$ for all i, and so

$$\overline{B} \supseteq \cup_{i=1}^\infty \overline{A}_i.$$

If we let $A_i = \{r_i\}$, where $\{r_1, r_2, \ldots, r_n, \ldots\}$ is an enumeration of the rational numbers, then B is the full set of rational numbers. Hence $\overline{B} = R^1$, while $\overline{A}_i = A_i$ for each i, i.e., $\cup \overline{A}_i$ is the set of rational numbers.

Exercise 2.8 Is every point of every open set $E \subset R^2$ a limit point of E. Answer the same question for closed sets in R^2.

Answer. Yes. Every point of an open set E is a limit point of E. To see this, let E be an open set in R^2, let $(x_1, x_2) \in E$, let s be such that $(y_1, y_2) \in E$ if $\sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} < s$, and let $r > 0$. Then the point $(z_1, z_2) = (x_1 + \frac{1}{2}\min(r,s), x_2)$ belongs to E and satisfies $0 < \sqrt{(z_1 - x_1)^2 + (z_2 - x_2)^2} < r$.

There are closed sets for which this statement is not true. For example, any finite set E is closed, yet $E' = \varnothing$ for a finite set.

Exercise 2.9 Let E° denote the set of all interior points of a set E.

(a) Prove that E° is always open.

(b) Prove that E is open if and only if $E^\circ = E$.

(c) If $G \subset E$ and G is open, prove that $G \subset E^\circ$.

(d) Prove that the complement of E° is the closure of the complement of E.

(e) Do E and $\overline{E}$ always have the same interiors?

(*f*) Do E and E° always have the same closures?

Solution. (*a*) Let $x \in E^\circ$. Then there exists $r > 0$ such that $y \in E$ if $d(x, y) < r$. We claim that in fact $y \in E^\circ$ if $d(x, y) < r$, so that $x \in (E^\circ)^\circ$. Indeed if $d(x, y) < r$, let $s = r - d(x, y)$, so that $s > 0$. Then if $d(z, y) < s$, we have (by the triangle inequality) $d(x, z) < r$, and so $z \in E$. By definition this means $y \in E^\circ$. Since y was any point with $d(x, y) < r$, it follows that all such points are in E°, and so $x \in (E^\circ)^\circ$.

(*b*) By definition E is open if and only if each of its points is an interior point, which says precisely that $E = E^\circ$.

(*c*) If $G \subset E$ and G is open, then $G = G^\circ \subseteq E^\circ$.

(*d*) Part (*c*) shows that E° is the largest open set contained in E, i.e., the union of all open sets contained in E. Hence its complement is the intersection of all closed sets containing the complement of E, and this, by Theorem 2.27 (*c*), is the closure of the complement of E.

(*e*) Emphatically not. If E is the rational numbers in the space R^1, then $E^\circ = \varnothing$, while $\overline{E} = R^1$, so that the interior of $\overline{E}$ is R^1.

(*f*) Emphatically not. If E is the rational numbers in the space R^1, then $\overline{E} = R^1$, while $E^\circ = \varnothing$, so that $\overline{E^\circ} = \varnothing$.

Exercise 2.10 Let X be an infinite set. For $p \in X$ and $q \in X$, define

$$d(p, q) = \begin{cases} 1, & (\text{if } p \neq q), \\ 0, & (\text{if } p = q). \end{cases}$$

Prove that this is a metric. Which subsets of the resulting metric space are open? Which are closed? Which are compact?

Solution. It is obvious that $d(p, q) > 0$ if $p \neq q$ and $d(p, p) = 0$; likewise it is obvious that $d(p, q) = d(q, p)$. To show the triangle inequality $d(x, z) \leq d(x, y) + d(y, z)$, note that the maximal value of the left-hand side is 1, and can be attained only if $x \neq z$. In that case y cannot be equal to both x and z, so that at least one term on the right-hand side is also 1.

Each one-point set is open in this metric, since $B_{\frac{1}{2}}(x) \subseteq \{x\}$. Therefore every set, being the union of all its one-point subsets, is open. Hence every set, being the complement of its complement, is also closed. Only finite sets are compact, since any infinite subset has an open covering (by the union of its one-point subsets) that cannot be reduced to a finite subcovering.

Exercise 2.11 For $x \in R^1$ and $y \in R^1$, define

$$\begin{aligned} d_1(x, y) &= (x - y)^2, \\ d_2(x, y) &= \sqrt{|x - y|}, \end{aligned}$$

$$\begin{aligned} d_3(x,y) &= |x^2 - y^2|, \\ d_4(x,y) &= |x - 2y|, \\ d_5(x,y) &= \frac{|x-y|}{1+|x-y|}, \end{aligned}$$

Determine, for each of these, whether it is a metric or not.

Solution. The function $d_1(x,y)$ fails the triangle inequality condition, since

$$d_1(0,1) + d_1(1,2) = 1 + 1 = 2 < 4 = d_1(0,2).$$

The function $d_2(x,y)$ meets the triangle inequality condition, since

$$\sqrt{|x-z|} \le \sqrt{|x-y|} + \sqrt{|y-z|},$$

as one can easily see by squaring both sides. Hence d_2 is a metric.

The function $d_3(x,y)$ fails the positivity condition, since $d_2(1,-1) = 0$. (Restricted to $[0,\infty)$, d_3 would be a metric.)

Since $d_4(1,\frac{1}{2}) = 0$, the function $d_4(x,y)$ likewise fails the positivity condition. It also fails the symmetry condition, since $d_4(x,y) \neq d_4(y,x)$ in general.

The function $d_5(x,y)$ is a metric. In fact we can prove more generally that if $d(x,y)$ is a metric, so is $\rho(x,y) = \frac{d(x,y)}{1+d(x,y)}$. It is obvious that ρ meets the nonnegativity and symmetry requirements, and we need only verify the triangle inequality, which in this case says that

$$\frac{d(x,z)}{1+d(x,z)} \le \frac{d(x,y)}{1+d(x,y)} + \frac{d(y,z)}{1+d(y,z)}.$$

To do this, let $a = d(x,z)$, $b = d(x,y)$, and $c = d(y,z)$. We need to show that if $a \le b + c$, then

$$\frac{a}{1+a} \le \frac{b}{1+b} + \frac{c}{1+c}.$$

Clearing out the denominators, we find this inequality to be equivalent to

$$a + ab + ac + abc \le b + c + ab + ac + 2bc + 2abc,$$

which is clearly true.

Exercise 2.12 Let $K \subset R^1$ consist of 0 and the numbers $1/n$, for $n = 1, 2, 3, \ldots$. Prove that E is compact directly from the definition without using the Heine-Borel theorem.

Solution. Suppose $K \subset U_\alpha$, where U_α is open. Then 0 must be in some set U_{α_0}. Since U_{α_0} is open, there exists $\delta > 0$ such that $(-\delta,\delta) \subset U_{\alpha_0}$. In particular $1/n \in U_{\alpha_0}$ if $n > \frac{1}{\delta}$. Let N be the largest integer in $\frac{1}{\delta}$, and let α_j, $j = 1, \ldots, N$, be such that $\frac{1}{j} \in U_{\alpha_j}$. Then $K \subset \bigcup_{j=0}^{N} U_{\alpha_j}$.

Exercise 2.13 Construct a compact set of real numbers whose limit points form a countable set.

Solution. Let $K = \{0\} \cup \{\frac{1}{n} : n = 1, 2, \ldots\} \cup \{\frac{1}{m} + \frac{1}{n} : n = m, m+1, \ldots; m = 1, 2, \ldots\}$. It is clear that 0 and the points $\frac{1}{m}$ are limit points of K. We need only show that these are all the limit points. Since $x \geq 0$ for all $x \in K$ and for any positive number ε there is only a finite set of numbers in K larger than $1 + \varepsilon$, it is clear that no negative number and no number larger than 1 can be a limit point of K. Hence we need only consider positive numbers x satisfying $0 < x < 1$. If x is such a number and x is not one of the points $\frac{1}{m}$, let p be such that $\frac{1}{p+1} < x < \frac{1}{p}$, and let $\varepsilon = \frac{1}{2}\min(x - \frac{1}{p+1}, \frac{1}{p} - x)$. The intersection of the set K with the interval $(x - \varepsilon, x + \varepsilon)$ is contained in the set of points $\{\frac{1}{p+1} + \frac{1}{k} : p+1 \leq k < \frac{1}{\varepsilon}\} \cup \{\frac{1}{m} + \frac{1}{n} : m \leq n < \frac{1}{p+1} - \frac{1}{p+2}; m = p+2, \ldots, 2p+2\}$, which is a finite set. Therefore x cannot be a limit point of K.

Exercise 2.14 Give an example of an open cover of the segment $(0, 1)$ which has no finite subcover.

Solution. Let $A_n = (\frac{1}{n}, \frac{n-1}{n})$, $n = 3, 4, \ldots$. If $0 < x < 1$, then $x \in A_n$ if $n > 1/\min(x, 1-x)$, so that $\bigcup\limits_{n=3}^{\infty} A_n$ covers $(0, 1)$. However, the union any finite collection $\{A_1, \ldots, A_N\}$ is an interval $(\frac{1}{k}, \frac{k-1}{k})$, which fails to contain the point $\frac{1}{2k}$.

Exercise 2.15 Show that Theorem 2.36 and its Corollary become false (in R^1, for example) if the word "compact" is replaced by "closed" or "bounded."

Solution. Theorem 2.36 asserts that if a family of closed subsets has the finite intersection property (any finite collection of the sets has a non-empty intersection), then the entire family has a non-empty intersection. To see why this fails for sets that are merely bounded or merely closed, let $A_n = (0, \frac{1}{n})$ and $B_n = [n, \infty)$. The sets A_n are bounded, and the sets B_n are closed. Any finite intersection of the $A_n's$ is nonempty, and any finite intersection of the $B_n's$ is nonempty, yet $\bigcap\limits_{n=1}^{\infty} A_n = \varnothing = \bigcap\limits_{n=1}^{\infty} B_n$.

The corollary asserts that a nested sequence of nonempty compact sets has a nonempty intersection, and the examples just given show that compactness cannot be replaced by either closedness or boundedness.

Exercise 2.16 Regard Q, the set of all rational numbers, as a metric space, with $d(p, q) = |p - q|$. Let E be the set of all $p \in Q$ such that $2 < p^2 < 3$. Show that E is closed and bounded in Q, but that E is not compact. Is E open in Q?

Solution. Suppose $x \in Q \setminus E$. We claim that x is an interior point of the complement of E (which by definition means E is closed). In fact if $x^2 \le 2$, then $x^2 < 2$, since there is no rational number whose square is 2. If $x = 0$, let $\delta = 1$; otherwise let $\delta = \min(\sqrt{\frac{2-x^2}{3}}, \frac{2-x^2}{3|x|})$. Then if $y \in (x - \delta, x + \delta)$, we have $y^2 < 2$. This is obvious if $x = 0$ and $\delta = 1$. In the other case let $y = x+h$, where $|h| < \delta$. Then $y^2 = x^2 + 2xh + h^2 < x^2 + 2|x|\delta + \delta^2 < x^2 + \frac{2}{3}(2 - x^2) + \frac{2-x^2}{3} = 2$. Hence x is an interior point of the complement of E.

Similarly suppose $x^2 \ge 3$. Since there is no rational number whose square is 3, we must have $x^2 > 3$. Since $x \ne 0$, we let $\delta = \frac{x^2-3}{2|x|}$. Then if $y \in (x - \delta, x + \delta)$, we have $y^2 > 3$. For since $y = x + h$, with $|h| < \delta$, and so $y^2 = x^2 + 2xh + h^2 > x^2 - 2|x|\delta = 3$. Thus again x is an interior point of the complement of E.

Hence in all cases $Q \setminus E$ is open, so that E is closed.

That E is bounded is obvious, since $E \subset [-2, 2]$.

To show that E is not compact, let $U_n = \{p : 2 < p^2 < 3 - \frac{1}{n}\}$, $n = 2, 3, \ldots$. The argument that will be used below to show that E is open shows that U_n is open. The sets U_n cover E, but no finite collection of them covers E. Thus E is not compact.

The set E is also open, since if $2 < x^2 < 3$, we can let δ be the minimum of $\sqrt{\frac{3-x^2}{3}}$, $\frac{3-x^2}{3|x|}$, and $\frac{x^2-2}{2|x|}$. Then if $y \in (x - \delta, x + \delta)$, we must have $2 < y^2 < 3$, by the same set of inequalities that was used above.

Exercise 2.17 Let E be the set of all $x \in [0, 1]$ whose decimal expansion contains only the digits 4 and 7. Is E countable? Is E dense in $[0, 1]$? Is E compact? Is E perfect?

Solution. The set E is not countable, since for any hypothetical list of its elements $a_1, a_2, \ldots, a_n, \ldots$ we can always produce an element a of E not in the list by taking the nth digit of a to be 4 if the nth digit of a_n is 7 and equal to 7 if the nth digit of a_n is 4.

The set E is not dense in $[0, 1]$, since $E \subset [0.4, 0.8]$

The set E is closed and bounded, and therefore compact. To show that E is closed, let $x \in [0, 1] \setminus E$, i.e., the decimal expansion of x contains a digit different from 4 and 7. Let the first such digit occur in the nth place (x_n). Let y be any element of E, and let the first digit in which x and y differ be the mth digit ($m \le n$, $x_m \ne y_m$). Then $|x - y| \ge 10^{-m} - \varepsilon$, $\varepsilon \le \sum\limits_{k=m+1}^{\infty} 10^{-k}|x_k - y_k|$. Since $y_k \in \{4, 7\}$ and $x_k \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, it follows that $|x_k - y_k| \le 7$. Hence $\varepsilon \le \frac{7}{9}10^{-m}$, and it follows that $|x - y| \ge \frac{2}{9 \cdot 10^m} \ge \frac{1}{9 \cdot 10^n}$. Thus x is an interior point of $[0, 1] \setminus E$, and so E is closed.

The set E is perfect. For each $x \in E$ and each $\varepsilon > 0$ we can find a point $y \in E$ with $0 < |x - y| < \varepsilon$ by changing the nth digit of x from 4 to 7 or from 7 to 4 in the nth place for any $n > 1 - \log_{10} \varepsilon$. Hence $x \in E'$, i.e., $E \subseteq E'$. Since we already know E is closed, it follows that $E = E'$.

Exercise 2.18 Is there a non-empty perfect set in R^1 which contains no rational number?

Answer. Yes. Let $\{r_1, r_2, \ldots, r_n, \ldots\}$ be the rational numbers in the interval $[-\pi, \pi]$. Let $E_0 = [-\pi, \pi]$. Now assume that E_k has been chosen for $k < n$ in such a way that E_k is a pairwise disjoint union of at most $2^{k+1} - 1$ closed intervals with irrational endpoints, each of *positive* length at most $(\frac{2}{3})^k\pi$ and that E_k does not contain r_j if $j \le k$. (All of these conditions hold trivially for $k = 0$.) Define a set F_{k+1}, which is obtained from E_k by removing first the middle third of each of the intervals that constitute E_k. The result is a set of at most $2^{k+2} - 2$ pairwise disjoint intervals having irrational endpoints, each interval being of length at most $(\frac{2}{3})^{k+1}\pi$. If $r_{k+1} \notin F_{k+1}$, let $E_{k+1} = F_{k+1}$. If $r_{k+1} \in F_{k+1}$, then r_{k+1} is not the endpoint of the interval $I = [a, b]$ of F_{k+1} that it belongs to. Hence let δ be an irrational positive number less than the minimum of $r_{k+1} - a$ and $b - r_{k+1}$, and let E_{k+1} be obtained from F_{k+1} by removing the interval $(r_{k+1} - \delta, r_{k+1} + \delta)$ (which has irrational endpoints). Then E_{k+1} consists of at most $2^{k+2} - 1$ pairwise disjoint closed intervals, each of positive length at most $(\frac{2}{3})^{k+1}\pi$, and each having irrational endpoints.

The sets E_k form a nested sequence of nonempty compact sets. Hence the intersection $E = \bigcap\limits_{k=0}^{\infty}$ is a nonempty compact set. By construction it contains no rational numbers. To show that it is perfect, we merely observe that if $x \in E$, then for each k there is a unique interval $I_k = [a_k, b_k]$, among the finite set of closed intervals constituting the set E_k such that $x \in I_k$. Let $y_k = a_k$ if $a_k \ne x$, otherwise let $y_k = b_k$. In either case $y_k \in E$ (since in our construction no endpoint of any E_k is ever removed) and $|y_k - x| < 2 \cdot 3^{-k}\pi$. Therefore $x \in E'$.

Exercise 2.19 (a) if A and B are disjoint closed sets in some metric space X, prove that they are separated.

(b) Prove the same for disjoint open sets.

(c) Fix $p \in X$, $\delta > 0$, define A to be the set of all $q \in X$ for which $d(p, q) < \delta$, define B similarly with $>$ in place of $<$. Prove that A and B are separated.

(d) Prove that every connected space with at least two points is uncountable. *Hint:* Use (c).

Solution. (a) We are given that $A \cap B = \varnothing$. Since A and B are closed, this means $A \cap \overline{B} = \varnothing = \overline{A} \cap B$, which says that A and B are separated.

(b) Since $X \setminus B$ is a closed set containing A, it follows from Theorem 2.27 (c) that $X \setminus B \supseteq \overline{A}$, i.e., that $\overline{A} \cap B = \varnothing$. Similarly $A \cap \overline{B} = \varnothing$.

(c) The sets A and B are disjoint open sets, hence by part (b) they are separated.

(d) Let $x \in X$ and $y \in X$, and let $d(x, y) = d > 0$. Then for every $\delta \in (0, d)$, there must be a point z such that $d(x, z) = \delta$. (If not, the sets A and B defined in part (c) would separate X.) Hence there is a subset of X that can be placed in one-to-one correspondence with the interval $[0, d]$, and so X is uncountable.

Exercise 2.20 Are closures and interiors of connected sets always connected? (Look at subsets of R^2.)

Answer. The closure of a connected set is connected. Indeed if E is connected and $E \subseteq F \subseteq \overline{E}$, then F is connected. For, suppose $F = G \cup H$, where G and H are separated, nonempty sets. The set E cannot be contained entirely in G. (If it were, since H is nonempty, H would contain a limit point of E, hence a limit point of G, contrary to hypothesis.) For the same reason E cannot be contained entirely in H. Hence $G_1 = E \cap G$ and $H_1 = E \cap H$ are nonempty separated sets such that $E = G_1 \cup H_1$, and E is not connected.

The interior of a connected set may fail to be connected, as we see by letting E be the union of two closed disks in R^2 that are tangent to each other.

Exercise 2.21 Let A and B be separated subsets of some R^k, suppose $\mathbf{a} \in A$, $\mathbf{b} \in B$, and define

$$\mathbf{p}(t) = (1-t)\mathbf{a} + t\mathbf{b}$$

for $t \in R^1$. Put $A_0 = \mathbf{p}^{-1}(A)$, $B_0 = \mathbf{p}^{-1}(B)$. [Thus $t \in A_0$ if and only if $\mathbf{p}(t) \in A$.]

(*a*) Prove that A_0 and B_0 are separated subsets of R^1.

(*b*) Prove that there exists $t_0 \in (0,1)$ such that $\mathbf{p}(t_0) \notin A \cup B$.

(*c*) Prove that every convex subset of R^k is connected.

Solution. (*a*) The definition shows that A_0 and B_0 are disjoint. We need only show that neither contains a limit point of the other. Let x be a limit point of A_0, and suppose $x \in B_0$. This means that for any $\delta > 0$ there exists $t \in A_0$ with $0 < |x-t| < \delta$, $\mathbf{p}(t) = (1-t)\mathbf{b} + t\mathbf{b} \in A$ and $\mathbf{p}(x) = (1-x)\mathbf{a} + x\mathbf{b} \in B$. Now $d(\mathbf{p}(t), \mathbf{p}(x)) = |\mathbf{p}(t) - \mathbf{p}(x)| = |x-t|\,|\mathbf{a}-\mathbf{b}| \le |x-t|\,(|\mathbf{a}| + |\mathbf{b}|) < M\delta$, where $M = |\mathbf{a}| + |\mathbf{b}|$. Since δ is arbitrary, this means that B contains a limit point of A, contrary to hypothesis. This contradiction shows that B_0 contains no limit points of A_0 Likewise A_0 contains no limit points of B_0, and so A_0 and B_0 are separated.

(*b*) If $\mathbf{p}(t) \in A \cup B$ for all $t \in [0,1]$, then $[0,1] \subseteq A_0 \cup B_0$. Hence $[0,1] = G \cup H$, where $G = [0,1] \cap A_0$ and $H = [0,1] \cap B_0$ are both nonempty ($0 \in G$ and $1 \in H$) and separated. This would mean $[0,1]$ is not connected. Therefore $\mathbf{p}(t_0) \notin A \cup B$ for some $t_0 \in [0,1]$, and necessarily $t_0 \in (0,1)$, since $\mathbf{p}(0) = \mathbf{a} \in A$ and $\mathbf{p}(1)\mathbf{b} \in B$.

(*c*) By definition a convex set C is one for which the mapping $\mathbf{p}$ has the property $\mathbf{p}(t) \in C$ for all $t \in [0,1]$ provided $\mathbf{p}(0) = \mathbf{a} \in C$ and $\mathbf{p}(1) = \mathbf{b} \in C$. Hence by part (*b*) there cannot be separated nonempty sets A and B such that $C = A \cup B$.

Exercise 2.22 A metric space is called *separable* if it contains a countable dense subset. Show that R^k is separable. *Hint:* Consider the set of points which have only rational coordinates.

Solution. We need to show that every non-empty open subset E of R^k contains a point with all coordinates rational. Now E contains a ball $B_r(\mathbf{x})$, and this ball contains all points $\mathbf{y}$ such that $(x_j - y_j)^2 < \frac{1}{k}$ for $j = 1, 2, \ldots, k$. Each interval $(x_j - \frac{1}{k}, x_j + \frac{1}{k})$ contains a rational number r_j, and so the point $\mathbf{r} = (r_1, \ldots, r_k)$ belongs to E. Thus E contains a point with only rational coordinates.

Exercise 2.23 A collection $\{V_\alpha\}$ of open sets of X is said to be a *base* for X if the following is true: For every $x \in X$ and every open set $G \subset X$ such that $x \in G$, we have $x \in V_\alpha \subset G$ for some α. In other words, every open set in X is the union of a subcollection of $\{V_\alpha\}$.

Prove that every separable metric space has a *countable* base. *Hint:* Take all neighborhoods with rational radius and center in some countable dense subset of X.

Solution. Let $\{x_1, x_2, \ldots, x_n, \ldots\}$ be a countable dense subset of X. For each positive integer m and each positive rational number r let $V_{m,r} = \{y : d(y, x_m) < r\}$. The collection $V_{m,r}$ is countable.

Let $x \in X$, and let G be any open subset of X with $x \in G$. Then there exists $\delta > 0$ such that $B_\delta(x) \subset G$. The open ball $B_{\frac{\delta}{2}}(x)$ contains a point x_k for some k. Let r be a rational number such that $d(x_k, x) < r < \frac{\delta}{2}$. Then $x \in B_r(x_k) \subset B_\delta(x) \subset G$, and we are done.

Exerçise 2.24 Let X be a metric space in which every infinite subset has a limit point. Prove that X is separable. *Hint:* Fix $\delta > 0$, and pick $x_1 \in X$. Having chosen $x_1, \ldots, x_j \in X$, choose $x_{j-1} \in X$, if possible, so that $d(x_j, x_{j-1}) \geq \delta$ for $i = 1, \ldots, j$. Show that this process must stop after a finite number of steps, and that X can therefore be covered by finitely many neighborhoods of radius δ. Take $\delta = 1/n$ ($n = 1, 2, 3, \ldots$), and consider the centers of the corresponding neighborhoods.

Solution. Following the hint, we observe that if the process of constructing x_j did not terminate, the result would be an infinite set of points x_j, $j = 1, 2, \ldots$, such that $d(x_i, x_j) \geq \delta$ for $i \neq j$. It would then follow that for any $x \in X$, the open ball $B_{\frac{\delta}{2}}(x)$ contains at most one point of the infinite set, hence that no point could be a limit point of this set, contrary to hypothesis. Hence X is *totally bounded*, i.e., for each $\delta > 0$ there is a *finite* set $x_1, \ldots, x_{N_\delta}$ such that $X = \bigcup_{j=1}^{N_\delta} B_\delta(x_j)$.

Let $x_{n1}, \ldots, x_{nN_n}$ be such that $X = \bigcup_{j=1}^{N_n} B_{\frac{1}{n}}(x_{nj})$, $n = 1, 2, \ldots$. We claim that $\{x_{nj} : 1 \leq j \leq N_n; n = 1, 2, \ldots\}$ is a countable dense subset of X. Indeed if $x \in X$ and $\delta > 0$, then $x \in B_{\frac{1}{n}}(x_{nj})$ for some x_{nj} for some $n > \frac{1}{\delta}$, and hence $d(x, x_{nj}) < \delta$. By definition, this means that $\{x_{nj}\}$ is dense in X.

Exercise 2.25 Prove that every compact metric space K has a countable base, and that K is therefore separable. *Hint:* For every positive integer n, there are finitely many neighborhoods of radius $1/n$ whose union covers K.

Solution. It is easier simply to refer to the previous problem. The hint shows that K can be covered by a finite union of neighborhoods of radius $1/n$, and the previous problem shows that this implies that K is separable.

It is not entirely obvious that a metric space with a countable base is separable. To prove this, let $\{V_n\}_{n=1}^{\infty}$ be a countable base, and let $x_n \in V_n$. The points V_n must be dense in X. For if G is any non-empty open set, then G contains V_n for some n, and hence $x_n \in G$. (Thus for a metric space, having a countable base and being separable are equivalent.)

Exercise 2.26 Let X be a metric space in which every infinite subset has a limit point. Prove that X is compact. *Hint:* By Exercises 23 and 24, X has a countable base. It follows that every open cover of X has a *countable* subcover $\{G_n\}_{n=1}$, $n = 1, 2, 3, \ldots$. If no finite subcollection of $\{G_n\}$ covers X, then the complement F_n of $G_1 \cup \cdots \cup G_n$ is nonempty for each n, but $\cap F_n$ is empty. If E is a set which contains a point from each F_n, consider a limit point of E, and obtain a contradiction.

Solution. Following the hint, we consider a set E consisting of one point from the complement of each finite union, i.e., $x_n \notin G_1 \cup \cdots \cup G_n$. Since there are infinitely many finite unions and every point is in *some* set of the covering, the set E cannot be finite. (If $\{x_{i_1}, \ldots, x_{i_n}\}$ is any finite subset of E, there are sets $G_{j_1}, \ldots, G_{j_n}$ such that $x_{i_k} \in G_{j_k}$ for each k. Since E contains a point not in $G_{j_1} \cup \cdots \cup G_{j_n}$, it contains a point different from $x_1, \ldots, x_n$. Hence E is not finite.)

Now by hypothesis E must have a limit point z. The point z must belong to some set G_n; and since G_n is open, there is a number $\delta > 0$ such that $B_\delta(z) \subseteq G_n$. But then $B_\delta(z)$ cannot contain x_m if $m \geq n$, and so z cannot be a limit point of $\{x_m\}$. We have now reached a contradiction.

Exercise 2.27 Define a point p in a metric space X to be a *condensation point* of a set $E \subset X$ if every neighborhood of p contains uncountably many points of E.

Suppose $E \subset R^k$, E is uncountable, and let P be the set of all condensation points of E. Prove that P is perfect and that at most countably many points of E are not in P. In other words, show that $P^c \cap E$ is at most countable. *Hint:*

Let $\{V_n\}$ be a countable base of R^k, let W be the union of those V_n for which $E \cap V_n$ is at most countable, and show that $P = W^c$.

Solution. Following the hint, we see that $E \cap W$ is at most countable, being a countable union of at-most-countable sets. It remains to show that $P = W^c$, and that P is perfect.

If $x \in W^c$, and O is any neighborhood of x, then $x \in V_n \subseteq O$ for some n. Since $x \notin W$, $V_n \cap E$ is uncountable. Hence O contains uncountably many points of E, and so x is a condensation point of E. Thus $x \in P$, i.e., $W^c \subseteq P$.

Conversely if $x \in W$, then $x \in V_n$ for some V_n such that $V_n \cap E$ is countable. Hence x has a neighborhood (any neighborhood contained in V_n) containing at most a countable set of points of E, and so $x \notin P$, i.e., $W \subseteq P^c$. Hence $P = W^c$.

It is clear that P is closed (since its complement W is open), so that we need only show that $P \subseteq P'$. Hence suppose $x \in P$, and O is any neighborhood of x. (By definition of P this means $O \cap E$ is uncountable.) We need to show that there is a point $y \in P \cap (O \setminus \{x\})$. If this is not the case, i.e., if every point y in $O \setminus \{x\}$ is in P^c, then for each such point y there is a set V_n containing y such that $V_n \cap E$ is at most countable. That would mean that $y \in W$, i.e., that $O \setminus \{x\}$ is contained in W. It would follow that $O \cap E \subseteq \{x\} \cup (W \cap E)$, and so $O \cap E$ contains at most a countable set of points, contrary to the hypothesis that $x \in P$. Hence O contains a point of P different from x, and so $P \subseteq P'$. Thus P is perfect.

Remark: This result has now been proved to be true in any separable metric space, not just R^k.

Exercise 2.28 Prove that every closed set in a separable metric space is the union of a (possibly empty) perfect set and a set which is at most countable. (*Corollary:* Every countable closed set in R^k has isolated points.) *Hint:* Use Exercise 27.

Solution. If E is closed, it contains all its limit points, and hence certainly all its condensation points. Thus $E = P \cup (E \setminus P)$, where P is perfect (the set of all condensation points of E), and $E \setminus P$ is at most countable.

Since a perfect set in a separable metric space has the same cardinality as the real numbers, the set P must be empty if E is countable. The at-most-countable set $E \setminus P$ cannot be perfect, hence must have isolated points if it is nonempty.

Exercise 2.29 Prove that every open set in R^1 is the union of an at most countable collection of disjoint segments. *Hint:* Use Exercise 22.

Solution. Let O be open. For each pair of points $x \in O$, $y \in O$, we define an equivalence relation $x \sim y$ by saying $x \sim y$ if and only if $[\min(x,y), \max(x,y)] \subset O$. This is an equivalence relation, since $x \sim x$ ($[x,x] \subset O$ if $x \in O$); if $x \sim y$, then $y \sim x$ (since $\min(x,y) = \min(y,x)$ and $\max(x,y) = \max(y,x)$); and if $x \sim y$ and $y \sim z$, then $x \sim z$ ($[\min(x,z), \max(x,z)] \subseteq [\min(x,y), \max(x,y)] \cup [\min(y,z), \max(y,z)] \subseteq O$). In fact it is easy to prove that

$$\min(x,z) \geq \min\big(\min(x,y), \min(y,z)\big)$$

and

$$\max(x,z) \leq \max\big(\max(x,y), \max(y,z)\big).$$

It follows that O can be written as a disjoint union of pairwise disjoint equivalence classes. We claim that each equivalence class is an open interval.

To show this, for each $x \in O$, let $A = \{z : [z,x] \subseteq O\}$ and $B = \{z : [x,z] \subseteq O\}$, and let $a = \inf A$, $b = \sup B$. We claim that $(a,b) \subset O$. Indeed if $a < z < b$, there exists $c \in A$ with $c < z$ and $d \in B$ with $d > z$. Then $z \in [c,x] \cup [x,d] \subseteq O$. We now claim that (a,b) is the equivalence class containing x. It is clear that each element of (a,b) is equivalent to x by the way in which a and b were chosen. We need to show that if $z \notin (a,b)$, then z is not equivalent to x. Suppose that $z < a$. If z were equivalent to x, then $[z,x]$ would be contained in O, and so we would have $z \in A$. Hence a would not be a lower bound for A. Similarly if $z > b$ and $z \sim x$, then b could not be an upper bound for B.

We have now established that O is a union of pairwise disjoint open intervals. Such a union must be at most countable, since each open interval contains a rational number not in any other interval.

Exercise 2.30 Imitate the proof of Theorem 2.43 to obtain the following result:

> If $R^k = \cup_1^\infty F_n$, where each F_n is a closed subset of R^k, then at least one F_n has a nonempty interior.
>
> *Equivalent statement:* If G_n is a dense open subset of R^k, for $n = 1, 2, 3, \ldots$, then $\cap_1^\infty G_n$ is not empty (in fact, it is dense in R^k).

(This is a special case of Baire's theorem; see Exercise 22, Chap. 3, for the general case.)

Solution. The equivalence of the two statements is easily established. Suppose the first statement is true, and G_n is a dense open subset of R^k for $n = 1, 2, 3, \ldots$. Let $F_n = R^k \setminus G_n$. Then F_n is a closed subset of R^k having empty interior (if the interior of F_n were non-empty, G_n would not be dense). Hence by the first statement, the union of the set F_n cannot be all of R^k, and hence the intersection of their complements is not empty.

Conversely, if the second statement holds and F_n are closed subsets of R^k whose union is all of R^k, let G_n be the complement of F_n. Since the intersection of the G_n's is empty, at least one of them must fail to be dense in R^k, which means that its complement contains a non-empty open set.

We now prove the second statement, including the parenthetical remark. Let G_n be a sequence of dense open sets in R^k, and let O be any non-empty open set in R^k. Since O is an open set and G_1 is dense, it must intersect G_1 in a non-empty open set O_1. Let $x_1 \in O_1$, and choose $r_1 > 0$ such that the *closed* ball $\overline{B_{r_1}(x_1)}$ is contained in O_1. Then the *open* ball $B_{r_1}(x_1)$ is non-empty, and hence must intersect G_2 in a non-empty open set O_2. Let $x_2 \in O_2$, and choose $r_2 > 0$ such that the *closed* ball $\overline{B_{r_2}(x_2)}$ is contained in O_2. In this way we obtain a nested sequence of nonempty compact sets (closed balls) $\overline{B}_1 \supseteq \overline{B}_2 \supseteq \cdots \supseteq \overline{B}_n \supseteq \cdots$. If $x \in \cap \overline{B}_n$, then $x \in O_n$ for each n, and hence $x \in O \cap G_n$ for each n. Thus $\cap G_n$ intersects each non-empty open set O in at least one point, which says precisely that $\cap G_n$ is dense in R^k. Notice that the whole proof works exactly the same way if R^k is replaced by O, since $G_n \cap O$ is dense in O.

Remark: The stronger form of the second statement that we have proved shows that the first statement can also be strengthened. If $\{F_n\}$ is a sequence of closed sets whose union is all of R^k and O is any non-empty open set, then the interior of $F_n \cap O$ is non-empty for at least one n. (Simply apply the original statement with R^k replaced by O and F_k by $F_k \cap O$.)

Chapter 3

Numerical Sequences and Series

Exercise 3.1 Prove that convergence of $\{s_n\}$ implies convergence of $\{|s_n|\}$. Is the converse true?

Solution. Let $\varepsilon > 0$. Since the sequence $\{s_n\}$ is a Cauchy sequence, there exists N such that $|s_m - s_n| < \varepsilon$ for all $m > N$ and $n > N$. We then have $\big||s_m| - |s_n|\big| \le |s_m - s_n| < \varepsilon$ for all $m > N$ and $n > N$. Hence the sequence $\{|s_n|\}$ is also a Cauchy sequence, and therefore must converge.

The converse is not true, as shown by the sequence $\{s_n\}$ with $s_n = (-1)^n$.

Exercise 3.2 Calculate $\lim\limits_{n\to\infty} (\sqrt{n^2+n} - n)$.

Solution. Multiplying and dividing by $\sqrt{n^2+n}+n$ yields

$$\sqrt{n^2+n} - n = \frac{n}{\sqrt{n^2+n}+n} = \frac{1}{\sqrt{1+\frac{1}{n}}+1}.$$

It follows that the limit is $\frac{1}{2}$.

Exercise 3.3 If $s_1 = \sqrt{2}$ and

$$s_{n+1} = \sqrt{2+\sqrt{s_n}} \quad (n = 1, 2, 3 \ldots),$$

prove that $\{s_n\}$ converges, and that $s_n < 2$ for $n = 1, 2, 3 \ldots$.

Solution. Since $\sqrt{2} < 2$, it is manifest that if $s_n < 2$, then $s_{n+1} < \sqrt{2+2} = 2$. Hence it follows by induction that $\sqrt{2} < s_n < 2$ for all n. In view of this fact, it also follows that $(s_n - 2)(s_n + 1) < 0$ for all $n > 1$, i.e., $s_n > s_n^2 - 2 = s_{n-1}$. Hence the sequence is an increasing sequence that is bounded above (by 2) and so converges. Since the limit s satisfies $s^2 - s - 2 = 0$, it follows that the limit is 2.

Exercise 3.4 Find the upper and lower limits of the sequence $\{s_n\}$ defined by

$$s_1 = 0; \quad s_{2m} = \frac{s_{2m-1}}{2}; \quad s_{2m+1} = \frac{1}{2} + s_{2m}.$$

Solution. We shall prove by induction that

$$s_{2m} = \frac{1}{2} - \frac{1}{2^m} \text{ and } s_{2m+1} = 1 - \frac{1}{2^m}$$

for $m = 1, 2, \ldots$. The second of these equalities is a direct consequence of the first, and so we need only prove the first. Immediate computation shows that $s_2 = 0$ and $s_3 = \frac{1}{2}$. Hence assume that both formulas hold for $m \leq r$. Then

$$s_{2r+2} = \frac{1}{2} s_{2r+1} = \frac{1}{2}\left(1 - \frac{1}{2^r}\right) = \frac{1}{2} - \frac{1}{2^{r+1}}.$$

This completes the induction. We thus have $\limsup_{n\to\infty} s_n = 1$ and $\liminf_{n\to\infty} s_n = \frac{1}{2}$.

Exercise 3.5 For any two real sequences $\{a_n\}$, $\{b_n\}$ prove that

$$\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n,$$

provided the sum on the right is not of the form $\infty - \infty$.

Solution. Since the case when $\limsup_{n\to\infty} a_n = +\infty$ and $\limsup_{n\to\infty} b_n = -\infty$ has been excluded from consideration, we note that the inequality is obvious if $\limsup_{n\to\infty} a_n = +\infty$. Hence we shall assume that $\{a_n\}$ is bounded above.

Let $\{n_k\}$ be a subsequence of the positive integers such that $\lim_{k\to\infty}(a_{n_k} + b_{n_k}) = \limsup_{n\to\infty}(a_n + b_n)$. Then choose a subsequence of the positive integers $\{k_m\}$ such that

$$\lim_{m\to\infty} a_{n_{k_m}} = \limsup_{k\to\infty} a_{n_k}.$$

The subsequence $a_{n_{k_m}} + b_{n_{k_m}}$ still converges to the same limit as $a_{n_k} + b_{n_k}$, i.e., to $\limsup_{n\to\infty}(a_n + b_n)$. Hence, since a_{n_k} is bounded above (so that $\limsup_{k\to\infty} a_{n_k}$ is finite), it follows that $b_{n_{k_m}}$ converges to the difference

$$\lim_{m\to\infty} b_{n_{k_m}} = \lim_{m\to\infty}(a_{n_{k_m}} + b_{n_{k_m}}) - \lim_{m\to\infty} a_{n_{k_m}}.$$

Thus we have proved that there exist subsequences $\{a_{n_{k_m}}\}$ and $\{b_{n_{k_m}}\}$ which converge to limits a and b respectively such that $a+b = \limsup_{n\to\infty}(a_n+b_n)$. Since a is the limit of a subsequence of $\{a_n\}$ and b is the limit of a subsequence of $\{b_n\}$, it follows that $a \leq \limsup_{n\to\infty} a_n$ and $b \leq \limsup_{n\to\infty} b_n$, from which the desired inequality follows.

Exercise 3.6 Investigate the behavior (convergence or divergence) of $\sum a_n$ if

(a) $a_n = \sqrt{n+1} - \sqrt{n}$;

(b) $a_n = \dfrac{\sqrt{n+1}-\sqrt{n}}{n}$;

(c) $a_n = (\sqrt[n]{n}-1)^n$;

(d) $a_n = \dfrac{1}{1+z^n}$ for complex values of z.

Solution. (a) Multiplying and dividing a_n by $\sqrt{n+1}+\sqrt{n}$, we find that $a_n = \frac{1}{\sqrt{n+1}+\sqrt{n}}$, which is larger than $\frac{1}{2\sqrt{n+1}}$. The series $\sum a_n$ therefore diverges by comparison with the p series $(p = \frac{1}{2})$.

Alternatively, since the sum telescopes, the nth partial sum is $\sqrt{n+1}-1$, which obviously tends to infinity.

(b) Using the same trick as in part (a), we find that $a_n = \frac{1}{n[\sqrt{n+1}+\sqrt{n}]}$, which is less than $\frac{1}{n^{3/2}}$. Hence the series converges by comparison with the p series $(p = \frac{3}{2})$.

(c) Using the root test, we find that $a_n^{\frac{1}{n}} = \sqrt[n]{n}-1$, which tends to zero as $n \to \infty$. Hence the series converges. (Alternatively, since by part (c) of Theorem 3.20 $\sqrt[n]{n}$ tends to 1 as $n \to \infty$, we have $a_n \leq 2^{-n}$ for all large n, and the series converges by comparison with a geometric series.)

(d) If $|z| \leq 1$, then $|a_n| \geq \frac{1}{2}$, so that a_n does not tend to zero. Hence the series diverges. If $|z| > 1$, the series converges by comparison with a geometric series with $r = \frac{1}{|z|} < 1$.

Exercise 3.7 Prove that the convergence of Σa_n implies the convergence of

$$\sum \frac{\sqrt{a_n}}{n},$$

if $a_n \geq 0$.

Solution. Since $(\sqrt{a_n} - \frac{1}{n})^2 \geq 0$, it follows that

$$\frac{\sqrt{a_n}}{n} \leq \frac{1}{2}\Big(a_n^2 + \frac{1}{n^2}\Big).$$

Now Σa_n^2 converges by comparison with Σa_n (since Σa_n converges, we have $a_n < 1$ for large n, and hence $a_n^2 < a_n$). Since $\Sigma \frac{1}{n^2}$ also converges (p series, $p = 2$), it follows that $\Sigma \frac{\sqrt{a_n}}{n}$ converges.

Exercise 3.8 If Σa_n converges, and if $\{b_n\}$ is monotonic and bounded, prove that $\Sigma a_n b_n$ converges.

Solution. We shall show that the partial sums of this series form a Cauchy sequence, i.e., given $\varepsilon > 0$ there exists N such that $\left| \sum\limits_{k=m+1}^{n} a_k b_k \right| < \varepsilon$ if $n > m \geq N$. To do this, let $S_n = \sum_{k=1}^{n} a_k$ ($S_0 = 0$), so that $a_k = S_k - S_{k-1}$ for $k = 1, 2, \ldots$. Let M be an uppper bound for both $|b_n|$ and $|S_n|$, and let $S = \sum a_n$ and $b = \lim b_n$. Choose N so large that the following three inequalities hold for all $m > N$ and $n > N$:

$$|b_n S_n - bS| < \frac{\varepsilon}{3}; \quad |b_m S_m - bS| < \frac{\varepsilon}{3}; \quad |b_m - b_n| < \frac{\varepsilon}{3M}.$$

Then if $n > m > N$, we have, from the formula for summation by parts,

$$\sum_{k=m+1}^{n} a_n b_n = b_n S_n - b_m S_m + \sum_{k=m}^{n-1} (b_k - b_{k+1}) S_k.$$

Our assumptions yield immediately that $|b_n S_n - b_m S_m| < \frac{2\varepsilon}{3}$, and

$$\left| \sum_{k=m}^{n-1} (b_k - b_{k+1}) S_k \right| \leq M \sum_{k=m}^{n-1} |b_k - b_{k+1}|.$$

Since the sequence $\{b_n\}$ is monotonic, we have

$$\sum_{k=m}^{n-1} |b_k - b_{k+1}| = \left| \sum_{k=m}^{n-1} (b_k - b_{k+1}) \right| = |b_m - b_n| < \frac{\varepsilon}{3M},$$

from which the desired inequality follows.

Exercise 3.9 Find the radius of convergence of each of the following power series

$$(a)\ \sum n^3 z^n, \qquad (b)\ \sum \frac{2^n}{n!} z^n,$$

$$(c)\ \sum \frac{2^n}{n^2} z^n, \qquad (d)\ \sum \frac{n^3}{3^n} z^n.$$

Solution. (*a*) The radius of convergence is 1, since $a_n = n^3$ satisfies $\lim_{n\to\infty} \dfrac{a_n}{a_{n+1}} = 1$.
(*b*) The radius of convergence is infinite, since $a_n = \frac{2^n}{n!}$ satisfies $\lim_{n\to\infty} \dfrac{a_n}{a_{n+1}} = \lim_{n\to\infty} \dfrac{n+1}{2} = \infty$.
(*c*) The radius of convergence is $\frac{1}{2}$, since $a_n = \frac{2^n}{n^2}$ satisfies

$$\lim_{n\to\infty} \frac{a_n}{a_{n+1}} = \lim_{n\to\infty} \frac{1}{2}\left(1+\frac{1}{n}\right)^2 = \frac{1}{2}.$$

(*d*) The radius of convergence is 3, since $a_n = \frac{n^3}{3^n}$ satisfies

$$\lim_{n\to\infty} \frac{a_n}{a_{n+1}} = \lim_{n\to\infty} 3\left(\frac{n}{n+1}\right)^3 = 3.$$

Exercise 3.10 Suppose that the coefficients of the power series $\Sigma a_n z^n$ are integers, infinitely many of which are distinct from zero. Prove that the radius of convergence is at most 1.

Solution. The series diverges if $|z| > 1$, since its general term does not tend to zero. (Infinitely many terms are larger than 1 in absolute value.)

Exercise 3.11 Suppose $a_n > 0$, $s_n = a_1 + \cdots + a_n$, and Σa_n diverges.
(*a*) Prove that $\sum \frac{a_n}{1+a_n}$ diverges.
(*b*) Prove that

$$\frac{a_{N+1}}{s_{N+1}} + \cdots + \frac{a_{N+k}}{s_{N+k}} \geq 1 - \frac{s_N}{s_{N+k}}$$

and deduce that $\sum \frac{a_n}{s_n}$ diverges.
(*c*) Prove that

$$\frac{a_n}{s_n^2} \leq \frac{1}{s_{n-1}} - \frac{1}{s_n}$$

and deduce that $\sum \frac{a_n}{s_n^2}$ converges.
(*d*) What can be said about

$$\sum \frac{a_n}{1+na_n} \text{ and } \sum \frac{a_n}{1_n^2 a_n}?$$

Solution. (*a*) If a_n does not remain bounded, then $\frac{a_n}{1+a_n}$ does not tend to zero, and hence the series $\sum \frac{a_n}{1+a_n}$ diverges. If $a_n \leq M$ for all n, then $\dfrac{a_n}{1+a_n} \geq \dfrac{1}{1+M} a_n$, and hence again the series is divergent.

(b) Replacing each denominator on the left by s_{N+k}, we have

$$\frac{a_{N+1}}{s_{N+1}} + \cdots + \frac{a_{N+k}}{s_{N+k}} \geq \frac{1}{s_{N+k}}(a_{N+1} + a_{N+2} + \cdots + a_{N+k}) =$$
$$= \frac{1}{s_{N+k}}(s_{N+k} - s_N) = 1 - \frac{s_N}{s_{N+k}}.$$

It follows that the partial sums of the series $\sum \frac{a_n}{s_n}$ do not form a Cauchy sequence. For, no matter how large N is taken, if N is held fixed, the right hand side can be made larger than $\frac{1}{2}$ by taking k sufficiently large (since $S_{N+k} \to \infty$).
(c) We observe that if $n \geq 2$, then

$$\frac{1}{s_{n-1}} - \frac{1}{s_n} = \frac{s_n - s_{n-1}}{s_{n-1}s_n} = \frac{a_n}{s_{n-1}s_n} \geq \frac{a_n}{s_n^2}.$$

Since the series $\sum\limits_{n=2}^{\infty} \frac{1}{s_{n-1}} - \frac{1}{s_n}$ converges to $\frac{1}{a_1}$, it follows by comparison that $\sum \frac{a_n}{s_n^2}$ converges.
(d) The series $\sum \frac{a_n}{1+na_n}$ may be either convergent or divergent. If the sequence $\{na_n\}$ is bounded above or has a positive lower bound, it definitely diverges. Thus if $na_n \leq M$, each term is at least $\frac{1}{1+M}a_n$, and so the series diverges. If $na_n \geq \varepsilon > 0$ for all n, then each term is at least $\frac{\varepsilon}{1+\varepsilon}\frac{1}{n}$, and once again the series is divergent.

In general, however, the series $\sum \frac{a_n}{1+na_n}$ may converge. For example let $a_n = \frac{1}{n^2}$ if n is not a perfect square and $a_n = \frac{1}{\sqrt{n}}$ if n is a perfect square. The sum of $\frac{a_n}{1+na_n}$ over the nonsquares obviously converges by comparison with the p series, $p = 2$. As for the sum over the square integers it is $\sum \frac{1}{n+n^2}$, which converges by comparison with the p series, $p = 2$.

Finally, the series $\sum \frac{a_n}{1+n^2a_n}$ is obviously majorized by the p series with $p = 2$, hence converges.

Exercise 3.12 Suppose $a_n > 0$ and $\sum a_n$ converges. Put

$$r_n = \sum_{m=n}^{\infty} a_n.$$

(a) Prove that

$$\frac{a_m}{r_m} + \cdots + \frac{a_n}{r_n} > 1 - \frac{r_n}{r_m}$$

if $m < n$, and deduce that $\sum \frac{a_n}{r_n}$ diverges.
(b) Prove that

$$\frac{a_n}{\sqrt{r_n}} < 2(\sqrt{r_n} - \sqrt{r_{n+1}})$$

and deduce that $\sum \frac{a_n}{\sqrt{r_n}}$ converges.

Solution. (a) Replacing all the denominators on the left-hand side by the largest one (r_m), we find

$$\frac{a_m}{r_m} + \cdots + \frac{a_n}{r_n} > \frac{a_m + \cdots + a_n}{r_m} = \frac{r_m - r_{n+1}}{r_m} > 1 - \frac{r_n}{r_m},$$

since $r_n > r_{n+1}$.

As in the previous problem, this keeps the partial sums of the series $\sum \frac{a_n}{r_n}$ from forming a Cauchy sequence. No matter how large m is taken, one can choose n larger so that the difference $\sum_{k=m}^{n} \frac{a_k}{r_k}$ is at least $\frac{1}{2}$, since $r_n \to 0$ as $n \to \infty$.

(b) We have

$$\frac{a_n}{\sqrt{r_n}}(\sqrt{r_n} + \sqrt{r_{n+1}}) = a_n + a_n \frac{\sqrt{r_{n+1}}}{\sqrt{r_n}} < 2a_n = 2(r_n - r_{n+1}).$$

Dividing both sides by $\sqrt{r_n} + \sqrt{r_{n+1}}$ now yields the desired inequality.

Since the series $\sum(\sqrt{r_n} - \sqrt{r_{n+1}})$ converges to $\sqrt{r_1}$, it follows by comparison that $\sum \frac{a_n}{\sqrt{r_n}}$ converges.

Exercise 3.13 Prove that the Cauchy product of two absolutely convergent series converges absolutely.

Solution. Since both the hypothesis and conclusion refer to absolute convergence, we may assume both series consist of nonnegative terms. We let $S_n = \sum_{k=0}^{n} a_n$, $T_n = \sum_{k=0}^{n} b_n$, and $U_n = \sum_{k=0}^{n} \sum_{l=0}^{k} a_l b_{k-l}$. We need to show that U_n remains bounded, given that S_n and T_n are bounded. To do this we make the convention that $a_{-1} = T_{-1} = 0$, in order to save ourselves from having to separate off the first and last terms when we sum by parts. We then have

$$\begin{aligned}
U_n &= \sum_{k=0}^{n} \sum_{l=0}^{k} a_l b_{k-l} \\
&= \sum_{k=0}^{n} \sum_{l=0}^{k} a_l (T_{k-l} - T_{k-l-1}) \\
&= \sum_{k=0}^{n} \sum_{j=0}^{k} a_{k-j} (T_j - T_{j-1}) \\
&= \sum_{k=0}^{n} \sum_{j=0}^{k} (a_{k-j} - a_{k-j-1}) T_j \\
&= \sum_{j=0}^{n} \sum_{k=j}^{n} (a_{k-j} - a_{k-j-1}) T_j
\end{aligned}$$

$$\begin{aligned} &= \sum_{j=0}^{n} a_{n-j}T_j \\ &\leq T\sum_{m=0}^{n} a_m \\ &= TS_n \\ &\leq ST. \end{aligned}$$

Thus U_n is bounded, and hence approaches a finite limit.

Exercise 3.14 If $\{s_n\}$ is a complex sequence, define its arithmetic mean σ_n by

$$\sigma_n = \frac{s_0 + s_1 + \cdots + s_n}{n+1} \quad (n = 0, 1, 2, \ldots).$$

(*a*) If $\lim s_n = s$, prove that $\lim \sigma_n = s$.
(*b*) Construct a sequence $\{s_n\}$ which does not converge, although $\lim \sigma_n = 0$.
(*c*) Can it happen that $s_n > 0$ for all n and that $\limsup s_n = \infty$, even though $\lim \sigma_n = 0$?
(*d*) Put $a_n = s_n - s_{n-1}$ for $n \geq 1$. Show that

$$s_n - \sigma_n = \frac{1}{n+1}\sum_{k=1}^{n} ka_k.$$

Assume that $\lim(na_n) = 0$ and that $\{\sigma_n\}$ converges. Prove that $\{s_n\}$ converges. [This gives a converse of (*a*), but under the additional assumption that $na_n \to 0$.]
(*e*) Derive the last conclusion from a weaker hypothesis: Assume $M < \infty$, $|na_n| \leq M$ for all n, and $\lim \sigma_n = \sigma$. Prove that $\lim s_n = \sigma$ by completing the following outline:
If $m < n$, then

$$s_n - \sigma_n = \frac{m+1}{n-m}(\sigma_n - \sigma_m) + \frac{1}{n-m}\sum_{i=m+1}^{n}(s_n - s_i).$$

For these i,

$$|s_n - s_i| \leq \frac{(n-i)M}{i+1} \leq \frac{(n-m-1)M}{m+2}.$$

Fix $\varepsilon > 0$ and associate with each n the integer m that satisfies

$$m \leq \frac{n-\varepsilon}{1+\varepsilon} < m+1.$$

then $(m+1)/(n-m) \leq 1/\varepsilon$ and $|s_n - s_i| < M\varepsilon$. Hence

$$\limsup_{n\to\infty} |s_n - \sigma| \leq M\varepsilon.$$

Since ε was arbitrary, $\lim s_n = \sigma$.

Solution. Let $\varepsilon > 0$. Let $M = \sup\{|s_n|\}$, and let N_0 be the first integer such that $|s_n - s| < \frac{\varepsilon}{2}$ for all $n > N_0$. Let $N = \max\left(N_0, \left[\frac{2(N_0+1)(M+|s|)}{\varepsilon}\right]\right)$. Then if $n > N$, we have

$$\begin{aligned}
|\sigma_n - s| &= \left|\frac{(s_0 - s) + (s_1 - s) + \cdots + (s_n - s)}{n+1}\right| \\
&\le \left|\frac{(s_0 - s) + \cdots + (s_{N_0} - s)}{n+1}\right| + \\
&\qquad + \left|\frac{(s_{N_0+1}) - s) + \cdots + (s_n - s)}{n+1}\right|.
\end{aligned}$$

The first sum on the right-hand side here is at most $\frac{(N_0+1)(M+|s|)}{n+1}$, and since $n+1 > \frac{2(N_0+1)(M+|s|)}{\varepsilon}$, this sum is at most $\frac{\varepsilon}{2}$. The second sum is at most $\frac{(n-N_0)\frac{\varepsilon}{2}}{n+1}$, which is at most $\frac{\varepsilon}{2}$. Thus $|\sigma_n - s| < \varepsilon$ if $n > N$, which was to be proved.

(*b*) Let $s_n = (-1)^n$. Here σ_n is 0 if n is odd and $\frac{1}{n+1}$ if n is even. Thus $\sigma_n \to 0$, though s_n has no limit.

(*c*) Let $s_n = \frac{1}{n}$ if n is not a perfect cube and $s_n = \sqrt[3]{n}$ if n is a perfect cube. Then if $k^3 \le n < (k+1)^3$ we have

$$\begin{aligned}
\sigma_n &\le \frac{1}{n+1}\sum_{m=1}^{n}\frac{1}{m} + \frac{1}{n+1}\sum_{j=1}^{k} j \\
&= \frac{1}{n+1}\left(\sum_{m=1}^{n}\frac{1}{m}\right) + \frac{1}{n+1}\cdot\frac{k(k+1)}{2}.
\end{aligned}$$

The first sum on the right tends to zero by part (*a*) applied to the sequence $s_0 = 0$, $s_n = \frac{1}{n}$ for $n \ge 1$. As for the last term, since $n \ge k^3$, it is less than $\frac{1}{2k} + \frac{1}{2k^2}$, which tends to zero as $k \to \infty$. Since $(k+1)^3 > n$, it follows that k tends to infinity as n tends to infinity, and hence we have $\sigma_n \to 0$, even though $s_{n^3} \to \infty$.

(*d*) If we set $a_0 = s_0$, we have $s_n = \sum_{k=0}^{n} a_k$. Then

$$\begin{aligned}
s_n - \sigma_n &= s_n - \frac{s_0 + s_1 + \cdots + s_n}{n+1} \\
&= (a_0 + a_1 + \cdots + a_{n-1} + a_n) - \\
&\qquad \frac{(n+1)a_0 + na_1 + \cdots + 2a_{n-1} + a_n}{n+1} \\
&= \frac{a_1 + 2a_2 + \cdots + (n-1)a_{n-1} + na_n}{n+1},
\end{aligned}$$

which was to be proved. If $na_n \to 0$, then the expression on the right-hand side tends to zero by part (*a*) with s_n replaced by na_n. Hence $s_n - \sigma_n \to 0$.

(*e*) If $m < n$ we have

$$\begin{aligned}\sigma_n - \sigma_m &= \frac{s_0 + \cdots + s_n}{n+1} - \frac{s_0 + \cdots + s_m}{m+1} \\ &= (s_0 + \cdots + s_n)\Big(\frac{1}{n+1} - \frac{1}{m+1}\Big) + \sum_{i=m+1}^{n} \frac{s_i}{m+1} \\ &= \frac{m-n}{m+1}\sigma_n + \frac{1}{m+1}\sum_{i=m+1}^{n} s_i.\end{aligned}$$

If we multiply both sides of this equation by $\frac{m+1}{m-n}$, and then transpose the left-hand side to the right and the term σ_n to the left, we obtain

$$-\sigma_n = \frac{m+1}{n-m}(\sigma_n - \sigma_m) - \frac{1}{n-m}\sum_{i=m+1}^{n} s_i.$$

Adding $s_n = \frac{1}{n-m}\sum_{i=m+1}^{n} s_n$ to both sides then yields the result.

We then have

$$|s_n - s_i| = |a_{i+1} + \cdots + a_n| \le M\Big(\frac{1}{i+1} + \cdots + \frac{1}{n}\Big) \le \frac{(n-i)M}{i+1}.$$

Since the function $\frac{n-x}{x+1} = \frac{n+1}{x+1} - 1$ is decreasing, the maximal value of the right-hand side here is reached with $i = m+1$, so that $|s_n - s_i| \le \frac{(n-m-1)M}{m+2}$, as asserted.

When we choose m to be the largest integer in $\frac{n-\varepsilon}{1+\varepsilon}$, we clearly have $m < n$. Since ε is fixed, we can assume $m > \varepsilon$. The inequality $\frac{n-\varepsilon}{1+\varepsilon} < m+1$ can easily be converted to $\frac{n-m-1}{m+2} < \varepsilon$, and the inequality $m \le \frac{n-\varepsilon}{1+\varepsilon}$ likewise becomes $\frac{m+1}{n-m} \le \frac{1}{\varepsilon}$. The first of these implies that $m \to \infty$ as $n \to \infty$, and we have

$$|s_n - \sigma_n| \le \frac{1}{\varepsilon}|\sigma_n - \sigma_m| + M\varepsilon$$

for all n. This implies that the limit of any subsequence of $|s_n - \sigma_n|$ is at most $M\varepsilon$, and since ε is arbitrary, every convergent subsequence of $|s_n - \sigma_n|$ converges to zero. This, of course, implies that $s_n - \sigma_n$ tends to zero, so that if $\sigma_n \to s$, then $s_n \to s$.

Exercise 3.15 Definition 3.21 can be extended to the case in which the a_n lie in some fixed R^k. Absolute convergence is defined as convergence of $\sum |\mathbf{a}_n|$. Show that Theorems 3.22, 3.23, 3.25(*a*), 3.33, 3.34, 3.42, 3.45, 3.47, and 3.55 are true in this more general setting. (Only slight modifications are required in any of the proofs.)

Solution. (Theorem 3.22). $\sum \mathbf{a}_n$ *converges if and only if for every* $\varepsilon > 0$ *there is an integer* N *such that*

$$\left| \sum_{k=n}^{m} \mathbf{a}_k \right| \leq \varepsilon$$

if $m \geq n \geq N$.

It is a trivial remark that, since $|a_j - b_j| \leq |\mathbf{a} - \mathbf{b}| \leq |a_1 - b_1| + \cdots + |a_k - b_k|$, the sequence $\{\mathbf{a}_n\}$ converges if and only if each sequence of components $\{a_{nj}\}$ converges, $j = 1, \ldots, k$. Hence the sequence of vector-valued functions converges if and only if each sequence of its components is a Cauchy sequence, and by the same inequalities, this is equivalent to saying that the vector-valued sequence is a Cauchy sequence.

(Theorem 3.23) *If* $\sum \mathbf{a}_n$ *converges, then* $\lim_{n\to\infty} \mathbf{a}_n = \mathbf{0}$.

Using the remark made in the previous paragraph, if $\sum \mathbf{a}_n$ converges, then each sum of components $\sum a_{nj}$ converges. Hence for each j we have $a_{nj} \to 0$, which, again by the remark, means $\mathbf{a}_n \to \mathbf{0}$.

(Theorem 3.25 (a)) *If* $|\mathbf{a}_n| \leq c_n$ *for* $n \geq N_0$, *where* N_0 *is some fixed integer, and if* $\sum c_n$ *converges, then* $\sum \mathbf{a}_n$ *converges.*

Again, the hypothesis implies that $|a_{nj}| \leq c_n$ for $n \geq N_0$, so that $\sum a_{nj}$ converges for each $j = 1, 2, \ldots, k$. Once again, by the remark, this means that $\sum \mathbf{a}_n$ converges.

(Theorem 3.33) *Given* $\sum \mathbf{a}_n$, *put* $\alpha = \limsup\limits_{n\to\infty} \sqrt[n]{|\mathbf{a}_n|}$. *Then*

(a) *if* $\alpha < 1$, $\sum \mathbf{a}_n$ *converges;*

(b) *if* $\alpha > 1$, $\sum \mathbf{a}_n$ *diverges;*

(c) *if* $\alpha = 1$, *the test gives no information.*

Part (a) follows from the remarks made above, since $\sqrt[n]{|a_{nj}|} \leq \sqrt[n]{|\mathbf{a}_n|}$. (If $\alpha < 1$, then each component series converges.)

As for part (b), if $\alpha > 1$, then $|\mathbf{a}_n| > 1$ for infinitely many n, and hence the series diverges.

(Theorem 3.34) *The series* $\sum \mathbf{a}_n$

(a) *converges if* $\limsup\limits_{n\to\infty} \dfrac{|\mathbf{a}_{n+1}|}{|\mathbf{a}_n|} < 1$,

(b) *diverges if* $\dfrac{|\mathbf{a}_{n+1}|}{|\mathbf{a}_n|} \geq 1$ *for* $n \geq n_0$, *where* n_0 *is some fixed integer.*

(a) The inequality implies that for some constant A and some fixed $r < 1$ we have $|\mathbf{a}_n| < Ar^n$, so that $\sum |\mathbf{a}_n|$ converges. Therefore by 3.25 the series $\sum \mathbf{a}_n$ also converges.

(b) As in the numerical case, this inequality implies that $\mathbf{a}_n$ does not tend to zero, so that the series must diverge.

(Theorem 3.42) *Suppose*

(*a*) *the partial sums* $\mathbf{A}_n$ *of* $\sum \mathbf{a}_n$ *form a bounded sequence;*
(*b*) $b_0 \geq b_1 \geq b_2 \geq \cdots$;
(*c*) $\lim_{n\to\infty} b_n = 0$.
Then $\sum b_n \mathbf{a}_n$ *converges.*

We reduce this to Theorem 3.22 by showing that the partial sums of the series $\sum b_n \mathbf{a}_n$ form a Cauchy sequence. In fact

$$\begin{aligned}\left|\sum_{n=p}^{q} b_n \mathbf{a}_n\right| &= \left|\sum_{n=p}^{q-1}(b_n - b_{n+1})\mathbf{A}_n + b_q\mathbf{A}_q - b_p\mathbf{A}_{p-1}\right| \\ &\leq M\Big(\sum_{n=p}^{q-1}|b_n - b_{n+1}| + b_q + b_p\Big) \\ &\leq 2Mb_p.\end{aligned}$$

Now, given $\varepsilon > 0$ choose N so large that $b_p < \frac{\varepsilon}{2M}$ for all $p > N$. Then if $q \geq p > N$, we have

$$\left|\sum_{n=p}^{q} b_n \mathbf{a}_n\right| \leq 2Mb_p < \varepsilon.$$

This proves that the partial sums form a Cauchy sequence, as required.

(Theorem 3.45) *If* $\sum \mathbf{a}_n$ *converges absolutely, then* $\sum \mathbf{a}_n$ *converges.*
Again this is a consequence of 3.25, with $c_n = |\mathbf{a}_n|$.

(Theorem 3.47) *If* $\sum \mathbf{a}_n = \mathbf{A}$ *and* $\sum \mathbf{b}_n = \mathbf{B}$, *then* $\sum(\mathbf{a}_n + \mathbf{b}_n) = \mathbf{A} + \mathbf{B}$ *and* $\sum c\mathbf{a}_n = c\mathbf{A}$ *for any fixed c.*
This theorem holds for each component of the vectors involved, hence it holds for the vectors themselves.

(Theorem 3.55) *If* $\sum \mathbf{a}_n$ *is a series of vectors which converges absolutely, then every rearrangement of* $\sum \mathbf{a}_n$ *converges, and they all converge to the same sum.*
Let $\mathbf{A}$ be the sum of the series in its original arrangement, and let $\varepsilon > 0$. Choose N so large that $\sum_{k=m}^{n} |\mathbf{a}_k| < \frac{\varepsilon}{2}$ if $n \geq m > N$. Then of course $\left|\sum_{k=1}^{n} \mathbf{a}_k - \mathbf{A}\right| \leq \frac{\varepsilon}{2}$ if $n > N$. For any arrangement of the series $\sum \mathbf{a}_{n_k}$, Choose N_1 so large that $\{1, 2, \ldots, N\} \subseteq \{n_1, n_2, \ldots, n_{N_1}\}$. Then if $m > N_1$ and N_2 is such that $\{n_1, \ldots, n_m\} \subseteq \{1, \ldots, N_2\}$ have,

$$\begin{aligned}\left|\sum_{k=1}^{m} \mathbf{a}_{n_k} - \mathbf{A}\right| &\leq \left|\sum_{k=1}^{m} \mathbf{a}_{n_k} - \sum_{k=1}^{m} \mathbf{a}_k\right| + \left|\sum_{k=1}^{m} \mathbf{a}_k - \mathbf{A}\right| \\ &\leq \sum_{k=N+1}^{m} |\mathbf{a}_k| + \frac{\varepsilon}{2} \\ &< \varepsilon\end{aligned}$$

Exercise 3.16 Fix a positive number α. Choose $x_1 > \sqrt{\alpha}$, and define x_1, x_2, $x_3, \ldots$, by the recursion formula

$$x_{n+1} = \frac{1}{2}\Big(x_n + \frac{\alpha}{x_n}\Big).$$

(a) Prove that $\{x_n\}$ decreases monotonically and that $\lim x_n = \sqrt{\alpha}$.
(b) Put $\varepsilon = x_n - \sqrt{\alpha}$, and show that

$$\varepsilon_{n+1} = \frac{\varepsilon_n^2}{2x_n} < \frac{\varepsilon_n^2}{2\sqrt{\alpha}}$$

so that, setting $\beta = 2\sqrt{\alpha}$,

$$\varepsilon_{n+1} < \beta\Big(\frac{\varepsilon_1}{\beta}\Big)^{2^n} \quad (n = 1, 2, 3, \ldots,).$$

(c) This is a good algorithm for computing square roots, since the recursion formula is simple and the convergence is extremely rapid. For example, if $\alpha = 3$ and $x_1 = 2$, show that $\varepsilon_1/\beta < \frac{1}{10}$, and that therefore

$$\varepsilon_5 < 4 \cdot 10^{-16}, \quad \varepsilon_6 < 4 \cdot 10^{-32}.$$

Solution. (a) We note that x_n is always positive, and that if $x_n > \sqrt{\alpha}$, then $x_{n+1}^2 - \alpha = \frac{1}{4}\big(x_n - \frac{\alpha}{x_n}\big)^2 > 0$. Thus $x_n > \sqrt{\alpha}$ for all n. Since $x_n > \sqrt{\alpha}$, it follows that $\frac{\alpha}{x_n} < \sqrt{\alpha} < x_n$. Hence $x_n - x_{n+1} = \frac{1}{2}\Big(x_n - \frac{\alpha}{x_n}\Big) > 0$, and so $\{x_n\}$ decreases to a limit $\lambda \geq \sqrt{\alpha}$, which must satisfy $\lambda = \frac{\alpha}{\lambda}$, i.e., $\lambda = \sqrt{\alpha}$.
(b) We have $\frac{\varepsilon_n^2}{2x_n} = \frac{x_n^2 - 2x_n\sqrt{\alpha}+\alpha}{2x_n} = \frac{1}{2}\big(x_n + \frac{\alpha}{x_n}\big) - \sqrt{\alpha} = x_{n+1} - \sqrt{\alpha} = \varepsilon_{n+1}$. The inequality then results from the simple fact that $x_n > \sqrt{\alpha}$. Thus $\varepsilon_2 < \frac{\varepsilon_1^2}{\beta} = \beta\big(\frac{\varepsilon_1}{\beta}\big)^2$. By induction, if we suppose that $\varepsilon_n < \beta\big(\frac{\varepsilon_1}{\beta}\big)^{2^{n-1}}$, we find $\varepsilon_{n+1} < \frac{\varepsilon_n^2}{\beta} < \beta\big(\frac{\varepsilon_1}{\beta}\big)^{2^n}$.
(d) Taking $x_1 = 2$, $\alpha = 3$, we certainly have $\beta < 4$. And, since $\sqrt{3} > \frac{5}{3}$, we deduce that $12\sqrt{3} > 20$, so that $2\sqrt{3} > 10(2 - \sqrt{3})$, i.e., $\varepsilon_1 = 2 - \sqrt{3}$ and $\beta = 2\sqrt{3}$ satisfy $\varepsilon_1/\beta < \frac{1}{10}$, as asserted. It follows that $\varepsilon_n < 4 \cdot 10^{-2^{n-1}}$. In particular $\varepsilon_5 < 4 \cdot 10^{-16}$ and $\varepsilon_6 < 4 \cdot 10^{-32}$.

Exercise 3.17 Fix $\alpha > 1$. Take $x_1 > \sqrt{\alpha}$, and define

$$x_{n+1} = \frac{\alpha + x_n}{1 + x_n} = x_n + \frac{\alpha - x_n^2}{1 + x_n}.$$

(a) Prove that $x_1 > x_3 > x_5 > \cdots$.

(b) Prove that $x_2 < x_4 < x_6 < \cdots$.

(c) Prove that $\lim x_n = \sqrt{\alpha}$.

(d) Compare the rapidity of convergence of this process with the one described in Exercise 16.

Solution. Most of the work in this problem is done by the following three identities, whose proofs are routine computations:

$$(1+x_n)(1+x_{n+1}) = 2(1+x_n)+(\alpha-1),$$

$$x_{n+1}^2-\alpha = -\left[\frac{(\alpha-1)}{(1+x_n)^2}\right](x_n^2-\alpha),$$

$$x_{n+1}^2-\alpha = \frac{(\alpha-1)^2}{(1+x_n)^2(1+x_{n-1})^2}(x_{n-1}^2-\alpha) = $$

$$= \left[\frac{\alpha-1}{(\alpha-1)+2(1+x_{n-1})}\right]^2(x_{n-1}^2-\alpha).$$

The second of these identities shows that x_n and x_{n+1} lie on opposite sides of $\sqrt{\alpha}$. The third shows that x_{n+1} is closer to $\sqrt{\alpha}$ than x_{n-1}. Hence, since $x_1 > \sqrt{\alpha}$ by hypothesis, parts (a) and (b) are proved. As for (c), the third relation shows that $|x_{n+1}^2-\alpha| \le r^2|x_{n-1}^2-\alpha|$, where $r = \frac{\alpha-1}{2+\alpha-1} < 1$. It follows that $|x_{n+2k}^2-\alpha| \le r^{2k}|x_n^2-\alpha|$, and the right-hand side of this expression tends to zero as $k \to \infty$. Thus $\lim\limits_{k\to\infty} x_{n+2k} = \sqrt{\alpha}$ whether n is odd or even, and so $\lim\limits_{n\to\infty} x_n = \sqrt{\alpha}$.

The convergence in this case is geometric, but not quadratically geometric, as in Exercise 16. The rate of convergence will depend on the size of α. For $1 < \alpha \le 2$ we certainly have $x_n \ge \alpha - 2$ for all n, and so in this case $r < \frac{1}{3}$, i.e., $|x_{n+1}^2-\alpha| < \frac{1}{9}|x_{n-1}^2-\alpha|$. This implies that $|x_{n+1}-\sqrt{\alpha}| < \frac{1}{9}\frac{x_{n-1}+\sqrt{\alpha}}{x_{n+1}+\sqrt{\alpha}}|x_{n-1}-\sqrt{\alpha}|$. If n is odd, we have $x_{n-1} < x_{n+1}$, and so $|x_{n+1}-\sqrt{\alpha}| < \frac{1}{9}|x_{n-1}-\sqrt{\alpha}|$. If n is even, we can at least assume $x_1 < 1.5$ (since $\alpha \le 2$), and so $\frac{x_{n-1}+\sqrt{\alpha}}{x_{n+1}+\sqrt{\alpha}} < 1.5$, so that $|x_{n+1}-\sqrt{\alpha}| < \frac{1.5}{9}|x_{n-1}-\sqrt{\alpha}|$.

Exercise 3.18 Replace the recursion formula of Exercise 16 by

$$x_{n+1} = \frac{p-1}{p}x_n + \frac{\alpha}{p}x_n^{-p+1},$$

where p is a fixed positive integer, and describe the behavior of the resulting sequences $\{x_n\}$.

Solution. (Exercise 16 is the case $p = 2$, of course.) The main work is done by the following easily derived formulas, which hold if $x_n > \alpha^{\frac{1}{p}}$.

$$x_{n+1}-\alpha^{\frac{1}{p}} = (x_n-\alpha^{\frac{1}{p}})\left[\left(\frac{p-1}{p}\right)-\frac{1}{p}\left(\left(\frac{\alpha^{\frac{1}{p}}}{x_n}\right)+\cdots+\left(\frac{\alpha^{\frac{1}{p}}}{x_n}\right)^{p-1}\right)\right]$$

$$
\begin{aligned}
&< (x_n - \alpha^{\frac{1}{p}})\Big(\frac{p-1}{p}\Big)\Big(1 - \Big(\frac{\alpha^{\frac{1}{p}}}{x_n}\Big)^{p-1}\Big) \\
&= (x_n - \alpha^{\frac{1}{p}})\Big(\frac{p-1}{px_n^{p-1}}\Big)(x_n^{p-1} - (\alpha^{\frac{1}{p}})^{p-1}) \\
&= (x_n - \alpha^{\frac{1}{p}})^2 \cdot \frac{p-1}{px_n^{p-1}} \cdot [x_n^{p-2} + x_n^{p-3}\alpha^{\frac{1}{p}} + \cdots + \alpha^{\frac{p-2}{p}}] \\
&< (x_n - \alpha^{\frac{1}{p}})^2 \cdot \frac{(p-1)^2}{px_n} \\
&< (x_n - \alpha^{\frac{1}{p}})^2 \cdot \frac{(p-1)^2}{p\alpha^{\frac{1}{p}}}.
\end{aligned}
$$

Thus we can guarantee quadratic-geometric convergence if we start with $x_1 - \alpha^{\frac{1}{p}} = \varepsilon_1 < \beta = \frac{p\alpha^{\frac{1}{p}}}{(p-1)^2}$. In that case we obtain the same inequalities as in Exercise 16, and $x_n \to \alpha^{\frac{1}{p}}$.

Exercise 3.19 Associate to each sequence $a = \{\alpha_n\}$, in which α_n is 0 or 2, the real number

$$x(a) = \sum_{n=1}^{\infty} \frac{\alpha_n}{3^n}.$$

Prove that the set of all $x(a)$ is precisely the Cantor set described in Sec. 2.44.

Solution. We note that the open middle third removed at the first stage of the construction is precisely the set of points whose ternary expansions *must* have a 1 as their first digit. (The numbers $\frac{1}{3}$ and $\frac{2}{3}$ *can* be written with a 1 in this place, since

$$
\begin{aligned}
\frac{1}{3} &= \frac{1}{3} + \frac{0}{9} + \cdots + \frac{0}{3^n} + \cdots \\
\frac{2}{3} &= \frac{1}{3} + \frac{2}{9} + \cdots + \frac{2}{3^n} + \cdots .
\end{aligned}
$$

However, these numbers can also be written as

$$
\begin{aligned}
\frac{1}{3} &= \frac{0}{3} + \frac{2}{9} + \cdots + \frac{2}{3^n} + \cdots \\
\frac{2}{3} &= \frac{2}{3} + \frac{0}{9} + \cdots + \frac{0}{3^n} + \cdots .
\end{aligned}
$$

Thus the points retained in the Cantor set after the first dissection are precisely those whose ternary expansions *may* be written without a 1 in the first digit. The same argument shows that the points retained in the Cantor set after the nth dissection are precisely those whose ternary expansions *may* be written without using a 1 in any of the first n digits. It then follows that the Cantor set is the set of points in $[0, 1]$ whose ternary expanions can be written without using any 1's, i.e., it is precisely the set of numbers $x(a)$ just described.

Exercise 3.20 Suppose $\{p_n\}$ is a Cauchy sequence in a metric space X, and some subsequence $\{p_n\}$ converges to a point $p \in X$. Prove that the full sequence $\{p_n\}$ converges to p.

Solution. Let $\varepsilon > 0$. Choose N_1 so large that $d(p_m, p_n) < \frac{\varepsilon}{2}$ if $m > N_1$ and $n > N_1$. Then choose $N \geq N_1$ so large that $d(p_{n_k}, p) < \frac{\varepsilon}{2}$ if $k > N$. Then if $n > N$, we have

$$d(p_n, p) \leq d(p_n, p_{n_{N+1}}) + d(p_{n_{N+1}}, p) < \varepsilon.$$

For the first term on the right is less than $\frac{\varepsilon}{2}$ since $n > N_1$ and $n_{N+1} > N+1 > N_1$. The second term is less than $\frac{\varepsilon}{2}$ by the choice of N.

Exercise 3.21 Prove the following analogue of Theorem 3.10(*b*): If $\{E_n\}$ is a sequence of closed and bounded sets in a *complete* metric space X, if $E_n \supset E_{n+1}$, and if

$$\lim_{n\to\infty} \operatorname{diam} E_n = 0,$$

then $\cap_1^\infty E_n$ consists of exactly one point.

Solution. Choose $x_n \in E_n$. (We use the axiom of choice here.) The sequence $\{x_n\}$ is a Cauchy sequence, since the diameter of E_n tends to zero as n tends to infinity and E_n contains E_{n+1}. Since the metric space X is complete, the sequence x_n converges to a point x, which must belong to E_n for all n, since E_n is closed and contains x_m for all $m \geq n$. There cannot be a second point y in all of the E_n, since for any point $y \neq x$ the diameter of E_n is less than $d(x, y)$ for large n.

Exercise 3.22 Suppose X is a complete metric space, and $\{G_n\}$ is a sequence of dense open subsets of X. Prove Baire's theorem, namely that $\cap_1^\infty G_n$ is not empty. (In fact, it is dense in X.) *Hint:* Find a shrinking sequence of neighborhoods E_n such that $\overline{E}_n \subset G_n$, and apply Exercise 21.

Solution. Let F_n be the complement of G_n, so that F_n is closed and contains no open sets. We shall prove that any nonempty open set U contains a point not in any F_n, hence in all G_n. To this end, we note that U is not contained in F_1, so that there is a point $x_1 \in U \setminus F_1$. Since $U \setminus F_1$ is open, there exists $r_1 > 0$ such that B_1, defined as the open ball of radius r_1 about x_1, is contained in $U \setminus F_1$. Let E_1 be the open ball of radius $\frac{r_1}{2}$ about x_1, so that the closure of E_1 is contained in B_1. Now F_2 does not contain E_1, and so we can find a point $x_2 \in E_1 \setminus F_2$. Since $E_1 \setminus F_2$ is an open set, there exists a positive number r_2 such that B_2, the open ball of radius R_2 about x_2, is contained in $E_1 \setminus F_2$, which in turn is contained in $U \setminus (F_1 \cup F_2)$. We let E_2 be the open ball of radius $\frac{r_2}{2}$ about x_2, so that $\overline{E}_2 \subseteq B_2$. Proceeding in this way, we construct a sequence of open balls E_j, such that $E_j \supseteq \overline{E}_{j+1}$, and the diameter of E_j tends to zero. By the previous exercise, there is a point x belonging to all the sets $\overline{E}_j$, hence to all the sets $U \setminus (F_1 \cup F_2 \cup \cdots \cup F_n)$. Thus the point x belongs to $U \cap \left(\cap_1^\infty G_n\right)$.

Exercise 3.23 Suppose $\{p_n\}$ and $\{q_n\}$ are Cauchy sequences in a metric space X. Show that the sequence $\{d(p_n, q_n)\}$ converges. *Hint:* For any m, n,

$$d(p_n, q_n) \leq d(p_n, p_m) + d(p_m, q_m) + d(q_m, q_n);$$

it follows that

$$|d(p_n, q_n) - d(p_m, q_m)|$$

is small if m and n are large.

Solution. The inequality in the hint, which is an extension of the triangle inequality, shows that

$$d(p_n, q_n) - d(p_m, q_m) \leq d(p_n, p_m) + d(q_m, q_n);$$

and since the same inequality holds with m and n reversed, it follows that

$$|d(p_n, q_n) - d(p_m, q_m)| \leq d(p_n, p_m) + d(q_m, q_n).$$

Now if $\varepsilon > 0$, choose N_1 and N_2 so that $d(p_n, p_m) < \frac{\varepsilon}{2}$ if $m > N_1$, $n > N_1$, and $d(q_n, q_m) < \frac{\varepsilon}{2}$ if $m > N_2$, $n > N_2$. Then let $N = \max(N_1, N_2)$. It follows immediately that $|d(p_n, q_n) - d(p_m, q_m)| < \varepsilon$ if $m > N$ and $n > N$. Since the real numbers are a complete metric space, it follows that $\{d(p_n, q_n)\}$ converges.

Exercise 3.24 Let X be a metric space.

(*a*) Call two Cauchy sequences $\{p_n\}$, $\{q_n\}$ in X *equivalent* if

$$\lim_{n\to\infty} d(p_n, q_n) = 0.$$

Prove that this is an equivalence relation.

(*b*) Let X^* be the set of all equivalence classes so obtained. If $P \in X^*$ and $Q \in X^*$, $\{p_n\} \in P$, $\{q_n\} \in Q$, define

$$\Delta(P, Q) = \lim_{n\to\infty} d(p_n, q_n);$$

by Exercise 23, this limit exists. Show that the number $\Delta(P, Q)$ is unchanged if $\{p_n\}$ and $\{q_n\}$ are replaced by equivalent sequences, and hence that Δ is a distance function in X^*.

(*c*) Prove that the resulting metric space X^* is complete.

(*d*) For each $p \in X$, there is a Cauchy sequence all of whose terms are p; let P_p be the element of X^* which contains this sequence. Prove that

$$\Delta(P_p, P_q) = d(p, q)$$

for all $p, q \in X$. In other words, the mapping φ defined by $\varphi(p) = P_p$ is an isometry (i.e., a distance-preserving mapping) of X into X^*.

(*e*) Prove that $\varphi(X)$ is dense in X, and that $\varphi(X) = X^*$ if X is complete. By (*d*), we may identify X and $\varphi(X)$ and thus regard X as embedded in the complete metric space X^*. We call X^* the *completion* of X

Solution. (*a*) We need to show that: 1) $\{p_n\}$ is equivalent to itself; 2) if $\{p_n\}$ is equivalent to $\{q_n\}$, then $\{q_n\}$ is equivalent to $\{p_n\}$; and 3) if $\{p_n\}$ is equivalent to $\{q_n\}$ and $\{q_n\}$ is equivalent to $\{r_n\}$, then $\{p_n\}$ is equivalent to $\{r_n\}$. These follow from the properties of any metric. Thus 1) follows, since $d(p_n, p_n) = 0$ for all n; 2) follows since $d(p_n, q_n) = d(q_n, p_n)$; and 3) follows from the triangle inequality, i.e., $d(p_n, r_n) \le d(p_n, q_n) + d(q_n, r_n)$, so that if $d(p_n, q_n) \to 0$ and $d(q_n, r_n) \to 0$, then $d(p_n, r_n) \to 0$.

(*b*) Let $\{p_n\}$ be equivalent to $\{p'_n\}$ and $\{q_n\}$ equivalent to $\{q'_n\}$. Then, since we know in advance that all the limits exist, we have

$$\lim_{n\to\infty} d(p'_n, q'_n) \le \lim_{n\to\infty} \left(d(p'_n, p_n) + d(p_n, q_n) + d(q_n, q'_n)\right) = \lim_{n\to\infty} d(p_n, q_n).$$

By symmetry, however, we must also have the opposite inequality, so that the two limits are actually equal.

Now X^* is a metric space; for $\Delta(P, Q) \ge 0$, by definition $\Delta(P, Q) = 0$ means $P = Q$, and symmetry and the triangle inequality on X^* follow from the same properties on X.

(*c*) Suppose $\{P_k\}$ is a Cauchy sequence in X^*. Choose Cauchy sequences $\{p_{kn}\}$ in X such that $\{p_{kn}\} \in P_k$, $k = 1, 2, \ldots$. For each k, let N_k be the first positive integer such that $d(p_{kn}, p_{km}) < 2^{-k}$ if $m \ge N_k$ and $n \ge N_k$. Let $p_k = p_{kN_k}$. Observe that $d(p_k, p_{kn}) < 2^{-k}$ for any $n \ge N_k$, so that $\lim_{n\to\infty} d(p_k, p_{kn}) \le 2^{-k}$. (This limit exists since the sequence all of whose terms equal p_k is a Cauchy sequence.) Also, for any k, l, and n we have

$$d(p_k, p_l) \le d(p_k, p_{kn}) + d(p_{kn}, p_{ln}) + d(p_{ln}, p_l).$$

Hence, taking n sufficiently large and assuming $k < l$, we obtain

$$d(p_k, p_l) \le 2^{-k} + \Delta(P_k, P_l) + 2^{-k} + 2^{-l} < 3 \cdot 2^{-k} + \Delta(P_k, P_l).$$

It follows that $\{p_k\}$ is a Cauchy sequence. Let P be the element of X^* containing $\{p_k\}$. We claim $P_k \to P$ in X^*. For

$$\begin{aligned}
\Delta(P_k, P) &= \lim_{n\to\infty} d(p_{kn}, p_n) \\
&\le \lim_{n\to\infty} \left(d(p_{kn}, p_k) + d(p_k, p_n)\right) \\
&\le 2^{-k} + \limsup_{n\to\infty} \Delta(P_k, P_n) + 3 \cdot 2^{-k}.
\end{aligned}$$

Thus if $\varepsilon > 0$, choose $N_1 = 2 + \left[\dfrac{-\log \varepsilon}{\log 2}\right]$, and N_2 such that $\Delta(P_k, P_l) < \frac{\varepsilon}{2}$ if $k > N_2$ and $l > N_2$. Let $N = \max(N_1, N_2)$. We claim that if $k > N$, then $d(P_k, P) < \varepsilon$. Indeed this follows, since we then have $2^{-k+2} < \frac{\varepsilon}{2}$ and $\limsup_{n\to\infty} \Delta(P_k, P_n) \le \frac{\varepsilon}{2}$. We have thus finally proved that X^* is complete.

(d) The assertion $\Delta(P_p, P_q) = d(p,q)$ is the trivial assertion that if $p_n = p$ and $q_n = q$ for all n, then

$$\lim_{n\to\infty} d(p_n, q_n) = d(p,q).$$

(e) Let P be any element of X^*, and let $\varepsilon > 0$. We shall find $p \in X$ such that $\Delta(P, P_p) < \varepsilon$. To this end, let $\{p_n\} \in P$ and let N be such that $d(p_n, p_m) < \frac{\varepsilon}{2}$ if $n > N$ and $m > N$. Let $p = p_{N+1}$. Then $\Delta(P, P_p) = \lim d(p_n, p) \le \frac{\varepsilon}{2}$, and we are done.

If X is already complete, then for each $P \in X^*$ and $\{p_n\} \in P$ there exists $p \in X$ such that $p_n \to p$. This p is obviously the same for any sequence equivalent to $\{p_n\}$, and it is clear that $P = P_p$. Hence $\varphi(X) = X^*$ when X is complete.

It should be remarked that X^* is unique, in the sense that if Y and Z are any two complete metric spaces, each containing a dense subset isometric to X, then Y is isometric to Z. Indeed let φ and ψ be isometries of X into Y and Z respectively, such that $\varphi(X)$ is dense in Y and $\psi(X)$ is dense in Z. We construct an isometry of Y onto Z as follows. For each $y \in Y$, there is a sequence $\{x_n\} \subset X$ such that $\varphi(x_n) \to y$. The sequence $\{x_n\}$ is a Cauchy sequence in X, and hence $\{\psi(x_n)\}$ is a Cauchy sequence in Z (since ψ preserves distance). Since Z is complete, there is an element z such that $\psi(x_n) \to z$. We define $\theta(y) = z$. We claim first of all that this definition is unambiguous. For if y is given and some other sequence $\{x'_n\}$ in X is such that $\{\varphi(x'_n)\}$ converges to y, then $d_Z(\psi(x_n), \psi(x'_n)) = d_X(x_n, x'_n) = d_Y(\varphi(x_n), \varphi(x'_n) \to 0$, and hence $\psi(x'_n) \to z$ also. The mapping θ is an isometry, since if $y_1 = \lim \varphi(x_{1n})$ and $y_2 = \lim \varphi(x_{2n})$, then

$$\begin{aligned} d_Z(\theta(y_1), \theta(y_2)) &= \lim d_Z(\psi(x_{1n}), \psi(x_{2n})) \\ &= \lim d_X(x_{1n}, x_{2n}) \\ &= \lim d_Y(\varphi(x_{1n}), \varphi(x_{2n})) \\ &= d_Y(y_1, y_2). \end{aligned}$$

(Here we have used the fact that if $p_n \to p$ and $q_n \to q$, then $d(p_n, q_n) \to d(p,q)$, which in turn follows from the inequality

$$|d(p,q) - d(p_n, q_n)| \le d(p, p_n) + d(q, q_n)$$

proved in Exercise 23 above.)

Finally $\theta(Y) = Z$, since one can easily define an inverse mapping $\eta : Z \to Y$ by merely reversing the steps used to define θ.

Exercise 3.25 Let X be the metric space whose points are the rational numbers, with the metric $d(x,y) = |x - y|$. What is the completion of this space? (Compare Exercise 24.)

Answer. By the remarks at the end of Exercise 24, the completion of a metric space X is any complete metric space containing a dense subset isometric to the space X. Since the real numbers have this property, the completion of the rational numbers is the real numbers. A Cauchy sequence of rational numbers converges to a unique real number, of course, and two sequences are equivalent if and only if they converge to the same real number. Hence we have also a more direct reason for claiming that the completion of the rational numbers is the real numbers.

Chapter 4

Continuity

Exercise 4.1 Suppose f is a real function defined on R^1 which satisfies

$$\lim_{h\to 0}[f(x+h)-f(x-h)]=0$$

for every $x \in R^1$. Does this imply that f is continuous?

Solution. No. In fact even the stronger statement

$$\lim_{h\to 0}\frac{f(x+h)-f(x-h)}{h^n}=0$$

for every $x \in R^1$, where n is an arbitrary positive number, does not imply that f is continuous, since this property is possessed by the function

$$f(x)=\begin{cases}1 & \text{if } x \text{ is an integer,}\\ 0 & \text{if } x \text{ is not an integer.}\end{cases}$$

(If x is an integer, then $f(x+h)-f(x-h)\equiv 0$ for all h; while if x is not an integer, $f(x+h)-f(x-h)=0$ for $|h|<\min(x-[x],1+[x]-x)$.

Exercise 4.2 If f is a continuous mapping of a metric space X into a metric space Y, prove that

$$f(\overline{E})\subset\overline{f(E)}$$

for every set $E\subset X$. ($\overline{E}$ denotes the closure of E.) Show, by an example, that $f(\overline{E})$ can be a proper subset of $\overline{f(E)}$.

Solution. Let $x \in \overline{E}$. We need to show that $f(x)\in\overline{f(E)}$. To this end, let O be any neighborhood of $f(x)$. Since f is continuous, $f^{-1}(O)$ contains (is) a neighborhood of x. Since $x\in\overline{E}$, there is a point u of E in $f^{-1}(O)$. Hence $f(u)\in O\cap f(E)$. Since O was any neighborhood of $f(x)$, it follows that $f(x)\in\overline{f(E)}$.

Consider $f: R^1\to R^1$ given by $f(x)=\dfrac{x}{1+x^2}$, and let $E=\overline{E}=[1,\infty)$, so that $f(E)=f(\overline{E})=(0,\frac{1}{2}]$, yet $\overline{f(E)}=[0,\frac{1}{2}]$.

Exercise 4.3 Let f be a continuous real function on a metric space X. Let $Z(f)$ (the *zero set* of f) be the set of all $p \in X$ at which $f(p) = 0$. Prove that $Z(f)$ is closed.

Solution. $Z(f) = f^{-1}(\{0\})$, which is the inverse image of a closed set. Hence $Z(f)$ is closed.

Exercise 4.4 Let f and g be continuous mappings of a metric space X into a metric space Y, and let E be a dense subset of X. Prove that $f(E)$ is dense in $f(X)$. If $g(p) = f(p)$ for all $p \in E$, prove that $g(p) = f(p)$ for all $p \in X$. (In other words, a continuous mapping is determined by its values on a dense subset of its domain.)

Solution. To prove that $f(E)$ is dense in $f(X)$, simply use Exercise 2 above: $f(X) = f(\overline{E}) \subseteq \overline{f(E)}$.

The function $\varphi : X \to R^1$ given by

$$\varphi(p) = d_Y(f(p), g(p))$$

is continuous, since

$$|d_Y(f(p), g(p)) - d_Y(f(q), g(q))| \leq d_Y(f(p), f(q)) + d_Y(g(p), g(q)).$$

(This inequality follows from the triangle inequality, since

$$d_Y(f(p), g(p)) \leq d_Y(f(p), f(q)) + d_Y(f(q), g(q)) + d_Y(g(q), g(p)),$$

and the same inequality holds with p and q interchanged. The absolute value $|d_Y(f(p), g(p)) - d_Y(f(q), g(q))|$ must be either $d_Y(f(p), g(p)) - d_Y(f(q), g(q))$ or $d_Y(f(q), g(q)) - d_Y(f(p), g(p))$, and the triangle inequality shows that both of these numbers are at most $d_Y(f(p), f(q)) + d_Y(g(p), g(q))$.)

By the previous problem, the zero set of φ is closed. But by definition

$$Z(\varphi) = \{p : f(p) = g(p)\}.$$

Hence the set of p for which $f(p) = g(p)$ is closed. Since by hypothesis it is dense, it must be X.

Exercise 4.5 If f is a real continuous function defined on a closed set $E \subset R^1$, prove that there exist continuous real functions g on R^1 such that $g(x) = f(x)$ for all $x \in E$. (Such functions g are called *continuous extensions* of f from E to R^1.) Show that the result becomes false if the word "closed" is omitted. Extend the result to vector-valued functions. *Hint:* Let the graph of g be a straight line on each of the segments which constitute the complement of E (compare Exercise 29, Chap. 2). The result remains true if R^1 is replaced by any metric space, but the proof is not so simple.

Solution. Following the hint, let the complement of E consist of a countable collection of finite open intervals (a_k, b_k) together with possibly one or both of the the semi-infinite intervals $(b, +\infty)$ and $(-\infty, a)$. The function $f(x)$ is already defined at a_k and b_k, as well as at a and b (if these last two points exist). Define $g(x)$ to be $f(b)$ for $x > b$ and $f(a)$ for $x < a$ if a and b exist. On the interval (a_k, b_k) let

$$g(x) = f(a_k) + \frac{x - a_k}{b_k - a_k}(f(b_k) - f(a_k)).$$

Of course we let $g(x) = f(x)$ for $x \in E$. It is now fairly clear that $g(x)$ is continuous. A rigorous proof proceeds as follows. Let $\varepsilon > 0$. To choose $\delta > 0$ such that $|x - u| < \delta$ implies $|g(x) - g(u)| < \varepsilon$, we consider three cases.

i. If $x > b$, let $\delta = x - b$. Then if $|x - u| < \delta$, it follows that $u > b$ also, so that $g(u) = f(b) = g(x)$, and $|g(u) - g(x)| = 0 < \varepsilon$. Similarly if $x < a$, let $\delta = a - x$.

ii. If $a_k < x < b_k$ and $f(a_k) = f(b_k)$, let $\delta = \min(x - a_k, b_k - x)$. Since $|x - u| < \delta$ implies $a_k < u < b_k$, so that $g(u) = f(a_k) = f(b_k) = g(x)$, we again have $|g(x) - g(u)| = 0 < \varepsilon$. If $a_k < x < b_k$ and $f(a_k) \neq f(b_k)$, let $\delta = \min\left(x - a_k, b_k - x, \dfrac{(b_k - a_k)\varepsilon}{|f(b_k) - f(a_k)|}\right)$. Then if $|x - u| < \delta$, we again have $a_k < u < b_k$ and so

$$|g(x) - g(u)| = \frac{|x - u|}{b_k - a_k}|f(b_k) - f(a_k)| < \varepsilon.$$

iii. If $x \in E$, let δ_1 be such that $|f(u) - f(x)| < \varepsilon$ if $u \in E$ and $|x - u| < \delta_1$.

(Subcase a). If there are points $x_1 \in E \cap (x - \delta_1, x)$ and $x_2 \in E \cap (x, x + \delta_1)$, let $\delta = \min(x - x_1, x_2 - x)$. If $|u - x| < \delta$ and $u \in E$, then $|f(u) - f(x)| < \varepsilon$ by definition of δ_1. if $u \notin E$, then, since x_1, x, and x_2 are all in E, it follows that $u \in (a_k, b_k)$, where $a_k \in E$, $b_k \in E$, and $|a_k - x| < \delta$ and $|b_k - x| < \delta$, so that $|f(a_k) - f(x)| < \varepsilon$ and $|f(b_k) - f(x)| < \varepsilon$. If $f(a_k) = f(b_k)$, then $f(u) = f(a_k)$ also, and we have $|f(u) - f(x)| < \varepsilon$. If $f(a_k) \neq f(b_k)$, then

$$\begin{aligned}
|f(u) - f(x)| &= \left|f(a_k) - f(x) + \frac{u - a_k}{b_k - a_k}(f(b_k) - f(a_k))\right| \\
&= \left|\frac{b_k - u}{b_k - a_k}(f(a_k) - f(x)) + \frac{u - a_k}{b_k - a_k}(f(b_k) - f(x))\right| \\
&< \frac{b_k - u}{b_k - a_k}\varepsilon + \frac{u - a_k}{b_k - a_k}\varepsilon \\
&= \varepsilon
\end{aligned}$$

(Subcase b). Suppose x_2 does not exist, i.e., either $x = a_k$ or $x = a_k$ and $b_k > a_k + \delta_1$. Let us consider the second of these cases and show how to get $|f(u) - f(x)| < \varepsilon$ for $x < u < x + \delta$. If $f(a_k) = f(b_k)$, let $\delta = \delta_1$. If $u > x$ we have $a_k < u < b_k$ and $f(u) = f(a_k) = f(x)$. If $f(a_k) \neq f(b_k)$, let $\delta = \min\left(\delta_1, \dfrac{(b_k - a_k)\varepsilon}{|f(b_k) - f(a_k)|}\right)$. Then, just as in Subcase a, we have $|f(u) - f(x)| < \varepsilon$.

The case when $x = b_k$ for some k and $a_k < x - \delta_1$ is handled in exactly the same way.

If $x = b$, let $\delta = \delta_1$. If $u > x$ we have $f(x) - f(u)$; and if $u < x$ and $u \notin E$, we use the same argument as in Subcases a and b.

The case $x = a$ is handled similarly.

The extension of this result to vector-valued functions is immediate: Simply extend each component of the function. A vector-valued function is continuous if and only if each of its components is continuous.

Exercise 4.6 If f is defined on E, the *graph* of f is the set of points $(x, f(x))$ for $x \in E$. In particular, if E is the set of real numbers and f is real-valued, the graph of f is a subset of the plane.

Suppose E is compact, and prove that f is continuous on E if and only if its graph is compact.

Solution. Let Y be the co-domain of the function f. We invent a new metric space $E \times Y$ as the set of pairs of points (x, y), $x \in E$, $y \in Y$, with the metric $\rho((x_1, y_1), (x_2, y_2)) = d_E(x_1, x_2) + d_Y(y_1, y_2)$. The function $\varphi(x) = (x, f(x))$ is then a mapping of E into $E \times Y$.

We claim that the mapping φ is continuous if f is continuous. Indeed, let $x \in X$ and $\varepsilon > 0$ be given. Choose $\eta > 0$ so that $d_Y(f(x), f(u)) < \dfrac{\varepsilon}{2}$ if $d_E(x, y) < \eta$. Then let $\delta = \min\left(\eta, \dfrac{\varepsilon}{2}\right)$. It is easy to see that $\rho(\varphi(x), \varphi(u)) < \varepsilon$ if $d_E(x, u) < \delta$. Conversely if φ is continuous, it is obvious from the inequality $\rho(\varphi(x), \varphi(u)) \geq d_Y(f(x), f(u))$ that f is continuous.

From these facts we deduce immediately that the graph of a continuous function f on a compact set E is compact, being the image of E under the continuous mapping φ. Conversely, if f is not continuous at some point x, there is a sequence of points x_n converging to x such that $f(x_n)$ does not converge to $f(x)$. If no subsequence of $f(x_n)$ converges, then the sequence $\{(x_n, f(x_n))\}_{n=1}^{\infty}$ has no convergent subsequence, and so the graph is not compact. If some subsequence of $f(x_n)$ converges, say $f(x_{n_k}) \to z$, but $z \neq f(x)$, then the graph of f fails to contain the limit point (x, z), and hence is not closed. A fortiori it is not compact.

Exercise 4.7 If $E \subset X$ and if f is a function defined on X, the *restriction* of f to E is the function g whose domain of definition is E, such that $g(p) = f(p)$ for $p \in E$. Define f and g on R^2 by $f(0,0) = g(0,0) = 0$, $f(x, y) = xy^2/(x^2 + y^4)$, $g(x, y) = xy^2/(x^2 + y^6)$ if $(x, y) \neq (0, 0)$. Prove that f is bounded on R^2, that g is unbounded in every neighborhood of $(0, 0)$, and that f is not continuous at $(0, 0)$; nevertheless, the restrictions of both f and g to every straight line in R^2 are continuous!

Solution. The fact that $|f(x,y)| \le \frac{1}{2}$ is an easy consequence of the inequality $(x-y^2)^2 \ge 0$. The fact that $\lim_{y\to 0} g(y^3,y) = \lim_{y\to 0} \dfrac{y^5}{2y^6} = \lim_{y\to 0} \dfrac{1}{2y} = \infty$ shows that g is unbounded on every neighborhood of infinity. The fact that $\lim_{y\to 0} f(y^2,y) = \lim_{y\to 0} \dfrac{y^4}{2y^4} = \dfrac{1}{2}$ shows that f is not continuous at $(0,0)$.

Since f and g are continuous except at $(0,0)$, it is obvious that their restrictions to any line that does not pass through $(0,0)$ are continuous. Now a line that *does* pass through $(0,0)$ has an equation that is either $x = 0$ or $y = ax$ for some a. Both f and g are constantly 0 on the first of these, and on the second we have $f(x,ax) = a^2x^3/(x^2+a^4x^4) = a^2x/(1+a^4x^2)$, while $g(x,ax) = a^2x^3/(x^2+a^6x^6) = a^2x/(1+a^6x^4)$. Both of the latter are obviously continuous functions.

Exercise 4.8 Let f be a real uniformly continuous function on the bounded set E in R^1. Prove that f is bounded on E.

Show that the conclusion is false if boundedness of E is omitted from the hypothesis.

Let $a = \inf E$ and $b = \sup E$, and let $\delta > 0$ be such that $|f(x) - f(y)| < 1$ if $x, y \in E$ and $|x-y| < \delta$. Now choose a positive integer N larger than $(b-a)/\delta$, and consider the N intervals $I_k = \left[a + \dfrac{k-1}{b-a}, a + \dfrac{k}{b-a}\right]$, $k = 1, 2, \ldots, N$. For each k such that $I_k \cap E \ne \varnothing$ let $x_k \in E \cap I_k$. Then let $M = 1 + \max\{|f(x_k)|\}$. If $x \in E$, we have $|x - x_k| < \delta$ for some k, and hence $|f(x)| < M$.

The function $f(x) = x$ is uniformly continuous on the entire line, but not bounded.

Exercise 4.9 Show that the requirement in the definition of uniform continuity can be rephrased as follows, in terms of diameters of sets: To every $\varepsilon > 0$ there exists a $\delta > 0$ such that $\operatorname{diam} f(E) < \varepsilon$ for all $E \subset X$ with $\operatorname{diam} E < \delta$.

Solution. Suppose f is uniformly continuous and $\varepsilon > 0$ is given. Choose any positive number α smaller than ε. Then there exists $\delta > 0$ such that $d_Y(f(x), f(u)) < \alpha$ if $d_X(x,u) < \delta$. Hence if E is any set of diameter less than δ and x and u are any two points in E we have $d_Y(f(x), f(u)) < \alpha$, so that $\operatorname{diam} f(E) \le \alpha < \varepsilon$.

Conversely if f satisfies the condition stated in the problem, it is obvious that for any $\varepsilon > 0$ there exists $\delta > 0$ such that $d_Y(f(x), f(u)) < \varepsilon$ whenever $d_X(x,u) < \delta$. (Choose $\delta > 0$ corresponding to ε in the condition of the problem and then let E be the two-point set $\{x, u\}$.)

Exercise 4.10 Complete the details of the following alternate proof of Theorem 4.19: If f is not uniformly continuous, then for some $\varepsilon > 0$ there are sequences $\{p_n\}$, $\{q_n\}$ in X such that $d_X(p_n, q_n) \to 0$ but $d_Y(f(p_n), f(q_n)) > \varepsilon$. Use Theorem 2.37 to obtain a contradiction.

Solution. Theorem 4.19 asserts that a continuous function on a compact set is uniformly continuous. By Theorem 2.37 there are subsequences $\{p_{n_k}\}$ and $\{q_{n_k}\}$ that converge to points p and q respectively. Since $d_X(p_n, q_n) \to 0$, it follows that $p = q$. However, since f is continuous, it follows from Theorem 4.2 that $f(p_{n_k})$ and $f(q_{n_k})$ converge to $f(p)$, which, since $d_Y(f(p_{n_k}), f(q_{n_k}) \leq d_Y(f(p_{n_k}), f(p)) + d_Y(f(p), f(q_{n_k}))$, implies that $d_Y(f(p_{n_k}), f(q_{n_k})) \to 0$, contradicting the inequality $d_Y(f(p_{n_k}), f(q_{n_k})) > \varepsilon$.

Exercise 4.11 Suppose f is a uniformly continuous mapping of a metric space X into a metric space Y and prove that $\{f(x_n)\}$ is a Cauchy sequence in Y for every Cauchy sequence $\{x_n\}$ in X. Use this result to give an alternative proof of the theorem stated in Exercise 13.

Solution. Suppose $\{x_n\}$ is a Cauchy sequence in X. Let $\varepsilon > 0$ be given. Let $\delta > 0$ be such that $d_Y(f(x), f(u)) < \varepsilon$ if $d_X(x, u) < \delta$. Then choose N so that $d_X(x_n, x_m) < \delta$ if $n, m > N$. Obviously $d_Y(f(x_n), f(x_m)) < \varepsilon$ if $m, n > N$, showing that $\{f(x_n)\}$ is a Cauchy sequence.

Now let f be a uniformly continuous function defined on a dense subset E of X, mapping E into a *complete* metric space Y (for example, Y could be the real numbers). To prove that f has a unique continuous extension to all of X, proceed as follows. For each $x \in X \setminus E$ let $\{x_n\}$ be a sequence of points in E converging to x. Define $f(x)$ to be the limit of the Cauchy sequence $\{f(x_n)\}$. This definition is unambiguous; for if $\{u_n\}$ also converges to x, then the sequence $\{y_n\}$ defined by

$$y_n = \begin{cases} x_{n/2} & \text{if } n \text{ is even,} \\ u_{(n+1)/2} & \text{if } n \text{ is odd,} \end{cases}$$

also converges to x. Hence $\{f(y_n)\}$ is a Cauchy sequence in Y, and so all subsequences of $\{f(y_n)\}$ converge to the same limit. In particular $\{f(x_n)\}$ and $\{f(u_n)\}$ both converge to the same value.

The extended function is also uniformly continuous. For if $\varepsilon > 0$, let $\delta > 0$ be such that $d_Y(f(x), f(u)) < \dfrac{\varepsilon}{3}$ if $x, u \in E$ and $d_X(x, u) < \delta$. Then if $x \in E$, $u \in X \setminus E$, and $d_X(x, u) < \delta$, choose $v \in E$ with $d_X(v, u) < \delta - d_X(x, u)$ and $d_Y(f(v), f(u)) < \dfrac{\varepsilon}{3}$ (this is possible because of the definition of $f(u)$). We then have $d_X(x, v) \leq d_X(x, u) + d_X(u, v) < \delta$, and so

$$d_Y(f(x), f(u)) \leq d_Y(f(x), f(v)) + d_Y(f(v), f(u)) < \frac{2\varepsilon}{3} < \varepsilon.$$

Similarly if $x \in X \setminus E$, $u \in X \setminus E$, and $d_X(x, u) < \delta$, choose $v, w \in E$ with $d_X(v, u) < \dfrac{1}{2}(\delta - d_X(x, u))$, $d_X(x, w) < \dfrac{1}{2}(\delta - d_X(x, u))$, $d_Y(f(v), f(u)) < \dfrac{\varepsilon}{3}$, and $d_Y(f(w), f(x)) < \frac{\varepsilon}{3}$. We then have

$$d_X(v, w) \leq d_X(v, u) + d_X(u, x) + d_X(x, w) < \delta$$

and hence

$$d_Y(f(x), f(u)) \leq d_Y(f(x), f(w)) + d_Y(f(w), f(v)) + d_Y(f(v), f(u)) < \varepsilon.$$

The uniqueness of this extension follows from Exercise 4 above.

Exercise 4.12 A uniformly continuous function of a uniformly continuous function is uniformly continuous.

State this more precisely and prove it.

Solution. Let $f : X \to Y$ and $g : Y \to Z$ be uniformly continuous. Then $g \circ f : X \to Z$ is uniformly continuous, where $g \circ f(x) = g\big(f(x)\big)$ for all $x \in X$.

To prove this fact, let $\varepsilon > 0$ be given. Then, since g is uniformly continuous, there exists $\eta > 0$ such that $d_Z(g(u), g(v)) < \varepsilon$ if $d_Y(u, v) < \eta$. Since f is uniformly continuous, there exists $\delta > 0$ such that $d_Y(f(x), f(y)) < \eta$ if $d_X(x, y) < \delta$.

It is then obvious that $d_Z(g(f(x)), g(f(y))) < \varepsilon$ if $d_X(x, y) < \delta$, so that $g \circ f$ is uniformly continuous.

Exercise 4.13 Let E be a dense subset of a metric space X, and let f be a uniformly continuous *real* function defined on E. Prove that f has a continuous extension from E to X (see Exercise 5 for terminology). (Uniqueness follows from Exercise 4.) *Hint:* For each $p \in X$ and each positive ineger n, let $V_n(p)$ be the set of all $q \in E$ with $d(p, q) < 1/n$. Use Exercise 9 to show that the intersection of the closures of the sets $f(V_1(p)), f(V_2(p)), \ldots$, consists of a single point, say $g(p)$, of R^1. Prove that the function g so defined is the desired extension of f.

Could the range space R^1 be replaced by R^n. By any compact metric space? By any complete metric space? By any metric space?

Solution. We shall carry out the proof in the context of any complete metric space, showing that the range space could be R^n or any compact metric space.

The diameter of the closure of $f(V_i(p))$ is the same as the diameter of $f(V_i(p))$ itself. Hence by Exercise 9 above these diameters tend to zero. Since they form a nested sequence of nonempty closed sets, their intersection must consist of a single point, which can be defined to be $g(p)$. If $p \in E$, the intersection of these sets is just $f(p)$ (since $f(p)$ is in all the sets, and only one point belongs to all of them), so that g coincides with f on E. It remains to show that g is continuous. This proof is identical to the proof given in Exercise 11 above, which depends only on the fact that for $u \in X \setminus E$ and $\varepsilon > 0$, $\delta > 0$ there is a point $v \in E$

with $d_X(v,u) < \delta$ and $d_Y(f(v), f(u)) < \varepsilon$. This condition clearly holds in the present context as well.

In general this theorem fails on an incomplete metric space. For example, take X to be the real numbers, Y and E the rational numbers, and let $f : E \to Y$ be given by $f(x) = x$. There is no possible extension of f to a mapping from X into Y. (There is a unique extension of f to a mapping from X into X, but its range is not contained in Y. If there were an extension of f to a mapping from X into Y, there would be two extensions of f to mappings from X into X, contradicting the uniqueness of the extension.)

Exercise 4.14 Let $I = [0,1]$ be the closed unit interval. Suppose f is a continuous mapping of I into I. Prove that $f(x) = x$ for at least one $x \in I$.

Solution. If $f(0) = 0$ or $f(1) = 1$, we are done. If not, then $0 < f(0)$ and $f(1) < 1$. Hence the continuous function $g(x) = x - f(x)$ satisfies $g(0) < 0 < g(1)$. By the intermediate value theorem, there must be a point $x \in (0,1)$ where $g(x) = 0$.

Exercise 4.15 Call a mapping from X into Y *open* if $f(V)$ is an open set in Y whenever V is an open set in X.

Prove that every continuous open mapping of R^1 into R^1 is monotonic.

Solution. Suppose f is continuous and not monotonic, say there exist points $a < b < c$ with $f(a) < f(b)$, and $f(c) < f(b)$. Then the maximum value of f on the closed interval $[a,c]$ is assumed at a point u in the open interval (a,c). If there is also a point v in the open interval (a,c) where f assumes its minimum value on $[a,c]$, then $f(a,c) = [f(v), f(u)]$. If no such point v exists, then $f(a,c) = (d, f(u)]$, where $d = \min(f(a), f(c))$. In either case, the image of (a,c) is not open.

Exercise 4.16 Let $[x]$ denote the largest integer contained in x, that is $[x]$ is the integer such that $x - 1 < [x] \le x$; and let $(x) = x - [x]$ denote the fractional part of x. What discontinuities do the functions $[x]$ and (x) have?

Solution. The two functions have the same discontinuities, since each can be written as the difference of the continuous function $f(x) = x$ and the other function. Now the function $[x]$ is constant on each open interval $(k, k+1)$; hence its only possible discontinuities are the integers. These are of course real discontinuities, since if $\varepsilon = 1$, there is no $\delta > 0$ such that $|[x] - [k]| < \varepsilon$ whenever $|x - k| < \delta$. (For if any δ is given, let $\eta = \min(1, \delta)$. Then $[k] - [k - \frac{\eta}{2}] = 1$.)

Exercise 4.17 Let f be a real function defined on (a, b). Prove that the set of points at which f has a simple discontinuity is at most countable. *Hint:* Let E be the set on which $f(x-) < f(x+)$. With each point x of E associate a triple (p, q, r) of rational numbers such that
(a) $f(x-) < p < f(x+)$,
(b) $a < q < t < x$ implies $f(t) < p$,
(c) $x < t < r < b$ implies $f(t) > p$.

The set of such triples is countable. Show that each triple is associated with at most one point of E. Deal similarly with the other possible types of simple discontinuities.

Solution. The existence of three such rational numbers (p, q, r) for each simple discontinuity of this type follows from the assumption $f(x-) < f(x+)$, and the definition of $f(x-)$ and $f(x+)$. We need to show that a given triple (p, q, r) cannot be associated with any other discontinuity of this type. To that end, suppose $y > x$ and $f(y-) < f(y+)$. If we do not have $f(y-) < p < f(y+)$, then the triple chosen for y will differ from (p, q, r) in its first element. Hence suppose $f(y-) < p < f(y+)$. In this case we definitely cannot have $r > y$, since there are points $t \in (x, y)$ such that $f(t) < p$ (if there weren't, we would have $f(y-) \geq p$).

We have thus shown that the set of points $x \in (a, b)$ at which $f(x-) < f(x+)$ is at most countable. The proof that the set of points at which $f(x-) > f(x+)$ is at most countable is, of course, nearly identical.

Now consider the set of points x at which $\lim_{t \to x} f(t)$ exists, but is not equal to $f(x)$. For each point $x \in (a, b)$ such that $\lim_{t \to x} f(t) < f(x)$, we take a triple (p, q, r) of rational numbers such that
(a) $\lim_{t \to x} f(t) < p < f(x)$,
(b) $a < q < t < x$ or $x < t < r < b$ implies $f(t) < p$.

As before, if $y > x$ and $\lim_{t \to y} f(t) < f(y)$, the triple associated with y will be different from that associated with x. For even if $\lim_{t \to y} f(t) < p < f(y)$, we cannot have $r > y$, since $f(y) > p$ and $x < y$.

The proof that the set of points $x \in (a, b)$ at which $\lim_{t \to x} f(t) > f(x)$ is countable is nearly identical.

Hence, the number of points in $[a, b]$ at which f has a discontinuity of first kind is countable.

Exercise 4.18 Every rational x can be written in the form $x = m/n$, where $n > 0$ and m and n are integers without any common divisors. When $x = 0$, we take $n = 1$. Consider the function f defined on R^1 by

$$f(x) = \begin{cases} 0 & (x \text{ irrational}), \\ \dfrac{1}{n} & \left(x = \dfrac{m}{n}\right). \end{cases}$$

Prove that f is continuous at every irrational point, and that f has a simple discontinuity at every rational point.

Solution. We shall show that $\lim_{t\to x} f(t) = 0$ for every t. Both assertions follow immediately from this fact. To this end, let $\varepsilon > 0$ be given, and let x be any real number. Let N be the unique positive integer such that $N \le 1/\varepsilon < N+1$, and for each positive integer $n = 1, 2, \ldots, N$, let k_n be the unique integer such that

$$\frac{k_n}{n} \le x < \frac{k_n+1}{n}$$

Then for each such n let $\delta_n = \frac{1}{n}$ if $x = \frac{k_n}{n}$, otherwise let $\delta_n = \min\left(x - \frac{k}{n}, \frac{k_n+1}{n} - x\right)$. Finally let $\delta = \min(\delta_1, \ldots, \delta_N)$. We claim that $|f(t)| < \varepsilon$ if $0 < |x-t| < \delta$. This is obvious if t is irrational, while if t is rational and $t = \frac{m}{n}$, we necessarily have $n > N$ by the choice of the numbers δ_n for $n \le N$. Hence if t is rational, then $f(t) \le \frac{1}{N+1} < \varepsilon$. The proof is now complete.

Exercise 4.19 Suppose f is a real function with domain R^1 which has the intermediate-value property: If $f(a) < c < f(b)$, then $f(x) = c$ for some x between a and b.

Suppose also, for every rational r, that the set of all x with $f(x) = r$ is closed.

Prove that f is continuous.

Hint: If $x_n \to x_0$ but $f(x_n) > r > f(x_0)$ for some r and all n, then $f(t_n) = r$ for some t_n between x_0 and x_n; thus $t_n \to x_0$. Find a contradiction. (N. M. Fine, *Amer. Math. Monthly*, vol. 73, 1966, p. 782.)

Solution. The contradiction is evidently that x_0 is a limit point of the set of t such that $f(t) = r$, yet, x_0 does not belong to this set. This contradicts the hypothesis that the set is closed.

Exercise 4.20 If E is a nonempty subset of a metric space X, define the distance from $x \in X$ to E by

$$\rho_E(x) = \inf_{z\in E} d(x,z).$$

(a) Prove that $\rho_E(x) = 0$ if and only if $x \in \overline{E}$.
(b) Prove that ρ_E is a uniformly continuous function on X by showing that

$$|\rho_E(x) - \rho_E(y)| \le d(x,y)$$

for all $x \in X$ and $y \in X$.

Hint: $\rho_E(x) \le d(x,z) \le d(x,y) + d(y,z)$, so that

$$\rho_E(x) \le d(x,y) + \rho_E(y).$$

Solution. (*a*) For each positive integer n, let $z_n \in E$ be such that $\rho_E(x) \le d(x, z_n) < \rho_E(x) + \frac{1}{n}$. It follows that $d(x, z_n) \to \rho_E(x)$. If $\rho_E(x) = 0$, this means $z_n \to x$, i.e., $x \in \overline{E}$. Conversely, if $x \in \overline{E}$, there exists a sequence $\{z_n\}_{n=1}^{\infty} \subseteq E$ such that $z_n \to x$, and this means $d(z_n, x) \to 0$, so that $\rho_E(x) = 0$.

(*b*) The last inequality given in the hint follows form the first by taking the infimum over z on the right-hand side. This inequality immediately implies that

$$\rho_E(x) - \rho_E(y) \le d(x,y).$$

By interchanging x and y, we also obtain

$$\rho_E(y) - \rho_E(x) \le d(y,x) = d(x,y).$$

Since $|\rho_E(x) - \rho_E(y)|$ must be either $\rho_E(x) - \rho_E(y)$ or $\rho_E(y) - \rho_E(x)$, it follows that

$$|\rho_E(x) - \rho_E(y)| \le d(x,y).$$

Exercise 4.21 Suppose K and F are disjoint sets in a metric space X, K is compact, F is closed. Prove that there exists $\delta > 0$ such that $d(p,q) > \delta$ if $p \in K$, $q \in F$. *Hint:* ρ_F is a continuous positive function on K.

Show that the conclusion may fail for two disjoint closed sets if neither is compact.

Solution. Following the hint, we observe that $\rho_F(x)$ must attain its minimum value on K, i.e., there is some point $r \in K$ such that

$$\rho_F(r) = \min_{q \in K} \rho_F(q).$$

Since F is closed and $r \notin F$, it follows from Exercise 4.20 that $\rho_F(r) > 0$. Let δ be any positive number smaller than $\rho_F(r)$. Then for any $p \in F$, $q \in K$, we have

$$d(p,q) \ge \rho_F(q) \ge \rho_F(r) > \delta.$$

This proves the positive assertion.

As for closed sets in general, one could let $F = \{1, 2, 3, \ldots\}$ and $K = \{1 + \frac{1}{2}, 2 + \frac{1}{3}, 3 + \frac{1}{4} \ldots\}$ in R^1, or one could let $F = \{(x,y) : y = 0\}$ and $K = \{(x,y) : y = \frac{1}{1+x^2}\}$ in R^2. In both cases there are sequences of points $p_n \in F$, $q_n \in K$ such that $d(p_n, q_n) \to 0$.

Exercise 4.22 Let A and B be disjoint nonempty closed sets in a metric space X, and define

$$f(p) = \frac{\rho_A(p)}{\rho_A(p) + \rho_B(p)} \quad (p \in X).$$

Show that f is a continuous function on X whose range lies in $[0,1]$, that $f(p) = 0$ precisely on A and $f(p) = 1$ precisely on B. This establishes a converse of Exercise 3: Every closed set $A \subset X$ is $Z(f)$ for some continuous real f on X. Setting

$$V = f^{-1}\left(\left[0, \frac{1}{2}\right)\right), \quad W = f^{-1}\left(\left(\frac{1}{2}, 1\right]\right),$$

show that V and W are open and disjoint, and that $A \subset V$, $B \subset W$. (Thus pairs of disjoint closed sets in a metric space can be covered by pairs of disjoint open sets. This property of metric spaces is called *normality*.)

Solution. The continuity of f follows from the fact that the quotient of two continuous real-valued continuous functions is continuous wherever the denominator is non-zero. Now the denominator of the fraction that defines f cannot be zero, since the first term is zero only on A and the second is zero only on B, while A and B are disjoint. The fact that $f(p) = 0$ if and only if $p \in A$ follows from Exercise 20 and the fact that A is closed. Likewise the fact that $f(p) = 1$ if and only if $p \in B$ follows from Exercise 20 and the fact that B is closed. The assertion about V and W is immediate, since V and W are the inverse images of disjoint open sets containing 0 and 1 respectively.

Exercise 4.23 A real-valued function f defined in (a, b) is said to be *convex* if

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$$

whenever $a < x < b$, $a < y < b$, $0 < \lambda < 1$. Prove that every convex function is continuous. Prove that every increasing convex function of a convex function is convex. (For example, if f is convex, so is e^f.)

If f is convex in (a, b) and if $a < s < t < u < b$, show that

$$\frac{f(t) - f(s)}{t - s} \le \frac{f(u) - f(s)}{u - s} \le \frac{f(u) - f(t)}{u - t}.$$

Solution. Fix any points c, d with $a < c < d < b$, let $\eta > 0$ be any fixed positive number with $\eta < \dfrac{d-c}{2}$ and consider any two points x, y satisfying $c + \eta \le x < y \le d - \eta$. The inequality in the definition implies that $f(t)$ is bounded above on $[c, d]$. Indeed, if $c < t < d$, taking $\lambda = \dfrac{t-c}{d-c}$, we have $t = (1-\lambda)c + \lambda d$, and so, if $M = \max(f(c), f(d))$, we have

$$f(t) \le (1-\lambda) f(c) + \lambda f(d) \le (1-\lambda)M + \lambda M = M.$$

It is less obvious that f is also bounded below on $[c,d]$. In fact if $\frac{c+d}{2} < t < d$, we have

$$\frac{c+d}{2} = (1-\lambda)c + \lambda t,$$

where $\lambda = \frac{d-c}{2(t-c)}$, so that

$$f\Big(\frac{c+d}{2}\Big) \le \Big(\frac{2t-(c+d)}{2(t-c)}\Big) f(c) + \Big(\frac{d-c}{2(t-c)}\Big) f(t),$$

which implies

$$f(t) \ge \Big(\frac{2(t-c)}{d-c}\Big) f\Big(\frac{c+d}{2}\Big) - \frac{2t-(c+d)}{d-c}\Big) f(c) \ge -2\Big|f\Big(\frac{c+d}{2}\Big)\Big| - |f(c)|.$$

The proof that f is bounded below on $\left[c, \frac{c+d}{2}\right]$ is similar. Hence there exists M such that $|f(t)| \le M$ for all $t \in [c,d]$.

We can also write

$$x = (1-\lambda)c + \lambda y,$$

where $\lambda = \frac{x-c}{y-c} \in (0,1)$. Accordingly we have

$$\begin{aligned} f(x) - f(y) &\le (1-\lambda)\big(f(c) - f(y)\big) = \\ &= \frac{y-x}{y-c}\big(f(c) - f(y)\big) \le \frac{y-x}{\eta}\big|f(c) - f(y)\big|. \end{aligned}$$

Thus

$$f(x) - f(y) \le \frac{2M}{\eta}(y-x).$$

Similarly, writing $y = \lambda x + (1-\lambda)d$, where $\lambda = \frac{d-y}{d-x} \in (0,1)$, we find

$$\begin{aligned} f(y) - f(x) &\le (1-\lambda)\big(f(d) - f(x)\big) = \\ &= \frac{y-x}{d-x}\big(f(d) - f(x)\big) \le \frac{y-x}{\eta}\big|f(d) - f(x)\big|. \end{aligned}$$

Hence we also have

$$f(y) - f(x) \le \frac{2M}{\eta}(y-x).$$

Therefore

$$|f(y) - f(x)| \le \frac{2M}{\eta}|y-x|$$

for all $x, y \in [c+\eta, d-\eta]$. Since c, d, and η are arbitrary, it follows that f is continuous on (a,b).

If $f(x)$ is convex on (a,b), and $g(x)$ is an increasing convex function on $f((a,b))$, we have

$$g(f(\lambda x+(1-\lambda)y)) \le g(\lambda f(x)+(1-\lambda)f(y)) \le \lambda g(f(x))+(1-\lambda)g(f(y)).$$

The inequality

$$\frac{f(t)-f(s)}{t-s} \le \frac{f(u)-f(s)}{u-s}$$

can be rewritten as

$$f(t) \le \frac{t-s}{u-s}f(u)+\Big(1-\frac{t-s}{u-s}\Big)f(s),$$

which is precisely the definition of convexity if we note that

$$t=\lambda u+(1-\lambda)s$$

when $\lambda = \dfrac{t-s}{u-s}$.

The other inequality is proved in exactly the same way.

Exercise 4.24 Assume that f is a continuous real function defined in (a,b) such that

$$f\Big(\frac{x+y}{2}\Big) \le \frac{f(x)+f(y)}{2}$$

for all $x, y \in (a,b)$. Prove that f is convex.

Solution. We shall prove that

$$f(\lambda x+(1-\lambda)y) \le \lambda f(x)+(1-\lambda)f(y)$$

for all "dyadic rational" numbers, i.e., all numbers of the form $\lambda = \dfrac{k}{2^n}$, where k is a nonnegative integer not larger than 2^n. We do this by induction on n. The case $n=0$ is trivial (since $\lambda=0$ or $\lambda=1$). In the case $n=1$ we have $\lambda=0$ or $\lambda=1$ or $\lambda=\frac{1}{2}$. The first two cases are again trivial, and the third is precisely the hypothesis of the theorem. Suppose the result is proved for $n \le r$, and consider $\lambda=\dfrac{k}{2^{r+1}}$. If k is even, say $k=2l$, then $\dfrac{k}{2^{r+1}}=\dfrac{l}{2^r}$, and we can appeal to the induction hypothesis. Now suppose k is odd. Then $1 \le k \le 2^{r+1}-1$, and so the numbers $l=\dfrac{k-1}{2}$ and $m=\dfrac{k+1}{2}$ are integers with $0 \le l < m \le 2^r$. We can now write

$$\lambda = \frac{s+t}{2},$$

where $s=\dfrac{k-1}{2^{r+1}}=\dfrac{l}{2^r}$ and $t=\dfrac{k+1}{2^{r+1}}=\dfrac{m}{2^r}$. We then have

$$\lambda x+(1-\lambda)y = \frac{[sx+(1-s)y]+[tx+(1-t)y]}{2}.$$

Hence by the hypothesis of the theorem and the induction hypothesis we have

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\le \frac{f(sx+(1-s)y) + f(tx+(1-t)y)}{2} \\ &\le \frac{sf(x)+(1-s)f(y)+tf(x)+(1-t)f(y)}{2} \\ &= \Big(\frac{s+t}{2}\Big)f(x) + \Big(1-\frac{s+t}{2}\Big)f(y) \\ &= \lambda f(x) + (1-\lambda)f(y). \end{aligned}$$

This completes the induction.

Now for each fixed x and y both sides of the inequality

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y)$$

are continuous functions of λ. Hence the set on which this inequality holds (the inverse image of the closed set $[0,\infty)$ under the mapping $\lambda \mapsto \lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y)$) is a closed set. Since it contains all the points $\frac{k}{2^n}$, $0 \le k \le n$, $n = 1, 2, \ldots$, it must contain the closure of this set of points, i.e., it must contain all of $[0,1]$. Thus f is convex.

Exercise 4.25 If $A \subset R^k$ and $B \subset R^k$, define $A+B$ to be the set of all sums $\mathbf{x}+\mathbf{y}$ with $\mathbf{x} \in A$, $\mathbf{y} \in B$.

(*a*) If K is compact and C is closed in R^k, prove that $K+C$ is closed.

Hint: Take $\mathbf{z} \notin K+C$, put $F = \mathbf{z} - C$, the set of all $\mathbf{z}-\mathbf{y}$ with $\mathbf{y} \in C$. Then K and F are disjoint. Choose δ as in Exercise 21. Show that the open ball with center $\mathbf{z}$ and radius δ does not intersect $K+C$.

(*b*) Let α be an irrational number. Let C_1 be the set of all integers. Let C_2 be the set of all $n\alpha$ with $n \in C_1$. Show that C_1 and C_2 are closed subsets of R^1 whose sum $C_1 + C_2$ is not closed, by showing that $C_1 + C_2$ is a countable dense subset of R^1.

Solution. (*a*) It is clear that the set F defined in the hint is a closed set. It is disjoint from K, since $\mathbf{z} \notin K+C$. Let δ be such that $|\mathbf{p}-\mathbf{q}| > \delta$ if $\mathbf{p} \in F$ and $\mathbf{q} \in K$. We claim that there is no point of $K+C$ inside the ball of radius δ about $\mathbf{z}$. For suppose $\mathbf{w}$ were such a point. By definition we would have $\mathbf{w} = \mathbf{u}+\mathbf{v}$, where $\mathbf{u} \in K$ and $\mathbf{v} \in C$. But then we would have

$$|\mathbf{u} - (\mathbf{z}-\mathbf{v})| = |\mathbf{w}-\mathbf{z}| < \delta,$$

which is a contradiction, since $\mathbf{u} \in K$ and $\mathbf{z}-\mathbf{v} \in F$. Thus $K+C$ is closed.

(*b*) Neither of the sets C_1 and C_2 has any limit points; hence both are closed sets. For each fixed integer $N \ge 2$, consider the fractional parts $\beta_1 = \alpha - [\alpha]$, $\beta_2 = 2\alpha - [2\alpha], \ldots, \beta_N = N\alpha - [N\alpha]$. There must be some half-open interval

$\left[\frac{k-1}{N-1}, \frac{k}{N-1}\right)$, $k = 1, 2, \ldots, N-1$ containing two of the numbers $\beta_1, \ldots, \beta_N$, since there are N numbers and only $N-1$ intervals. (Note: No two of these numbers are equal, since $\beta_i = \beta_j$, $i \neq j$, would imply

$$\alpha = \frac{[i\alpha] - [j\alpha]}{i - j},$$

i.e., α would be a rational number.) Now the inequalities

$$0 < (i\alpha - [i\alpha]) - (j\alpha - [j\alpha]) < \frac{1}{N-1}$$

say that $(i-j)\alpha + ([j\alpha] - [i\alpha]) \in \left(0, \frac{1}{N-1}\right)$, that is, there is certainly a point of $C_1 + C_2$ in $\left(0, \frac{1}{N-1}\right)$ for any $N \geq 2$. We shall now prove that there is a point of $C_1 + C_2$ in $\left(\frac{k}{n}, \frac{k+1}{n}\right)$ for any integer k and any positive integer n. To do so, fix the integer q such that $qn \leq k < (q+1)n$, and choose $y \in C_1 + C_2$ such that $0 < y < \frac{1}{n}$. Then $x = ny \in C_1 + C_2$ and $0 < x < 1$. Hence there is a positive integer p such that $k < px + qn < k+1$. This says precisely that

$$\frac{k}{n} < py + q < \frac{k+1}{n},$$

and certainly $py + q \in C_1 + C_2$. Now let O be any nonempty open subset of R^1. Then O contains an interval (a, b). If $n > \frac{2}{b-a}$, there is an integer k such that $\left(\frac{k}{n}, \frac{k+1}{n}\right) \subset (a, b)$. This interval, as just shown, contains a point of $C_1 + C_2$, and hence O contains such a point. Therefore $C_1 + C_2$ is dense in R^1. Since it is a countable set, it is not all of R^1, and hence not closed.

Exercise 4.26 Suppose X, Y, Z are metric spaces and Y is compact. Let f map X into Y, let g be a continuous one-to-one mapping of Y into Z, and put $h(x) = g(f(x))$ for $x \in X$.

Prove that f is uniformly continuous if h is uniformly continuous.

Hint: g^{-1} has compact domain $g(Y)$, and $f(x) = g^{-1}(h(x))$.

Prove also that f is continuous if h is continuous.

Show (by modifying Example 4.21, or by finding a different example) that the compactness of Y cannot be omitted from the hypotheses, even when X and Z are compact.

Solution. Theorem 4.17 asserts that g^{-1} is continuous, and since its domain is compact, it is uniformly continuous. Exercise 12 above then implies that f is uniformly continuous. The same argument, with the word "uniformly" omitted, shows that f is continuous if h is continuous.

To get a counterexample when Y is not compact, let $X = [0,1] = Z$, $Y = \{0\} \cup [1, \infty)$, and let $f : X \to Y$ and $g : Y \to Z$ be given by

$$f(x) = \begin{cases} \frac{1}{x}, & 0 < x \leq 1, \\ 0, & x = 0, \end{cases}$$

$$g(y) = \begin{cases} \frac{1}{y}, & 1 \leq y < \infty, \\ 0, & y = 0. \end{cases}$$

Then $h(x) = g(f(x)) = x$, so that h is uniformly continuous, and g is continuous and one-to-one, yet f is not even continuous.

Chapter 5

Differentiation

Exercise 5.1 Let f be defined for all real x, and suppose that

$$|f(x) - f(y)| \le (x - y)^2$$

for all real x and y. Prove that f is constant.

Solution. Dividing by $x - y$, and letting $x \to y$, we find that $f'(y) = 0$ for all y. Hence f is constant.

Exercise 5.2 Suppose $f'(x) > 0$ in (a, b). Prove that f is strictly increasing in (a, b), and let g be its inverse function. Prove that g is differentiable, and that

$$g'(f(x)) = \frac{1}{f'(x)} \quad (a < x < b).$$

Solution. For any c, d with $a < c < d < b$ there exists a point $p \in (c, d)$ such that $f(d) - f(c) = f'(p)(d - c) > 0$. Hence $f(c) < f(d)$.

We know from Theorem 4.17 that the inverse function g is continuous. (Its restriction to each closed subinterval $[c, d]$ is continuous, and that is sufficient.) Now observe that if $f(x) = y$ and $f(x + h) = y + k$, we have

$$\frac{g(y+k) - g(y)}{k} - \frac{1}{f'(x)} = \frac{1}{\frac{f(x+h)-f(x)}{h}} - \frac{1}{f'(x)}.$$

Since we know $\lim \frac{1}{\varphi(t)} = \frac{1}{\lim \varphi(t)}$ provided $\lim \varphi(t) \neq 0$, it follows that for any $\varepsilon > 0$ there exists $\eta > 0$ such that

$$\left| \frac{1}{\frac{f(x+h)-f(x)}{h}} - \frac{1}{f'(x)} \right| < \varepsilon$$

if $0 < |h| < \eta$. Since $h = g(y+k) - g(y)$, there exists $\delta > 0$ such that $0 < |h| < \eta$ if $0 < |k| < \delta$. The proof is now complete.

Exercise 5.3 Suppose g is a real function on R^1 with bounded derivative (say $|g'| \leq M$). Fix $\varepsilon > 0$, and define $f(x) = x + \varepsilon g(x)$. Prove that f is one-to-one if ε is small enough. (A set of admissible values of ε can be determined which depends only on M.)

Solution. If $0 < \varepsilon < \dfrac{1}{M}$, we certainly have

$$f'(x) \geq 1 - \varepsilon M > 0,$$

and this implies that $f(x)$ is one-to-one, by the preceding problem.

Exercise 5.4 If

$$C_0 + \frac{C_1}{2} + \cdots + \frac{C_{n-1}}{n} + \frac{C_n}{n+1} = 0,$$

where $C_0, \ldots, C_n$ are real constants, prove that the equation

$$C_0 + C_1 x + \cdots + C_{n-1} x^{n-1} + C_n x^n = 0$$

has at least one real root between 0 and 1.

Solution. Consider the polynomial

$$p(x) = C_0 x + \frac{C_1}{2} x^2 + \cdots + \frac{C_{n-1}}{n} x^n + \frac{C_n}{n+1} x^{n+1},$$

whose derivative is

$$p'(x) = C_0 + C_1 x + \cdots + C_{n-1} x^{n-1} + C_n x^n.$$

It is obvious that $p(0) = 0$, and the hypothesis of the problem is that $p(1) = 0$. Hence Rolle's theorem implies that $p'(x) = 0$ for some x between 0 and 1.

Exercise 5.5 Suppose f is defined and differentiable for every $x > 0$, and $f'(x) \to 0$ as $x \to +\infty$. Put $g(x) = f(x+1) - f(x)$. Prove that $g(x) \to 0$ as $x \to +\infty$.

Solution. Let $\varepsilon > 0$. Choose x_0 such that $|f'(x)| < \varepsilon$ if $x > x_0$. Then for any $x \geq x_0$ there exists $x_1 \in (x, x+1)$ such that

$$f(x+1) - f(x) = f'(x_1).$$

Since $|f'(x_1)| < \varepsilon$, it follows that $|f(x+1) - f(x)| < \varepsilon$, as required.

Exercise 5.6 Suppose

(a) f is continuous for $x \geq 0$,

(b) $f'(x)$ exists for $x > 0$,

(c) $f(0) = 0$,

(d) f' is monotonically increasing.

Put

$$g(x) = \frac{f(x)}{x} \quad (x > 0)$$

and prove that g is monotonically increasing.

Solution. By the mean-value theorem

$$f(x) = f(x) - f(0) = f'(c)x,$$

for some $c \in (0, x)$. Since f' is monotonically increasing, this result implies that $f(x) < xf'(x)$. It therefore follows that

$$g'(x) = \frac{xf'(x) - f(x)}{x^2} > 0,$$

so that g is also monotonically increasing.

Exercise 5.7 Suppose $f'(x)$ and $g'(x)$ exist, $g'(x) \neq 0$, and $f(x) = g(x) = 0$. Prove that

$$\lim_{t \to x} \frac{f(t)}{g(t)} = \frac{f'(x)}{g'(x)}.$$

(This holds also for complex functions.)

Solution. Since $f(x) = g(x) = 0$, we have

$$\begin{aligned}
\lim_{t \to x} \frac{f(t)}{g(t)} &= \lim_{t \to x} \frac{\dfrac{f(t) - f(x)}{t - x}}{\dfrac{g(t) - g(x)}{t - x}} \\
&= \frac{\lim\limits_{t \to x} \dfrac{f(t) - f(x)}{t - x}}{\lim\limits_{t \to x} \dfrac{g(t) - g(x)}{t - x}} \\
&= \frac{f'(x)}{g'(x)}.
\end{aligned}$$

Exercise 5.8 Suppose f' is continuous on $[a, b]$ and $\varepsilon > 0$. Prove that there exists $\delta > 0$ such that

$$\left|\frac{f(t)-f(x)}{t-x} - f'(x)\right| < \varepsilon$$

whenever $0 < |t-x| < \delta$, $a \leq x \leq b$, $a \leq t \leq b$. (This could be expressed by saying that f is *uniformly differentiable* on $[a, b]$ if f' is continuous on $[a, b]$.) Does this hold for vector-valued functions too?

Solution. Let δ be such that $|f'(x)-f'(u)| < \varepsilon$ for all $x, u \in [a, b]$ with $|x-u| < \delta$. Then if $0 < |t-x| < \delta$ there exists u between t and x such that

$$\frac{f(t)-f(x)}{t-x} = f'(u),$$

and hence, since $|u-x| < \delta$,

$$\left|\frac{f(t)-f(x)}{t-x} - f'(x)\right| = |f'(u) - f'(x)| < \varepsilon.$$

Since this result holds for each component of a vector-valued function $\mathbf{f}(x)$, it must hold also for $\mathbf{f}$.

Exercise 5.9 Let f be a continuous real function on R^1, of which it is known that $f'(x)$ exists for all $x \neq 0$ and that $f'(x) \to 3$ as $x \to 0$. Does it follow that $f'(0)$ exists?

Solution. Yes. By L'Hospital's rule

$$\lim_{t\to 0} \frac{f(t)-f(0)}{t} = \lim_{t\to 0} f'(t) = 3,$$

and this by definition means that $f'(0) = 3$.

Exercise 5.10 Suppose f and g are complex differentiable functions on $(0, 1)$, $f(x) \to 0$, $g(x) \to 0$, $f'(x) \to A$, $g'(x) \to B$ as $x \to 0$, where A and B are complex numbers, $B \neq 0$. Prove that

$$\lim_{x\to 0} \frac{f(x)}{g(x)} = \frac{A}{B}.$$

Compare with Example 5.18. *Hint:*

$$\frac{f(x)}{g(x)} = \left\{\frac{f(x)}{x} - A\right\} \cdot \frac{x}{g(x)} + A \cdot \frac{x}{g(x)}.$$

Apply Theorem 5.13 to the real and imaginary parts of $f(x)/x$ and $g(x)/x$.

Solution. We can make f and g continuous on $[0, 1)$ by simply defining $f(0) = 0 = g(0)$. Then Exercise 9 applied to the real and imaginary parts of f and g show that $f'(0) = A$ and $g'(0) = B$. (These are one-sided derivatives, since f and g are not defined for negative values of x; however, we could extend them as odd functions, since both are 0 at 0). We could then apply Exercise 7, whose proof does not use anything but the definition of the derivative and some general facts about limits. In this way we get the result without resorting to the combinatorial trick referred to in the hint. This result shows that many of the facts ordinarily proved for real functions by use of the mean-value theorem and L'Hospital's rule remain true for complex-valued functions, even though, as Example 5.18 shows, these theorems are not true for complex-valued functions.

Exercise 5.11 Suppose f is defined in a neighborhood of x, and suppose $f''(x)$ exists. Show that

$$\lim_{h\to 0} \frac{f(x+h) + f(x-h) - 2f(x)}{h^2} = f''(x).$$

Solution. For a real-valued function this is a routine application of L'Hospital's rule:

$$\begin{aligned}
\lim_{h\to 0} \frac{f(x+h) + f(x-h) - 2f(x)}{h^2} &= \lim_{h\to 0} \frac{f'(x+h) - f'(x-h)}{2h} \\
&= \frac{1}{2}\lim_{h\to 0} \frac{f'(x+h) - f'(x)}{h} + \\
&\qquad + \frac{f'(x) - f'(x-h)}{h} \\
&= f''(x).
\end{aligned}$$

For complex-valued functions the result follows from separate consideration of real and imaginary parts.

The limit will be zero at $x = 0$ for any odd function f whatsoever, even if the function is not continuous. For example we could take $f(x) = \operatorname{sgn}(x)$, which is $+1$ for $x > 0$, 0 for $x = 0$, and -1 for $x < 0$

Exercise 5.12 If $f(x) = |x|^3$, compute $f'(x)$, $f''(x)$ for all real x, and show that $f^{(3)}(0)$ does not exist.

Solution. For $x > 0$ we have $f'(x) = 3x^2$, $f''(x) = 6x$, and for $x < 0$ $f'(x) = -6x^2$, $f''(x) = -6x$, i.e., $f'(x) = 3x|x|$, and $f''(x) = 6|x|$ for $x \neq 0$. By Exercise 9, it therefore follows that $f'(0)$ exists and equals 0, and then another application of Exercise 9 shows that $f''(0)$ also exists and equals 0. However

$$\frac{f''(x) - f''(0)}{x} = 6\operatorname{sgn}(x),$$

which has no limit at 0. Hence $f^{(3)}(0)$ does not exist.

Exercise 5.13 Suppose a and c are real numbers, $c > 0$, and f is defined on $[-1, 1]$ by

$$f(x) = \begin{cases} x^a \sin(x^{-c}) & (\text{if } x \neq 0), \\ 0 & (\text{if } x = 0). \end{cases}$$

Prove the following statements:

(*a*) f is continuous if and only if $a > 0$.

(*b*) $f'(0)$ exists if and only if $a > 1$.

(*c*) f' is bounded if and only if $a \geq 1 + c$.

(*d*) f' is continuous if and onlyl if $a > 1 + c$.

(*e*) $f''(0)$ exists if and only if $a > 2 + c$.

(*f*) f'' is bounded if and only if $a \geq 2 + 2c$.

(*g*) f'' is continuous if and only if $a > 2 + 2c$.

Solution. We remark editorially that there are two difficulties with this problem. One is that we haven't yet introduced the function sin. To overcome that problem we can rely on our intuitive notion or use the Taylor series if we have to. The second problem is more serious, however: What do x^a and x^{-c} mean when $x < 0$? In general these will be complex-valued functions. It might be better to use absolute values in both cases. Thus we shall amend the problem by defining $f(x) = |x|^a \sin(|x|^{-c})$ when $x \neq 0$.

(*a*) Since f is infinitely differentiable except at $x = 0$, the only question of continuity is at $x = 0$. Let $t_n = 2\pi(n + \frac{1}{8})$, $x_n = t_n^{-\frac{1}{c}}$ and $y_n = \dfrac{1}{\sqrt{2}} t_n^{-\frac{a}{c}}$. Notice that $f(x_n) = y_n$ and that y_n tends to $\dfrac{1}{\sqrt{2}}$ if $a = 0$ and to $+\infty$ if $a < 0$. Hence the function cannot be continuous if $a \leq 0$. On the other hand, we have

$$|f(x) - f(0)| = |f(x)| \leq |x|^a,$$

so that if $a > 0$ and ε is given, we can choose $\delta = \varepsilon^{\frac{1}{a}}$, and then $|x - 0| < \delta$ implies $|f(x) - f(0)| < \varepsilon$, i.e., $f(x)$ is continuous at $x = 0$.

(*b*) If $f'(0)$ exists, then f is continuous at 0, so that $a > 0$. Notice that

$$\frac{f(x_n) - f(0)}{x_n} = \frac{y_n}{x_n} = \frac{1}{\sqrt{2}} t_n^{\frac{1-a}{c}}.$$

which tends to $\dfrac{1}{\sqrt{2}}$ if $a = 1$ and to $+\infty$ if $0 < a < 1$. Hence $f'(0)$ does not exist if $a \leq 1$. On the other hand if $a > 1$ we have

$$0 \leq \frac{f(x) - f(0)}{x} < |x|^{a-1} \to 0,$$

and so $f'(0) = 0$.

(*c*) For $x \neq 0$ we have

$$f'(x) = \operatorname{sgn}(x)|x|^{a-1}\big[a \sin(|x|^{-c}) - c|x|^{-c} \cos(|x|^{-c})\big].$$

Hence $f'(x_n) = \frac{1}{\sqrt{2}}\left[ax_n^{a-1} - cx_n^{-c+a-1}\right]$, which tends to $-\infty$ if $a < 1 + c$. On the other hand we have

$$|f'(x)| \leq |a||x|^{a-1} + c|x|^{a-1-c},$$

which is certainly bounded on $[-1, 1]$ if $a \geq 1 + c$.
(*d*) If f' is continuous, it is bounded, and so $a \geq 1 + c$. However if $a = 1 + c$, then

$$f'(x_n) = \frac{1}{\sqrt{2}}\left[(1+c)t_n^{-1} - c\right]$$

which tends to $-\frac{c}{\sqrt{2}}$ as $n \to \infty$, while $x_n \to 0$. Hence f' is not continuous at 0 unless $a > 1 + c$. If $a > 1 + c$, the inequality

$$|f'(x)| \leq |a||x|^{a-1} + c|x|^{a-1-c},$$

implies that $f(x) \to 0$ as $x \to 0$, and so f' is continuous.
(*e*) If $f''(0)$ exists, then f' must be continuous at 0, and so $a \geq 1 + c$. Now for $x \neq 0$

$$\frac{f'(x) - f'(0)}{x} = \operatorname{sgn}(x)\left[a|x|^{a-2}\sin(|x|^{-c}) - c|x|^{a-c-2}\cos(|x|^{-c}\right).$$

Taking $x = x_n$, we find that this difference quotient equals

$$\frac{1}{\sqrt{2}}\left[at_n^{\frac{2-a}{c}} - ct_n^{\frac{c+2-a}{c}}\right],$$

which tends to $\frac{1}{\sqrt{2}}$ if $a = c+2$ and to $-\infty$ if $a < c+2$. Hence $f''(0)$ exists only if $a > c + 2$.

On the other hand, if $a > c + 2$, we have the inequality

$$\left|\frac{f'(x) - f'(0)}{x}\right| \leq a|x|^{a-2} + c|x|^{a-c-2},$$

from which it follows immediately that $f''(0) = 0$.
(*f*) For $x \neq 0$ we have

$$\begin{aligned} f''(x) = \operatorname{sgn}(x)[a(a-1)|x|^{a-2} - c^2|x|^{a-2c-2}]\sin(|x|^{-c}] \\ - c(2a - c - 1)|x|^{a-c-1}\cos(|x|^{-c}]. \end{aligned}$$

In particular

$$f''(x_n) = \frac{1}{\sqrt{2}}[a(a-1)t_n^{\frac{2-a}{c}} - c^2 t_n^{\frac{2+2c-a}{c}} - c(2a-c-1)t_n^{\frac{c+1-a}{c}}],$$

which tends to $-\infty$ if $a < 2 + 2c$. On the other hand, we have the inequality

$$|f''(x)| \leq |a||a-1||x|^{a-2} + c^2|x|^{a-2c-2} + c|2a - c - 1||x|^{a-c-1},$$

and the right-hand side is certainly bounded if $a \geq 2+2c$.
(*g*) If f'' is continuous, then it is bounded, and hence $a \geq 2+2c$. If $a = 2+2c$, we have

$$f''(x_n) = \frac{1}{\sqrt{2}}[(2c+2)(2c+1)t_n^{-2} - c^2 - c(3+3c)t_n^{\frac{-c-1}{c}}],$$

which tends to $-\frac{c^2}{\sqrt{2}}$, so that f'' is not continuous at 0. On the other hand, if $a > 2+2c$, the inequality

$$|f''(x)| \leq |a||a-1||x|^{a-2} + c^2|x|^{a-2c-2} + c|2a-c-1||x|^{a-c-1},$$

shows that $f''(x) \to 0$ as $x \to 0$, and hence f'' is continuous.

Exercise 5.14 Let f be a differentiable real function defined in (a,b). Prove that f is convex if and only if f' is monotonically increasing. Assume next that $f''(x)$ exists for every $x \in (a,b)$, and prove that f is convex if and only if $f''(x) \geq 0$ for all $x \in (a,b)$.

Suppose first that f' is montonically increasing, and that $x < y$. We wish to show that if $0 < \lambda < 1$, then

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

Letting $z = \lambda x + (1-\lambda)y$, we have $\lambda = \frac{y-z}{y-x}$, $1-\lambda = \frac{z-x}{y-x}$, and $x < z < y$. Now the required inequality can be written

$$(1-\lambda)[f(y)-f(z)] \geq \lambda[f(z)-f(x)],$$

which, when we insert the values of λ and $1-\lambda$, and multiply by the positive number $\frac{y-x}{(z-x)(y-z)}$, becomes

$$\frac{f(y)-f(z)}{y-z} \geq \frac{f(z)-f(x)}{z-x}.$$

Since the left-hand side is $f'(d)$ for some $d \in (z,y)$, the right-hand side is $f'(c)$ for some $c \in (x,z)$, and f' is nondecreasing, we have the required inequality.

By Exercise 23 of Chapter 4 we know that if f is convex on (a,b) and $a < c < d < p < q < b$, then

$$\frac{f(d)-f(c)}{d-c} \leq \frac{f(p)-f(d)}{p-d} \leq \frac{f(q)-f(p)}{q-p}.$$

Hence, if f' exists, letting $d \to c$ and $q \to p$, we find

$$f'(c) \leq f'(p),$$

so that f' is nondecreasing.

Finally if f'' exists, we know that f' is nondecreasing if and only if $f''(x) \geq 0$ for all $x \in (a,b)$. Hence f is convex if and only if $f''(x) \geq 0$ for all $x \in (a,b)$.

Exercise 5.15 Suppose $a \in R^1$, f is a twice-differentiable real function on (a, ∞), and M_0, M_1, M_2 are the least upper bounds of $|f(x)|$, $|f'(x)|$, $|f''(x)|$, respectively, on (a, ∞). Prove that

$$M_1^2 \leq 4M_0M_2.$$

Hint: If $h > 0$, Taylor's theorem shows that

$$f'(x) = \frac{1}{2h}[f(x+2h) - f(x)] - hf''(\xi)$$

for some $\xi \in (x, x+2h)$. Hence

$$|f(x)| \leq hM_2 + \frac{M_0}{h}.$$

To show that $M_1^2 = 4M_0M_2$ can actually happen, take $a = -1$, define

$$f(x) = \begin{cases} 2x^2 - 1, & (-1 < x < 0), \\ \dfrac{x^2 - 1}{x^2 + 1}, & (0 \leq x < \infty), \end{cases}$$

and show that $M_0 = 1$, $M_1 = 4$, $M_2 = 4$.

Does $M_1^2 \leq 4M_1M_2$ hold for vector-valued functions too?

Solution. The inequality is obvious if $M_0 = +\infty$ or $M_2 = +\infty$, so we shall assume that M_0 and M_2 are both finite. We need to show that

$$|f'(x)| \leq 2\sqrt{M_0M_2}$$

for all $x > a$. We note that this is obvious if $M_2 = 0$, since in that case $f'(x)$ is constant, $f(x)$ is a linear function, and the only bounded linear function is a constant, whose derivative is zero. Hence we shall assume from now on that $0 < M_2 < +\infty$ and $0 < M_0 < +\infty$.

Following the hint, we need only choose $h = \sqrt{\dfrac{M_0}{M_2}}$, and we obtain

$$|f'(x)| \leq 2\sqrt{M_0M_2},$$

which is precisely the desired inequality.

The case of equality follows, since the example proposed satisfies

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

for $x \geq 0$. We see easily that $|f(x)| \leq 1$ for all $x > -1$. Now $f'(x) = \dfrac{4x}{(x^2+1)^2}$ for $x > 0$ and $f'(x) = 4x$ for $x < 0$. It thus follows from Exercise 9 above that $f'(0) = 0$, and that $f'(x)$ is continuous. Likewise $f''(x) = 4$ for $x < 0$

and $f''(x) = \dfrac{4-4x^2}{(x^2+1)^3} = -4\dfrac{x^2-1}{(x^2+1)^3}$. This shows that $|f''(x)| < 4$ for $x > 0$ and also that $\lim\limits_{x\to 0} f''(x) = 4$. Hence Exercise 9 again implies that $f''(x)$ is continuous and $f''(0) = 4$.

On n-dimensional space let $\mathbf{f}(x) = (f_1(x), \ldots, f_n(x))$, $M_0 = \sup|\mathbf{f}(x)|$, $M_1 = \sup|\mathbf{f}'(x)|$, and $M_2 = \sup|\mathbf{f}''(x)|$. Just as in the numerical case, there is nothing to prove if $M_2 = 0$ or $M_0 = +\infty$ or $M_2 = +\infty$, and so we assume $0 < M_0 < +\infty$ and $0 < M_2 < \infty$. Let a be any positive number less than M_1, let x_0 be such that $|\mathbf{f}'(x_0)| > a$, and let $\mathbf{u} = \dfrac{1}{|\mathbf{f}'(x_0)|}\mathbf{f}'(x_0)$. Consider the real-valued function $\varphi(x) = \mathbf{u}\cdot\mathbf{f}(x)$. Let N_0, N_1, and N_2 be the suprema of $|\varphi(x)|$, $|\varphi'(x)|$, and $|\varphi''(x)|$ respectively. By the Schwarz inequality we have (since $|\mathbf{u}| = 1$) $N_0 \le M_0$ and $N_2 \le M_2$, while $N_1 \ge \varphi(x_0) = |\mathbf{f}'(x_0)| > a$. We therefore have $a^2 < 4N_0N_2 \le 4M_0M_2$. Since a was any positive number less than M_1, we have $M_1^2 \le 4M_0M_2$, i.e., the result holds also for vector-valued functions.

Equality can hold on any R^n, as we see by taking $\mathbf{f}(x) = (f(x), 0, \ldots, 0)$ or $\mathbf{f}(x) = (f(x), f(x), \ldots, f(x))$, where $f(x)$ is a real-valued function for which equality holds.

Exercise 5.16 Suppose f is twice-differentiable on $(0,\infty)$, f'' is bounded on $(0,\infty)$, and $f(x) \to 0$ as $x \to \infty$. Prove that $f'(x) \to 0$ as $x \to \infty$.

Solution. We shall prove an even stronger statement. If $f(x) \to L$ as $x \to \infty$ and $f'(x)$ is uniformly continuous on $(0,\infty)$, then $f'(x) \to 0$ as $x \to \infty$.

For, if not, let $x_n \to \infty$ be a sequence such that $f(x_n) \ge \varepsilon > 0$ for all n. (We can assume $f(x_n)$ is positive by replacing f with $-f$ if necessary.) Let δ be such that $|f'(x) - f'(y)| < \dfrac{\varepsilon}{2}$ if $|x-y| < \delta$. We then have $f'(y) > \dfrac{\varepsilon}{2}$ if $|y - x_n| < \delta$, and so

$$|f(x_n+\delta) - f(x_n-\delta)| \ge 2\delta\cdot\frac{\varepsilon}{2} = \delta\varepsilon.$$

But, since $\delta\varepsilon > 0$, there exists X such that

$$|f(x) - L| < \frac{1}{2}\delta\varepsilon$$

for all $x > X$. Hence for all large n we have

$$|f(x_n+\delta) - f(x_n-\delta)| \le |f(x_n+\delta) - L| + |L - f(x_n-\delta)| < \delta\varepsilon,$$

and we have reached a contradiction.

The problem follows from this result, since if f'' is bounded, say $|f''(x)| \le M$, then $|f'(x) - f'(y)| \le M|x-y|$, and f' is certainly uniformly continuous.

Exercise 5.17 Suppose f is a real, three times differentiable function on $[-1,1]$, such that

$$f(-1)=0, \quad f(0)=0, \quad f(1)=1, \quad f'(0)=0.$$

Prove that $f^{(3)}(x) \geq 3$ for some $x \in (-1,1)$.

Note that equality holds for $\frac{1}{2}(x^3+x^2)$.

Hint: Use Theorem 5.15 with $\alpha = 1$ and $\beta = \pm 1$, to to show that there are $s \in (0,1)$ and $t \in (-1,0)$ such that

$$f^{(3)}(s) + f^{(3)}(t) = 6.$$

Solution. Following the hint, we observe that Theorem 5.15 (Taylor's formula with remainder) implies that

$$\begin{aligned} f(1) &= f(0) + f'(0) + \frac{1}{2}f''(0) + \frac{1}{6}f^{(3)}(s) \\ f(-1) &= f(0) - f'(0) + \frac{1}{2}f''(0) - \frac{1}{6}f^{(3)}(t) \end{aligned}$$

for some $s \in (0,1)$, $t \in (-1,0)$. By subtracting the second equation from the first and using the given values of $f(1)$, $f(-1)$, and $f'(0)$, we obtain

$$1 = \frac{1}{6}\left(f^{(3)}(s) + f^{(3)}(t)\right),$$

which is the desired result. Note that we made no use of the hypothesis $f(0) = 0$.

Exercise 5.18 Suppose f is a real function on $[a,b]$, n is a positive integer, and $f^{(n-1)}$ exists for every $t \in [a,b]$. Let α, β, and P be as in Taylor's theorem (5.15). Define

$$Q(t) = \frac{f(t) - f(\beta)}{t - \beta}$$

for $t \in [a,b]$, $t \neq \beta$, differentiate

$$f(t) - f(\beta) = (t - \beta)Q(t)$$

$n-1$ times at $t = \alpha$, and derive the following version of Taylor's theorem:

$$f(\beta) = P(\beta) + \frac{Q^{n-1}(\alpha)}{(n-1)!}(\beta - \alpha)^n.$$

Solution. The function $Q(t)$ is differentiable $n-1$ times except possibly at $t = \beta$, so we don't have to worry when differentiating $n-1$ times at $t = \alpha$. It is easy to prove by induction that

$$f^{(k)}(t) = (t - \beta)Q^{(k)}(t) + kQ^{(k-1)}(t)$$

for $0 < k \leq n-1$. Hence

$$\frac{1}{k!}f^{(k)}(\alpha)(\beta-\alpha)^k = -\frac{(\beta-\alpha)^{k+1}}{k!}Q^{(k)}(\alpha) + \frac{(\beta-\alpha)^k}{(k-1)!}Q^{(k-1)}(\alpha).$$

Then, because the sum telescopes, we find

$$\sum_{k=0}^{n-1}\frac{f^{(k)}(\alpha)}{k!}(\beta-\alpha)^k = f(\beta) - \frac{Q^{n-1}(\alpha)}{(n-1)!}(\beta-\alpha)^n,$$

which can be rewritten as

$$f(\beta) = P(\beta) + \frac{Q^{(n-1)}(\alpha)}{(n-1)!}(\beta-\alpha)^n.$$

Exercise 5.19 Suppose f is defined in $(-1,1)$ and $f'(0)$ exists. Suppose $-1 < \alpha_n < \beta_n < 1$, $\alpha_n \to 0$, and $\beta_n \to 0$ as $n \to \infty$. Define the difference quotients

$$Dn = \frac{f(\beta_n) - f(\alpha_n)}{\beta_n - \alpha_n}.$$

Prove the following statements:

(a) If $\alpha_n < 0 < \beta_n$, then $\lim D_n = f'(0)$.

(b) If $0 < \alpha_n < \beta_n$ and $\beta_n/(\beta_n - \alpha_n)$ is bounded, then $\lim D_n = f'(0)$.

(c) if f' is continuous in $(-1,1)$, then $\lim D_n = f'(0)$.

Give an example in which f is differentiable in $(-1,1)$ (but f' is not continuous at 0) and in which α_n, β_n tends to 0 in such a way that $\lim D_n$ exists but is different from $f'(0)$.

Solution. We assume that $\alpha_n\beta_n \neq 0$ throughout, i.e., that neither α_n nor β_n is zero.

(a) Write

$$\begin{aligned} D_n &= \frac{f(\beta_n)-f(0)}{\beta_n-\alpha_n} + \frac{f(0)-f(\alpha_n)}{\beta_n-\alpha_n} \\ &= \frac{\beta_n}{\beta_n-\alpha_n}\frac{f(\beta_n)-f(0)}{\beta_n} + \frac{-\alpha_n}{\beta_n-\alpha_n}\frac{f(\alpha_n)-f(0)}{\alpha_n}. \end{aligned}$$

Now let $\varepsilon > 0$. Choose $\delta > 0$ such that

$$\left|\frac{f(x)-f(0)}{x} - f'(0)\right| < \varepsilon$$

if $0 < |x| < \delta$. Then choose N so that $0 < \beta_n < \delta$ and $-\delta < \alpha_n < 0$ for $n > N$. Then for all $n > N$ we have

$$\begin{aligned}
|D_n - f'(0)| &\leq \frac{\beta_n}{\beta_n - \alpha_n}\left|\frac{f(\beta_n) - f(0)}{\beta_n} - f'(0)\right| + \\
&\qquad + \frac{-\alpha_n}{\beta_n - \alpha_n}\left|\frac{f(\alpha_n) - f(0)}{\alpha_n} - f'(0)\right| \\
&< \frac{\beta_n}{\beta_n - \alpha_n}\varepsilon + \frac{-\alpha_n}{\beta_n - \alpha_n}\varepsilon \\
&= \varepsilon.
\end{aligned}$$

(*b*) If $\dfrac{\beta_n}{\beta_n - \alpha_n} \leq M$ for all n, and $0 < \alpha_n < \beta_n$, then surely $\dfrac{\alpha_n}{\beta_n - \alpha_n} < M$ for all n. Hence if $\varepsilon > 0$ is given, choose N so that

$$\left|\frac{f(x) - f(0)}{x} - f'(0)\right| < \frac{\varepsilon}{2M}$$

if $0 < |x| < \delta$. Then choose N so that $0 < \beta_n < \delta$ (hence also $0 < \alpha_n < \delta$) for $n > N$. Then for all $n > N$ we have

$$\begin{aligned}
|D_n - f'(0)| &\leq \frac{\beta_n}{\beta_n - \alpha_n}\left|\frac{f(\beta_n) - f(0)}{\beta_n} - f'(0)\right| + \\
&\qquad + \frac{\alpha_n}{\beta_n - \alpha_n}\left|\frac{f(\alpha_n) - f(0)}{\alpha_n} - f'(0)\right| \\
&< \frac{\beta_n}{\beta_n - \alpha_n}\frac{\varepsilon}{2M} + \frac{\alpha_n}{\beta_n - \alpha_n}\frac{\varepsilon}{2M} \\
&< \varepsilon.
\end{aligned}$$

(*c*) By the mean-value theorem there exists γ_n between α_n and β_n such that $D_n = f'(\gamma_n)$. Since $\gamma_n \to 0$ and f' is continuous, it follows that $D_n \to f'(0)$.

Let $f(x)$ be *any* function such that $f'(0)$ exists but $\lim\limits_{x\to 0} f'(x)$ does not exist. We know that $f'(x)$ does not tend to infinity as $x \to 0$, since if it did, we would have $|f'(x)| > 1 + |f'(0)|$ for all sufficiently small nonzero x, and this contradicts the intermediate-value property of derivatives. Hence there is a sequence $x_n \to 0$, $x_n \neq 0$, such that $\lim\limits_{n\to\infty} f'(x_n) = L \neq f'(0)$. Let $\beta_n = x_n$, and let y_n be such that $0 < |y_n - x_n| < \frac{1}{2}|x_n|$ and

$$\left|\frac{f(y_n) - f(x_n)}{y_n - x_n} - f'(x_n)\right| < \frac{|L - f'(0)|}{2n}$$

It is then immediate that

$$\lim_{n\to\infty} \frac{f(y_n) - f(x_n)}{y_n - x_n} = L \neq f'(0).$$

A suitable example of such a function $f(x)$ is

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right), & x \neq 0, \\ 0, & x = 0. \end{cases}$$

In this case we can get the counterexample in a slightly different form by taking $x_n = \dfrac{1}{2\pi n}$ and $y_n = \dfrac{1}{2\pi(n+\frac{1}{4})}$. We then have $f'(0) = 0$ and

$$\frac{f(y_n) - f(x_n)}{y_n - x_n} = \frac{2n}{\pi\left(n+\frac{1}{4}\right)} \to \frac{2}{\pi}.$$

Exercise 5.20 Formulate and prove an inequality which follows from Taylor's theorem and which remains valid for vector-valued functions.

Solution. There is a variety of possibilities, of which we choose just one: *Suppose $f(x)$ has continuous derivatives up to order n on $[a, b]$. Then there exists $c \in (a, b)$ such that*

$$|f(b) - P(b)| \leq \left|\frac{f^n(c)}{n!}\right|(b-a)^n.$$

To prove this assertion true for a vector-valued function $\mathbf{f}$, we merely observe that it holds for each scalar-valued function $\mathbf{u}\cdot\mathbf{f}$ if $\mathbf{u}$ is any fixed vector of length 1. It is obviously true if $|\mathbf{f}(b) - \mathbf{P}(b)| = 0$, and in all other cases it follows by taking $\mathbf{u} = \dfrac{1}{|\mathbf{f}(b) - \mathbf{P}(b)|}(\mathbf{f}(b) - \mathbf{P}(b))$.

Exercise 5.21 Let E be a closed subset of R^1. We saw in Exercise 22, Chap. 4, that there is a real continuous function f on R^1 whose zero set is E. Is it possible, for each closed set E, to find such an f which is differentiable on R^1, or one which is n times differentiable, or even one which has derivatives of all orders on R^1?

Solution. Yes, it is possible. The proof depends on the following lemma:

Let a and b be any real numbers with $a < b$, and let $f(x)$ be defined for all real numbers x by the formulas

$$f(x) = \begin{cases} e^{\frac{1}{(x-a)(x-b)}}, & a < x < b, \\ \\ 0, & x \leq a \text{ or } x \geq b. \end{cases}$$

Then f has derivatives of all orders on R^1.

It is obvious that f has derivatives of all orders at every point except possibly a and b. To prove that derivatives exist at these points we need two sublemmas: *For each nonnegative integer n there exists a polynomial $p_n(z, w)$ such that*

$$f^{(n)}(x) = p_n\left(\frac{1}{x-a}, \frac{1}{x-b}\right)e^{\frac{1}{(x-a)(x-b)}}$$

for $a < x < b$.

The proof of this sublemma uses only the partial-fraction decomposition

$$\frac{1}{(x-a)(x-b)} = \frac{1}{b-a}\left[\frac{1}{x-b} - \frac{1}{x-a}\right],$$

together with the chain rule and the fact that the partial derivative of a polynomial is again a polynomial. We omit the details.

The second sublemma is stated as a formula: *For every nonnegative integer n,*

$$\lim_{x \downarrow a} \frac{e^{\frac{1}{(x-a)(x-b)}}}{(x-a)^n} = 0.$$

Its proof is a consequence of Taylor's formula. To be specific, Taylor's formula with remainder implies the following result:

For each nonnegative integer k and each positive number t

$$e^t > \frac{1}{k!}t^k.$$

This last result follows easily since there is a point $t_k \in (0,t)$ for which

$$e^t = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^{k-1}}{(k-1)!} + \frac{e^{t_k}}{k!}t^k,$$

every term in this last sum is positive, and $e^{t_k} > 1$.

We now apply this result with $k = n$ and $t = \dfrac{1}{(x-a)(b-x)}$, to obtain

$$\begin{aligned} e^{\frac{1}{(x-a)(x-b)}} &= \frac{1}{e^{\frac{1}{(x-a)(b-x)}}} \\ &< n!(b-x)^n(x-a)^n \end{aligned}$$

for all $n = 0, 1, \ldots$. In particular

$$e^{\frac{1}{(x-a)(x-b)}} < n!(b-a)^n(x-a)^n = K_n(x-a)^n.$$

Since the kth derivative of $e^{\frac{1}{(x-a)(x-b)}}$ is a polynomial in $\dfrac{1}{x-a}$ and $\dfrac{1}{x-b}$, each derivative also satisfies such an estimate. It follows from this last result that

$$\lim_{x \downarrow a} p\left(\frac{1}{z-b}, \frac{1}{z-a}\right) e^{\frac{1}{(x-a)(x-b)}} = 0$$

for any polynomial $p(z,w)$, and hence that $f^{(n)}(a) = 0$ for all n. The proof that $f^{(n)}(b) = 0$ is similar. We observe that the zero set of $f(x)$ is the complement of the open interval (a,b).

Identical reasoning shows that the function

$$f(x) = \begin{cases} e^{\frac{1}{a-x}}, & x > a, \\ 0, & x \le a, \end{cases}$$

has derivatives of all orders, and its zero set is the complement of the semi-infinite open interval $(a, +\infty)$. A similar function can be constructed for a semi-infinite open interval $(-\infty, b)$.

Now let F be any non-empty closed set. The complement of F consists of a countable set of pairwise disjoint finite open intervals (a_k, b_k), together with possibly one or two semi-infinite open intervals. Define $f(x)$ to be zero on F, let $f(x) = e^{\frac{1}{(x-a_k)(x-b_k)}}$ in each finite open interval complementary to F with endpoints in F, $f(x) = e^{\frac{1}{a-x}}$ for $x > a$ if the complement of F contains a semi-infinite interval $(a, +\infty)$ with endpoint $a \in F$, and $f(x) = e^{\frac{1}{x-b}}$ if the complement of F contains a semi-infinite interval $(-\infty, b)$ with endpoint $b \in F$.

It is now obvious that f is zero precisely on F, and that F has derivatives of all orders at each point of the complement of F and at each interior point of F.

It remains to be shown that f has derivatives of all orders at each boundary point x of F. There are actually 4 cases to consider, but all are handled alike, and we shall settle for just one typical case, in which there is a decreasing sequence of points $x_p \in F$, $x_p \to x$, and a decreasing sequence of points $y_p \notin F$, $y_p \to x$, but no increasing sequence of points $z_p \in F$, $z_p \to x$. This means either $x = b_k$ for some k or $x = b$. Now for each y such that $x < y < x_1$ there is a complementary interval to F, say $(a_l, b_l) \subset (x, x_1)$, with $a_k < y < b_k$. Then for all nonnegative integers k and n we have

$$0 < f^{(k)}(y) < K_{n,k}(y - a_l)^n < K_{n,k}(y - x)^n$$

where $K_{n,k}$ is a positive constant independent of l, hence independent of y. It therefore follows, upon taking $n = 2$ that if $x_1 > y > x$, then

$$\left|\frac{f(y) - f(x)}{(y - x)}\right| \leq K_{2,0}(y - x)$$

(We have just proved this inequality for $y \notin F$, and $f(y) = f(x) = 0$ if $y \in F$.) Hence the right-handed derivative

$$f'_+(x) = \lim_{y \downarrow x} \frac{f(y) - f(x)}{y - x}$$

is zero. That the left-hand derivative is zero follows from the fact that $x = b_k$ or $x = b$. Hence $f'(x) = 0$. We now assume by induction that $f^{(k-1)}(x) = 0$ Then the inequality $f^{(k-1)}(y) \leq K_{2,k-1}(y - x)^2$ shows that

$$f_+^{(k)}(x) = \lim_{y \downarrow x} \frac{f^{(k-1)}(y)}{y - x} = 0.$$

Again, the left-hand kth derivative is zero since $x = b_k$ or $x = b$. It follows easily that $f^{(k)}(x)$ exists and equals zero for all k.

Exercise 5.22 Suppose f is a real function on $(-\infty, \infty)$. Call x a *fixed point* of f if $f(x) = x$.

(a) If f is differentiable and $f'(t) \neq 1$ for every real t, prove that f has at most one fixed point.

(b) Show that the function f defined by

$$f(t) = t + (1 + e^t)^{-1}$$

has no fixed point, although $0 < f'(t) < 1$ for all real t.

(c) However, if there is a constant $A < 1$ such that $|f'(t)| \leq A$ for all real t, prove that a fixed point x of f exists, and that $x = \lim x_n$, where x_1 is an arbitrary real number and

$$x_{n+1} = f(x_n)$$

for $n = 1, 2, 3, \ldots$.

(d) Show that the process described in (c) can be visualized by the zig-zag path

$$(x_1, x_2) \to (x_2, x_2) \to (x_2, x_3) \to (x_3, x_3) \to (x_3, x_4) \to \cdots.$$

Solution. (a) If a function $f(x)$ has two fixed points x and y, $x \neq y$, the mean-value theorem implies that there exists a point z between x and y such that

$$y - x = f(y) - f(x) = f'(z)(y - x),$$

so that $f'(z) = 1$.

(b) The equation $f(t) = t$ implies that $(1+e^t)^{-1} = 0$, which is clearly impossible, while $f'(t) = 1 - \dfrac{e^t}{(1+e^t)^2}$ always lies in $(0, 1)$.

(c) Since f' is bounded, f is uniformly continuous, and we observe that the sequence $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence. Indeed, if $n > m > N$, we have

$$|x_n - x_m| \leq |x_n - x_{n-1}| + |x_{n-1} - x_{n-2}| + \cdots + |x_{m+1} - x_m|.$$

Now it is easy to show by induction, using the mean-value theorem and the fact that $|f'(x)| \leq A$ for all x, that

$$|x_{n+1} - x_n| \leq A^{n-1}|x_2 - x_1|$$

for $n \geq 1$. We therefore have

$$\begin{aligned} |x_n - x_m| &\leq |x_2 - x_1|(A^{n-2} + A^{n-3} + \cdots + A^{m-1}) \\ &< \frac{1}{1-A} A^{m-1}|x_2 - x_1| \\ &\leq \frac{|x_2 - x_1|}{1-A} A^N. \end{aligned}$$

Since $0 \leq A < 1$, it follows that $A^N \to 0$ as $N \to \infty$, and so this is a Cauchy sequence. Let its limit be x. We claim that x is a fixed point. Indeed, $x =$

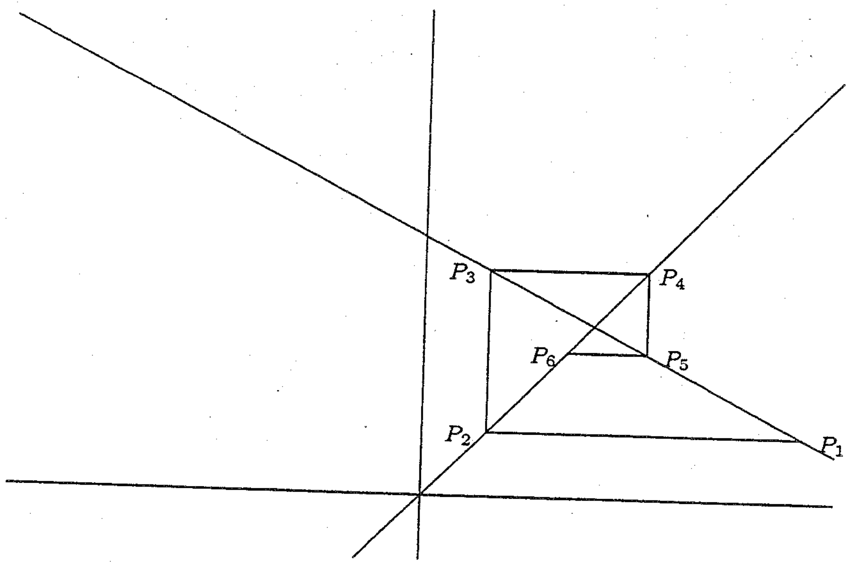

Figure 5.1: Finding a fixed point

$\lim_{n\to\infty} x_n = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n) = f(x)$, since f is continuous. There can of course be only one fixed point because of the result proved in (a).

(d) The procedure described can be depicted on the graph of the function f, i.e., the set of points $(x, f(x))$, as follows: Let x_1 be any abscissa; locate the point $(x_1, f(x_1))$ on the graph. Thereafter, for each point (x_n, y_n) located on the graph, let the abscissa of (x_{n+1}, y_{n+1}) be the ordinate of (x_n, y_n), i.e., $x_{n+1} = y_n$. Thus, from a point (x_n, y_n) on the graph of f we move horizontally to the line $y = x$, then vertically back to the graph of f. It is clear visually that this process leads to the point of intersection of the graph of f with the line $y = x$, as illustrated in Fig. 1 for the case of $f(x) = 2 - \frac{1}{2}x$, where $P_1 = (2, 1)$, $P_2 = (1, 1)$, $P_3 = (1, \frac{3}{2})$, $P_4 = (\frac{3}{2}, \frac{3}{2})$, $P_5 = (\frac{3}{2}, \frac{5}{4})$, and $P_6 = (\frac{5}{4}, \frac{5}{4})$. (The fixed point is $(\frac{4}{3}, \frac{4}{3})$, which is the point of intersection of the graph of f and the line $y = x$.)

Exercise 5.23 The function f defined by

$$f(x) = \frac{x^3 + 1}{3}$$

has three fixed points, say α, β, γ, where

$$-2 < \alpha < -1, \quad 0 < \beta < 1, \quad 1 < \gamma < 2.$$

For arbitrarily chosen x_1, define $\{x_n\}$ by setting $x_{n+1} = f(x_n)$.

(a) If $x_1 < \alpha$, prove that $x_n \to -\infty$ as $n \to \infty$.

(b) If $\alpha < x_1 < \gamma$, prove that $x_n \to \beta$ as $n \to \infty$.

(c) If $\gamma < x_1$, prove that $x_n \to +\infty$ as $n \to \infty$.

Thus β can be located by this method, but α and γ cannot.

Solution. We shall make use of the auxiliary functions

$$g(x) = f(x) - x = \frac{x^3+1}{3} - x$$

and

$$h(x) = \begin{cases} \dfrac{g(x)-g(\beta)}{x-\beta}, & x \neq \beta, \\ g'(\beta) & x = \beta, \end{cases}$$

i.e., $g(x) = \dfrac{x^2+\beta x+\beta^2}{3} - 1$. We observe that the fixed points of f are the zeros of g. Since $g(-2) = -\frac{1}{3} < 0$, $g(-1) = 1 > 0$, $g(0) = \frac{1}{3} > 0$, $g(1) = -\frac{1}{3} < 0$, and $g(2) = 1 > 0$, the intermediate value theorem shows that α, β, and γ are located in the intervals they are asserted to be in.

Since $g(\alpha) = g(\beta) = g(\gamma) = 0$, it follows that $h(\alpha) = h(\gamma) = 0$. Since h is a quadratic function, it has only the two zeros α and γ, and in particular $h(x)$ is negative for $\alpha < x < \gamma$. Now the minimum value of $h(x)$ is attained at $x = -\frac{\beta}{2}$, and this minimum value is c, where $c = \dfrac{\beta^2}{4} - 1$. Thus $-1 < c < 0$. In particular, for $\alpha < x < \gamma$ there is a number $r \in (0,1)$ such that

$$f(x) - x = r(\beta - x),$$

i.e.,

$$f(x) - \beta = s(x - \beta),$$

where $s = 1 - r$ is also in the interval $(0,1)$. This means that $f(x) - \beta$ and $x - \beta$ both have the same sign, but that $|f(x) - \beta| < |x - \beta|$. Thus $f(x)$ is always between β and x. Therefore the sequence $\{x_n\}$ is monotonic and converges to a fixed point in the interval whose endpoints are x_1 and β. Since the only fixed point in this interval is β, the sequence must converge to β.

If $x < \alpha$ (resp. $x > \gamma$), it is easy to see that $f(x) < x$ (resp. $f(x) > x$). Thus the sequence $\{x_n\}$ is monotonically decreasing (resp. increasing), and hence either tends to $-\infty$ (resp. $+\infty$) or converges to a fixed point δ in the interval $(-\infty, x_1)$ (resp. $(x_1, +\infty)$). Since there are no fixed points in this interval, it follows that $x_n \to -\infty$ (resp. $x_n \to +\infty$).

Exercise 5.24 The process described in part (*c*) of Exercise 22 can of course also be applied to functions that map $(0,\infty)$ to $(0,\infty)$.

Fix some $\alpha > 1$, and put

$$f(x) = \frac{1}{2}\Big(x + \frac{\alpha}{x}\Big), \quad g(x) = \frac{\alpha + x}{1 + x}.$$

Both f and g have $\sqrt{\alpha}$ as their only fixed point in $(0,\infty)$. Try to explain, on the basis of properties of f and g, why the convergence in Exercise 16, Chap. 3, is so much more rapid than it is in Exercise 17. (Compare f' and g', draw the zig-zag suggested in Exercise 22.)

Do the same when $0<\alpha<1$.

Solution. We recall that in Chap. 3 we proved that the first function leads to $|x_n-\sqrt{\alpha}|\leq Ar^{2^n}$ for some $r\in(0,1)$, while the second leads only to $|x_n-\sqrt{\alpha}|\leq Ar^n$. The exact values of A and r depend on α and x_1.

The best explanation of the difference between the two methods is that

$$\begin{aligned} f(x)-\sqrt{\alpha} &= \frac{1}{2}\Big(1-\frac{\sqrt{\alpha}}{x}\Big)(x-\sqrt{\alpha}), \\ g(x)-\sqrt{\alpha} &= \frac{1-\sqrt{\alpha}}{1+x}(x-\sqrt{\alpha}). \end{aligned}$$

The first of these makes it plain that if $x>\sqrt{\alpha}$, the same will be true of $f(x)$, though $f(x)$ will be closer to α than x by a factor that is at most $\frac{1}{2}$ and tends to zero as x tends to $\sqrt{\alpha}$, i.e., the relative improvement in accuracy itself improves as the recursion proceeds. The second equality shows that $g(x)-\alpha$ is on the opposite side of $\sqrt{\alpha}$ from x if $\alpha>1$, though closer by a factor that is at least the absolute value of $\dfrac{1-\sqrt{\alpha}}{1+x_1}$. Hence the relative improvement in accuracy as the recursion proceeds is limited.

In terms of the zigzag pattern, when we use g, the zigzag keeps circulating around the point of intersection of the graph of g and the line $y=x$ instead of moving steadily toward it in a staircase pattern.

When $0<\alpha<1$, the zigzag does stay on one side of the point of intersection of the two curves. However, the relative improvement is still at best a factor of $\dfrac{1-\sqrt{\alpha}}{2}$ when x is close to $\sqrt{\alpha}$.

Exercise 5.25 Suppose f is twice differentiable in $[a,b]$, $f(a)<0$, $f(b)>0$, $f'(x)\geq\delta>0$, and $0\leq f''(x)\leq M$ for all $x\in[a,b]$. Let ξ be the unique point in (a,b) at which $f(\xi)=0$.

Complete the details in the following outline of *Newton's method* for computing ξ.

(*a*) Choose $x_1\in(\xi,b)$, and define x_n by

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}.$$

Interpret this geometrically, in terms of a tangent to the graph of f.

(*b*) Prove that $x_{n+1}<x_n$, and that

$$\lim_{n\to\infty} x_n=\xi.$$

(*c*) Use Taylor's theorem to show that

$$x_{n+1}-\xi=\frac{f''(t_n)}{2f'(x_n)}(x_n-\xi)^2$$

for some $t_n \in (\xi, x_n)$.
(*d*) If $A = M/2\delta$, deduce that

$$0 \le x_{n+1}-\xi \le \frac{1}{A}[A(x_1-\xi)]^{2^n}.$$

(Compare with Exercises 16 and 18, Chap. 3.)
(*e*) Show that Newton's method amounts to finding a fixed point of the function g defined by

$$g(x)=x-\frac{f(x)}{f'(x)}.$$

How does $g'(x)$ behave for x near ξ?
(*f*) Put $f(x)=x^{1/3}$ on $(-\infty,\infty)$ and try Newton's method. What happens?

Solution. We remark at the outset that x_1 can be found by trying $z_0 = \frac{a+b}{2}$. If $f(z_0) > 0$, take $x_1 = z_0$. Otherwise let $z_{n+1} = (b+z_n)/2$, and let x_1 be the first z_n for which $f(z_n) > 0$. (In a finite number of steps we must reach such a point since $z_n \uparrow b$ and $f(b) > 0$.)

(*a*) The tangent line to the graph of f at the point x_n has the equation $y - f(x_n) = f'(x_n)(x - x_n)$. Setting $y = 0$ in this equation and solving for x gives $x = x_{n+1}$. Thus the interpretation of Newton's method is that we approximate the point where the graph of f intersects the x-axis by the point at which its tangent line at $(x_n, f(x_n))$ intersects the x-axis.

(*b*) We can assume by induction that $f(x_n) > 0$, and hence, since $f'(x_n) > 0$, it follows immediately that $x_{n+1} < x_n$. Notice that there exists c between x_n and x_{n+1} such that $f(x_{n+1}) = f(x_n) - f'(c)(x_n - x_{n+1}) > f(x_n) - f'(x_n)(x_n - x_{n+1}) = 0$ since $f'(c) < f'(x_n)$ and $x_n - x_{n+1} > 0$. Thus it follows that $\xi < x_{n+1} < x_n$. Hence $\{x_n\}$ converges to a limit η satisfying $\eta \ge \xi$. Now, however, we have

$$\eta=\eta-\frac{f(\eta)}{f'(\eta)},$$

from which it follows that $f(\eta) = 0$, i.e., $\eta = \xi$.

(*c*) The required equality can be written as

$$x_n-\xi-\frac{f(x_n)}{f'(x_n)}=\frac{f''(t_n)}{2f'(x_n)}(x_n-\xi)^2,$$

while Taylor's theorem can be written as

$$f(\xi)=f(x_n)+f'(x_n)(\xi-x_n)+\frac{f''(t_n)}{2}(\xi-x_n)^2.$$

Since $f(\xi) = 0$, it is clear that these two equations are equivalent.

(*d*) Since $0 \le f''(t_n) \le M$ and $f'(x_n) > \delta$, we have

$$0 \le x_{n+1} - \xi \le A(x_n - \xi)^2.$$

In particular

$$0 \le x_2 - \xi \le A(x_1 - \xi)^2 = \frac{1}{A}[A(x_1 - \xi)]^2,$$

and then an easy induction gets the general result.

We found this kind of convergence in Exercises 16 and 18 of Chap. 3 with the recursion relation

$$x_{n+1} = \frac{p-1}{p}x_n + \frac{\alpha}{p}x_n^{-p+1}.$$

We now recognize this recursion as Newton's method for the function $f(x) = x^p - \alpha$ on the interval $[1, \sqrt{\alpha} + 1]$. Exercise 16 of Chap. 2 was the special case $p = 2$.

(*e*) Obviously the equation $g(x) = x$ is equivalent to the equation $f(x) = 0$.

Since $g'(x) = \dfrac{f(x)f''(x)}{[f'(x)]^2}$, we see that $g'(x)$ tends to zero as x tends to ξ, i.e., the graph of $g(x)$ meets the line $y = x$ at a 45° degree angle at the point (ξ, ξ).

(*f*) The fixed point of $f(x)$ is $x = 0$. However $f'(x) \to \infty$ as $x \to 0$, and $f'(0)$ does not exist. This destroys the convergence of Newton's method. In fact, if $x_n \ne 0$, then $x_{n+1} = -2x_n$, so that x_n oscillates wildly: $\limsup x_n = +\infty$, $\liminf x_n = -\infty$.

Exercise 5.26 Suppose f is differentiable on $[a, b]$, $f(a) = 0$, and there is a real number A such that $|f'(x)| \le A|f(x)|$ on $[a, b]$. Prove that $f(x) = 0$ for all $x \in [a, b]$. *Hint:* Fix $x_n \in [a, b]$, let

$$M_0 = \sup |f(x)|, \quad M_1 = \sup |f'(x)|$$

for $a \le x \le x_0$. For any such x,

$$|f(x)| \le M_1(x_0 - a) \le A(x_0 - a)M_n.$$

Hence $M_0 = 0$ if $A(x_0 - a) \le 1$. That is, $f = 0$ on $[a, x_0]$. Proceed.

Solution. If we anticipate the fundamental result that the function $f(x) = e^x$ satisfies $f'(x) = f(x)$, Exercise 2 above yields the result that $\ln x$ is differentiable and has derivative $\frac{1}{x}$. Hence by the chain rule for any positive differentiable function $f(x)$ the function $g(x) = \ln f(x)$ is differentiable and $g'(x) = \dfrac{f'(x)}{f(x)}$. (Unfortunately this fundamental result is not proved until Chapter 7, so we shall just have to wait. However, since certain other functions such as $\sin x$ and $\cos x$ have been introduced without any formal definition, and their derivatives have been assumed known, we might as well continue along this line of reasoning.)

Now suppose there is an interval $(c, d) \subset [a, b]$ such that $f(c) = 0$ but $f(x) \neq 0$ for $c < x < d$. By passing to consideration of $-f(x)$ if necessary, we can assume $f(x) > 0$ for $c < x < d$. The function $g(x) = \ln f(x)$ is then defined for $c < x < d$, and its derivative satisfies

$$|g'(x)| = \left|\frac{f'(x)}{f(x)}\right| \leq A.$$

The mean-value theorem them implies that

$$g(x) \geq g\Big(\frac{c+d}{2}\Big) - A\Big(\frac{d-c}{2}\Big)$$

for all $x \in (c, d)$. But this is a contradiction, since $g(x) \to -\infty$ as $x \to c$.

This finishes the proof, except that it assumes we know the derivative of e^x. If we don't assume that, we have to fall back on the hint. In that case, let $x_0 = a + \frac{1}{2A}$, and let $M_0 = \sup\{|f(x)| : a \leq x \leq x_0\}$. We then have

$$|f(x)| \leq M_1(x-a) \leq AM_0(x_0 - a) = \frac{1}{2}M_0$$

for all $x \in [a, x_0]$. But by definition of M_0 this implies $M_0 \leq \frac{1}{2}M_0$, so that $M_0 \leq 0$, i.e., $M_0 = 0$. We now start over with a replaced by x_0, $x_1 = x_0 + \dfrac{1}{2A}$. In a finite number of steps, we will have $b < x_n + \dfrac{1}{2A}$, so that $f(x) = 0$ for $a \leq x \leq b$.

Exercise 5.27 Let ϕ be a real function defined on a rectangle R in the plane, given by $a \leq x \leq b$, $\alpha \leq y \leq \beta$. A *solution* of the initial-value problem

$$y' = \phi(x, y), \quad y(a) = c \quad (\alpha \leq c \leq \beta)$$

is, by definition, a differentiable function f on $[a, b]$ such that $f(a) = c$, $\alpha \leq f(x) \leq \beta$, and

$$f'(x) = \phi(x, f(x)) \quad (a \leq x \leq b).$$

Prove that such a problem has at most one such solution if there is a constant A such that

$$|\phi(x, y_2) - \phi(x, y_1)| \leq A|y_2 - y_1|$$

whenever $(x, y_1) \in R$ and $(x, y_2) \in R$.

Hint: Apply Exercise 26 to the difference of two solutions. Note that this uniqueness theorem does not hold for the initial-value problem

$$y' = y^{1/2}, \quad y(0) = 0,$$

which has two solutions: $f(x) = 0$ and $f(x) = x^2/4$. Find all other solutions.

Solution. Following the hint, we observe that if $f(x) = f_2(x) - f_1(x)$, then

$$\begin{aligned} |g'(x)| &= |f_2'(x) - f_1'(x)| \\ &= |\phi(x, f_2(x)) - \phi(x, f_1(x))| \\ &\leq A|f_2(x) - f_1(x)| \\ &= A|g(x)|. \end{aligned}$$

By the initial condition $g(a) = f_2(a) - f_1(a) = c - c = 0$. Hence by the preceding exercise $g(x) = 0$ for all $x \in [a, b]$.

As for the equation $y' = \sqrt{y}$, if $f(x)$ is a solution and $f(x) > 0$ on an interval (a, b), while $f(a) = 0$, we observe that $g(x) = \sqrt{f(x)}$ satisfies $g'(x) = \frac{1}{2}(f(x))^{-1/2}f'(x) = \frac{1}{2}$, so that for some constant c we have $g(x) = \frac{1}{2}(x + c)$. Thus

$$f(x) = (g(x))^2 = \frac{1}{4}(x + c)^2.$$

Since $f(a) = 0$, it follows that $c = -a$, i.e., $f(x) = \dfrac{(x-a)^2}{4}$. Thus the only possible solutions are

$$f(x) = \begin{cases} 0, & 0 \leq x \leq a, \\ \dfrac{(x-a)^2}{4}, & a \leq x. \end{cases}$$

Here $a \geq 0$ is arbitrary.

Exercise 5.28 Formulate and prove an analogous uniqueness theorem for systems of differential equations of the form

$$y_j' = \phi_j(x, y_1, \ldots, y_k), \quad y_j = c_j \quad (j = 1, \ldots, k).$$

Note that this can be rewritten in the form

$$\mathbf{y}' = \phi(x, \mathbf{y}), \quad \mathbf{y}(a) = \mathbf{c},$$

where $\mathbf{y} = (y_1, \ldots, y_k)$ ranges over a k-cell, ϕ is the mapping of a $(k+1)$-cell into the Euclidean k-space whose components are the functions $\phi_1, \ldots, \phi_k$, and $\mathbf{c}$ is the vector $(c_1, \ldots, c_k)$. Use Exercise 26 for vector-valued functions.

Solution. The result is the following:

Let ϕ be a vector-valued function defined on a $(k+1)$-cell $D = [a, b] \times C$ in R^{k+1} whose range is contained in R^k, and suppose that there exists a constant A such that

$$|\phi(x, \mathbf{y}_2) - \phi(x, \mathbf{y}_1)| \leq A|\mathbf{y}_2 - \mathbf{y}_1|$$

for all $\mathbf{y}_1 \in C$, $\mathbf{y}_2 \in C$. *Then the initial-value problem*

$$\mathbf{y}' = \boldsymbol{\phi}(x, \mathbf{y}) \quad \mathbf{y}(a) = \mathbf{c}$$

has at most one solution $\mathbf{y} : [a, b] \to C$.

The main tool needed to prove this result is the analogue of Exercise 26 for vector-valued functions, which does hold. Indeed the proof is identical, considering that the original proof depends only on the inequality $|f(d)-f(c)| \le |f'(r)|(d-c)$ for some $r \in (c, d)$, and this inequality is certainly valid for vector-valued functions. Once that result is obtained, the preceding exercise can be applied verbatim.

Exercise 5.29 Specialize Exercise 28 by considering the system

$$\begin{aligned} y_j' &= y_{j+1} \quad (j = 1, \dots, k-1), \\ y_k' &= f(x) - \sum_{j=1}^{k} g_j(x) y_j, \end{aligned}$$

where $f, g_1, \dots, g_k$ are continuous real functions on $[a, b]$, and derive a uniqueness theorem for solutions of the equation

$$y^{(k)} + g_k(x) y^{(k-1)} + \cdots + g_2(x) y' + g_1(x) y = f(x),$$

subject to initial conditions

$$y(a) = c_1, \quad y'(a) = c_2, \quad \dots, \quad y^{(k-1)}(a) = c_k.$$

Solution. We let $\mathbf{y} = (y_1, y_2, y_3, \dots, y_k) = (y, y', y'', \dots, y^{(k-1)})$ and $\boldsymbol{\phi}(x, \mathbf{y}) = \left(y_2, y_3, \dots, y_k, f(x) - \sum_{j=1}^{k} g_j(x) y_j\right)$. We then observe that if $\mathbf{y}_i = (y_{i1}, \dots, y_{ik})$, then

$$|\boldsymbol{\phi}(x, \mathbf{y}_2) - \boldsymbol{\phi}(x, \mathbf{y}_1)| = \left|\left(y_{22} - y_{12}, y_{23} - y_{13}, \dots, \sum_{j=1}^{k} g_j(x)(y_{1j} - y_{2j})\right)\right|.$$

If $M = \sup\{|g_j(x)| : a \le x \le b, \ 1 \le j \le k\}$, we then have

$$|\boldsymbol{\phi}(x, \mathbf{y}_2) - \boldsymbol{\phi}(x, \mathbf{y}_1)| \le (M+1) \sum_{j=1}^{k} |y_{2j} - y_{1j}| \le k(M+1)|\mathbf{y}_2 - \mathbf{y}_1|.$$

This provides the hypothesis of the theorem for any $(k+1)$-cell $[a, b] \times C$ whatsoever in R^{k+1}. Hence there is at most one solution to this initial-value problem.

Chapter 6

The Riemann–Stieltjes Integral

Exercise 6.1 Suppose α increases on $[a,b]$, $a \le x_0 \le b$, α is continuous at x_0, $f(x_0) = 1$, and $f(x) = 0$ if $x \ne x_0$. Prove that $f \in \mathcal{R}(\alpha)$ and that $\int f\,d\alpha = 0$.

Solution. Let $\varepsilon > 0$, and let δ be such that $|\alpha(x) - \alpha(x_0)| < \varepsilon$ if $|x - x_0| < \delta$. Now consider any partiion $a = t_0 < t_1 < \cdots < t_n = b$ with $n \ge 2$ such that $|t_i - t_{i-1}| < \frac{\delta}{2}$. There exists an index i such that $t_{i-1} < x_0 < t_{i+1}$ (there may possibly be 2 such indices). We then have, for any choice of $t_0^*, t_1^*, \ldots, t_n^*$,

$$\begin{aligned}\left|\sum_{j=1}^{n} f(t_j^*)\big(\alpha(t_j) - \alpha(t_{j-1})\big)\right| &\le |f(t_i^*)|[\alpha(t_i) - \alpha(t_{i-1})| + \\ &\quad +|f(t_{i+1}^*)|[\alpha(t_{i+1}) - \alpha(t_i)] \\ &\le \alpha(t_{i+1}) - \alpha(t_{i-1}) < \varepsilon.\end{aligned}$$

By definition of the Riemann–Stieltjes integral, this means that $f \in \mathcal{R}(\alpha)$ and $\int f\,d\alpha = 0$.

Exercise 6.2 Suppose $f \ge 0$, f is continuous on $[a,b]$, and $\displaystyle\int_a^b f(x)\,dx = 0$. Prove that $f(x) = 0$ for all $x \in [a,b]$. (Compare this with Exercise 1.)

Solution. Suppose $f(x_0) \ne 0$ for some $x_0 \in [a,b]$. Since $f(x)$ is continuous on $[a,b]$ and $\dfrac{f(x_0)}{2} > 0$, there exists $\delta > 0$ such that $|f(x) - f(x_0)| < \dfrac{f(x_0)}{2}$ for all $x \in [a,b]$ such that $|x - x_0| < \delta$. Let $\eta = \min(\delta, \max(x_0 - a, b - x_0))$, so that $\eta > 0$. Let I be the interval $[x_0 - \eta, x_0]$ if it is contained in $[a,b]$; otherwise let $I = [x_0, x_0 + \eta]$. Whichever is the case, $I \subseteq [a,b]$ and $f(x) =$

$f(x_0) + \big(f(x) - f(x_0)\big) \geq f(x_0) - |f(x) - f(x_0)| > \dfrac{f(x_0)}{2}$ for all $x \in I$. The functions $f_1(x)$ and $f_2(x)$ defined as

$$f_1(x) = \begin{cases} f(x), & x \in I, \\ 0, & x \notin I, \end{cases} \qquad f_2(x) = \begin{cases} f(x), & x \notin I, \\ 0, & x \in I, \end{cases}$$

are both nonnegative, bounded, and continuous except possibly at the two end-points of the interval I. They are therefore both Riemann-integrable. Consideration of Riemann sums shows that

$$\int_a^b f_1(x)\,dx \geq \eta\frac{\varepsilon}{2},$$

and

$$\int_a^b f_2(x)\,dx \geq 0,$$

It therefore follows that

$$\int_a^b f(x)\,dx = \int_a^b f_1(x)\,dx + \int_a^b f_2(x)\,dx \geq \eta\frac{\varepsilon}{2} > 0,$$

contradicting the hypothesis that $\displaystyle\int_a^b f(x)\,dx = 0$.

Exercise 6.3 Define three functions $\beta_1, \beta_2, \beta_3$ as follows: $\beta_j(x) = 0$ if $x < 0$, $\beta_j(x) = 1$ if $x > 0$ for $j = 1, 2, 3$ and $\beta_1(0) = 0$, $\beta_2(0) = 1$, $\beta_3(0) = \frac{1}{2}$. Let f be a bounded function on $[-1, 1]$.

(a) Prove that $f \in \mathcal{R}(\beta_1)$ if and only if $f(0-) = f(0)$ and that then

$$\int f\,d\beta_i = f(0).$$

(b) State and prove a similar result for β_2.

(c) Prove that $f \in \mathcal{R}(\beta_3)$ if and only if f is continuous at 0.

(d) If f is continuous at 0, prove that

$$\int f\,d\beta_1 = \int f\,d\beta_2 = \int f\,d\beta_3 = f(0).$$

Solution. Let $t_0 < t_1 < \cdots < t_{n-1} < t_n$ be any partition of any interval containing 0. Since the upper Riemann-Stieltjes sums become smaller and the lower ones larger when a point is added to any partition, in deciding whether a function is integrable or not, we may assume that 0 is one of the points of the partition. Let k be the index such that $t_k = 0$, so that the upper and lower Riemann-Stieltjes sums

$$\sum_{i=1}^{n} M_i(\beta_j(t_i) - \beta_j(t_{i-1})), \quad j = 1, 2, 3,$$

and

$$\sum_{i=1}^{n} m_i(\beta_j(t_i) - \beta_j(t_{i-1})), \quad j = 1, 2, 3,$$

are respectively M_k and m_k, M_{k-1} and m_{k-1}, $\dfrac{M_{k-1} + M_k}{2}$ and $\dfrac{m_{k-1} + m_k}{2}$.

(a) Since $m_k \le f(x) \le M_k$ for $0 \le x \le t_{k+1}$ in the first case, the sets of upper and lower sums contain elements arbitrarily near to each other if and only if for each ε there is a partition with $M_k - m_k < \varepsilon$. If such a partition exists, let $\delta = t_{k+1}$. Then we have $|f(x) - f(0)| \le M_k - m_k < \varepsilon$ for $0 \le x \le \delta$, and hence $\lim\limits_{x \to 0+} = f(0)$. Conversely, if $\lim\limits_{x \to 0+} = f(0)$, then for any ε, let $\delta > 0$ be such that $|f(x) - f(0)| < \delta$ if $0 < x < \delta$, and let P be a partition with $t_k = 0$, $t_{k+1} < \delta$. It is then clear that both upper and lower Riemann sums differ from $f(0)$ by less than ε, i.e., $\displaystyle\int f\,d\beta_1 = f(0)$.

(b) $f \in \mathcal{R}(\beta_2)$ if and only if $\lim\limits_{x \to 0-} f(x) = f(0)$ and if this condition holds, then $\displaystyle\int f\,d\beta_2 = f(0)$. The proof is identical to the proof just given, except that "+" is replaced by "−."

(c) In the third case, the upper and lower Riemann-Stieltjes sums differ by $\dfrac{(M_k - m_k) + (M_{k-1} + m_{k-1}}{2}$. If, given ε, there exists a partition containing 0 for which this difference is less than $\dfrac{\varepsilon}{2}$, let $\delta = \min(t_{k+1}, -t_{k-1})$. Then for $-\delta \le x \le \delta$ we certainly have

$$|f(x) - f(0)| \le \max\left(\frac{M_k - m_k}{2}, \frac{M_{k-1} - m_{k-1}}{2}\right) \le M_k - m_k + M_{k-1} - m_{k-1} < \varepsilon,$$

so that f is continuous at 0. The same argument shows that in this case

$$\int f\,d\beta_3 = f(0).$$

(d) This result is contained in (a)–(c).

Exercise 6.4 If $f(x) = 0$ for all irrational x, $f(x) = 1$ for all rational x, prove that $f \notin \mathcal{R}$ on $[a, b]$ for any $a < b$.

Solution. Every upper Riemann sum equals $b - a$, and every lower Riemann sum equals 0. Hence the set of upper sums and the set of lower sums do not have a common bound.

Exercise 6.5 Suppose f is a bounded real function on $[a, b]$ and $f^2 \in \mathcal{R}$ on $[a, b]$. Does it follow that $f \in \mathcal{R}$? Does the answer change if we assume that $f^3 \in \mathcal{R}$?

Solution. The integrability of f^2 does not imply the integrability of f. For example, one could let $f(x) = -1$ if x is irrational and $f(x) = 1$ if x is rational. Then every upper Riemann sum of f is $b - a$ and every lower sum is $a - b$. However, f^2, being the constant function 1, is integrable.

The integrability of f^3 does imply the integrability of f, by Theorem 6.11 with $\varphi(u) = \sqrt[3]{u}$.

Exercise 6.6 Le P be the Cantor set constructed in Sec. 2.44. Let f be a bounded real function on $[0, 1]$ which is continuous at every point outside P. Prove that $f \in \mathcal{R}$ on $[0, 1]$. [*Hint:* P can be covered by finitely many segments whose total length can be made as small as desired. Proceed as in Theorem 6.10.]

Solution. Let $M = \sup\{|f(x)| : a \le x \le b\}$, and let $\varepsilon > 0$ be given. Cover P by a finite collection of open intervals $O = \bigcup_{i=1}^{k} (a_i, b_i)$ such that $\sum (b_i - a_i) < \frac{\varepsilon}{4M}$. Let $\theta = \inf\{|x - y| : x \in P, y \in [a, b] \setminus O\}$. Since x and y range over disjoint compact sets, θ is a positive number. On the compact set $E = \{x : d(x, P) \ge \frac{1}{2}\theta\}$ the function f is uniformly continuous. Let $\delta > 0$ be such that $|f(x) - f(y)| < \frac{\varepsilon}{2(b - a)}$ if $x, y \in E$ and $|x - y| < \delta$. Then consider any partition $\{t_j\}$ of $[a, b]$ with $\max(t_j - t_{j-1}) < \min(\delta, \frac{1}{2}\theta)$. The difference between the upper and lower Riemann sums for this partition can be expressed as two sums:

$$\sum (M_j - m_j)(t_j - t_{j-1}) = \Sigma_1 + \Sigma_2,$$

where Σ_1 contains all the terms for which $[t_{j-1}, t_j]$ is contained in E and Σ_2 all the other terms. It is then obvious that

$$\Sigma_1 < \frac{\varepsilon}{2(b - a)} \sum (t_j - t_{j-1}) \le \frac{\varepsilon}{2},$$

and, since each interval $[t_{j-1}, t_j]$ that occurs in Σ_2 is contained in O,

$$\Sigma_2 < 2M \frac{\varepsilon}{4M} = \frac{\varepsilon}{2}.$$

Therefore the upper and lower Riemann sums for any such partition differ by less than ε, and so f is Riemann integrable.

Exercise 6.7 Suppose f is a real function on $[0, 1]$ and $f \in \mathcal{R}$ on $[c, 1]$ for every $c > 0$. Define

$$\int_0^1 f(x)\,dx = \lim_{c \to 0+} \int_c^1 f(x)\,dx$$

if this limit exists (and is finite).

(a) If $f \in \mathcal{R}$ on $[0,1]$ show that this definition of the integral agrees with the old one.

(b) Construct a function f such that the above limit exists, although it fails to exist with $|f|$ in place of f.

Solution. (a) Suppose $f \in \mathcal{R}$ on $[0,1]$. Let $\varepsilon > 0$ be given, and let $M = \sup\{|f(x)| : 0 \le x \le 1\}$. Let $c \in \left(0, \frac{\varepsilon}{4M}\right]$ be fixed, and consider any partition of $[0,1]$ containing c for which the upper and lower Riemann sums $\sum M_j(t_j - t_{j-1})$ and $\sum m_j(t_j - t_{j-1})$ of f differ by less than $\frac{\varepsilon}{4}$. Then the partition of $[c,1]$ formed by the points of this partition that lie in this interval certainly has the property that its upper and lower Riemann sums $\sum' M_j(t_j - t_{j-1})$ and $\sum' m_j(t_j - t_{j-1})$ differ by less than $\frac{\varepsilon}{4}$. Moreover, the terms of the original upper and lower Riemann sums *not* found in the sums for the smaller interval amount to less than $\frac{\varepsilon}{4}$. In short, we have shown that for $c < \frac{\varepsilon}{4M}$ and a suitable partition containing c,

$$\sum M_j(t-j-t_{j-1}) - \frac{\varepsilon}{4} < \int_0^1 f(x)\,dx \le< \sum m_j(t_j - t_{j-1}) + \frac{\varepsilon}{4}$$

and

$$\sum{}' M_j(t_j - t_{j-1}) - \frac{\varepsilon}{4} < \int_c^1 f(x)\,dx < \sum{}' m_j(t_j - +\frac{\varepsilon}{4}.$$

Moreover, we have also shown that

$$\left|\sum M_j(t_j - t_{j-1}) - \sum{}' M_j(t_j - t_{j-1})\right| < \frac{\varepsilon}{4}$$

and

$$\left|\sum m_j(t_j - t_{j-1}) - \sum{}' m_j(t_j - t_{j-1})\right| < \frac{\varepsilon}{4}.$$

combining these inequalities, we find that

$$\left|\int_0^1 f(x)\,dx - \int_c^1 f(x)\,dx\right| < \varepsilon$$

if $0 < c < \frac{\varepsilon}{4M}$.u

(b) Let

$$f(x) = (-1)^n(n+1)$$

for $\frac{1}{n+1} < x \le \frac{1}{n}$, $n = 1, 2, \ldots$. Then if $\frac{1}{N+1} \le c \le \frac{1}{N}$ we have

$$\int_c^1 f(x)\,dx = (-1)^N(N+1)\left(\frac{1}{N} - c\right) + \sum_{k=1}^{N-1} \frac{(-1)^k}{k}.$$

Since $0 \le \dfrac{1}{N} - c \le \dfrac{1}{N} - \dfrac{1}{N+1} = \dfrac{1}{N(N+1)}$, the first term on the right-hand side tends to zero as $c \downarrow 0$, while the sum approaches $\ln 2$. Hence this integral approaches a limit. However,

$$\int_c^1 |f(x)|\,dx = (N+1)\Big(\frac{1}{N} - c\Big) + \sum_{k=1}^{N-1} \frac{1}{k},$$

and in this case the first term on the right-hand side tends to zero as $c \downarrow 0$, while the sum tends to infinity.

Exercise 6.8 Suppose $f \in \mathcal{R}$ on $[a,b]$ for every $b > a$, where a is fixed. Define

$$\int_a^\infty f(x)\,dx = \lim_{x\to\infty} \int_a^b f(x)\,dx$$

if this limit exists (and is finite). In that case, we say that the integral on the left *converges*. If it also converges after f has been replaced by $|f|$, it is said to converge *absolutely*.

Assume that $f(x) \ge 0$ and that f decreases monotonically on $[1,\infty)$. Prove that

$$\int_1^\infty f(x)\,dx$$

converges if and only if

$$\sum_{n=1}^\infty f(n)$$

converges. (This is the so-called "integral test" for convergence of series.)

Solution. Since both the series and the integral are increasing functions of their upper limits, it suffices to show that they are bounded together. Define $f(x) = f(1)$ for $0 \le x \le 1$. Then consider a partition of $[0,n]$ consisting of the $n+1$ points $0, 1, 2, \ldots, n$. The upper Riemann sum for this partition is $\sum_{k=0}^{n-1} f(k)$ and the lower Riemann sum is $\sum_{k=1}^{n} f(k)$. Hence we have

$$\sum_{k=1}^{n} f(k) \le \int_0^n f(x)\,dx = f(0) + \int_1^n f(x)\,dx \le \sum_{k=0}^{n-1} f(k) = f(0) + \sum_{k=1}^{n-1} f(k).$$

This shows that

$$-f(0) + \sum_{k=1}^{n} f(k) \le \int_1^n f(x)\,dx \le \sum_{k=1}^{n-1} f(x),$$

and hence the sum and the integral converge or diverge together.

Exercise 6.9 Show that integration by parts can sometimes be applied to the "improper" integrals defined in Exercises 7 and 8. (State appropriate hypotheses, formulate a theorem, and prove it.) For instance show that

$$\int_0^\infty \frac{\cos x}{1+x}\,dx = \int_0^\infty \frac{\sin x}{(1+x)^2}\,dx.$$

Solution. Without striving for ultimate generality we can get the main ideas in the following theorem:

Theorem. *Let $f(x)$ and $g(x)$ be continuously differentiable functions defined on $[a,\infty)$ such that $\lim_{b\to\infty} f(b)g(b)$ exists and the integral $\int_a^\infty f(x)g'(x)\,dx$ converges. Then $\int_a^\infty f'(x)g(x)\,dx$ converges and*

$$\int_a^\infty f'(x)g(x)\,dx = \lim_{b\to\infty}[f(b)g(b) - f(a)g(a)] - \int_a^\infty f(x)g'(x)\,dx.$$

Proof. For each finite value of b larger than a the standard rule for integration by parts gives

$$\int_a^b f'(x)g(x)\,dx = [f(b)g(b) - f(a)g(a)] - \int_a^b f(x)g'(x)\,dx.$$

The hypotheses of the theorem guarantee that the limit on the right exists. Therefore, by definition, the integral on the left converges.

Applying this result with $f(x) = \sin x$, $g(x) = \frac{1}{1+x}$, we find, since $f(0)g(0) = 0$ and $\lim_{b\to\infty} f(b)g(b) = 0$, while $\int_0^\infty f(x)g'(x)\,dx$ converges absolutely, that

$$\int_0^\infty \frac{\cos x}{1+x}\,dx = \int_0^\infty \frac{\sin x}{(1+x)^2}\,dx.$$

Exercise 6.10 Let p and q be positive real number ssuch that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Prove the following statements.

(*a*) If $u \geq 0$ and $v \geq 0$, then

$$uv \leq \frac{u^p}{p} + \frac{v^q}{q}.$$

Equality holds if and only if $u^p = v^q$.

(*b*) if $f \in \mathcal{R}(\alpha)$, $g \in \mathcal{R}(\alpha)$, $f \geq 0$, $g \geq 0$, and

$$\int_a^b f^p \, d\alpha = 1 = \int_a^b g^q \, d\alpha,$$

then

$$\int_a^b fg \, d\alpha \leq 1.$$

(*c*) If f and g are complex functions in $\mathcal{R}(\alpha)$, then

$$\left| \int_a^b fg \, d\alpha \right| \leq \left\{ \int_a^b |f|^p \, d\alpha \right\}^{1/p} \left\{ \int_a^b |g|^q \, d\alpha \right|^{1/q}.$$

This is *Hölder's inequality.* When $p = q = 2$ it is usually called the Schwarz inequality. (Note that Theorem 1.35 is a very special case of this.)

(*d*) Show that Hölder's inequality is also true for the "improper" integrals described in Exercises 7 and 8.

Solution. (*a*) The inequality is obvious if either $u = 0$ or $v = 0$, and equality holds in that case if and only if $u = v = 0$. Hence assume $v > 0$. Keep v fixed. The inequality implies that $p > 1$ and $q > 1$, and hence the function $\varphi(u) = \dfrac{u^p}{p} + \dfrac{v^q}{q} - uv$ satisfies

$$\lim_{u \to +\infty} \varphi(u) = +\infty.$$

We also have $\varphi'(0) = -v < 0$. Hence the function $\varphi(u)$ has a minimum at some point u_0 on $(0, \infty)$ at which $0 = \varphi'(u_0) = u_0^{p-1} - v$, i.e., $u_0 = v^{\frac{1}{p-1}} = v^{q-1}$ and $u_0^p = v^q$. Note that $\varphi(u_0) = \dfrac{v^q}{p} + \dfrac{v^q}{q} - v^{q-1}v = v^q - v^q = 0$. Since this point is the only critical point for φ, we have $\varphi(u) > 0$ for all $u \neq u_0$, as required.

(*b*) Simply integrate the inequality

$$f(x)g(x) \leq \frac{f(x)^p}{p} + \frac{g(x)^q}{q}.$$

(*c*) The inequality is obviously equality if either of the two integrals on the right-hand side is zero. For the vanishing of, say $\displaystyle\int_a^b |f|^p \, d\alpha$ implies the vanishing of $\displaystyle\int_a^b M|f| \, d\alpha$ and hence the vanishing of $\displaystyle\int_a^b |g|\,|f| \, d\alpha$ if $|g(x)| \leq M$ for all x.

Hence we now assume that $\displaystyle\int_a^b |f|^p \, d\alpha > 0$ and $\displaystyle\int_a^b |g|^q \, d\alpha > 0$. In part (*b*) we replace $f(x)$ by $\dfrac{|f(x)|}{\left(\int_a^b |f|^p \, d\alpha\right)^{1/p}}$ and $g(x)$ by $\dfrac{|g(x)|}{\left(\int_a^b |g|^q \, d\alpha\right)^{1/q}}$. We then need only invoke the inequality $\left| \displaystyle\int_a^b h \, d\alpha \right| \leq \displaystyle\int_a^b |h| \, d\alpha$.

(*d*) The inequality holds on each finite interval. If either of the factors on the right-hand side diverges as $b \to \infty$, the inequality is obvious. If they both converge, it follows that the left-hand side converges absolutely, and to a limit not larger than the limit of the right-hand side.

Exercise 6.11 Let α be a fixed increasing function on $[a, b]$. For $u \in \mathcal{R}(\alpha)$ define

$$\|u\|_2 = \left\{ \int_a^b |u|^2 \, d\alpha \right\}^{1/2}.$$

Suppose f, g, and $h \in \mathcal{R}(\alpha)$, and prove the triangle inequality

$$\|f - h\|_2 \le \|f - g\|_2 + \|g - h\|_2$$

as a consequence of the Schwarz inequality, as in the proof of Theorem 1.37.

Solution. We have

$$\begin{aligned}
\|f - h\|_2^1 &= \int_a^b |f - h|^2 \, d\alpha \\
&= \int_a^b |(f - g) + (g - h)|^2 \, d\alpha \\
&= \int_a^b |f - g|^2 \, d\alpha + 2\int_a^b |f - g| \, |g - h| \, d\alpha + \int_a^b |g - h|^2 \, d\alpha \\
&\le \|f - g\|_2^2 + 2\|f - g\|_2 \|g - h\|_2 + \|g - h\|_2^2 \\
&= \left(\|f - g\|_2 + \|g - h\|_2\right)^2,
\end{aligned}$$

from which the desired inequality follow when square roots are taken.

Exercise 6.12 With the notations of Exercise 11, suppose $f \in \mathcal{R}(\alpha)$ and $\varepsilon > 0$. Prove that there exists a continuous function g on $[a, b]$ such that $\|f - g\|_2 < \varepsilon$.

Hint: Let $P = \{x_0, \ldots, x_n\}$ be a suitable partition of $[a, b]$, define

$$g(t) = \frac{x_i - t}{\Delta x_i} f(x_{i-1}) + \frac{t - x_{i-1}}{\Delta x_i} f(x_i)$$

if $x_{i-1} \le t \le x_i$.

Solution. Since $g(t)$ is defined on $[x_{i-1}, x_i]$ as the weighted average of the values of $f(x)$ at the endpoints, the weights being proportional to the distances from t to the endpoints, it is clear that $g(t)$ is piecewise linear, hence continuous. For the same reason the maximum value of the function $h = |g - f|$ on the interval $[x_{i-1}, x_i]$ will be at most $M_i - m_i$ where M_i and m_i are the maximum and minimum values of f on this interval. Let M be the maximum of $|f(x)|$ for $a \leq x \leq b$. If the partition is chosen so that

$$\sum (M_i - m_i)[\alpha(t_i) - \alpha(t_{i-1}] < \frac{\varepsilon^2}{2M},$$

then we will have

$$\sum (M_i - m_i)^2[\alpha(t_i) - \alpha(t_{i-1}] \leq 2M \sum (M_i - m_i)[\alpha(t_i) - \alpha(t_{i-1})] < \varepsilon^2,$$

and hence the upper Riemann integral for $|g - f|^2$ for this partition will also be less than ε^2. Therefore $\|g - f\|_2 < \varepsilon$, as required.

Exercise 6.13 Define

$$f(x) = \int_x^{x+1} \sin(t^2)\, dt.$$

(a) Prove that $|f(x)| < 1/x$ if $x > 0$.
Hint: Put $t^2 = u$ and integrate by parts to show that $f(x)$ is equal to

$$\frac{\cos(x^2)}{2x} - \frac{\cos[(x+1)^2]}{2(x+1)} - \int_{x^2}^{(x+1)^2} \frac{\cos u}{4u^{3/2}}\, du.$$

Replace $\cos u$ by -1.

(b) Prove that

$$2xf(x) = \cos(x^2) - \cos[(x+1)^2] + r(x),$$

where $|r(x)| < c/x$, and c is constant.

(c) Find the upper and lower limits of $xf(x)$ as $x \to \infty$.

(d) Does $\int_0^\infty \sin(t^2)\, dt$ converge?

Solution. (a) This inequality is obvious if $0 < x \leq 1$. Hence we assume $x > 1$. Following the hint, we observe that

$$\begin{aligned} f(x) &< \frac{\cos(x^2)}{2x} - \frac{\cos[(x+1)^2]}{2(x+1)} + \frac{1}{2x} - \frac{1}{2(x+1)} \\ &= \frac{1+\cos(x^2)}{2x} - \frac{1+\cos[(x+1)^2]}{2(x+1)} \\ &\leq \frac{1+\cos(x^2)}{2x} \\ &\leq \frac{1}{x}. \end{aligned}$$

A similar argument shows that

$$\begin{aligned}
f(x) &> \frac{\cos(x^2)}{2x} - \frac{\cos[(x+1)^2]}{2(x+1)} - \frac{1}{2x} + \frac{1}{2(x+1)} \\
&= \frac{-1+\cos(x^2)}{2x} - \frac{-1+\cos[(x+1)^2]}{2(x+1)} \\
&= \frac{-1+\cos(x^2)}{2x} + \frac{1-\cos[(x+1)^2]}{2(x+1)} \\
&\geq \frac{-1+\cos(x^2)}{2x} \\
&\geq \frac{-1}{x}.
\end{aligned}$$

(*b*) The expression just written for $f(x)$ shows that

$$2xf(x) = \cos(x^2) - \cos[(x+1)^2] + r(x),$$

where

$$r(x) = \left(\frac{1}{x+1}\right)\cos[(x+1)^2] - \frac{x}{2}\int_{x^2}^{(x+1)^2} \frac{\cos u}{u^{3/2}}\,du.$$

If we integrate by parts again, we find that

$$\int_{x^2}^{(x+1)^2} \frac{\cos u}{u^{3/2}}\,du = \frac{\sin[(x+1)^2]}{(x+1)^3} - \frac{\sin(x^2)}{x^3} + \frac{3}{2}\int_{x^2}^{(x+1)^2} \frac{\sin u}{x^{5/2}}\,du.$$

We now observe that the absolute value of this last integral is at most

$$\frac{3}{2}\int_{x^2}^{\infty} \frac{1}{u^{5/2}}\,du = -u^{-3/2}\Big|_{x^2}^{\infty} = x^{-3}.$$

It then follows by collecting the terms that

$$|r(x)| < \frac{3}{x}.$$

(*c*) Since $r(x) \to 0$, the upper and lower limits of $xf(x)$ will be the corresponding limits of

$$\frac{\cos(x^2) - \cos[(x+1)^2]}{2} = \sin\left(x^2 + x + \frac{1}{2}\right)\sin\left(x + \frac{1}{2}\right).$$

We can write this last expression as $\sin s \sin\left(s^2 + \frac{1}{4}\right)$, where $s = x + \frac{1}{2}$. We claim that the upper limit of this expression is 1 and the lower limit is -1. Indeed, let $\varepsilon > 0$ be given. Choose n to be any positive integer larger than $\dfrac{2-\varepsilon}{8\varepsilon}$. Then the interval $\left(\frac{1}{4} + \left(\left(2n+\frac{1}{2}\right)\pi - \varepsilon\right)^2, \frac{1}{4} + \left(\left(2n+\frac{1}{2}\right)\pi + \varepsilon\right)^2\right)$ is longer than 2π, and hence there exists a point $t \in \left(\left(2n+\frac{1}{2}\right)\pi - \varepsilon, \left(2n+\frac{1}{2}\right)\pi + \varepsilon\right)$

at which $\sin\left(t^2+\frac{1}{4}\right)=1$ and also a point u in the same interval at which $\sin\left(u^2+\frac{1}{4}\right)=-1$. But then $tf(t)>1-\varepsilon$ and $uf(u)<-1+\varepsilon$. It follows that the upper limit is 1 and the lower limit is -1. (This argument actually shows that the limit points of $xf(x)$ fill up the entire interval $[-1,1]$.)
(d) The integral does converge. We observe that for integers N we have

$$\begin{aligned}
\int_0^N \sin(t^2)\,dt &= \sum_{k=0}^{N} f(k) \\
&= f(0)+\sum_{k=1}^{N}\frac{r(k)}{k}+\sum_{k=1}^{N}\frac{\cos(k^2)-\cos[(k+1)^2]}{k} \\
&= f(0)+\sum_{k=1}^{N}\frac{r(k)}{k}+\left[\frac{\cos 1}{2}-\frac{\cos[(N+1)^2]}{N}\right]+\sum_{k=2}^{N}\frac{\cos(k^2)}{k(k-1)}.
\end{aligned}$$

The first sum on the right converges since $|r(k)|<\frac{3}{k}$, and the rest obviously converges. Hence we will be finished if we show that

$$\lim_{x\to\infty}\int_{[x]}^{x}\sin(t^2)\,dt=0,$$

where $[x]$ is the integer such that $[x]\le x<[x]+1$. But this is easily done using integration by parts. The integral equals

$$\frac{\cos([x]^2)}{2[x]}-\frac{\cos(x^2)}{x^2}-\int_{[x]^2}^{x^2}\frac{\cos u}{4u^{3/2}}\,du,$$

and this expression obviously tends to zero as $x\to\infty$.

Exercise 6.14 Deal similarly with

$$f(x)=\int_x^{x+1} sin(e^t)\,dt.$$

Show that

$$e^x|f(x)|<2$$

and that

$$e^x f(x)=\cos(e^x)-e^{-1}\cos(e^{x+1})+r(x),$$

where $|r(x)|<Ce^{-x}$ for some constant C.

Solution. The arguments are completely analogous to the preceding problem. The substitution $u=e^t$ changes $f(x)$ into

$$f(x)=\int_{e^x}^{e^{x+1}}\frac{\sin u}{u}\,du,$$

and then integration by parts yields

$$f(x) = \frac{\cos(e^x)}{e^x} - \frac{\cos(e^{x+1})}{e^{x+1}} - \int_{e^x}^{e^{x+1}} \frac{\cos u}{u^2}\, du$$

from which it then follows that

$$-\frac{1-\cos(e^x)}{e^x} \le f(x) \le \frac{1+\cos(e^x)}{e^x}.$$

We have the equality

$$e^x f(x) = \cos(e^x) - e^{-1}\cos(e^{x+1}) - e^x \int_{e^x}^{e^{x+1}} \frac{\cos u}{u^2}\, du,$$

and one more integration by parts shows that

$$\left| e^x \int_{e^x}^{e^{x+1}} \frac{\cos u}{u^2}\, du \right| < \frac{3}{e^x}.$$

In this case $f(x)$ decreases so rapidly that there is no difficulty at all proving the convergence of the integral.

Exercise 6.15 Suppose f is a real, continuously differentiable function on $[a, b]$, $f(a) = f(b) = 0$, and

$$\int_a^b f^2(x)\, dx = 1.$$

Prove that

$$\int_a^b x f(x) f'(x)\, dx = -\frac{1}{2}$$

and that

$$\int_a^b [f'(x)]^2\, dx \cdot \int_a^b x^2 f^2(x)\, dx \ge \frac{1}{4}.$$

Solution. To prove the first assertion we merely integrate by parts, taking $u = x$, $dv = f(x)f'(x)\, dx$, so that $du = dx$ and $v = \frac{1}{2}f^2(x)$. Since v vanishes at both endpoints, the result is

$$\int_a^b x f(x) f'(x)\, dx = -\frac{1}{2}\int_a^b f^2(x)\, dx = -\frac{1}{2}.$$

The second inequality is an immediate consequence of the Schwarz inequality applied to the two functions $xf(x)$ and $f'(x)$.

Exercise 6.16 For $1 < s < \infty$, define

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

(This is Riemann's zeta function, of great importance in the study of the distribution of prime numbers.) Prove that

$$(a) \qquad \zeta(s) = s\int_1^{\infty} \frac{[x]}{x^{s+1}}\,dx$$

and that

$$(b) \qquad \zeta(s) = \frac{s}{s-1} - s\int_1^{\infty} \frac{x-[x]}{x^{s+1}}\,dx,$$

where $[x]$ denotes the greatest integer $\leq x$.

Prove that the integral in (b) converges for all $x > 0$.

Hint: To prove (a) compute the difference between the integral over $[1, N]$ and the Nth partial sum of the series that defines $\zeta(s)$.

Solution. (a) Ignoring the author's advice, we note that

$$\begin{aligned} s\int_1^{\infty} \frac{[x]}{x^{s+1}}\,dx &= s\sum_{n=1}^{\infty} n\int_n^{n+1} \frac{1}{x^{s+1}}\,dx \\ &= \sum_{n=1}^{\infty} n\Big[\frac{1}{n^s} - \frac{1}{(n+1)^s}\Big] \\ &= 1\Big[\frac{1}{1^s} - \frac{1}{2^s}\Big] + 2\Big[\frac{1}{2^s} - \frac{1}{3^s}\Big] + \cdots \\ &= \sum_{n=1}^{\infty} \frac{1}{n^s} \\ &= \zeta(s). \end{aligned}$$

(b) This result is a trivial consequence of (a) and the identity

$$\frac{s}{s-1} = \int_1^{\infty} \frac{x}{x^{s+1}}\,dx.$$

Exercise 6.17 Suppose α increases monotonically on $[a, b]$, g is continuous, and $g(x) = G'(x)$ for $a \leq x \leq b$. Prove that

$$\int_a^b \alpha(x)g(x)\,dx = G(b)\alpha(b) - G(a)\alpha(a) - \int_z^b G\,d\alpha.$$

Hint: Take g real, without loss of generality. Given $P = \{x_0, x_1, \ldots, x_n\}$, choose $t_i \in (x_{i-1}, x_i)$ so that $g(t_i)\Delta x_i = G(x_i) - G(x_{i-1})$. Show that

$$\sum_{i=1}^{n} \alpha(x_i)g(t_i)\Delta x_i = G(b)\alpha(b) - G(a)\alpha(a) - \sum_{i=1}^{n} G(x_{i-1})\Delta\alpha_i.$$

Solution. The identity just given is a trivial consequence of Abel's method of rearranging the sums:

$$\begin{aligned}\sum_{i=1}^{n} \alpha(x_i)g(t_i)\Delta x_i &= \sum_{i=1}^{n} \alpha(x_i)\big(G(x_i) - G(x_{i-1})\big)\\ &= G(x_n)\alpha(x_n) - G(x_0)\alpha(x_0) - \sum_{i=1}^{G}(x_{i-1})\big(\alpha(x_i) - \alpha_{i-1})\big).\end{aligned}$$

Now the fact that $G(x)$ is continuous and α is nondecreasing means that the right-hand side can be made arbitrarily close to

$$G(b)\alpha(b) - G(a)\alpha(a) - \int_a^b G\,d\alpha,$$

whenever the partition is sufficiently fine. It does not follow immediately that the function $\alpha(x)g(x)$ is integrable on $[a, b]$. However, since α is nondecreasing, its only discontinuties are jumps, and for any given $\varepsilon > 0$ there can be only a finite number of jumps larger than ε. These can be enclosed in a finite number of open intervals of arbitrarily small length. We can then argue, as in Exercise 6 above, that *any* partition that is sufficiently fine will have upper and lower Riemann sums that differ by less than ε. Hence $\alpha(x)g(x)$ is integrable, and its integral is given by the stated relation.

Exercise 6.18 Let γ_1, γ_2, γ_3 be curves in the complex plane defined on $[0, 2\pi]$ by

$$\gamma_1(t) = e^{it}, \quad \gamma_2(t) = e^{2it}, \quad \gamma_3(t) = e^{2\pi it\sin(1/t)}.$$

Show that these curves have the same range, that γ_1 and γ_2 are rectifiable, that the length of γ_1 is 2π, that the length of γ_2 is 4π, and that γ_3 is not rectifiable.

Solution. Since e^{it} has period 2π it is obvious that γ_1 and γ_2 have the same range, namely the set of all complex numbers of absolute value 1. To show that this is also the range of γ_3, we need to show that the mapping $t \mapsto 2\pi t\sin(1/t)$, $0 \le t \le 2pi$, covers an interval of length 2π, i.e., that the mapping $t \mapsto t\sin(1/t)$, $0 \le t \le 2\pi$ covers an interval of length 1. (We naturally take the value to be zero when $t = 0$.) Since this range is connected, it suffices to find two points a and b in the range with $a - b > 1$. We choose those points to be $a = \dfrac{3}{\pi}$ (the image of $t = \frac{6}{\pi}$) and $b = \dfrac{-2}{3\pi}$, (the image of $t = \frac{2}{3\pi}$). We have $a - b = \frac{11}{3\pi} > 1$.

The rectification of γ_1 and γ_2 is straightforward:

$$l(\gamma_1) = \int_0^{2\pi} |\gamma_1'(t)|\,dt = 2\pi,$$

$$l(\gamma_2) = \int_0^{2\pi} |\gamma_2'(t)|\,dt = \int_0^{2\pi} 2\,dt = 4\pi.$$

To show that γ_3 is not rectifiable, we observe that its length would be

$$\int_0^{2\pi} \left| \sin(1/t) - \frac{1}{t}\cos(1/t) \right| dt \geq \int_0^{2\pi} \left| \frac{\cos(1/t)}{t} \right| dt - 2\pi.$$

By making the substitution $u = \frac{1}{t}$ in this last integral we get

$$\int_{\frac{1}{2\pi}}^{\infty} \left| \frac{\cos u}{u} \right| du.$$

But we already know that this integral diverges, since

$$\sum_{n=1}^{\infty} \int_{2n\pi}^{(2n+\frac{1}{2})\pi} \frac{\cos u}{u}\, du \geq \sum_{n=1}^{\infty} \frac{1}{(2n+\frac{1}{2})\pi} = \infty.$$

Exercise 6.19 Let γ_1 be a curve in R^k defined on $[a,b]$; let ϕ be a continuous 1-1 mapping of $[c,d]$ onto $[a,b]$ such that $\phi(c) = a$, and define $\gamma_2(x) = \gamma_1(\phi(x))$. Prove that γ_2 is an arc, a closed curve, or a rectifiable curve if and only if the same is true of γ_1. Prove that γ_1 and γ_2 have the same length.

Solution. We know that ϕ has a continuous 1-1 inverse φ, and that the composition of one-to-one functions is one-to-one. Hence, since $\gamma_1(x) = \gamma_2(\varphi(x))$, we see that γ_1 and γ_2 are both arcs (one-to-one) if either is. Since necessarily $\phi(d) = b$, we see that $\gamma_1(a) = \gamma_1(b)$ if and only if $\gamma_2(c) = \gamma_2(d)$. Hence both are closed curves if either is. Finally, since ϕ and φ establish a one-to-one correspondence between partitions $\{s_i\}$ of $[a,b]$ and $\{t_i\}$ of $[c,d]$ such that $\sum |\gamma_1(s_i) - \gamma_1(s_{i-1})| = \sum |\gamma_2(t_i) - \gamma_2(t_{i-1})|$, it follows that the two curves have the same length.

Chapter 7

Sequences and Series of Functions

Exercise 7.1 Prove that every uniformly convergent sequence of bounded functions is uniformly bounded.

Solution. Let $\{f_n(x)\}_{n=1}^{\infty}$ be a uniformly convergent sequence of bounded functions, say $|f_n(x)| \le M_n$ for all x and all n. Since the sequence converges uniformly, it is a uniformly Cauchy sequence. Hence there exists N such that $|f_m(x) - f_n(x)| < 1$ for all $m, n \ge N$. In particular if $m \ge N$, we have $|f_m(x)| \le |f_N(x)| + |f_m(x) - f_N(x)| \le M_N + 1$, and therefore if $M = 1 + \max(M_1, \ldots, M_N)$ we have $|f_n(x)| \le M$ for all n and x.

Exercise 7.2 If $\{f_n\}$ and $\{g_n\}$ converge uniformly on a set E, prove that $\{f_n + g_n\}$ converges uniformly on E. If, in addition, $\{f_n\}$ and $\{g_n\}$ are sequences of bounded functions, prove that $\{f_n g_n\}$ converges uniformly on E.

Solution. Let f and g denote the limits of the two sequences. Let $\varepsilon > 0$. There exist N_1 and N_2 such that $|f_n(x) - f(x)| < \frac{\varepsilon}{2}$ for all x if $n > N_1$ and $|g_n(x) - g(x)| < \frac{\varepsilon}{2}$ for all x if $n > N_2$. Let $N = \max(N_1, N_2)$. Then for $n > N$ we have, for all x,

$$|(f_n + g_n)(x) - (f + g)(x)| \le |f_n(x) - f(x)| + |g_n(x) - g(x)| < \varepsilon.$$

Hence $\{f_n + g_n\}$ converges uniformly.

Suppose now that each of the functions f_n and g_n is bounded. By the previous problem, both sequences are uniformly bounded. Hence there exists M such that $|f_n(x)| \le M$ and $|g_n(x)| \le M$ for all n and all x. It follows that $|g(x)| \le M$ also. Then, given $\varepsilon > 0$, choose N_1 and N_2 such that $|f_n(x) - f(x)| \le \frac{\varepsilon}{2M}$ for all x and all $n > N_1$ and $|g_n(x) - g(x)| < \frac{\varepsilon}{2M}$ for all x and $n > N_2$.

Again let $N = \max(N_1, N_2)$. We then have, for all x and all $n > N$,

$$\begin{aligned}
|f_n(x)g_n(x) - f(x)g(x)| &\leq |f_n(x)g_n(x) - f_n(x)g(x)| + \\
&\quad +|f_n(x)g(x) - f(x)g(x)| \\
&\leq M|g_n(x) - g(x)| + M|f_n(x) - f(x)| \\
&< M\frac{\varepsilon}{2M} + M\frac{\varepsilon}{2M} \\
&= \varepsilon.
\end{aligned}$$

Exercise 7.3 Construct sequences $\{f_n\}$, $\{g_n\}$ which converge uniformly on some set E, but such that $\{f_n g_n\}$ does not converge uniformly on E (of course, $\{f_n g_n\}$ must converge on E).

Solution. Let $f_n(x) = x$ for all x and all n, and let $g_n(x) = \frac{1}{n}$ for all x and all n. Then $f_n(x)$ converges uniformly to x, and $g_n(x)$ converges uniformly to 0. Therefore $f_n(x)g_n(x)$ converges to 0, but not uniformly. In fact for every n there is an x, namely $x = n$, such that $f_n(x)g_n(x) = 1$. Hence, no matter how large n is taken, the inequality $|f_n(x)g_n(x)| < 1$ will never hold for all x.

Exercise 7.4 Consider

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{1+n^2x}.$$

For what values of x does the series converge absolutely? On what intervals does it converge uniformly? On what intervals does it fail to converge uniformly? Is f continuous wherever the series converges? Is f bounded?

Solution. The series converges for all x except 0 and $x = \frac{-1}{n^2}$, $n = 1, 2, \ldots$. For $x = 0$ all the terms of the series are defined, but the terms do not tend to 0. For $x = \frac{-1}{n^2}$ the nth term is not defined. For all other values of x the series converges. By Theorem 7.10 (the Weierstrass M-test) the series converges uniformly on the interval $[\delta, \infty)$ if $\delta > 0$, since on that interval

$$\frac{1}{1+n^2x} \leq \frac{1}{n^2\delta}.$$

Likewise, the series converges uniformly on $(-\infty, -\delta]$ except at the points $x = -\frac{1}{n^2}$, since for $n \geq \sqrt{\frac{2}{\delta}}$ we have

$$\left|\frac{1}{1+n^2x}\right| \leq \frac{1}{n^2} \cdot \frac{1}{\delta - \frac{1}{n^2}} \leq \frac{2}{\delta n^2}.$$

The series does not converge uniformly on any interval having 0 as an endpoint. This is easy to see in the case when 0 is the left-hand endpoint. For each of the terms of the series is a bounded function on $[0,\infty)$. If the series converged uniformly, the limit would be bounded by Problem 1 above. But we have

$$f\Big(\frac{1}{m^2}\Big) \geq \sum_{n=1}^{m} \frac{1}{1+\frac{n^2}{m^2}} \geq \frac{m}{2}.$$

Likewise the series cannot be a uniformly Cauchy series (i.e., the sequence of partial sums cannot be a uniformly Cauchy sequence) on any interval $(-\delta, 0)$, since, no matter how large n is taken, there is a point x in this interval, namely $x = -\frac{1}{2n^2}$, at which the nth term has the value 2. Hence, if S_n denotes the sum of the first n terms, then $|S_n(x) - S_{n-1}(x)| = 2$.

The uniform convergence shows that the limiting function $f(x)$ is continuous wherever it is defined on $(-\infty, \delta] \cup [\delta, +\infty)$. Since δ is arbitrary, $f(x)$ is continuous wherever it is defined. The argument given above shows that $f(x)$ is not bounded.

Exercise 7.5 Let

$$f_n(x) = \begin{cases} 0 & (x < \frac{1}{n+1}), \\ \sin^2 \frac{\pi}{x} & (\frac{1}{n+1} \leq x \leq \frac{1}{n}), \\ 0 & (\frac{1}{n} < x). \end{cases}$$

Show that $\{f_n\}$ converges to a continuous function, but not uniformly. Use the series $\sum f_n$ to show that absolute convergence, even for all x does not imply uniform convergence.

Solution. The limit of $f_n(x)$ is zero. If $x \leq 0$ or $x \geq 1$, then $f_n(x) = 0$ for all n, and so this assertion is obvious. If $0 < x < 1$, then $f_n(x) = 0$ for all $n \geq \frac{1}{x}$, and so once again the assertion is obvious.

The convergence is not uniform, since, no matter how large n is taken, there is a point x, namely $x = \frac{1}{2n+\frac{1}{2}}$, for which $f_n(x) = 1$.

The series $\sum f_n(x)$ converges to 0 for $x \leq 0$ and $x \geq 1$, and to $\sin^2 \frac{\pi}{x}$ for $0 < x < 1$. Since the terms are nonnegative, the series obviously converges absolutely. Since the sum is not continuous at 0, the series does not converge uniformly on any interval containing 0.

Exercise 7.6 Prove that the series

$$\sum_{n=1}^{\infty} (-1)^n \frac{x^2+n}{n^2}$$

converges uniformly in every bounded interval, but does not converge absolutely for any value of x.

Solution. The series is the sum of two series:

$$x^2 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} + \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}.$$

The first of these converges both uniformly and absolutely on any bounded interval $[a, b]$ by the M-test (with $M_n = \frac{M^2}{n^2}$, where $M = \max(|a|, |b|)$). The second is independent of x and converges, hence it converges uniformly in x. By Exercise 2 above, the sum of the two series converges uniformly.

The series does not converge absolutely since the absolute value of each term is at least $\frac{1}{n}$ for any x.

Exercise 7.7 For $n = 1, 2, 3, \ldots$, x real, put

$$f_n(x) = \frac{x}{1 + nx^2}.$$

Show that $\{f_n\}$ converges uniformly to a function f, and that the equation

$$f'(x) = \lim_{n\to\infty} f_n'(x)$$

is correct if $x \neq 0$, but false if $x = 0$.

Solution. The Schwarz inequality, which implies that $|f_n(x)| \leq \frac{|x|}{2\sqrt{n}|x|} = \frac{1}{2\sqrt{n}}$ for $x \neq 0$, shows that $f_n(x)$ tends uniformly to 0. Now $f_n'(x) = \frac{1-nx^2}{(1+nx^2)^2}$, which tends to 0 if $x \neq 0$, though $f_n'(0) = 1$ for all n.

Exercise 7.8 If

$$I(x) = \begin{cases} 0 & (x \leq 0), \\ 1 & (x > 0), \end{cases}$$

if $\{x_n\}$ is a sequence of distinct points of (a, b), and if $\sum |c_n|$ converges, prove that the series

$$f(x) = \sum_{n=1}^{\infty} c_n I(x - x_n) \quad (a \leq x \leq b)$$

converges uniformly, and that f is continuous for every $x \neq x_n$.

Solution. The uniform convergence is a consequence of the M-test with $M_n = |c_n|$. Hence f is continuous wherever each of the individual terms is continuous, in particular, at least for $x \neq x_n$.

Exercise 7.9 Let $\{f_n\}$ be a sequence of continuous functions which converges uniformly to a function f on a set E. Prove that

$$\lim_{n\to\infty} f_n(x_n) = f(x)$$

for every sequence of points $x_n \in E$ such that $x_n \to x$, and $x \in E$. Is the converse of this true?

Solution. Let $\varepsilon > 0$. Choose N_1 so large that $|f_m(x) - f(x)| < \frac{\varepsilon}{2}$ for all $m > N_1$. Then, since f is continuous at x, choose $\delta > 0$ so small that $|f(y) - f(x)| < \frac{\varepsilon}{2}$ if $|y - x| < \delta$. Finally, choose N_2 so large that $|x_n - x| < \delta$ if $n > N_2$. Then if $n > \max(N_1, N_2)$ we have

$$|f_n(x_n) - f(x)| \le |f_n(x_n) - f(x_n)| + |f(x_n) - f(x)| < \varepsilon.$$

The converse is not true in general. For example let $f_n(x)$ be given by $f_n(x) = \sin^2 \pi x$ for $n \le |x| \le n+1$ and $f_n(x) = 0$ for $|x| \le n$ or $|x| \ge n+1$. Thus $f_n(x)$ tends to zero, since $f_n(x) = 0$ if $n \ge |x|$, but $f_n(x)$ does not converge uniformly, since $f_n(n+\frac{1}{2}) = 1$ Then for any convergent sequence, say $x_n \to x$, let $N \ge \max(|x|, |x_1|, |x_2|, \ldots, |x_n|, \ldots)$. We then have $f_n(x_n) = 0$ for all $n \ge N$, and so $f_n(x_n) \to f(x)$.

This condition does guarantee uniform convergence on any *compact* set, however. For if $\{f_n(x)\}$ is *not* a uniformly Cauchy sequence, then for some $\varepsilon_0 > 0$ there is a sequence of integers $n_1 < n_2 < \cdots$ and a sequence of points $x_1, x_2, \ldots$ such that

$$|f_{n_{2k-1}}(x_k) - f_{n_{2k}}(x_k)| \ge \varepsilon_0$$

for $k = 1, 2, \ldots$. Since K is compact, some subsequence of $\{x_k\}$ converges, say $x_{k_r} \to x$ as $r \to \infty$. Now define $y_n = x$ for all $n \ne n_{2k_r}$, $n \ne n_{2k_r-1}$, and let $y_{n_{2k_r-1}} = y_{n_{2k}} = x_{k_r}$, so that so that $y_n \to x$. Then the sequence $\{z_n\} = \{f_n(y_n)\}$ is not a Cauchy sequence, since $|z_{n_{2k_r}} - z_{n_{2k_r-1}}| \ge \varepsilon_0$.

Exercise 7.10 Let $(x) = x - n$, where n is the unique integer such that $n \le x < n+1$. Prove that

$$f(x) = \sum_{n=1}^{\infty} \frac{(nx)}{n^2}$$

is discontinuous at a dense set of points.

Solution. We shall prove that $f(x)$ is discontinuous at every rational number. Since $f(x)$ has period 1, it suffices to prove this for $0 \le x < 1$. To that end, let $x = \dfrac{p}{q}$ where p and q are relatively prime integers, $0 \le p < q$. We "stratify" the sum that defines $f(x)$ by grouping all the indices n that are congruent modulo q, i.e., we let $n = kq + r$, where $1 \le r \le q$:

$$f(x) = \sum_{k=0}^{\infty} \sum_{r=1}^{q} \frac{((kq+r)x)}{(kq+r)^2}.$$

Reversing the order of summation, we find

$$f(x) = f_1(x) + f_2(x) + \cdots + f_{q-1}(x) + f_q(x),$$

where

$$f_r(x) = \sum_{k=0}^{\infty} \frac{((kq+r)x)}{(kq+r)^2}.$$

Now it is easy to see that $f_1(x), \ldots, f_{q-1}(x)$ are continuous at $x = \frac{p}{q}$. For if $1 \leq r < q$, then $((kq+r)x)$ is continuous at that point, since (x) is continuous at the point $x = (kq+r)\frac{p}{q} = kp + \frac{rp}{q}$. (This point is not an integer, since p and q are relatively prime.) Since the series defining $f_r(x)$ converges uniformly, its limit is continuous at each point where all of the terms are continuous. In particular $f_r(x)$ is continuous at $x = \frac{p}{q}$, for $1 \leq r < q$.

We shall now show that $f_q(x)$ is discontinuous at $x = \frac{p}{q}$. It will then follow that $f(x)$ is discontinuous at that point. Observe that

$$f_q(x) = \frac{1}{q^2} \sum_{k=0}^{\infty} \frac{((k+1)qx)}{(k+1)^2} = \frac{1}{q^2} \sum_{k=1}^{\infty} \frac{(kqx)}{k^2},$$

so that

$$f_q\left(\frac{p}{q}\right) = \frac{1}{q^2} \sum_{k=1}^{1} \frac{(kp)}{k^2} = 0.$$

We shall prove that $\lim_{x\omega \frac{p}{q}+} f_q(x) > 0$, and this will show that $f_q(x)$ is discontinuous at $x = \frac{p}{q}$. Since all the terms of the series for f_q are nonnegative, it suffices to show that the limit of the first term is positive. To that end, let $\delta = \frac{1}{2q}$. If $\frac{p}{q} - \delta < x < \frac{p}{q}$, then $p - \frac{1}{2} < qx < p$, and hence $(qx) > \frac{1}{2}$, from which it follows that $f_q(x) \geq \frac{1}{2q^2}$. Therefore the lower left-hand limit of $f_q(x)$ at $x = \frac{p}{q}$ is at least $\frac{1}{2q^2}$.

Since, by the M-test with $M_n = \frac{1}{n^2}$, this series converges uniformly and each of its terms is Riemann-integrable, it follows from Theorem 7.16 that the sum of the series is Riemann-integrable.

Exercise 7.11 Suppose $\{f_n\}$, $\{g_n\}$ are defined on E and

(*a*) $\sum f_n$ has uniformly bounded partial sums;

(*b*) $g_n \to 0$ uniformly on E;

(c) $g_1(x) \geq g_2(x) \geq g_3(x) \geq \cdots$ for every $x \in E$.

Prove that $\sum f_n g_n$ converges uniformly on E. *Hint:* Compare with Theorem 3.42.

Solution. Following the hint, we let $S_N(x) = \sum\limits_{n=1}^{N} f_n(x)g_n(x)$ and $F_N(x) = \sum_{n=1}^{N} f_n(x)$ ($F_0(x) = 0$), so that $|F_N(x)| \leq B$ for all x. Then if $N > M$, we have

$$\begin{aligned}
|S_N(x) - S_M(x)| &= \Big| \sum_{n=M+1}^{N} [F_n(x) - F_{n-1}(x)]g_n(x)\Big| \\
&= \Big|F_N(x)g_N(x) - F_M(x)g_{M+1}(x) + \\
&\qquad + \sum_{n=M+1}^{N-1} F_n(x)[g_n(x) - g_{n+1}(x)\Big| \\
&\leq B\Big\{|g_N(x)| + |g_{M+1}(x)| + \sum_{n=M+1}^{N-1} [g_n(x) - g_{n+1}(x)]\Big\} \\
&= B\big[|g_N(x)| + |g_{M+1}(x)| + g_{M+1}(x) - g_N(x)\big],
\end{aligned}$$

and this last expression can be made uniformly small by choosing M sufficiently large by hypothesis (b). Hypothesis (c) was used in moving the summation sign outside the absolute value.

Exercise 7.12 Suppose g and f_n ($n = 1, 2, 3, \ldots$) are defined on $(0, \infty)$, are Riemann-integrable on $[t, T]$ whenever $0 < t < T < \infty$, $|f_n| \leq g$, $f_n \to f$ uniformly on every compact subset of $(0, \infty)$, and

$$\int_0^\infty g(x)\,dx < \infty.$$

Prove that

$$\lim_{n\to\infty} \int_0^\infty f_n(x)\,dx = \int_0^\infty f(x)\,dx.$$

(See Exercises 7 and 8 of Chap. 6 for the relevant definitions.)

This is a rather weak form of Lebesgue's dominated convergence theorem (Theorem 11.32). Even in the context of the Riemann integral, uniform convergence can be replaced by pointwise convergence if it is assumed that $f \in \mathcal{R}$. (See the articles by F. Cunningham in *Math. Mag.*, vol. 40, 1967, pp. 179–186, and by H. Kestelman in *Amer. Math. Monthly*, vol. 77, 1970, pp. 182–187.)

Solution. We shall prove that $\displaystyle\int_0^\infty f_n(x)\,dx$ converges for each n, that the limit $\displaystyle\lim_{n\to\infty} \int_0^\infty f_n(x)\,dx$ exists, that $\displaystyle\int_0^\infty f(x)\,dx$ converges and that these last two quantities are equal.

Since we obviously have $|f(x)| \le g(x)$ also, it follows that for any interval $[r, s] \subset (0, \infty)$ we have

$$\left|\int_r^s f_n(x)\,dx\right| \le \int_r^s g(x)\,dx,$$
$$\left|\int_r^s f(x)\,dx\right| \le \int_r^s g(x)\,dx,$$
$$\left|\int_r^s f_n(x) - f(x)\,dx\right| \le 2\int_r^s g(x)\,dx.$$

Now let $\varepsilon > 0$. Choose a and b with $0 < a < b < \infty$ so that if $0 < c < a < b < d < \infty$, then

$$\left|\int_c^d g(x)\,dx - \int_0^\infty g(x)\,dx\right| < \frac{\varepsilon}{2}.$$

It follows in particular that if $d > e > b$ we have

$$\begin{aligned}\int_e^d g(x)\,dx &= \int_{\frac{a}{2}}^d g(x)\,dx - \int_{\frac{a}{2}}^e g(x)\,dx\\ &\le \left|\int_0^\infty g(x)\,dx - \int_{\frac{a}{2}}^d g(x)\,dx)\right|\\ &\quad +\left|\int_0^\infty g(x)\,dx - \int_{\frac{a}{2}}^e g(x)\,dx\right|\\ &< \varepsilon.\end{aligned}$$

Then for any $d > e > b > r$ and any n we certainly have

$$\left|\int_r^d f_n(x)\,dx - \int_r^e f_n(x)\,dx\right| = \left|\int_d^e f_n(x)\,dx\right| \le \int_d^e g(x)\,dx < \varepsilon.$$

Thus by the Cauchy criterion $\lim\limits_{d\to\infty}\int_r^d f_n(x)\,dx$ exists. A similar argument shows that all the improper integrals in question converge. Moreover the argument shows that

$$\left|\int_c^d \varphi(x)\,dx - \int_0^\infty \varphi(x)\,dx\right| < \varepsilon$$

when $0 < c < a < b < d$, whether $\varphi(x) = f_n(x)$, $\varphi(x) = f(x)$, or $\varphi(x) = g(x)$.

We now merely observe that

$$\begin{aligned}\left|\int_0^\infty f_n(x)\,dx - \int_0^\infty f(x)\,dx\right| &\le \left|\int_0^\infty f_n(x)\,dx - \int_c^d f_n(x)\,dx\right| +\\ &\quad +\left|\int_c^d [f_n(x) - f(x)]\,dx\right| +\\ &\quad +\left|\int_c^d f(x)\,dx - \int_0^\infty f(x)\,dx\right|.\end{aligned}$$

Given $\varepsilon > 0$ we can choose c and d so that the first and last terms on the right are less than $\frac{\varepsilon}{3}$ (for all n, in the case of the first term). Then, since $f_n(x) \to f(x)$ uniformly on c, d, we can choose n_0 so large that the middle term is less than $\frac{\varepsilon}{3}$ if $n > n_0$.

Exercise 7.13 Assume that $\{f_n\}$ is a sequence of monotonically increasing functions on R^1 with $0 \le f_n(x) \le 1$ for all x and all n.

(a) Prove that there is a function f and a sequence $\{n_k\}$ such that

$$f(x) = \lim_{k\to\infty} f_{n_k}(x)$$

for every $x \in R^1$. (The existence of such a pointwise convergent subsequence is usually called *Helly's selection theorem.*)

(b) If, moreover f is continuous, prove that $f_{n_k} \to f$ uniformly on R^1.

Hint: (i) Some subsequence $\{f_{n_i}\}$ converges at all rational points r, say to $f(r)$. (ii) Define $f(x)$ for any $x \in R^1$ to be $\sup f(r)$, the sup being taken over all $r \le x$. (iii) Show that $f_{n_i}(x) \to f(x)$ at every x at which f is continuous. (This where monotonicity is strongly used.) (iv) A subsequence of $\{f_{n_i}\}$ converges at every point of discontinuity of f since there are at most countably many such points. This proves (a). To prove (b), modify your proof of (iii) appropriately.

Solution. (a) Following the hint, we enumerate the rational numbers (or any countable dense set) as $\{r_n\}$ and use the well-known diagonal procedure to get first a subsequence that converges at r_1, then a further subsequence that converges at r_2, etc. The sequence formed by taking the nth term of the nth subsequence is itself a subsequence and converges at each r_n. (Note that we have not used the fact that $0 \le f_n(x) \le 1$ for all x and n, only the much weaker fact that for each x there is an $M(x)$ such that $|f_n(x)| \le M(x)$ for all x and n.) Let the function $f(x)$ be defined as $f(r_k) = \lim_{n_i} f_{n_i}(r_k)$ and $f(x) = \sup\{f(r_k) : r_k \le x\}$ for all other x. The second definition could be taken as the general one if we wished, since it is consistent with the definition already given at the points $x = r_k$.

Since each of the functions is nondecreasing, it is clear that the function $f(x)$ is nondecreasing. By its definition it is continuous from the left. Suppose $f(x)$ is continuous at x_0. Let $\varepsilon > 0$ be given. Choose rational numbers r and s with $r \le x_0 \le s$, $f(x_0) - \frac{\varepsilon}{4} \le f(r) \le f(x_0) \le f(s) \le f(x_0) + \frac{\varepsilon}{4}$. Then choose i_0 so large that $|f_{n_i}(t) - f(t)| < \frac{\varepsilon}{4}$ for all $i > i_0$, $t = r$ or $t = s$. We then have

$$f(x_0) - \frac{\varepsilon}{2} \le f(r) - \frac{\varepsilon}{4} < f_{n_i}(r) \le f_{n_i}(x_0) \le f_{n_i}(s) < f(s) + \frac{\varepsilon}{4} \le f(x_0) + \frac{\varepsilon}{2}.$$

Hence $|f(x_0) - f_{n_i}(x_0)| < \varepsilon$ if $i > i_0$, which proves the convergence at points of continuity. One more application of the diagonal procedure now allows us to assure that some subsequence converges at every point (since the set of of discontinuities of a nondecreasing function is countable). We can then modify the definition of $f(x)$ at these points.

The claim that the convergence is uniform if the limit is continuous is not true. Let

$$f_n(x) = \begin{cases} \frac{1}{3} + \frac{x}{1+3|x|}, & \text{if } x \le n, \\ 1, & \text{if } x > n. \end{cases}$$

It is clear that $\lim_{n\to\infty} f_n(x) = \dfrac{1}{3} + \dfrac{x}{1+3|x|}$ for each x, yet $f_n(y) - f(y) \ge \frac{1}{3}$ if $y > n$, so that the convergence is not uniform. Here the functions $f_n(x)$ are not continuous, but they could easily be made so without violating the conditions of the problem.

To get uniform convergence we must assume in addition that $f(x) \to 1$ as $x \to \infty$ and $f(x) \to 0$ as $x \to -\infty$. Let us grant these relations and assume that $f(x)$ is continuous at all points x. To simplify the notation we shall write f_k instead of f_{n_k}. Given $\varepsilon > 0$, choose an interval $[a, b]$ such that $f(x) < \dfrac{\varepsilon}{2}$ if $x \le a$ and $f(x) > 1 - \dfrac{\varepsilon}{2}$ if $x > b$. Then, since $f(x)$ is uniformly continuous on $[a, b]$, let $a = t_0 < t_1 < \cdots < t_n = b$ be such that $f(t_i) - f(t_{i-1}) < \frac{\varepsilon}{5}$. Choose k so large that $|f_l(t_i) - f(t_i)| < \frac{\varepsilon}{5}$ for all $i = 1, \ldots, n$ and all $l > k$.Then for all $y \ge b = t_n$ we have

$$1 \ge f_l(y) \ge f_l(t_k) > 1 - \frac{4\varepsilon}{5}$$

and

$$1 \ge f(y) > 1 - \frac{\varepsilon}{2} > 1 - \frac{4\varepsilon}{5}.$$

Hence certainly

$$|f_l(y) - f(y)| \le 1 - \left(1 - \frac{4\varepsilon}{5}\right) < \varepsilon$$

for all $l > k$ and all $y \ge b$.

A similar argument shows that f_l converges uniformly to f on $(-\infty, a]$. The argument that f_l converges uniformly to f on $[t_{i-1}, t_i]$ is identical to that given above.

Exercise 7.14 Let f be a continuous real function on R^1 with the following properties: $0 \le f(t) \le 1$, $f(t+2) = f(t)$ for every t, and

$$f(t) = \begin{cases} 0 & (0 \le t \le \frac{1}{3}) \\ 1 & (\frac{2}{3} \le t \le 1). \end{cases}$$

Put $\Phi(t) = (x(t), y(t))$, where

$$x(t) = \sum_{n=1}^{\infty} 2^{-n} f(3^{2n-1}t), \quad y(t) = \sum_{n=1}^{\infty} 2^{-n} f(3^{2n}t).$$

Prove that Φ is *continuous* and that Φ maps $I = [0,1]$ *onto* the unit square $I^2 \subset R^2$. In fact, show that Φ maps the Cantor set onto I^2.

Hint: Each $(x_0, y_0) \in I^2$ has the form

$$x_0 = \sum_{n=1}^{\infty} 2^{-n} a_{2n-1}, \quad y_0 = \sum_{n=1}^{\infty} 2^{-n} a_{2n},$$

where each a_i is 0 or 1. If

$$t_0 = \sum_{i=1}^{\infty} 3^{-i-1}(2a_i),$$

show that $f(3^k t_0) = a_k$, and hence that $x(t_0) = x_0$, $y(t_0) = y_0$.

(This simple example of a so-called "space-filling curve" is due to I. J. Schoenberg, *Bull. A.M.S.*, vol. 44, 1938, p. 519.)

Solution. We note that $3^k t_0$ is the sum of the even integer $2(3^{k-2}a_1 + \cdots + 3a_{k-2} + a_{k-1})$ and a fractional part $\sum_{i=k}^{\infty} \frac{2a_i}{3^{i-k+1}}$. This fractional part lies in $\left[\frac{2}{3}, 1\right]$ if $a_k = 1$, while if $a_k = 0$ it is at least 0 and at most $\frac{2}{9}$. Thus it lies in the interval $\left[0, \frac{1}{3}\right]$ if $a_k = 0$. In either case $f(3^k t_0) = a_k$, as claimed. We therefore have

$$x(t_0) = \sum_{n=1}^{\infty} 2^{-n} a_{2n-1} = x_0, \quad y(t_0) = \sum_{n=1}^{\infty} 2^{-n} a_{2n} = y_0,$$

as asserted.

Exercise 7.15 Suppose f is a real continuous function on R^1, $f_n(t) = f(nt)$ for $n = 1, 2, 3, \ldots$, and $\{f_n\}$ is equicontinuous on $[0,1]$. What conclusion can you draw about f?

Solution. The function $f(t)$ must be constant on $[0, \infty)$. For if $f(x) \neq f(y)$ and $0 \le x < y < \infty$, say $|f(x) - f(y)| = \varepsilon > 0$, it follows that $|f_n\left(\frac{x}{n}\right) - f_n\left(\frac{y}{n}\right)| = \varepsilon$ for all n. Since $\frac{x-y}{n} \to 0$, it follows that the family $\{f_n\}$ cannot be equicontinuous on $[0,1]$, or, indeed, on any neighborhood of 0.

Exercise 7.16 Suppose $\{f_n\}$ is an equicontinuous sequence of functions on a compact set K, and $\{f_n\}$ converges pointwise on K. Prove that $\{f_n\}$ converges uniformly on K.

Solution. Let $\varepsilon > 0$. Choose $\delta > 0$ such that $|f_n(x) - f_n(y)| < \frac{\varepsilon}{3}$ for all $n \neq m$ if $x, y \in K$ and $|x - y| < \delta$. Choose a finite number of points $x_1, \ldots, x_N$ such that for every $x \in K$ there exists j with $|x - x_j| < \delta$. (Such a finite set exists; otherwise we could inductively select a sequence $\{x_n\}$ such that $|x_m - x_n| \ge \delta$

for all n, and this sequence would have no Cauchy subsequence, contradicting the compactness of K.) Then choose n_0 so large that $|f_m(x_j) - f_n(x_j)| < \frac{\varepsilon}{3}$ for all $m, n > n_0$ and all $j = 1, 2, \ldots, N$. Then for any point $x \in K$, fix j so that $|x - x_j| < \delta$. If $m, n > n_0$ we have

$$|f_m(x) - f_n(x)| \leq |f_m(x) - f_m(x_j)| + |f_m(x_j) - f_n(x_j)| + |f_n(x_j) - f_n(x)|.$$

The first and last terms are smaller than $\frac{\varepsilon}{3}$ because $|x - x_j| < \delta$; the middle term is smaller than $\frac{\varepsilon}{3}$ since $m, n > n_0$. Thus the sequence is a uniformly Cauchy sequence.

Exercise 7.17 Define the notions of uniform convergence and equicontinuity for mappings into any metric space. Show that Theorems 7.9 and 7.12 are valid for mappings into any metric space, that Theorems 7.8 and 7.11 are valid for mappings into any complete metric space, and that Theorems 7.10, 7.16, 7.17, 7.24, and 7.25 hold for vector-valued functions, that is, for mappings into any R^n.

Solution. Let X and Y be any metric spaces. The sequence $\{f_n\}$, where $f_n : X \to Y$, *converges uniformly* to $f : X \to Y$ if for every $\varepsilon > 0$ there exists N such that $d_Y(f_n(x), f(x)) < \varepsilon$ for all $x \in X$ and all $n > N$. A family of functions $\mathcal{F}$ is *equicontinuous* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $d_Y(f(x_1), f(x_2)) < \varepsilon$ for all $f \in \mathcal{F}$ whenever $d_X(x_1, x_2) < \delta$.

An immediate consequence of this definition is that $\{f_n\}$ converges uniformly to f if and only if $M_n \to 0$, where $M_n = \sup\limits_{x \in X} d_Y(f_n(x), f(x))$ (Theorem 7.9).

The same $\frac{\varepsilon}{3}$ argument that proves Theorem 7.12 shows that the uniform limit of a sequence of continuous functions is continuous.

The Cauchy convergence criterion accepts the additional word *uniformly* without any change, provided Y is complete. Suppose for every $\varepsilon > 0$ there exists N such that $d_Y(f_m(x), f_n(x)) < \varepsilon$ for all $m, n > N$ and all x. Then, in particular, for each $x \in X$, the sequence $\{f_n(x)\}$ is a Cauchy sequence in Y. Since Y is complete, this sequence converges to a value that we shall call $f(x)$. We now claim that $\{f_n\}$ converges uniformly to f. Indeed, given $\varepsilon > 0$ choose N so that $d_Y(f_m(x), f_n(x)) < \frac{\varepsilon}{2}$ if $m, n > N$. Since a metric is a continuous function, it follows that $d_Y(f(x), f_n(x)) \leq \frac{\varepsilon}{2} < \varepsilon$ if $n > N$, that is $\{f_n\}$ converges uniformly to f. This is Theorem 7.8.

Suppose now $\{f_n\}$ converges uniformly to f, Y is complete, $x_0 \in X$, and $\lim\limits_{x \to x_0} f_n(x) = A_n$ for $n = 1, 2, \ldots$. Then $\{A_n\}$ converges, and $\lim\limits_{x \to x_0} f(x) = \lim\limits_{n \to \infty} A_n$. (This is Theorem 7.11.) The proof is as follows. Given $\varepsilon > 0$ choose N so that $d_Y(f(x), f_n(x)) < \frac{\varepsilon}{3}$ for all x if $n \geq N$. Let $n > N$ be fixed. Choose $\delta > 0$ (depending on n and ε in general) such that $d_Y(f_n(x), A_n) < \frac{\varepsilon}{3}$ if $0 < d_X(x, x_0) < \delta$. (The fact that $\lim_{x \to x_0} f_n(x) = A_n$ implies that there must exist x satisfying these inequalities, i.e., that x_0 is an accumulation point of X.) We then have, for $m, n > N$,

$$d_Y(A_m, A_n) \leq d_Y(A_m, f_m(x)) + d_Y(f_m(x), f_n(x)) + d_Y(f_n(x), f(x))$$

The middle term is less than $\frac{\varepsilon}{3}$ for all $m, n > N$ and all $x \in X$. If m and n are then fixed integers larger than N, the first and last terms can be made smaller than $\frac{\varepsilon}{3}$ by choosing x sufficiently close to x_0. Hence we have $d_Y(A_m, A_n) < \varepsilon$ if $m, n > N$. Since Y is complete, the sequence $\{A_n\}$ converges, say to A. Now observe that

$$d_Y(f(x), A) \leq d_Y(f(x), f_n(x)) + d_Y(f_n(x), A_n) + d_Y(A_n, A).$$

If N is chosen sufficiently large, the first and last terms on the right-hand side will be less than $\frac{\varepsilon}{3}$ (for all x, in the case of the first term). For a fixed n satisfying these conditions, if $\delta > 0$ is sufficiently small, the second term will be less than $\frac{\varepsilon}{3}$ whenever $0 < d_X(x, x_0) < \delta$, and hence $d_Y(f(x), A) < \varepsilon$ if $0 < d_X(x, x_0) < \delta$.

The proof of the stated theorems for vector-valued functions is a consequence of the obvious facts that a vector-valued function $\mathbf{f}$ is integrable, differentiable or continuous if and only if each of its components has the corresponding property, and that a series of vector-valued functions $\{\mathbf{f}_n\}$ is Cauchy, bounded, convergent, uniformly convergent, majorized by a convergent sequence, equicontinuous, etc., if and only if each component has those properties. A typical proof proceeds as follows (Theorem 7.25). Suppose $\{\mathbf{f}_n\}$ is a bounded equicontinuous sequence of vector-valued functions on a compact set K. Let $\|\mathbf{f}_n(x)\| \leq M$ for all $x \in K$ and all n, and given $\varepsilon > 0$ choose $\delta > 0$ such that $\|\mathbf{f}_n(x) - \mathbf{f}_n(y)\| < \varepsilon$ whenever $d(x, y) < \delta$. Then for each component f_n^i of $\mathbf{f}_n$ we have $|f_n^i(x)| \leq \|\mathbf{f}_n\| \leq M$ and $|f_n^i(x) - f_n^i(y)| \leq \|\mathbf{f}_n(x) - \mathbf{f}_n(y)\| < \varepsilon$ whenever $d(x, y) < \delta$. Hence each sequence of components $\{f_n^i\}_n$, $i = 1, \ldots, k$, is bounded and equicontinuous. Therefore for each i there is a subsequence $\{n_r\}$ such that $f_{n_r}^i$ converges uniformly. By refining to subsubsequences, we can obtain a single subsequence $\{n_r\}$ such that $\{f_{n_r}^i\}$ converges uniformly for all i, say to $f^i(x)$. Then, given $\varepsilon > 0$, choose r_0 so large that $|f_{n_r}^i(x) - f^i(x)| < \frac{\varepsilon}{k}$ for $i = 1, 2, \ldots, k$ and $r > r_0$. It then follows that $\|\mathbf{f}_{n_r}(x) - \mathbf{f}(x)\| < \varepsilon$ if $r > r_0$. The proofs of the other results all follow this model argument.

Exercise 7.18 Let $\{f_n\}$ be a uniformly bounded sequence of functions which are Riemann integrable on $[a, b]$, and put

$$F_n(x) = \int_a^x f_n(t)\, dt \quad (a \leq x \leq b).$$

Prove that there exists a subsequence $\{F_{n_k}\}$ which converges uniformly on $[a, b]$.

Solution. Let M be such that $|f_n(x)| \leq M$ for all n and x. Then clearly $|F_n(x)| \leq M(b - a)$ for all n, so that $\{F_n\}$ is uniformly bounded. Also, given $\varepsilon > 0$, let $\delta = \frac{\varepsilon}{M}$. Then if $x < y$ and $|x - y| < \delta$, we have

$$|F_n(y) - F_n(x)| = \left| \int_x^y f_n(t)\, dt \right| < M|x - y| < \varepsilon.$$

Hence $\{F_n\}$ is also uniformly equicontinuous. Therefore by Ascoli's Theorem (Theorem 7.25), there exists a uniformly convergent subsequence of $\{F_n\}$.

Exercise 7.19 Let K be a compact metric space, let S be a subset of $\mathcal{C}(K)$. Prove that S is compact (with respect to the metric defined in Section 7.14) if and only if S is uniformly closed, pointwise bounded, and equicontinuous. (If S is not equicontinuous, then S contains a sequence which has no equicontinuous subsequence, hence has no sequence that converges uniformly on K.)

Solution. First suppose S is compact. Then we know that S has to be closed and bounded (in this metric *bounded* means the same thing as *uniformly bounded*). If S is not equicontinuous, then there exists $\varepsilon > 0$ such that for all $\delta > 0$ there exist x, y and $g \in S$ such that $d(x,y) < \delta$ and $|g(x) - g(y)| \geq \varepsilon$. Let $x_n, y_n \in K$ and $g_n \in S$ be such that $d(x_n, y_n) < \frac{1}{n}$ and $|g_n(x_n) - g_n(y_n)| \geq \varepsilon$. Then no subsequence of $\{g_n\}$ can be equicontinuous, since $|g_{n_k}(x_{n_k} - g_{n_k}(y_{n_k}) \geq \varepsilon$. Hence by Theorem 7.24 no subsequence of $\{g_n\}$ can converge in $\mathcal{C}(K)$, and so S cannot be compact. We conclude, then, that if S is compact, then S is closed, bounded, and equicontinuous.

Conversely, if S is closed, pointwise bounded, and equicontinuous, then every sequence $\{g_n\}$ contains a subsequence that converges uniformly, hence converges in the metric of $\mathcal{C}(K)$ (by Ascoli's theorem). Since S is closed, the limit belongs to S, and so S is compact by Exercise 26 of Chapter 2.

Exercise 7.20 If f is continuous on $[0,1]$ and if

$$\int_0^1 f(x)x^n \, dx = 0 \quad (n = 0, 1, 2 \ldots),$$

prove that $f(x) = 0$ on $[0,1]$. *Hint:* The integral of the product of f with any polynomial is zero. Use the Weierstrass theorem to show that $\int_0^1 f^2(x)\, dx = 0$.

Solution. There exists a sequence of polynomials $p_n(x)$ such that $p_n(x)$ converges uniformly to $f(x)$. Since f is bounded, $\{p_n\}$ is uniformly bounded, and hence $p_n f$ converges uniformly to f^2. Then by Theorem 7.16

$$\int_0^1 f^2(x)\, dx = \lim_{n\to\infty} \int_0^1 p_n(x) f(x)\, dx = 0.$$

But we know already (Exercise 2 of Chapter 6) that this implies $f^2(x) \equiv 0$.

Exercise 7.21 Let K be the unit circle in the complex plane (i.e., the set of all z with $|z| = 1$), and let $\mathcal{A}$ be the algebra of all functions of the form

$$f(e^{i\theta}) = \sum_{n=0}^{N} c_n e^{in\theta} \quad (\theta \text{ real}).$$

The $\mathcal{A}$ separates points on K, and $\mathcal{A}$ vanishes at no point of K, but nevertheless there are continuous functions on K which are not in the uniform closure of $\mathcal{A}$. *Hint:* For every $f \in \mathcal{A}$

$$\int_0^{2\pi} f(e^{i\theta})e^{i\theta}\,d\theta = 0,$$

and this is also true for every f in the closure of $\mathcal{A}$.

Solution. The function $f(z) = z \in \mathcal{A}$ separates points on K and never vanishes. The equality given in the hint is a straighforward computation. It implies that the continuous function $\frac{1}{z}$, which is $e^{-i\theta}$, is not in the uniform closure of $\mathcal{A}$, since

$$\int_0^{2\pi} e^{-i\theta}e^{i\theta}\,d\theta = 2\pi.$$

Exercise 7.22 Assume $f \in \mathcal{R}(\alpha)$ on $[a, b]$, and prove that there are polynomials P_n such that

$$\lim_{n\to\infty} \int_a^b |f - P_n|^2\,d\alpha = 0.$$

(Compare with Exercise 12, Chap. 6.)

Solution. The parenthetical remark refers to the proof that there is a sequence of *continuous* functions $\{f_n\}$ such that

$$\lim_{n\to\infty} \int_a^b |f - f_n|^2\,d\alpha = 0.$$

All that is now needed is to note that one can find polynomials P_n such that $|f_n(x) - P_n(x)| < \frac{1}{n}$ for all $x \in [a, b]$ and all n.

Exercise 7.23 Put $P_n = 0$, and define, for $n = 0, 1, 2, \ldots$,

$$P_{n+1}(x) = P_n(x) + \frac{x^2 - P_n^2(x)}{2}$$

Prove that

$$\lim_{n\to\infty} P_n(x) = |x|,$$

uniformly on $[-1, 1]$.

(This makes it possible to prove the Stone-Weierstrass theorem without first proving Theorem 7.26.

Hint: Use the identity

$$|x| - P_{n+1}(x) = [|x| - P_n(x)]\left[1 - \frac{|x| + P_n(x)}{2}\right]$$

to prove that $0 \leq P_n(x) \leq P_{n+1}(x) \leq |x|$ if $|x| \leq 1$, and that

$$|x| - P_n(x) \leq |x|\left(1 - \frac{|x|}{2}\right)^n < \frac{2}{n+1}$$

if $|x| \leq 1$.

Solution. The identity given in the hint is a trivial consequence of the identity $x^2 - P_n^2(x) = [|x| - P_n(x)][|x| + P_n(x)]$. Then, granting that $0 \leq P_n(x) \leq |x|$, we conclude that $0 \leq 1 - \dfrac{|x| + P_n(x)}{2} < 1$ for $|x| \leq 1$, and hence that $0 \leq |x| - P_{n+1}(x) \leq |x| - P_n(x)$, which gives all of the desired inequalities. An immediate corollary of the same identity (obtained by replacing $P_n(x)$ by 0 in the second factor on the right-hand side) is

$$|x| - P_{n+1}(x) \leq [|x| - P_n(x)]\left(1 - \frac{|x|}{2}\right),$$

and this inequality makes it possible to obtain the inequality

$$|x| - P_n(x) \leq |x|\left(1 - \frac{|x|}{2}\right)^n$$

by induction on n. Finally, by symmetry, the maximum of $|x|(1 - \frac{|x|}{2})^n$ on $[-1, 1]$ is its maximum on $[0, 1]$, and this can be found by simple calculus to occur at $x = \frac{2}{n+1}$. Since this function is always less than $|x|$, the final inequality now follows.

Exercise 7.24 Let X be a metric space, with metric d. Fix a point $a \in X$. Assign to each $p \in X$ the function f_p defined by

$$f_p(x) = d(x, p) - d(x, a) \quad (x \in X).$$

Prove that $|f_p(x)| \leq d(a, p)$ for all $x \in X$, and therefore $f_p \in \mathcal{C}(X)$.
Prove that

$$\|f_p - f_q\| = d(p, q)$$

for all $p, q \in X$.

If $\Phi(p) = f_p$, it follows that Φ is an *isometry* (a distance-preserving mapping) of X onto $\Phi(X) \subset \mathcal{C}(X)$.

Let Y be the closure of $\Phi(X)$ in $\mathcal{C}(X)$. Show that Y is complete.

Conclusion: X is isometric to a dense subset of a complete metric space Y. (Exercise 24, Chap. 3 contains a different proof of this.)

Solution. The inequality $|f_p(x)| \leq d(a, p)$ is well-known, i.e., the fact that

$$|d(x, p) - d(x, a)| \leq d(a, p)$$

and follows from the triangle inequality by merely transposing a term. (The left-hand side is either $d(x,p) - d(x,a)$ or $d(x,a) - d(x,p)$. Whichever is the case, if the subtracted term is moved to the other side, we have the ordinary triangle inequality.)

As for the isometry, we certainly have, for all x,

$$|f_q(x) - f_p(x)| = |d(x,q) - d(x,p)| \leq d(p,q)$$

and equality holds here if $x = q$ or $x = p$. Hence the supremum over all x is exactly $d(p,q)$.

As for the closure Y of $\Phi(X)$ being complete, it is a closed subset of a complete metric space, hence necessarily complete. By definition of closure, $\Phi(X)$ is dense in Y.

Exercise 7.25 Suppose ϕ is a continuous bounded real function in the strip defined by $0 \leq x \leq 1$, $-\infty < y < \infty$. Prove that the initial-value problem

$$y' = \phi(x,y), \quad y(0) = c$$

has a solution. (Note that the hypotheses of this existence theorem are less strigent than those of the corresponding uniqueness theorem; see Exercise 27, Chap. 5.)

Hint: Fix n. For $i = 0, \ldots, n$; put $x_i = i/n$. Let f_n be a continuous function on $[0,1]$ such that $f_n(0) = c$,

$$f_n'(t) = \phi(x_i, f_n(x_i)) \quad \text{if} \quad x_i < t < x_{i+1},$$

and put

$$\Delta_n(t) = f_n'(t) - \phi(t, f_n(t)),$$

except at the points x_i, where $\Delta_n(t) = 0$. Then

$$f_n(x) = c + \int_0^x [\phi(t, f_n(t)) + \Delta_n(t)]\, dt.$$

Choose $M < \infty$ so that $|\phi| \leq M$. Verify the following assertions.

(*a*) $|f_n'| \leq M$, $|\Delta_n| \leq 2M$, $\Delta_n \in \mathcal{R}$, and $|f_n| \leq |c| + M = M_1$, say, on $[0,1]$ for all n.

(*b*) $\{f_n\}$ is equicontinuous on $[0,1]$, since $|f_n'| \leq M$.

(*c*) Some $\{f_{n_k}\}$ converges to some f, uniformly on $[0,1]$.

(*d*) Since ϕ is uniformly continuous on the rectangle $0 \leq x \leq 1$, $|y| \leq M_1$,

$$\phi(t, f_{n_k}(t)) \to \phi(t, f(t))$$

uniformly on $[0,1]$.

(*e*) $\Delta_n(t) \to 0$ uniformly on $[0,1]$, since

$$\Delta_n(t) = \phi(x_i, f_n(x_i)) - \phi(t, f_n(t))$$

in (x_i, x_{i+1}).
(*f*) Hence

$$f(x) = c + \int_0^x \phi(t, f(t))\, dt.$$

This f is a solution of the given problem.

Solution. It will save trouble if we assume that ϕ is a bounded continuous mapping from $[0,1] \times R^k$ into R^k and that c is a vector in R^k. That way we can do Exercise 26 simultaneously with this one. Since we are defining the functions $f_n(t)$ to be piecewise-linear, there is no difficulty in doing this with vector-valued functions. We simply define $f_n(t) = c + t\phi(0,c)$ for $0 \le t \le x_1$, and then, by induction on i,

$$f_n(t) = f_n(x_i) + (t - x_i)\phi(x_i, f_n(x_i))$$

for $x_i < t \le x_{i+1}$.

Then, if $\Delta_n(t)$ is defined as indicated, we have $f_n'(t) = \Delta_n(t) + \phi(t, f_n(t))$ except at a finite set of points, and therefore

$$f_n(x) = f_n(0) + \int_0^x [\phi(t, f_n(t)) + \Delta_n(t)]\, dt.$$

(*a*) The assertions $|f_n'| \le M$ and $|\Delta_n| \le 2M$ are immediate consequences of the definitions of these two functions and the fact that $|\phi(x,y)| \le M$ for all x and y (here in general $y \in R^k$). Since $\Delta_n(t)$ is bounded and continuous except at x_i, it is Riemann-integrable. The inequality $|f_n| \le |c| + M = M_1$ is then immediate.

(*b*) $|f_n(x) - f_n(y)| \le \int_x^y |f_n'(t)|\, dt \le M|x-y|$.

(*c*) This is Ascoli's Theorem (Theorem 7.25).

(*d*) Given $\varepsilon > 0$ let $\delta > 0$ be such that $|\phi(t,y) - \phi(t,z)| < \varepsilon$ if $|y - z| < \delta$, for all $t \in [0,1]$, and $y, z \in R^k$. Then if $|f_{n_k}(t) - f(t)| < \delta$ for all t (which is the case if k is large), we have $|\phi(t, f_{n_k}(t)) - \phi(t, f(t))| < \varepsilon$ for all t.

(*e*) For each t and n let $i(n)$ be chosen so that $t \in [x_{i(n)}, x_{i(n)+1}]$, so that $|t - x_{i(n)}| \le \frac{1}{n}$. Since $f_{n_k}(t)$ converges uniformly to $f(t)$ and $x_{i(n)} \to t$, it follows that $\phi(x_{i(n)}, f_n(x_{i(n)})) - \phi(t, f_n(t)) \to 0$.

(*f*) We now invoke Theorem 7.16 to get

$$f(x) = c + \int_0^x \phi(t, f(t))\, dt.$$

Clearly $f(0) = c$, and since the right-hand side has a continuous derivative, so does the left-hand side, and $f'(x) = \phi(x, f(x))$.

Exercise 7.26 Prove an analogous existence theorem for the initial-value problem

$$\mathbf{y}' = \Phi(x, \mathbf{y}), \quad \mathbf{y}(0) = \mathbf{c},$$

where now $\mathbf{c} \in R^k$, $\mathbf{y} \in R^k$, and Φ is a continuous bounded mapping of the part of R^{k+1} defined by $0 \le x \le 1$, $\mathbf{y} \in R^k$ into R^k. (Compare Exercise 28, Chap. 5.) *Hint:* Use the vector-valued version of Theorem 7.25.

Solution. Since we were foresightful enough to make all the necessary notes in the solution of the previous problem, there is nothing to be done. Observe that an k-th order initial-value problem

$$y^{(k)} = \phi(x, y, y', y'', \dots, y^{(k-1)})$$

with $y(0) = c_0$, $y'(0) = c_1, \dots, y^{(k-1)}(0) = c_{k-1}$ falls under this theorem if we let

$$\Phi(x, y_1, y_2, \dots, y_k) = (y_2, y_3, \dots, y_k, \phi(x, y_1, \dots, y_{k-1})),$$

$\mathbf{y}(0) = (c_0, \dots, c_{k-1})$ Any solution of this problem provides a solution of the k-th order equation (namely y_1 if $\mathbf{y} = (y_1, \dots, y_k)$).

Chapter 8

Some Special Functions

Exercise 8.1 Define

$$f(x) = \begin{cases} e^{-1/x^2} & (x \neq 0), \\ 0 & (x = 0). \end{cases}$$

Prove that f has derivatives of all orders at $x = 0$ and that $f^{(n)}(0) = 0$ for $n = 1, 2, 3, \ldots$.

Solution. We have $\lim\limits_{x \to 0} x^k e^{-1/x^2} = 0$ for all $k = 0, \pm 1, \pm 2, \ldots$ by L'Hospital's rule. It is easily shown by induction that there is a polynomial p_n such that $f^{(n)}(x) = p_n\left(\frac{1}{x}\right) e^{-1/x^2}$ for $x \neq 0$. Assuming (by induction) that $f^{(n)}(0) = 0$, we then have $f^{(n+1)}(0) = \lim\limits_{x \to 0} q_n\left(\dfrac{1}{x}\right) e^{-1/x^2} = 0$, where $q_n(x) = x p_n(x)$.

Exercise 8.2 Let a_{ij} be the number in the ith row and jth column of the array

$$\begin{array}{ccccc} -1 & 0 & 0 & 0 & \ldots \\ \frac{1}{2} & -1 & 0 & 0 & \ldots \\ \frac{1}{4} & \frac{1}{2} & -1 & 0 & \ldots \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{2} & -1 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{array}$$

so that

$$a_{ij} = \begin{cases} 0 & (i < j), \\ -1 & (i = j), \\ 2^{j-i} & (i > j). \end{cases}$$

Prove that

$$\sum_i \sum_j a_{ij} = -2, \quad \sum_j \sum_i = 0.$$

Solution. This is a routine computation:

$$\begin{aligned}
\sum_i \sum_j a_{ij} &= \sum_{i=1}^{\infty} \left[-1 + \sum_{j=1}^{i-1} 2^{j-i} \right] \\
&= \sum_{i=1}^{\infty} [-1 + (1 - 2^{1-i})] \\
&= \sum_{i=1}^{\infty} -2^{1-i} = -2,
\end{aligned}$$

while

$$\begin{aligned}
\sum_j \sum_i a_{ij} &= \sum_{j=1}^{\infty} \left[-1 + \sum_{i=j+1}^{\infty} 2^{j-i} \right] \\
&= \sum_{j=1}^{\infty} [-1+1] \\
&= 0.
\end{aligned}$$

Exercise 8.3 Prove that

$$\sum_i \sum_j a_{ij} = \sum_j \sum_i a_{ij}$$

if $a_{ij} \geq 0$ for all i and j (the case $+\infty = +\infty$ may occur).

Solution. In fact the only case that we need to consider is the case when one of the two sums is infinite. If either sum is finite, we merely invoke Theorem 8.3, which explicitly states that the two sums are equal. Hence if either sum is infinite, then both are.

Exercise 8.4 Prove the following limit relations:

(a) $\displaystyle\lim_{x \to 0} \frac{b^x - 1}{x} = \log b \quad (b > 0).$

(b) $\displaystyle\lim_{x \to 0} \frac{\log(1+x)}{x} = 1.$

(c) $\displaystyle\lim_{x \to 0} (1+x)^{\frac{1}{x}} = e.$

(d) $\displaystyle\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x.$

Solution. (a) Consider the function $f(x) = b^x = e^{x \log b}$. The limit we are considering is $f'(0)$. By the chain rule

$$f'(x) = e^{x \log b} \log b.$$

Now take $x = 0$.

(*b*) Let $y = \log(1+x)$, so that $x = e^y - 1$. It is easy to justify the relation

$$\lim_{x\to 0} \frac{\log(1+x)}{x} = \lim_{y\to 0} \frac{y}{e^y - 1} = \frac{1}{\lim_{y\to 0} \frac{e^y - 1}{y}} = 1,$$

since $\lim\limits_{y\to 0} \dfrac{e^y - 1}{y} = E'(0)$.

(*c*) Consider the function $(1+x)^{1/x} = e^{\frac{\log(1+x)}{x}}$. By part (*b*) $\lim\limits_{x\to 0}(1+x)^{\frac{1}{x}} = e^1 = e$.

(*d*) As above, we have $\left(1 + \dfrac{x}{n}\right)^n = \left[\left(1 + \dfrac{x}{n}\right)^{1/(x/n)}\right]^x$, and by part (*c*) the limit of the expression inside the brackets is e.

Exercise 8.5 Find the following limits

(*a*) $\lim_{x\to 0} \frac{e-(1+x)^{1/x}}{x}$.

(*b*) $\lim_{n\to\infty} \frac{n}{\log n}[n^{1/n} - 1]$.

(*c*) $\lim_{x\to 0} \frac{\tan x - x}{x(1-\cos x)}$.

(*d*) $\lim_{x\to 0} \frac{x - \sin x}{\tan x - x}$.

Solution. (*a*) This limit is $f'(0)$, where $f(x) = (1+x)^{1/x}$ (by part (*c*) of the previous problem). Now for $x \neq 0$, we have

$$f'(x) = (1+x)^{1/x}\left[\frac{(1+x)\log(1+x) - x}{x^2(x+1)}\right].$$

Since we know that the limit of the first factor is e, we need only consider the limit inside the brackets. Since

$$(1+x)\log(1+x) = \left(x - \frac{x^2}{2} + \cdots\right) + x\left(x - \frac{x^2}{2} + \cdots\right),$$

we can cancel x^2 from the numerator and denominator of the expression in brackets, and we see that the limit of this expression is $\frac{1}{2}$. Hence the limit of $f'(x)$ as $x \to 0$ exists and equals $\frac{e}{2}$. It then follows from the mean-value theorem that $f'(0)$ equals this limit (see the corollary to Theorem 5.12).

(*b*) Write this expression as

$$\frac{e^{\frac{\log n}{n}} - 1}{\frac{\log n}{n}}.$$

Since $\frac{\log n}{n}$ tends to 0 as $n \to \infty$, this fraction tends to the derivative of e^x at 0, i.e., it tends to 1.

(*c*) Write this expression as

$$\frac{\sin x - x\cos x}{x\cos x(1-\cos x)}.$$

We can then use either Maclaurin series or L'Hospital's rule to prove that the limit is $\frac{2}{3}$.

(*d*) Write this expression as

$$\frac{(x-\sin x)\cos x}{\sin x - x\cos x}$$

and again either by Maclaurin series or L'Hospital's rule the limit is $\frac{1}{2}$.

Exercise 8.6 Suppose $f(x)f(y) = f(x+y)$ for all real x and y.

(*a*) Assuming that f is differentiable and not zero, prove that

$$f(x) = e^{cx}$$

where c is a constant.

(*b*) Prove the same thing, assuming only that f is continuous.

Solution. (*a*) Since f is not 0, it follows that $f(0) = 1$ (take $x = y = 0$ in the basic relation that defines f). It then follows that $f'(x) = f(x)f'(0)$, and hence that the function $g(x) = e^{-xf'(0)}f(x)$ satisfies $g'(0) = 0$ for all x. Therefore $g(x) = g(0) = f(0) = 1$ for all x, i.e., $f(x) = e^{cx}$, where $c = f'(0)$.

(*b*) The relation $f(x)f(y) = f(x+y)$ shows that either $f(x)$ is always zero, or it is never zero. In the latter case, since f is continuous, it cannot change sign, and therefore (since $f(0) = 1$) it is always positive. Let $g(x) = \log[f(x)]$. Then $g(x+y) = g(x) + g(y)$, and g is continuous. It suffices then to show that $g(x) = cx$ for some constant $c = g(1)$. To this end, we note that the additive property of g implies that $g(0) = 0$, $g(-x) = -g(x)$, and (by an easy induction) $g(nx) = ng(x)$ for all integers $n = 0, \pm 1, \pm 2, \ldots$. Consider the set of x such that $g(x) = g(1)x$. Obviously 0 and 1 belong to this set. If a belongs to this set, so does na for any n, since $g(na) = ng(a) = ng(1)a = g(1)(na)$. Finally, if a belongs to this set, so does $\frac{a}{n}$, $n = 1, 2, \ldots$, since $g(a) = g(n\frac{a}{n}) = ng(\frac{a}{n})$. That is, $g(\frac{a}{n}) = \frac{1}{n}g(a) = \frac{1}{n}g(1)a = g(1)\frac{a}{n}$. It now follows that r belongs to this set for all rational numbers r, that is, the two continuous functions $g(x)$ and $g(1)x$ have the same values at all rational numbers r. Since the rational numbers are dense, and the set of points at which two continuous functions are equal is a closed set, it follows that $g(x) = g(1)x$ for all x.

Exercise 8.7 If $0 < x < \frac{\pi}{2}$, prove that

$$\frac{2}{\pi} < \frac{\sin x}{x} < 1.$$

Solution. To show the left-hand inequality, consider the function $f(x) = \sin x - \frac{2x}{\pi}$ on the interval $0 \leq x \leq \frac{\pi}{2}$. We have $f(0) = f(\frac{\pi}{2}) = 0$. Since $f''(x) = -\sin x \leq 0$, the function $f'(x)$ is strictly decreasing on this interval. Therefore it has at most one zero on this interval; by Rolle's theorem, it has exactly one zero. Since $f''(x) < 0$ at that point, the function $f(x)$ has a maximum at that point. Therefore $f(x) > 0$ for $0 < x < \frac{\pi}{2}$.

The proof of the right-hand inequality is similar, but easier. The function $g(x) = x - \sin x$ has derivative $1 - \cos x$, which is nonnegative. Therefore $g(x)$ is strictly increasing, and so $g(x) > g(0) = 0$ for all $x > 0$ (the restriction $x < \frac{\pi}{2}$ is superfluous in this case).

Exercise 8.8 For $n = 0, 1, 2, \ldots$, and x real, prove that

$$|\sin nx| \leq n|\sin x|.$$

Note that this inequality may be false for other values of n. For instance,

$$|\sin \frac{1}{2}\pi| > \frac{1}{2}|\sin \pi|.$$

Solution. The inequality is obvious if $n = 0$ or $n = 1$. Then by induction we have

$$\begin{aligned}
|\sin nx| &= |\sin((n-1)x + x)| \\
&= |\sin((n-1)x)\cos x + \cos((n-1)x)\sin x| \\
&\leq |\sin((n-1)x)| + |\sin x| \\
&\leq (n-1)|\sin x| + |\sin x| = |n|\,|\sin x|.
\end{aligned}$$

A stronger remark can be made: If c is *not* an integer, then $|\sin c\pi| > |c|\,|\sin \pi|$. Hence this inequality fails for $x = \pi$ unless c is an integer.

Exercise 8.9 (a) Put $s_N = 1 + (\frac{1}{2}) + \cdots + (1/N)$. Prove that

$$\lim_{N \to \infty} (s_N - \log N)$$

exists. (The limit, often denoted by γ, is called Euler's constant. Its numerical value is $0.5772\ldots$. It is not known whether γ is rational or not.)

(b) Roughly how large must m be so that $N = 10^m$ satisfies $s_N > 100$?

Solution. (a) We observe that $\log(N+1) - \log N = \displaystyle\int_N^{N+1} \frac{1}{t}\,dt$, so that $(s_{N+1} - \log(N+1)) - (s_N - \log N) = \frac{1}{N+1} - \int_N^{N+1} \frac{1}{t}\,dt < 0$. Thus the sequence is a decreasing sequence. On the other hand, it consists of positive numbers, since

$\log N = \int_1^N \frac{1}{t}\,dt < 1 + \frac{1}{2} + \cdots + \frac{1}{N-1} < s_N$. It follows that the sequence must converge to a nonnegative number γ.

(*b*) The answer here depends on how "rough" an estimate is desired. We observe that $s_{10^{N+1}} - s_{10^N}$ lies between $9 \cdot 10^N\left(\frac{1}{10^{N+1}}\right)$ and $9 \cdot 10^N\left(\frac{1}{10^N}\right)$, i.e., between 0.9 and 9. Hence by an easy induction $0.9N < s_{10^N} < 9N$. Thus $m = 112$ will certainly work, and m must be at least 12.

Exercise 8.10 Prove that $\sum 1/p$ diverges; the sum extends over all primes.

(This shows that the primes form a fairly substantial subset of the positive integers.)

Hint: Given N, let $p_1, \ldots, p_k$ be those primes that divide at least one integer $\leq N$. Then

$$\begin{aligned}\sum_{n=1}^{N} \frac{1}{n} &\leq \prod_{j=1}^{k}\left(1 + \frac{1}{p_j} + \frac{1}{p_j^2} + \cdots\right)\\ &= \prod_{j=1}^{k}\left(1 - \frac{1}{p_j}\right)^{-1}\\ &\leq \exp\sum_{j=1}^{k}\frac{2}{p_j}.\end{aligned}$$

The last inequality holds because

$$(1-x)^{-1} \leq e^{2x}$$

if $0 \leq x \leq \frac{1}{2}$.

(There are many proofs of this result. See, for instance, the article by I. Niven in *Amer. Math. Monthly*, vol. 78, 1971, pp. 272–273, and the one by R. Bellman in *Amer. Math. Monthly*, vol. 50, 1943, pp. 318–319.)

Solution. We observe that the primes $p_1, \ldots, p_k$ form the set of all primes not greater than N. Each of them is at least 2, and therefore each integer from 1 to N is a unique product of the form $p_1^{e_1} \cdots p_k^{e_k}$ for nonnegative integers e_j, $0 \leq e_j \leq \log_2 N$. For simplicity let m be the greatest integer in $\log_2 N$. Then certainly

$$\begin{aligned}\sum_{n=1}^{N} \frac{1}{n} &\leq \sum_{e_1,\ldots,e_k=0}^{m} \frac{1}{p_1^{e_1}\cdots p_k^{e_k}}\\ &= \prod_{j=1}^{k}\left(1 + \frac{1}{p_j} + \cdots + \frac{1}{p_j^m}\right)\end{aligned}$$

$$
\begin{aligned}
&= \prod_{j=1}^{k} \Big(\frac{1 - p_j^{-m-1}}{1 - \frac{1}{p_j}} \Big) \\
&\le \prod_{j=1}^{k} \Big(\frac{1}{1 - \frac{1}{p_j}} \Big) \\
&= \prod_{j=1}^{k} \Big(1 - \frac{1}{p_j} \Big)^{-1} \\
&\le \exp \Big(\sum_{j=1}^{k} \frac{2}{p_j} \Big).
\end{aligned}
$$

To establish the inequality $(1-x)^{-1} \le e^{2x}$ on $[0, \frac{1}{2}]$, we simply observe that the function $f(x) = (1-x)e^{2x}$ has derivative $(1-2x)e^{2x}$, which is positive on this interval. Hence the smallest value this function has on the interval is its value at $x = 0$, which is 1.

We have now established the inequality

$$\sum_{j=1}^{k} \frac{1}{p_j} \ge \frac{1}{2} \log \Big(\sum_{n=1}^{N} \frac{1}{n} \Big)$$

for any integer N less than p_{k+1}. Since the right-hand side of this inequality tends to ∞, so does the left.

Exercise 8.11 Suppose $f \in \mathcal{R}$ on $[0, A]$ for all $A < \infty$, and $f(x) \to 1$ as $x \to +\infty$. Prove that

$$\lim_{t \to 0} \int_0^\infty e^{-tx} f(x)\, dx = 1 \quad (t > 0).$$

Solution. We first observe that the improper integral converges absolutely for all $t > 0$, since

$$\int_R^S e^{-tx} |f(x)|\, dx \le \frac{M}{t}(e^{-Rt} - e^{-St}) \to 0,$$

where $M = \sup\limits_{x \ge R} |f(x)|$, as $R, S \to \infty$.

We also note that

$$t \int_0^\infty e^{-tx} f(x)\, dx = \int_0^\infty e^{-u} f\big(\frac{u}{t}\big)\, du,$$

and this last improper integral also converges for all $t > 0$. Hence we have

$$
\begin{aligned}
\Big| t \int_0^\infty e^{-tx} f(x)\, dx - 1 \Big| &= \Big| \int_0^\infty e^{-u} f\Big(\frac{u}{t}\Big)\, du - 1 \Big| \\
&\le \int_0^\infty e^{-u} \Big| f\Big(\frac{u}{t}\Big) - 1 \Big|\, dx.
\end{aligned}
$$

Since $f(x)$ has a limit at infinity and $f(x)$ is Riemann- integrable on $[0,1]$, it follows that $f(x) \le K$ for some constant K and all x. Thus for any $\eta > 0$ we have

$$\begin{aligned}\left|t\int_0^\infty e^{-tx} f(x)\,dx - 1\right| &\le (K+1)\int_0^\eta e^{-u}\,du + \int_\eta^\infty \left|f\left(\frac{u}{t}\right) - 1\right| du \\ &\le \eta(K+1) + M_{\eta,t},\end{aligned}$$

where $M_{\eta,t} = \sup\limits_{z \ge \frac{\eta}{t}} |f(z) - 1|$. Hence, given $\varepsilon > 0$ we take $\eta = \frac{\varepsilon}{2(K+1)}$. We then choose $X > 0$ so large that $|f(z) - 1| < \frac{\varepsilon}{2}$ if $z > X$, and we let $\delta = \frac{\eta}{X}$. It then follows that $M_{\eta,t} < \frac{\varepsilon}{2}$ if $0 < t < \delta$.

Exercise 8.12 Suppose $0 < \delta < \pi$, $f(x) = 1$ if $|x| \le \delta$, $f(x) = 0$ if $\delta < |x| < \pi$, and $f(x + 2\pi) = f(x)$ for all x.

(a) Compute the Fourier coefficients of f.

(b) Conclude that

$$\sum_{n=1}^\infty \frac{\sin(n\delta)}{n} = \frac{\pi - \delta}{2} \quad (0 < \delta < \pi).$$

(c) Deduce from Parseval's theorem that

$$\sum_{n=1}^\infty \frac{\sin^2(n\delta)}{n^2\delta} = \frac{\pi - \delta}{2}.$$

(d) Let $\delta \to 0$, and prove that

$$\int_0^\infty \left(\frac{\sin x}{x}\right)^2 dx = \frac{\pi}{2}.$$

(e) Put $\delta = \pi/2$ in (c). What do you get?

Solution. (a) Since $f(x)$ is an even real-valued function, it makes sense to use the real form of the Fourier series, since symmetry shows that $b_n = 0$ for all n. Then $a_0 = \frac{2\delta}{\pi}$, and for $n \ge 1$ we have $a_n = \dfrac{1}{\pi}\displaystyle\int_{-\pi}^{\pi} f(x)\cos nx\,dx = \dfrac{2}{\pi}\displaystyle\int_0^\delta \cos nx\,dx = \dfrac{2\sin n\delta}{\pi n}$.

(b) Since $f(x)$ satisfies the Lipschitz condition of Theorem 8.14 at $x = 0$, it follows that the series actually converges to $f(0)$ at that point, i.e.,

$$\frac{\delta}{\pi} + \frac{2}{\pi}\sum_{n=1}^\infty \frac{\sin(n\delta)}{n} = 1,$$

so that

$$\sum_{n=1}^{\infty} \frac{\sin(n\delta)}{n} = \frac{\pi - \delta}{2}.$$

(*c*) Parseval's theorem now implies that

$$\frac{2\delta}{\pi} = \frac{1}{\pi}\int_{-\delta}^{\delta} |f(x)|^2\,dx = \frac{1}{2}\left(\frac{2\delta}{\pi}\right)^2 + \sum_{n=1}^{\infty} \frac{4\sin^2(n\delta)}{\pi^2 n^2}.$$

Now multiplying both sides by $\frac{\pi^2}{4\delta}$ gives the required result.

(*d*) Let R be any fixed number, N any positive integer, and let $\delta_N = \frac{R}{N}$. As $N \to \infty$ we have $\sum_{n=1}^{N} \frac{\sin^2(n\delta_N)}{n^2\delta_N} \to \int_0^R \left(\frac{\sin x}{x}\right)^2 dx$, since the left- hand side of this equality is a Riemann sum for this integral. Note that

$$\sum_{n=N+1}^{\infty} \frac{\sin^2(n\delta_N)}{n^2\delta_N} < \frac{1}{N\delta_N} = \frac{1}{R}.$$

(The inequality results from the fact that $\sum_{n=k}^{\infty} \frac{1}{n^2} < \int_{k-1}^{\infty} \frac{1}{t^2}\,dt = \frac{1}{k-1}$.) Given ε, choose $R > \frac{\varepsilon}{3}$ such that

$$\left|\int_0^{\infty} \left(\frac{\sin x}{x}\right)^2 dx - \int_0^{S} \left(\frac{\sin x}{x}\right)^2 dx\right| < \frac{\varepsilon}{3}$$

if $S > R$. Then choose $N_0 > \frac{3}{\varepsilon}$ so large that

$$\left|\sum_{n=1}^{N} \frac{\sin^2(n\delta_N)}{n^2\delta_N} - \int_0^{R} \left(\frac{\sin x}{x}\right)^2 dx\right| < \frac{\varepsilon}{3}$$

whenever $N > N_0$. Then for $N > N_0$, $\delta_N = \frac{R}{N}$ we have

$$\left|\sum_{n=1}^{\infty} \frac{\sin^2(n\delta)}{n^2\delta} - \int_0^{\infty} \left(\frac{\sin x}{x}\right)^2 dx\right| < \varepsilon.$$

Consequently

$$\int_0^{\infty} \left(\frac{\sin x}{x}\right)^2 dx = \lim_{N\to\infty} \frac{\pi - \delta_N}{2} = \frac{\pi}{2}.$$

(*e*) Taking $\delta = \pi/2$ yields

$$\sum_{k=1}^{\infty} \frac{1}{(2k-1)^2} = \frac{\pi^2}{8}.$$

Exercise 8.13 Put $f(x) = x$ if $0 \le x < 2\pi$, and apply Parseval's theorem to conclude that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Solution. By computation we see that $a_n = 0$ for $n > 0$, and $a_0 = 2\pi$. Computation shows that $b_n = \frac{-2}{n}$. Hence Parseval's relation gives

$$\frac{8\pi^2}{3} = \frac{1}{2}(2\pi)^2 + 4\sum_{n=1}^{\infty} \frac{1}{n^2},$$

so that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

There is another way of deriving this result. Since

$$\sum_{n=1}^{\infty} \frac{1}{(2n)^2} = \frac{1}{4}\sum_{n=1}^{\infty} \frac{1}{n^2},$$

denoting this last sum by X, we find that

$$X - \frac{1}{4}X = \sum_{k=1}^{\infty} \frac{1}{(2k-1)^2} = \frac{\pi^2}{8}$$

and hence, by part (e) of the previous problem

$$X = \frac{4}{3}\frac{\pi^2}{8} = \frac{\pi^2}{6}.$$

Exercise 8.14 If $f(x) = (\pi - |x|)^2$ on $[-\pi, \pi]$, prove that

$$f(x) = \frac{\pi^2}{3} + \sum_{n=1}^{\infty} \frac{4}{n^2} \cos nx$$

and deduce that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}, \quad \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}.$$

(A recent article by E. L. Stark contains many references to series of the form $\sum n^{-s}$, where s is a positive integer. See *Math. Mag.*, vol. 47, 1974, pp. 197–202.)

Solution. Since $f(x)$ is an even function, $b_n = 0$ for all n. The a_n's are computed in a straightforward manner:

$$a_0 = \frac{2}{\pi}\int_0^{\pi} f(x)\,dx = \frac{2}{\pi}\int_0^{\pi} (\pi - x)^2\,dx = \frac{2}{\pi}\int_0^{\pi} x^2\,dx = \frac{2}{3}\pi^2;$$

and

$$a_n = \frac{2}{\pi}\int_0^{\pi} (\pi - x)^2 \cos nx\, dx = (-1)^n \frac{2}{\pi}\int_0^{\pi} x^2 \cos nx\, dx,$$

so that, eventually, we find $a_n = \frac{4}{n^2}$.

This gives the stated Fourier series, and since $f(x)$ satisfies the Lipschitz condition of Theorem 8.14, the series converges to $f(x)$ at every point. Taking $x = 0$ gives the first of the two desired equalities:

$$\pi^2 = f(0) = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Parseval's theorem yields

$$\frac{2\pi^4}{5} = \frac{1}{\pi}\int_{-\pi}^{\pi} |f(x)|^2\, dx = \frac{2\pi^4}{9} + 16\sum_{n=1}^{\infty} \frac{1}{n^4},$$

which easily transforms to the desired relation.

Exercise 8.15 With D_n as defined in (77), put

$$K_n(x) = \frac{1}{N+1}\sum_{n=0}^{N} D_n(x).$$

Prove that

$$K_N(x) = \frac{1}{N+1}\frac{1-\cos(N+1)x}{1-\cos x}$$

and that

(a) $K_n \geq 0$,

(b) $\displaystyle\frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(x)\, dx = 1.$

(c) $\displaystyle K_n(x) \leq \frac{1}{N+1}\frac{2}{1-\cos\delta}$ if $0 < \delta \leq |x| \leq \pi$.

If $s_N = s_N(f;x)$ is the Nth partial sum of the Fourier series of f, consider the arithmetic means

$$\sigma_N = \frac{s_0 + s_1 + \cdots + s_N}{N+1}.$$

Prove that

$$\sigma_N(f;x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)K_N(t)\, dt$$

and hence prove Fejér's theorem:

If f is continuous, with period 2π, then $\sigma_N(f;x) \to f(x)$ uniformly on $[-\pi,\pi]$.

Hint: Use properties (a), (b), (c) to proceed as in Theorem 7.26.

Solution. Using the formula $1 - \cos\theta = 2\sin^2 \frac{1}{2}\theta$, and the formula $D_n(x) = \frac{\sin(n+\frac{1}{2})x}{\sin\frac{1}{2}x}$, we deduce that

$$(1 - \cos x)K_N(x) = \frac{1}{N+1}\sum_{n=0}^{N} \sin\frac{1}{2}x \sin(n+\frac{1}{2})x.$$

Now, however, $\sin\alpha\sin\beta = \frac{1}{2}\cos(\alpha-\beta) - \cos(\alpha+\beta)$, so that

$$(1 - \cos x)K_N(x) = \frac{1}{N+1}\sum_{n=0}^{N} \big(\cos(nx) - \cos((n+1)x)\big) =$$

$$= \frac{1}{N+1}\big(1 - \cos(N+1)x\big).$$

The formula is now established. Notice that it could also be written

$$K_N(x) = \frac{1}{N+1}\left[\frac{\sin\left(\frac{N+1}{2}x\right)}{\sin\frac{1}{2}x}\right]^2.$$

(a) The nonnegativity of $K_N(x)$ is an immediate consequence of either of the formulas just written.

(b) It was established in the text that $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(x)\,dx = 1$, and so the same result for $K_N(x)$, which is an average of the $D_n(x)$, must follow by routine computation.

(c) This inequality is an immediate consequence of the facts that $\cos(N+1)x \geq -1$ and that $\cos x$ is decreasing on $[0,\pi]$.

The formula for $\sigma_N(f;x)$ is an immediate consequence of the definition of $\sigma_N(f;x)$ and the corresponding formula for $s_n(f;x)$.

Now let $M = \sup|f(x)|$, the supremum being taken over all x. By (a) and (b) we have

$$\begin{aligned}
|\sigma_N(x) - f(x)| &= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi}[f(x-t) - f(x)]K_N(t)\,dt\right| \\
&\leq \frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x-t) - f(x)|K_N(t)\,dt \\
&\leq \frac{1}{2\pi}\int_{-\delta}^{\delta}|f(x-t) - f(x)|K_N(t)\,dt + \\
&\qquad + \frac{1}{\pi}(\pi-\delta)\frac{1}{N+1}\frac{2}{1-\cos\delta}2M \\
&\leq \sup_{|t|\leq\delta}|f(x-t) - f(x)| + \frac{Q_\delta}{N+1}
\end{aligned}$$

where $Q_\delta = \frac{4M(\pi-\delta)}{\pi(N+1)(1-\cos\delta)}$.

Given $\varepsilon > 0$, we first choose $\delta > 0$ so small that $\sup\limits_{|t|\le\delta} |f(x-t) - f(x)| < \frac{\varepsilon}{2}$ for all x. With this δ fixed, we then have $|\sigma_N(f;x) - f(x)| < \varepsilon$ for all $N > \frac{2Q_\delta}{\varepsilon}$ and all x.

Exercise 8.16 Prove a pointwise version of Fejér's Theorem:

If $f \in \mathcal{R}$ and $f(x+)$, $f(x-)$ exist for some x, then

$$\lim_{N\to\infty} \sigma_N(f;x) = \frac{1}{2}[f(x+) + f(x-)].$$

Solution. We need only a slight modification of the argument just given, namely the formula

$$\sigma_N(f;x) - \frac{1}{2}[f(x+) + f(x-)] =$$
$$= \frac{1}{2\pi}\int_0^\pi [f(x-t) - f(x-)]K_N(t)\,dt + \frac{1}{2\pi}\int_{-\pi}^0 [f(x-t) - f(x+)]K_N(t)\,dt.$$

Each of these two integrals can be broken up into an integral over a half-neighborhood of 0 and an integral outside that neighborhood. The first of the integrals can be made small if the neighborhood is taken small enough (independently of N). With that neighborhood fixed, the second integral in each case can be made small if N is large enough using the same inequalities just stated.

Exercise 8.17 Assume f is bounded and monotonic on $[-\pi, \pi)$ with Fourier coefficients c_n, as given by (62).
(a) Use Exercise 17 of Chap. 6 to prove that $\{nc_n\}$ is a bounded sequence.
(b) Combine (a) with Exercise 16 and with Exercise 14(e) of Chap. 3, to conclude that

$$\lim_{n\to\infty} s_N(f;x) = \frac{1}{2}[f(x+) + f(x-)]$$

for every x.
(c) Assume only that $f \in \mathcal{R}$ on $[-\pi, \pi]$ and that f is monotonic in some segment $(\alpha, \beta) \subset [-\pi, \pi]$. Prove that the conclusion of (b) holds for every $x \in (\alpha, \beta)$.

(This is an application of the localization theorem.)

Solution. (a) by Exercise 17 of Chap. 6 we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx = \frac{-1}{2\pi n}\int_{-\pi}^{\pi} e^{-inx}\,df(x),$$

from which it follows that

$$|nc_n| \le \frac{1}{2\pi}[f(\pi) - f(-\pi)].$$

(*b*) Since $f(x+)$ and $f(x-)$ exist at every point, it follows from the previous exercise that $\sigma_N(f;x) \to \frac{1}{2}[f(x+) + f(x-)]$. Then Exercise 14(*e*) of Chap. 3 assures us that $s_n(f;x)$ has the same limit.

(*c*) Let $g(x) = f(x)$ for $\alpha \le x \le \beta$, $g(x) = f(\alpha)$ for $0 \le x \le \alpha$ and $g(x) = f(\beta)$ for $\beta \le x \le 2\pi$. Then $s_N(g;x) \to \frac{1}{2}[g(x+) + g(x-)]$ for all x by part (*b*). Since $s_N(g;x) - s_N(f;x) \to 0$ for $\alpha < x < \beta$ by the Corollary to Theorem 8.14, it follows that $s_N(f;x) \to \frac{1}{2}[g(x+)+g(x-)] = \frac{1}{2}[f(x+)+f(x-)]$ for these values of x.

Exercise 8.18 Define

$$\begin{aligned} f(x) &= x^3 - \sin^2 x \tan x \\ g(x) &= 2x^2 - \sin^2 x - x \tan x. \end{aligned}$$

Find out, for each of these two functions, whether it is positive or negative for all $x \in (0, \pi/2)$, or whether it changes sign. Prove your answer.

Solution. Both functions tend to $-\infty$ as $x \to \frac{\pi}{2}$. Hence the only question is whether they ever become positive. We note that the derivative of the first function is $3x^2 - \sin^2 x - \tan^2 x$. By writing $\sin^2 x$ as $\frac{1}{2} - \frac{1}{2}\cos 2x$ and making repeated use of the formula $\frac{d}{dx}\tan^k x = k\tan^{k-1} x + k\tan^{k+1} x$, we find that the first six derivatives of this function vanish at 0, and that the sixth derivative is

$$-32\sin 2x - 272\tan x - 1232\tan^3 x - 1104\tan^5 x - 144\tan^7 x,$$

which is negative on $\left(0, \frac{\pi}{2}\right)$. Hence all of the first six derivatives are negative on this interval, and therefore the function itself is negative.

The same technique applies to the second function. All of its first five derivatives vanish at $x = 0$ and the fifth is

$$\begin{aligned} -[16\sin 2x + 16x + 80\tan x + 136x\tan^2 x + \\ + 200\tan^3 x + 240x\tan^4 x + 120\tan^5 x + 120x\tan^6 x], \end{aligned}$$

which is negative on $\left(0, \frac{\pi}{2}\right)$. Hence this function is always negative on that interval.

Exercise 8.19 Suppose f is a continuous function on R^1, $f(x + 2\pi) = f(x)$, and α/π is irrational. Prove that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{\infty} f(x + n\alpha) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\,dt$$

for every x. *Hint:* Do it first for $f(x) = e^{ikx}$.

Solution. Following the hint, we observe that both sides of the desired equality equal 1 trivially when $k = 0$. In any other case the right-hand side is zero, and the left-hand side is

$$\lim_{N\to\infty} e^{ikx}\frac{1-e^{i(n+1)k\alpha}}{N(1-e^{i\alpha})},$$

which tends to zero as $N \to \infty$.

Since both sides are linear functions of f, it now follows that the relation holds for all trigonometric polynomials. Finally, since both sides are bounded by the supremum of f, given ε, we can approximate f uniformly within ε by a trigonometric polynomial. It then follows that all the means on the left, for N sufficiently large, are within 2ε of the integral on the right. Since ε is arbitrary, it follows that the limit on the left equals the integral on the right.

Exercise 8.20 The following simple computation yields a good approximation to Stirling's formula.

For $m = 1, 2, 3, \ldots$, define

$$f(x) = (m + 1 - x)\log m - (x - m)\log(m + 1)$$

if $m \le x \le m + 1$, and define

$$g(x) = \frac{x}{m} - 1 + \log m$$

if $m-\frac{1}{2} \le x < m+\frac{1}{2}$. Draw the graphs of f and g. Note that $f(x) \le \log x \le g(x)$ if $x \ge 1$ and that

$$\int_1^n f(x)\,dx = \log(n!) - \frac{1}{2}\log n > -\frac{1}{8} + \int_1^n g(x)\,dx.$$

Integrate $\log x$ over $[1, n]$. Conclude that

$$\frac{7}{8} < \log(n!) - \left(n + \frac{1}{2}\right)\log n + n < 1$$

for $n = 2, 3, 4, \ldots$. (*Note:* $\log\sqrt{2\pi} \sim 0.918\ldots$.) Thus

$$e^{7/8} < \frac{n!}{(n/e)^n\sqrt{n}} < e.$$

Solution. We first draw the graphs of f and g in the range $x = 1$ to $x = 10$. We note that f is merely the broken set of chords joining the points on the graph of $\log x$ at integer values of x, and g is made up of segments of the tangents at these points (g is not continuous). Because the downward side of the graph of $\log x$ is convex, $f(x) \le \log x \le g(x)$ for all x. The estimate for the integral of f is straightforward: The integral is the sum of the areas of one triangle and $n-2$ trapezoids with base 1 and parallel sides $\log k$ and $\log(k+1)$ $(k = 2, \ldots, n-1)$.

We find it equal to $\frac{1}{2}\log 1 + \log 2 + \log 3 + \cdots + \log(n-1) + \frac{1}{2}\log n = \log(n!) - \frac{1}{2}\log n$, as asserted. Meanwhile the integral of $g(x)$ is also a sum of trapezoids and two triangles and equals $\frac{1}{8} + \log(n!) - \frac{1}{2}\log n - \frac{1}{8n}$. Hence we have

$$\log(n!) - \frac{1}{2}\log n < \int_1^n \log x\, dx < \frac{1}{8} + \log(n!) - \frac{1}{2}\log n - \frac{1}{8n} < \frac{1}{8} + \log(n!) - \frac{1}{2}\log n.$$

Now straightforward computation reveals that

$$\int_1^n \log x\, dx = (n\log n - n) - (1\log 1 - 1) = n\log n - n + 1.$$

The desired inequalities are now deduced by taking exponentials of the three expressions.

Exercise 8.21 Let

$$L_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} |D_n(t)|\, dt \quad (n = 1, 2, 3, \ldots).$$

Prove that there exists a constant $C > 0$ such that

$$L_n > C\log n \quad (n = 1, 2, 3, \ldots),$$

or, more precisely, that the sequence

$$\left\{L_n - \frac{4}{\pi^2}\log n\right\}$$

is bounded.

Solution. We observe that

$$L_n = \frac{1}{\pi}\int_0^{\frac{2\pi}{2n+1}} \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}\, dt +$$
$$+ \sum_{k=1}^{n-1} \frac{1}{\pi}\int_{\frac{2\pi k}{2n+1}}^{\frac{2\pi(k+1)}{2n+1}} \frac{(-1)^k \sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}\, dt + \frac{1}{\pi}\int_{\frac{2n\pi}{2n+1}}^{\pi} \frac{(-1)^n \sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}\, dt$$

The substitution $u = (n+\frac{1}{2})t$ changes the first and last terms into the sum

$$\frac{1}{\pi}\int_0^{\pi} \frac{\sin u}{(n+\frac{1}{2})\sin(\frac{u}{2n+1})}\, du + \int_{n\pi}^{(n+\frac{1}{2})\pi} \frac{(-1)^n \sin u}{(n+\frac{1}{2})\sin(\frac{u}{2n+1})}\, du.$$

The first of these terms tends to $\dfrac{1}{2\pi}\displaystyle\int_0^{\pi} \sin u\, du = \dfrac{1}{\pi}$ as $n \to \infty$. The second tends to 0 (for $u \in [n\pi, (n+\frac{1}{2})\pi]$ we have $\sin(\frac{u}{2n+1}) \geq \sin\frac{n\pi}{2n+1}$, which tends to 1 as $n \to \infty$).

Thus we find that

$$\frac{1}{\pi} + \varepsilon_n + \sum_{k=1}^{n-1} \frac{1}{\pi \sin(\frac{\pi(k+1)}{2n+1})} \left| \int_{\frac{2\pi k}{2n+1}}^{\frac{2\pi(k+1)}{2n+1}} \sin(n + \frac{1}{2})t \, dt \right| < L_n,$$

where $\varepsilon_n \to 0$ as $n \to \infty$.

If we take out the first two terms of the sum instead of just the first, we find similarly that

$$L_n = \frac{1}{\pi} \int_0^{\frac{4\pi}{2n+1}} \frac{|\sin(n+\frac{1}{2})t|}{\sin \frac{1}{2}t} \, dt +$$

$$+ \sum_{k=2}^{n-1} \frac{1}{\pi} \int_{\frac{2\pi k}{2n+1}}^{\frac{2\pi(k+1)}{2n+1}} \frac{(-1)^k \sin(n+\frac{1}{2})t}{\sin \frac{1}{2}t} \, dt + \frac{1}{\pi} \int_{\frac{2n\pi}{2n+1}}^{\pi} \frac{(-1)^n \sin(n+\frac{1}{2})t}{\sin \frac{1}{2}t} \, dt.$$

Again the substitution $u = (n + \frac{1}{2})t$ changes the first and last terms into the sum

$$\frac{1}{\pi} \int_0^{2\pi} \frac{\sin u}{(n+\frac{1}{2}) \sin(\frac{u}{2n+1})} \, du + \int_{n\pi}^{(n+\frac{1}{2})\pi} \frac{(-1)^n \sin u}{(n+\frac{1}{2}) \sin(\frac{u}{2n+1})} \, du.$$

The first of these terms tends to $\dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} |\sin u| \, du = \frac{2}{\pi}$ as $n \to \infty$, and once again the second tends to zero.

Thus we find that

$$L_n < \frac{2}{\pi} + \eta_n + \sum_{k=2}^{n-1} \frac{1}{\pi \sin(\frac{\pi k}{2n+1})} \left| \int_{\frac{2\pi k}{2n+1}}^{\frac{2\pi(k+1)}{2n+1}} \sin(n + \frac{1}{2})t \, dt \right|,$$

where $\eta_n \to 0$ as $n \to \infty$.

Once again, in each of the integrals under the sigma in the last two inequalities we make the substitution $u = (n + \frac{1}{2})t$. When we do so, we have

$$\frac{1}{\pi} + \varepsilon + \sum_{k=1}^{n-1} \frac{2}{(n+\frac{1}{2})\pi \sin(\frac{\pi(k+1)}{2n+1})} < L_n < \frac{2}{\pi} + \eta_n + \sum_{k=2}^{n-1} \frac{2}{(n+\frac{1}{2})\pi \sin(\frac{\pi k}{2n+1})},$$

where $\varepsilon_n \to 0$ and $\eta_n \to 0$. It therefore follows that

$$\frac{1}{\pi} + \varepsilon_n < L_n - \sum_{k=1}^{n-1} \frac{2}{\pi} \frac{1}{(n+\frac{1}{2}) \sin(\frac{\pi(k+1)}{2n+1})} < \frac{2}{\pi} + \eta_n - \frac{1}{(n+\frac{1}{2}) \sin(\frac{\pi n}{2n+1})}.$$

The extremes in these inequalities are both bounded. Hence we will be done if we can show that

$$\frac{2}{\pi} \log n - \sum_{k=1}^{n-1} \frac{1}{(n+\frac{1}{2}) \sin(\frac{\pi(k+1)}{2n+1})}$$

remains bounded. To do this, we use the fact that there is a constant K such that

$$\left|\frac{1}{\sin x}-\frac{1}{x}\right|\le Kx$$

for $0<x\le\frac{\pi}{2}$. This fact in turn is a consequence of the fact that, by L'Hospital's rule,

$$\lim_{x\to 0}\frac{x-\sin x}{x^2\sin x}=-\frac{1}{2}.$$

We thus have

$$\sum_{k=1}^{n-1}\frac{1}{(n+\frac{1}{2})\sin(\frac{\pi(k+1)}{2n+1})}=E_n+\sum_{k=1}^{n-1}\frac{1}{(n+\frac{1}{2})(\frac{\pi(k+1)}{2n+1})},$$

where

$$|E_n|\le K\frac{1}{n+\frac{1}{2}}\sum_{k=1}^{n-1}\frac{\pi(k+1)}{2n+1}=$$

$$=\frac{2K}{\pi(2n+1)^2}\sum_{k=1}^{n}k+1=\frac{2K}{\pi(2n+1)^2}\left[\frac{(n+1)(n+2)}{2}-1\right].$$

Since the right-hand side tends to $\frac{K}{4\pi}$ as $n\to\infty$, we see that E_n remains bounded as $n\to\infty$. We will be finished if we can show that

$$\log n-\sum_{k=1}^{n-1}\frac{1}{k+1}$$

remains bounded. But this was done in Exercise 9 above.

Exercise 8.22 If α is real and $-1<x<1$, prove Newton's binomial theorem

$$(1+x)^\alpha=1+\sum_{n=1}^{\infty}\frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n.$$

Hint: Denote the right side by $f(x)$. Prove that the series converges. Prove that

$$(1+x)f'(x)=\alpha f(x).$$

and solve this differential equation.

Show also that

$$(1-x)^{-\alpha}=\sum_{n=0}^{\infty}\frac{\Gamma(n+\alpha)}{n!\Gamma(\alpha)}x^n$$

if $-1<x<1$ and $\alpha>0$.

Solution. Following the hint, we use the ratio test to establish that the radius of convergence of the power series that defines $f(x)$ is 1. This amounts merely to the statement that

$$\lim_{n\to\infty}\left|\frac{\alpha-n}{n+1}\right|=1.$$

The differential equation then results from termwise operations on the series and the fact that

$$\frac{\alpha(\alpha-1)\cdots(\alpha-n)}{n!}+\frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{(n-1)!}=\alpha\frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}.$$

Then, given that $f(0)=1\neq 0$, it follows that for x near 0 we have

$$\frac{f'(x)}{f(x)}=\frac{\alpha}{1+x}.$$

so that $\log f(x)$ and $\log(1+x)^\alpha$ have the same derivative, and hence differ by a constant, which turns out to be zero, since both equal 1 at $x=0$. It thus follows that $f(x)=(1+x)^\alpha$.

To prove the other relation, we merely observe that

$$(-\alpha(-\alpha-1)\cdots(-\alpha-n+1)=(-1)^n\alpha(\alpha+1)\cdots(\alpha+n-1)=(-1)^n\frac{\Gamma(n+\alpha)}{\Gamma(\alpha)}.$$

Exercise 8.23 Let γ be a continuously differentiable *closed* curve in the complex plane with parameter interval $[a,b]$, and assume that $\gamma(t)\neq 0$ for every $t\in[a,b]$. Define the *index* of γ to be

$$\mathrm{Ind}\,(\gamma)=\frac{1}{2\pi i}\int_a^b\frac{\gamma'(t)}{\gamma(t)}\,dt.$$

Prove that $\mathrm{Ind}\,(\gamma)$ is always an integer.

Hint: There exists φ on $[a,b]$ with $\varphi'=\gamma'/\gamma$, $\varphi(a)=0$. Hence $\gamma\exp(-\varphi)$ is constant. Since $\gamma(a)=\gamma(b)$, it follows that $\exp(\varphi(a))=\exp(\varphi(b))=1$. Note that $\varphi(b)=2\pi i\mathrm{Ind}\,(\gamma)$.

Compute $\mathrm{Ind}\,(\gamma)$ when $\gamma(t)=e^{int}$, $a=0$, $b=2\pi$.

Explain why $\mathrm{Ind}\,(\gamma)$ is often called the *winding number* of γ around 0.

Solution. Again, following the hint leaves very little to do. We define

$$\varphi(x)=\int_a^x\frac{\gamma'(t)}{\gamma(t)}\,dt,$$

so that we automatically have $\varphi'(t)=\frac{\gamma'(t)}{\gamma(t)}$. The fact that $\gamma\exp(-\varphi)$ is constant is now a consequence of the chain rule. It then follows immediately that $\exp(\varphi(b))=1$, so that $\varphi(b)=2\pi in$ for some integer n.

Routine computation shows that $\mathrm{Ind}\,(\gamma)=n$ if $\gamma(t)=e^{int}$, $0\leq t\leq 2\pi$. Since this curve winds counterclockwise about 0 a total of n times, the name *winding number* is appropriate.

Exercise 8.24 Let γ be as in Exercise 23, and assume in addition that the range of γ does not intersect the negative real axis. Prove that $\text{Ind}(\gamma) = 0$. *Hint:* For $0 \leq c < \infty$, $\text{Ind}(\gamma + c)$ is a continuous integer-valued function of c. Also, $\text{Ind}(\gamma + c) \to 0$ as $c \to \infty$.

Solution. Following the hint, we observe that

$$f(c) = \frac{1}{2\pi i} \int_a^b \frac{\gamma'(t)}{\gamma(t) + c}\, dt,$$

is a continuous function of c on $[0, \infty)$, since

$$|f(c_1) - f(c_2)| = \frac{1}{2\pi} \left| \int_a^b \frac{\gamma'(t)(c_1 - c_2)}{(\gamma(t) + c_1)(\gamma(t) + c_2)}\, dt \right| \leq K|c_1 - c_2|,$$

where $K = \dfrac{1}{2\pi r^2} \displaystyle\int_a^b |\gamma'(t)|\, dt$ and r is the supremum of the integrand for $c_1, c_2 \geq 0$ and $0 \leq t \leq 2\pi$. (This supremum is finite, since the integrand tends to zero as either c_1 or c_2 tends to infinity.) Furthermore

$$|f(c)| \leq \frac{1}{2\pi c} \int_a^b \frac{|\gamma'(t)|}{|1 + \frac{\gamma(t)}{c}|}\, dt,$$

and this last expression tends to 0 as $c \to \infty$. It follows, since f assumes only integer values, that $f(c) \equiv 0$. In particular $f(0) = \text{Ind}(\gamma) = 0$.

Exercise 8.25 Suppose γ_1 and γ_2 are curves as in Exercise 23, and

$$|\gamma_1(t) - \gamma_2(t)| < |\gamma_1(t)| \quad (a \leq t \leq b)$$

Prove that $\text{Ind}(\gamma_1) = \text{Ind}(\gamma_2)$.

Hint: Put $\gamma = \gamma_2/\gamma_1$. Then $|1 - \gamma| < 1$. Hence $\text{Ind}(\gamma) = 0$ by Exercise 24. Also,

$$\frac{\gamma'}{\gamma} = \frac{\gamma_2'}{\gamma_2} - \frac{\gamma_1'}{\gamma_1}.$$

Solution. The hint leaves almost nothing to be done. The inequality established for γ shows that the real part of γ is always positive, so that the hypotheses of Exercise 24 are satisfied. The relation for $\frac{\gamma'}{\gamma}$ is a routine computation, and shows in general that $\text{Ind}(\gamma\delta) = \text{Ind}(\gamma) + \text{Ind}(\delta)$.

Exercise 8.26 Let γ be a *closed* curve in the complex plane (not necessarily differentiable) with parameter interval $[0, 2\pi]$, such that $\gamma(t) \neq 0$ for every $t \in [0, 2\pi]$.

Choose $\delta > 0$ such that $|\gamma(t)| > \delta$ for all $t \in [0, 2\pi]$. If P_1 and P_2 are trigonometric polynomials such that $|P_i(t) - \gamma(t)| < \delta/4$ for all $t \in [0, 2\pi]$, (their existence is assured by Theorem 8.15), prove that

$$\mathrm{Ind}\,(P_1) = \mathrm{Ind}\,(P_2)$$

by applying Exercise 25.

Define this common value to be $\mathrm{Ind}\,(\gamma)$.

Prove that the statements of Exercises 24 and 25 hold without any differentiability assumptions.

Solution. Since $|P_1(t) - P_2(t)| < \frac{\delta}{2} < |P_1(t)|$, (because $|P_1(t)| \geq |f(t)| - |f(t) - P_1(t)| > \frac{3\delta}{4}$), the equality of the indices follows from Exercise 25, as stated.

Exercise 24 remains valid, since if $\gamma(t)$ does not intersect the negative real axis, there is a positive number $\delta > 0$ such that $|\gamma(t) - x| \geq \delta$ for all $x \leq 0$. Then if $|P_j(t) - \gamma(t)| < \delta$ for all $t \in [0, 2\pi]$, it follows that $P_j(t)$ also does not intersect the negative real axis, hence has winding number 0.

Exercise 25 remains valid, since if $|\gamma_1(t) - \gamma_2(t)| < |\gamma_1(t)|$ for all t, we can let $\delta = \min_t |\gamma_1(t)| - |\gamma_1(t) - \gamma_2(t)|$. Then if $|P_i(t) - \gamma_i(t)| < \delta/4$ for all t, it follows that $|P_1(t) - P_2(t)| \leq |\gamma_1(t) - \gamma_2(t)| + (\delta/2) < |\gamma_1(t)| - (\delta/4) \leq |P_1(t)|$, and so $\mathrm{Ind}\,(P_1) = \mathrm{Ind}\,(P_2)$, by Exercise 25.

Exercise 8.27 Let f be a continuous complex function defined in the complex plane. Suppose there is a positive integer n and a complex number $c \neq 0$ such that

$$\lim_{|z| \to \infty} z^{-n} \gamma(z) = c.$$

Prove that $f(z) = 0$ for at least one complex number z.

Note that this is a generalization of Theorem 8.8.

Hint: Assume $f(z) \neq 0$ for all z, define

$$\gamma_r(t) = f(re^{it\theta})$$

for $0 \leq r < \infty$, $0 \leq t \leq 2\pi$, and prove the following statements about the curve γ.

(a) $\mathrm{Ind}\,(\gamma_0) = 0$.

(b) $\mathrm{Ind}\,(\gamma_r) = n$ for all sufficiently large r.

(c) $\mathrm{Ind}\,(\gamma_r)$ is a continuous function of r on $[0, \infty)$.

[In (b) and (c), use the last part of Exercise 26.]

Show that (a), (b), and (c) are contradictory, since $n > 0$.

Solution. (a) Since $\gamma_0(t) = f(0)$ for all t, we have $\gamma_0'(t) = 0$ for all t, and hence by definition $\mathrm{Ind}\,(\gamma_0) = 0$.

(b) Choose R so large that $|z^{-n} f(z) - c| < \frac{|c|}{2}$ whenever $|z| > R$. Then for all r we have $\mathrm{Ind}\,(\gamma_r) = \mathrm{Ind}\,(\gamma_{r1}) + \mathrm{Ind}\,(\gamma_{r2})$, where $\gamma_{r1}(t) = r^n e^{int}$ and

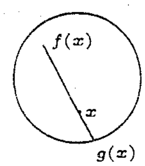

Figure 8.1: The Brouwer fixed-point theorem

$\gamma_{r2}(t) = r^{-n}e^{-int}f(re^{it})$. By Exercise 25 we have Ind $(\gamma_{r2}) = 0$ for $r > R$, and by direct computation we have Ind $(\gamma_{r1}) = n$ for all r.

(c) Fix $r_0 > 0$, and let $\varepsilon = \min\limits_{0\le t\le 2\pi} |f(r_0e^{it})|$. Then choose $\delta \in (0, r_0)$ such that $|f(r_0e^{it}) - f(re^{it})| < \varepsilon$ if $|r - r_0| < \delta$. Then by Exercise 25 we again have $\mathrm{Ind}(\gamma_r) = \mathrm{Ind}\,(\gamma_{r_0})$ for $|r - r_0| < \delta$. Hence Ind (γ_r) is a locally constant function of r. By the connectivity of $[0, \infty)$, it follows that it is globally constant, which contradicts (a) and (b).

Exercise 8.28 Let $\overline{D}$ be the closed unit disc in the complex plane. (Thus $z \in \overline{D}$ if and only if $|z| \le 1$.) Let g be a continuous mapping of $\overline{D}$ into the unit circle T. (Thus $|g(z)| = 1$ for every $z \in \overline{D}$.)

Prove that $g(z) = -z$ for at least one $z \in T$.

Hint: For $0 \le r \le 1$, $0 \le t \le 2\pi$, put

$$\gamma_r(t) = g(re^{it}),$$

and put $\psi(t) = e^{-it}\gamma_1(t)$. If $g(z) \ne -z$ for every $z \in T$, then $\psi(t) \ne -1$ for every $t \in [0, 2\pi]$. Hence Ind $(\psi) = 0$, by Exercises 24 and 25. It follows that Ind $(\gamma_1) = 1$. But Ind $(\gamma_0) = 0$. Derive a contradiction, as in Exercise 27.

Solution. The hint tells us that $\psi(t)$ does not meet the negative real axis, hence has index 0, by Exercise 24. Hence by Exercise 25, γ_1 has index 1. Again, since $\gamma_0 = g(0) \ne 0$ (since $g(0) \ne -0 = 0$), it follows that Ind $(\gamma_0) = 0$. But, as before, since $|g(z)| = 1$ for all z, it follows that Ind (γ_r) is locally constant and hence by the connectivity of $[0, 1]$, globally constant. Thus, once again, we have a contradiction.

Exercise 8.29 Prove that every continuous mapping f of $\overline{D}$ into $\overline{D}$ has a fixed point in $\overline{D}$.

(This is the 2-dimensional case of Brouwer's fixed-point theorem.)

Hint: Assume $f(z) \ne z$ for every $z \in \overline{D}$. Associate to each $z \in \overline{D}$ the point $g(z) \in T$ which lies on the ray that starts at $f(z)$ and passes through z. Then g maps $\overline{D}$ into T, $g(z) = z$ if $z \in T$, and g is continuous, because

$$g(z) = z - s(z)[f(z) - z],$$

where $s(z)$ is the unique nonnegative root of a certain quadratic equation whose coefficients are continuous functions of f and z. Apply Exercise 28.

Solution. The number $s = s(z)$ is a nonnegative real number because of the geometry of the situation (see figure). The quadratic equation in question is given by the relation $|g(z)|^2 = 1$, i.e.,

$$|f(z) - z|^2 s^2 + 2\big(|z|^2 - \operatorname{Re}(\bar{z}f(z))\big)s + |z|^2 - 1 = 0.$$

It is well-known that a quadratic equation $az^2 + bz + c = 0$ has one and only one nonnegative root if a, b, and c are real and $ac < 0$. We can write explicitly

$$s(z) = \frac{|z|^2 - \operatorname{Re}(\bar{z}f(z)) + \sqrt{(|z|^2 - \operatorname{Re}(\bar{z}f(z))^2 + |f(z) - z|^2(1 - |z|^2)}}{|f(z) - z|^2}.$$

which makes it clear that $s(z)$ is a continuous function of z. Hence $g(z)$ is continuous.

We now know that there must be a value at which $g(z) = -z$. But this is impossible, since $|g(z)| = 1$ for all z and $g(z) = z$ if $|z| = 1$.

Exercise 8.30 Use Stirling's formula to prove that

$$\lim_{x\to\infty} \frac{\Gamma(x+c)}{x^c\Gamma(x)} = 1$$

for every real constant c.

Solution. We need Stirling's formula in the form

$$\lim_{z\to\infty} \frac{\Gamma(z)}{\left(\frac{z-1}{e}\right)^{z-1}\sqrt{2\pi(z-1)}} = 1.$$

Applying this result with $z = x + c$ and $z = x$, we get

$$\lim_{x\to\infty} \frac{\Gamma(x+c)}{x^c\Gamma(x)} =$$

$$= \lim_{x\to\infty} f(x) \cdot \frac{\Gamma(x+c)}{(\frac{x+c-1}{e})^{x+c-1}\sqrt{2\pi(x+c-1)}} \cdot \frac{(\frac{x-1}{e})^{x-1}\sqrt{2\pi(x-1)}}{\Gamma(x)},$$

where

$$f(x) = \frac{1}{x^c} \cdot \frac{(\frac{x+c-1}{e})^{x+c-1}}{(\frac{x-1}{e})^{x-1}} \cdot \sqrt{\frac{x+c-1}{x-1}} = \frac{(1 + \frac{c-1}{x})^x(1 + \frac{c-1}{x})^{c-1}}{e^c(1 - \frac{1}{x})^x(1 - \frac{1}{x})^{-1}} \cdot \sqrt{\frac{x+c-1}{x-1}}.$$

Since $x^x \to 1$ as $x \to \infty$, it now follows that $\lim_{x\to\infty} f(x) = 1$, which, combined with Stirling's formula, gives the desired result.

Exercise 8.31 In the proof of Theorem 7.26 it was shown that

$$\int_{-1}^{1} (1-x^2)^n \, dx \geq \frac{4}{3\sqrt{\pi}}.$$

for $n = 1, 2, 3, \ldots$. Use Theorem 8.20 and Exercise 30 to show the more precise result

$$\lim_{n\to\infty} \sqrt{n} \int_{-1}^{1} (1-x^2)^n \, dx = \sqrt{\pi}.$$

Solution. Let $u = x^2$ in the integral, so that $dx = \frac{1}{2}u^{-\frac{1}{2}} \, du$. We then have

$$\sqrt{n} \int_{-1}^{1} (1-x^2)^n \, dx = \sqrt{n} \int_{0}^{1} (1-u)^n u^{-\frac{1}{2}} \, du = \frac{\sqrt{n}\Gamma(n+\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(n+1)},$$

and taking $c = \frac{1}{2}$ in Exercise 30, we find that this last expression tends to $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Chapter 9

Functions of Several Variables

Exercise 9.1 If S is a nonempty subset of a vector space X, prove (as asserted in Sec. 9.1) that the span of S is a vector space.

Solution. We need only verify that the span of S is closed under the two vector space operations. All the other properties of a vector space hold in the span of S, since it is contained in a vector space in which they hold.

To that end, let $\mathbf{x}$ and $\mathbf{y}$ be elements of the span of S, and let c be any real number. By definition there are elements $\mathbf{x}_1, \ldots, \mathbf{x}_m, \mathbf{y}_1, \ldots, \mathbf{y}_n$, and scalars $c_1, \ldots, c_m, d_1, \ldots, d_n$ such that $\mathbf{x} = c_1\mathbf{x}_1 + \cdots + c_m\mathbf{x}_m$ and $\mathbf{y} = d_1\mathbf{y}_1 + \cdots + d_n\mathbf{y}_n$. We then have

$$\mathbf{x} + \mathbf{y} = c_1\mathbf{x}_1 + \cdots + c_m\mathbf{x}_m + d_1\mathbf{y}_1 + \cdots + d_n\mathbf{y}_n,$$

which is a finite linear combination of elements of S, hence belongs to the span of S. Likewise, by the distributive law,

$$c\mathbf{x} = c(c_1\mathbf{x}_1 + \cdots + c_m\mathbf{x}_m) = (cc_1)\mathbf{x}_1 + \cdots + (cc_m)\mathbf{x}_m,$$

which belongs to the span of S.

Exercise 9.2 Prove (as asserted in Sec. 9.6) that BA is linear if A and B are linear transformations. Prove also that A^{-1} is linear and invertible.

Solution. Let $A : X \to Y$ and $B : Y \to Z$ be linear transformations, and let $\mathbf{x}$ and $\mathbf{y}$ be any elements of A and c any scalar. Then $BA : X \to Z$ satisfies

$$\begin{aligned} BA(\mathbf{x}+\mathbf{y}) &= B\big(A(\mathbf{x}+\mathbf{y})\big) \\ &= B\big(A(\mathbf{x}) + A(\mathbf{y})\big) \\ &= B\big(A(x)\big) + B\big(A(\mathbf{y})\big) \\ &= BA(\mathbf{x}) + BA(\mathbf{y}). \end{aligned}$$

Similarly,

$$\begin{aligned} BA(c\mathbf{x}) &= B(A(c\mathbf{x})) \\ &= B(cA(\mathbf{x})) \\ &= cB(A(\mathbf{x})) \\ &= cBA(\mathbf{x}). \end{aligned}$$

If A is a one-to-one mapping of X onto Y, and $\mathbf{z}$ and $\mathbf{w}$ are any elements of Y, let $\mathbf{x} = A^{-1}(\mathbf{z})$ and $\mathbf{y} = A^{-1}(\mathbf{w})$. Then by definition $A(\mathbf{x}) = \mathbf{z}$ and $A(\mathbf{y}) = \mathbf{w}$. It therefore follows from the linearity of A that $A(\mathbf{x}+\mathbf{y}) = \mathbf{z}+\mathbf{w}$. Again, by definition, this means that $A^{-1}(\mathbf{z}+\mathbf{w}) = \mathbf{x}+\mathbf{y} = A^{-1}(\mathbf{z}) + A^{-1}(\mathbf{w})$, so that A^{-1} preserves vector addition. Similarly, $A(c\mathbf{x}) = cA(\mathbf{x}) = c\mathbf{z}$, so that $A^{-1}(c\mathbf{z}) = c\mathbf{x} = cA^{-1}(\mathbf{z})$, and hence A^{-1} also preserves scalar multiplication.

Exercise 9.3 Assume $A \in L(X,Y)$ and $A\mathbf{x} = \mathbf{0}$ only when $\mathbf{x} = \mathbf{0}$. Prove that A is then 1-1.

Solution. Suppose $A(\mathbf{x}) = A(\mathbf{y})$. It then follows that $A(\mathbf{x}-\mathbf{y}) = A(\mathbf{x}) - A(\mathbf{y}) = \mathbf{0}$. Hence by assumption $\mathbf{x}-\mathbf{y} = \mathbf{0}$, and so $\mathbf{x} = \mathbf{y}$; therefore A is one-to-one.

Exercise 9.4 Prove (as asserted in Sec. 9.30) that null spaces and ranges of linear transformations are vector spaces.

Solution. Let N be the null space of the linear transformation $A : X \to Y$, let $\mathbf{x}$ and $\mathbf{y}$ be elements of N, and let c be any scalar. By definition $A(\mathbf{x}) = \mathbf{0} = A(\mathbf{y})$, and $A(\mathbf{x}+\mathbf{y}) = A(\mathbf{x}) + A(\mathbf{y}) = \mathbf{0}+\mathbf{0} = \mathbf{0}$, so that, by definition, $\mathbf{x}+\mathbf{y} \in N$. Likewise $A(c\mathbf{x}) = cA(\mathbf{x}) = c\mathbf{0} = \mathbf{0}$, and so $c\mathbf{x} \in N$. Therefore N is a subspace of X.

Let R be the range of A, let $\mathbf{z}$ and $\mathbf{w}$ be any elements of R, and let c be any scalar. By definition, there exist vectors $\mathbf{x} \in X$ and $\mathbf{y} \in X$ such that $z = A(\mathbf{x})$ and $w = A(\mathbf{y})$. Then $A(\mathbf{x}+\mathbf{y}) = A(\mathbf{x}) + A(\mathbf{y}) = \mathbf{z}+\mathbf{w}$, and hence $\mathbf{z}+\mathbf{w} \in R$. Likewise $A(c\mathbf{x}) = cA(\mathbf{x}) = c\mathbf{z}$, so that $c\mathbf{z} \in R$. Therefore R is a subspace of Y.

Exercise 9.5 Prove that to every $A \in L(R^n, R^1)$ corresponds a unique $\mathbf{y} \in R^n$ such that $A\mathbf{x} = \mathbf{x}\cdot\mathbf{y}$. Prove also that $\|A\| = |\mathbf{y}|$.

Hint: Under certain conditions, equality holds in the Schwarz inequality.

Solution. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis of R^n, and let $\mathbf{y} = A(\mathbf{e}_1)\mathbf{e}_1 + \cdots + A(\mathbf{e}_n)\mathbf{e}_n$. Then for any $\mathbf{x} = c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n$ we have

$$\begin{aligned} A(\mathbf{x}) &= c_1A(\mathbf{e}_1) + \cdots + c_nA(\mathbf{e}_n) \\ &= \mathbf{y}\cdot\mathbf{x}. \end{aligned}$$

There can be at most one such $\mathbf{y}$, since if $A(\mathbf{x}) = \mathbf{z} \cdot \mathbf{x}$, then $|\mathbf{y} - \mathbf{z}|^2 = \mathbf{y} \cdot \mathbf{y} - \mathbf{y} \cdot \mathbf{z} - \mathbf{z} \cdot \mathbf{y} + \mathbf{z} \cdot \mathbf{z} = A(\mathbf{y}) - A(\mathbf{y}) - A(\mathbf{z}) + A(\mathbf{z}) = 0$.

By the Schwarz inequality we have

$$|A(\mathbf{x})| = |\mathbf{y} \cdot \mathbf{x}| \leq |\mathbf{y}|\,|\mathbf{x}|$$

for all $\mathbf{x}$, so that $\|A\| \leq |\mathbf{y}|$. On the other hand $A(\mathbf{y}) = \mathbf{y} \cdot \mathbf{y} = |\mathbf{y}|^2$, so that $\|A\| \geq |\mathbf{y}|$.

Exercise 9.6 If $f(0,0) = 0$ and

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

prove that $(D_1 f)(x,y)$ and $(D_2 f)(x,y)$ exist at every point of R^2, although f is not continuous at $(0,0)$.

Solution. At any point (x,y) except $(0,0)$ the differentiability of $f(x,y)$ follows from the rules for differentiation and the principles of Chapter 5. At $(0,0)$ it is a routine computation to verify that both partial derivatives equal zero:

$$(D_1 f)(0,0) = \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} = 0.$$

However, $f(x,y)$ is not continuous at $(0,0$, since $f(x,x) = \frac{1}{2}$ for all $x \neq 0$, and hence $\lim\limits_{x \to 0} f(x,x) = \dfrac{1}{2} \neq f(0,0)$.

Exercise 9.7 Suppose that f is a real-valued function defined in an open set $E \subset R^n$, and that the partial derivatives $D_1 f, \ldots, D_n f$ are bounded in E. Prove that f is continuous in E

Hint: Proceed as in the proof of Theorem 9.21.

Solution. Let $\varepsilon > 0$ be given, and let $\mathbf{x}^0 = (x_1^0, \ldots, x_n^0)$ be any point of E. First choose $\delta_0 > 0$ so that $\mathbf{y} \in E$ if $|\mathbf{y} - \mathbf{x}^0| < 2\delta_0$. Then, if $M = \max\limits_{\mathbf{x} \in E} \left((D_1 f)(\mathbf{x}), \ldots, (D_n f)(\mathbf{x})\right)$, choose $\delta = \min\left(\delta_0, \dfrac{\varepsilon}{(n+1)M}\right)$. It then follows that if $|\mathbf{y} - \mathbf{x}^0| < \delta$, we have

$$\begin{aligned}
|f(\mathbf{y}) - f(\mathbf{x}^0)| &= |f(y_1, \ldots, y_n) - f(x_1^0, \ldots, x_n^0)| \\
&\leq |f(y_1, y_2, \ldots, y_n) - f(x_1^0, y_2, \ldots, y_n)| + \\
&\quad + |f(x_1^0, y_2), \ldots, y_n) - f(x_1^0, x_2^0, \ldots, y_n)| + \cdots \\
&\quad \cdots + |f(x_1^0, x_2^0, \ldots, x_{n-1}^0, y_n) - f(x_1^0, x_2^0, \ldots, x_{n-1}^0, x_n^0)|,
\end{aligned}$$

where the ellipsis indicates terms of the form

$$|f(x_1^0, x_2^0, \ldots, x_{k-1}^0, y_k, y_{k+1}, \ldots, y_n) - f(x_1^0, x_2^0, \ldots, x_{k-1}^0, x_k^0, y_{k+1}, \ldots, y_n)|.$$

By the mean-value theorem there is a number c_k between x_k^0 and y_k such that this last difference equals

$$|(D_k f)(x_1^0, x_2^0, \ldots, x_{k-1}^0, c_k, y_{k+1}, \ldots, y_n)(y_k - x_k^0)|,$$

which is at most $M\delta$. Since by definition $M\delta$ is at most $\dfrac{\varepsilon}{n+1}$, and there are only n such terms, it follows that $|f(\mathbf{x}^0) - f(\mathbf{y})| < \varepsilon$. Thus f is continuous.

Remark: We have actually shown that $f(\mathbf{x})$ satisfies a Lipschitz condition on any convex subset of E, i.e., that $|f(\mathbf{x}) - f(\mathbf{y})| \le nM|\mathbf{x} - \mathbf{y}|$ on each convex subset.

Exercise 9.8 Suppose that f is a differentiable real function in an open set $E \subset R^n$, and that f has a local maximum at a point $\mathbf{x} \in E$. Prove that $f'(\mathbf{x}) = 0$.

Solution. Let $\mathbf{y}$ be any element of R^n, and consider the real-valued function $\varphi(t) = f(\mathbf{x}+t\mathbf{y})$, defined near $t = 0$. This function is differentiable (by Theorem 9.15 $\varphi(t) = f'(\mathbf{x} + t\mathbf{y})(\mathbf{y})$). Since $\varphi(t)$ has a maximum at $t = 0$, it follows that $\varphi'(0) = 0$, i.e., that $f'(\mathbf{x})(\mathbf{y}) = 0$. Since $\mathbf{y}$ is arbitrary, it follows by definition of the zero linear transformation that $f'(\mathbf{x})$ is the zero linear transformation.

Exercise 9.9 If $\mathbf{f}$ is a differentiable mapping of a *connected* open set $E \subset R^n$ into R^m, and if $\mathbf{f}'(\mathbf{x}) = \mathbf{0}$ for for every $\mathbf{x} \in E$, prove that $\mathbf{f}$ is constant in E.

Solution. The mean-value argument given in Exercise 7 above, applied to each component of $\mathbf{f}$, shows that $\mathbf{f}$ is *locally* constant (the partial derivatives are all zero). Hence, if $\mathbf{x}^0$ is any point of E, the set of $\mathbf{x}$ such that $\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}^0)$ is an open set. Since this set is also closed in E, and E is connected, it follows that it must be all of E.

Exercise 9.10 If f is a real function defined in a convex open set $E \subset R^n$, such that $(D_1 f)(\mathbf{x}) = 0$ for every $\mathbf{x} \in E$, prove that $f(\mathbf{x})$ depends only on $x_2, \ldots, x_n$.

Show that the convexity of E can be replaced by a weaker condition, but that some condition is required. For example, if $n = 2$ and E is shaped like a horseshoe, the statement may be false.

Solution. We need to show that $f(x_1^0, x_2, \ldots, x_n) = f(x_1^1, x_2, \ldots, x_n)$ whenever $\mathbf{x}^0 = (x_1^0, x_2, \ldots, x_n)$ and $\mathbf{x}^1 = (x_1^1, x_2, \ldots, x_n)$ both belong to E. Since E is convex, the line segment joining $\mathbf{x}^0$ and $\mathbf{x}^1$ is contained in E. The mean-value theorem applies on this line segment, showing that $f(\mathbf{x}^0) - f(\mathbf{x}^1) = (x_1^0 - x_1^1)(D_1 f)(\mathbf{x})$ for some point $\mathbf{x}$ on this interval. Hence the result now follows from the hypothesis.

Note that convexity is needed only on each line segment through E parallel to the x_1-axis. Thus if the intersection of E with each line parallel to the x_1-axis is an interval and $(D_1 f)(\mathbf{x}) = 0$ for all $\mathbf{x} \in E$, then f is independent of x_1.

If we define $f(x, y)$ on all of R^2 except the nonnegative portion of the y-axis by specifying

$$f(x,y) = \begin{cases} 0 & \text{if } y < 0 \text{ or } x < 0, \\ y^2 & \text{if } y \geq 0 \text{ and } x > 0, \end{cases}$$

then $f(x, y)$ is continuously differentiable on its domain, $(D_1 f)(x, y) = 0$ everywhere on that domain, yet $f(-1, 1) = 0 \neq 1 = f(1, 1)$, so that f is not independent of x.

Exercise 9.11 If f and g are differentiable real functions in R^n, prove that

$$\nabla(fg) = f\nabla g + g\nabla f$$

and that $\nabla(1/f) = -f^{-2}\nabla f$ wherever $f \neq 0$.
Solution. This is a routine computation applied to the ith component of the various quantities.

Exercise 9.12 Fix two real numbers a and b, $0 < a < b$. Define a mapping $\mathbf{f} = (f_1, f_2, f_3)$ of R^2 into R^3 by

$$\begin{aligned} f_1(s,t) &= (b + a\cos s)\cos t \\ f_2(s,t) &= (b + a\cos s)\sin t \\ f_3(s,t) &= a\sin s \end{aligned}$$

Describe the range K of $\mathbf{f}$. (It is a certain compact subset of R^3.)
(a) Show that there are exactly 4 points $\mathbf{p} \in K$ such that

$$(\nabla f_1)(\mathbf{f}^{-1}(\mathbf{p})) = \mathbf{0}.$$

Find these points.
(b) Determine the set of all $\mathbf{q} \in K$ such that

$$(\nabla f_3)(\mathbf{f}^{-1}(\mathbf{q})) = \mathbf{0}.$$

(c) Show that one of the points $\mathbf{p}$ found in part (a) corresponds to a local maximum of f_1, one corresponds to a local minimum, and that the other two are neither (they are so-called "saddle points").

Which of the points $\mathbf{q}$ found in part (b) correspond to maxima or minima?
(d) Let λ be an irrational number, and define $\mathbf{g}(t) = \mathbf{f}(t, \lambda t)$. Prove that $\mathbf{g}$ is a 1-1 mapping of R^1 onto a dense subset of K. Prove that

$$|\mathbf{g}'(t)|^2 = a^2 + \lambda^2(b + a\cos t)^2.$$

Solution. The range K is a torus obtained by moving a circle of radius a with center on a circle of radius b, always keeping the planes of the two circles perpendicular and each plane passing through the center of the other circle. This can be seen by observing that in cylindrical coordinates the parametric equations say $r = b + a\cos s$, $z = a\sin x$, i.e., $(r-b)^2 + z^2 = a^2$, which, together with the equation $\theta = \text{const}$, gives the equation of a circle with center at $(b, 0)$ and radius a in the half-plane $\theta = \text{const}$.

(a) The equation $(\nabla f_1)(s,t) = \mathbf{0}$ says $-a\sin s\cos t = 0$ and $-(b + a\cos s)\sin t = 0$. This second equation requires $t = k\pi$, and since these functions have period 2π in both s and t, we may as well assume $t = 0$ or $t = \pi$. In that case the first equation implies $s = 0$ or $s = \pi$. Hence the only points $\mathbf{p}$ satisfying this equation are the images of the points $(0,0)$, $(0,\pi)$, $(\pi, 0)$, and (π,π), i.e., the points $(b+a, 0, 0)$, $(b-a, 0, 0)$, $(-b+a, 0, 0)$, and $(-b-a, 0, 0)$.

(b) The equation $(\nabla f_3)(s,t) = \mathbf{0}$ says only that $a\cos s = 0$, i.e., $s = \frac{\pi}{2}$ or $s = \frac{3\pi}{2}$. The image of these two conditions consists of the two circles of radius b about the z-axis in the planes $z = \pm a$.

(c) The point $(a+b, 0, 0)$ is the maximum possible value of $f_1(s,t)$, and occurs only when $\cos s = 1$ and $\cos t = 1$. Likewise the point $(-a-b, 0, 0)$ is the minimum possible value, and occurs only when $\cos s = 1$ and $\cos t = -1$. The other two points, which occur when $s = 0, t = \pi$ and when $s = \pi, t = 0$, lie near points of both larger and smaller values of $f_1(s,t)$. For example, when $s = 0$, the point $t = \pi$ is a minimum for the function $\varphi(t) = f_1(0,t) = b\cos t$; but when $t = \pi$, the point $s = 0$ is a maximum of $\psi(s) = f_1(s,\pi) = -(b + a\cos s)$. Hence the point $(0,\pi)$ is neither a maximum nor a minimum for $f_1(s,t)$.

The points with $z = +a$ are obviously absolute maxima of $f_3(s,t)$, while those with $z = -a$ are the absolute minima.

(d) Suppose $\mathbf{g}(t_1) = \mathbf{g}(t_2)$. Then because $a\sin t_1 = a\sin t_2$, and

$$\sqrt{(f_1(t_1,\lambda t_1))^2 + (f_2(t_1,\lambda t_1))^2} = \sqrt{(f_1(t_2,\lambda t_2))^2 + (f_2(t_2,\lambda t_2))^2}$$

(that is, $b + a\cos t_1 = b + a\cos t_2$), we have $\sin t_1 = \sin t_2$ and $\cos t_1 = \cos t_2$. Therefore $\sin(t_1 - t_2) = 0$, which means $t_2 = t_1 + k\pi$ for some integer k. Because $\sin t_1 = \sin t_2$, it follows that k is an even integer, say $k = 2m$. It then follows, since $f_i(t_1,\lambda t_1) = f_i(t_2,\lambda t_2)$, $\lambda = 1, 2$, that $\cos\lambda t_1 = \cos\lambda t_2$ and $\sin\lambda t_1 = \sin\lambda t_2$. This in turn implies that $\lambda t_2 = \lambda t_2 + 2r\pi$ for some integer r. Combining these two results, we find that $m\lambda = r$. Since λ is irrational, this means that $m = 0 = r$, i.e., $t_2 = t_1$. Thus $\mathbf{g}(t)$ is one-to-one.

To show that the range is dense in K, we need only show that the numbers $2\pi n\lambda$, $n = 0 \pm 1, \pm 2, \ldots$, are dense "modulo 2π," meaning that for any real number θ and any $\varepsilon > 0$ there is are integers m and n such that $|2\pi n\lambda - 2\pi m - \theta| < \varepsilon$. A proposition easily seen to be equivalent is that for any $\eta > 0$ and any real number c there exist integers m and n such that $|n\lambda - m - c| < \eta$. (This statement is obvious ($m = n = 0$) if $c = 0$.) To prove that, fix an integer r larger than $\frac{1}{\eta}$, and consider the numbers $0, \lambda - [\lambda], 2\lambda - [2\lambda], \ldots, r\lambda - [r\lambda]$. There are $r + 1$ such numbers, all lying in the interval $[0, 1)$. Hence two of them must be closer than $\frac{1}{r}$ to each other, say $0 < s\lambda - [s\lambda] - t\lambda + [t\lambda] < \frac{1}{r}$. In particular, the number $(s - t)\lambda$ lies within $\frac{1}{r}$ of an integer (namely $[s\lambda] - [t\lambda]$. Thus we have, say $(s - t)\lambda = k + \delta$, where $0 < \delta < \frac{1}{r}$. Let p be the unique integer such that $p\delta \le c < (p+1)\delta$. We then have $p(s-t)\lambda = pk + p\delta$, and hence, taking $n = p(s-t)$ and $m = pk$, we find $|n\lambda - m - c| = |p\delta - c| < \delta < \frac{1}{r} < \eta$.

This being established, consider any point in K, say the point $\mathbf{p} = (b + a\cos s_0)\cos t_0, (b + a\cos s_0)\sin t_0, a\sin s_0)$, and let $\varepsilon > 0$ be given. According to what was just established, there are integers m, n such that $|2\pi m\lambda - 2\pi n - (t_0 - s_0\lambda)| < \frac{\varepsilon}{3a+3b}$. It then follows that

$$\begin{aligned} |\cos((s_0 + 2\pi m)\lambda) - \cos t_0| &= |\cos((s_0 + 2\pi m)\lambda - 2\pi n) - \cos t_0| \\ &\le \frac{\varepsilon}{3a+3b}, \end{aligned}$$

where we have used the inequality $|\cos u - \cos v| \le |u - v|$, with $u = (s_0 + 2\pi m)\lambda - 2\pi n$ and $v = t_0$. A similar inequality applies with sin in place of cos. It then follows that $|\mathbf{g}(s_0 + 2\pi m) - \mathbf{p}| \le \frac{2\varepsilon}{3} < \varepsilon$. Therefore the range of $\mathbf{g}$ is dense in K.

The equation

$$|\mathbf{g}'(t)|^2 = a^2 + \lambda^2(b + a\cos t)^2$$

is a routine, though tedious, computation.

Exercise 9.13 Suppose $\mathbf{f}$ is a differentiable mapping of R^1 into R^3 such that $|\mathbf{f}(t)| = 1$ for every t. Prove that $\mathbf{f}'(t) \cdot \mathbf{f}(t) = 0$.

Interpret this result geometrically.

Solution. This result is obtained by merely differentiating the relation $\mathbf{f}(t)\cdot\mathbf{f}(t) = 1$. Geometrically it asserts that the velocity vector of a point moving over a sphere is tangent to the sphere (perpendicular to the radius vector from the center of the sphere to the point).

Exercise 9.14 Define $f(0,0) = 0$ and

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

(a) Prove that $D_1 f$ and $D_2 f$ are bounded functions in R^2. (Hence f is continuous.)
(b) Let $\mathbf{u}$ be any unit vector in R^2. Show that the directional derivative $(D_{\mathbf{u}} f)(0,0)$ exists, and that its absolute value is at most 1.
(c) Let γ be a differentiable mapping of R^1 into R^2 (in other words, γ is a differentiable curve in R^2), with $\gamma(0) = (0,0)$ and $|\gamma'(0)| > 0$. Put $g(t) = f(\gamma(t))$ and prove that g is differentiable for every $t \in R^1$.

If $\gamma \in C'$, prove that $g \in C'$.

(d) In spite of this, prove that f is not differentiable at $(0,0)$.
Hint: Formula (40) fails.

Solution. (a) For $(x,y) \neq (0,0)$ we have

$$D_1 f(x,y) = \frac{x^2(x^2+3y^2)}{(x^2+y^2)^2}, \quad D_2 f(x,y) = -\frac{2x^3 y}{(x^2+y^2)^2}.$$

It follows that

$$0 \leq D_1 f(x,y) \leq \frac{3x^2}{x^2+y^2} \leq 3$$

and

$$|D_2 f(x,y)| \leq \frac{x^2}{x^2+y^2} \leq 1.$$

Also $D_1 f(0,0) = \lim_{x\to 0} \dfrac{f(x,0)-f(0,0)}{x} = \lim_{x\to 0} \dfrac{x-0}{x} = 1$, and $D_2 f(0,0) = \lim_{y\to 0} \dfrac{f(0,y)-f(0,0)}{y} = \lim_{y\to 0} \dfrac{0}{y} = 0$. Hence, as asserted, $f(x,y)$ is continuous.

(b) Let $\mathbf{u} = (\cos\theta, \sin\theta)$. Then $D_{\mathbf{u}} f(0,0) = \lim_{t\to 0} \dfrac{f(t\cos\theta, t\sin\theta) - f(0,0)}{t} = \cos^3\theta$.

(c) Suppose $u(t)$ and $v(t)$ satisfy $u(0) = 0 = v(0)$, $u'(t)$ and $v'(t)$ exist for every t, and $u'(t)$ and $v'(t)$ do not both vanish at the same value of t. Setting $g(t) = f(u(t), v(t))$, we find that $g(t)$ is obviously differentiable at any value of t where $u(t)$ and $v(t)$ are not both zero. Now suppose $u(t_0) = v(t_0) = 0$. Then, since one of $u(t)$ and $v(t)$ is one-to-one on a neighborhood of t_0, it follows that, for small non-zero values of $t - t_0$ we have $(u(t))^2 + (v(t))^2 > 0$, and then

$$\begin{aligned} \frac{g(t)-g(t_0)}{t-t_0} &= \frac{f(u(t),v(t)) - f(u(t_0),v(t_0))}{t-t_0} \\ &= \frac{\left(\frac{u(t)-u(t_0)}{t-t_0}\right)^3}{\left(\frac{u(t)-u(t_0)}{t-t_0}\right)^2 + \left(\frac{v(t)-v(t_0)}{t-t_0}\right)^2}, \end{aligned}$$

so that

$$g'(t_0) = \lim_{t\to t_0} \frac{g(t)-g(t_0)}{t-t_0} = \frac{(u'(t_0))^3}{(u'(t_0))^2 + (v'(t_0))^2}.$$

Thus $g(t)$ is differentiable. Observe that if $\gamma(t) \neq (0,0)$, then

$$g'(t) = \frac{(u(t))^4 u'(t) + 3(u(t)v(t))^2 u'(t) - 2(u(t))^3 v(t) v'(t)}{((u(t))^2 + (v(t))^2)^2}.$$

The same argument used above to prove that $g'(t_0)$ exists shows that

$$\lim_{t\to t_0} g'(t) = \frac{(u'(t_0))^5 + (u'(t_0))^3 (v'(t_0))^2}{((u'(t_0))^2 + (v'(t_0))^2)^2} = \frac{(u'(t_0))^3}{(u'(t_0))^2 + (v'(t_0))^2} = g'(t_0),$$

so that g' is continuous at t_0 if u' and v' are. Continuity of g' at other points follows from the chain rule.

If f is differentiable at $(0,0)$, we necessarily have

$$f(x,y) = f(0,0) + \big[xD_1f(0,0) + yD_2f(0,0)\big] + \varepsilon(x,y),$$

where

$$\lim_{(x,y)\to(0,0)} \frac{\varepsilon(x,y)}{\sqrt{x^2+y^2}} = 0.$$

Since $D_1f(0,0) = 1$ and $D_2f(0,0) = 0$, it follows that

$$\varepsilon(x,y) = \frac{-xy^2}{x^2+y^2},$$

and so we must have

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

But this is clearly not the case, as we see by taking $x = y$. (The limit is then $-2^{-3/2}$.)

Exercise 9.15 Define $f(0,0) = 0$, and put

$$f(x,y) = x^2 + y^2 - 2x^2y - \frac{4x^6y^2}{(x^4+y^2)^2}$$

if $(x,y) \neq (0,0)$.
(a) Prove, for all $(x,y) \in R^2$, that

$$4x^4y^2 \leq (x^4+y^2)^2.$$

Conclude that f is continuous.
(b) For $0 \leq \theta \leq 2\pi$, $-\infty < t < \infty$, define

$$g_\theta(t) = f(t\cos\theta, t\sin\theta).$$

Show that $g_\theta(0) = 0$, $g'_\theta(0) = 0$, $g''_\theta(0) = 2$. Each g_θ has therefore a strict local minimum at $t = 0$.

In other words, the restriction of f to each line through $(0,0)$ has a strict local minimum at $(0,0)$.
(c) Show that $(0,0)$ is nevertheless not a local minimum for f, since $f(x,x^2) = -x^4$.

Solution. (a) This inequality follows by squaring the inequality $2x^2|y| \leq x^4+y^2$, which in turn is equivalent to the inequality $(x^2-|y|)^2 \geq 0$. Then, since $f(x,y)$

is obviously continuous except at $(0,0)$, the continuity at the remaining point follows from the inequality

$$|f(x,y) - f(0,0)| \leq 2x^2 + y^2 + 2x^2|y|,$$

which is easily derived from the inequality just proved and the definition of $f(x,y)$.
(*b*) We observe that for $t \neq 0$ we have

$$g_\theta(t) = t^2 - 2t^3 \cos^2\theta \sin\theta - 4t^4 \frac{\cos^6\theta \sin^2\theta}{(t^2\cos^4\theta + \sin^2\theta)^2},$$

from which it is routine computation to show that $g_\theta(0) = 0 = g'_\theta(0)$ and $g''_\theta(0) = 2$.
(*c*) The assertion that $f(x,x^2) = -x^4$ is routine computation. It implies that $f(x,y)$ assumes negative values in any neighborhood of $(0,0)$, and hence that the $f(x,y)$ does not have a local minimum at $(0,0)$.

Exercise 9.16 Show that the continuity of $\mathbf{f}'$ at the point $\mathbf{a}$ is needed in the inverse function theorem, even in the case $n = 1$: If

$$f(t) = t + 2t^2 \sin\left(\frac{1}{t}\right)$$

for $t \neq 0$, and $f(0) = 0$, then $f'(0) = 1$, f' is bounded in $(-1,1)$ but f is not one-to-one in any neighborhood of 0.

Solution. The assertion that $f'(0) = 1$ is proved by direct computation: $\dfrac{f(t)}{t} = 1 + 2t\sin\left(\dfrac{1}{t}\right) \to 1$ as $t \to 0$. Since $f'(t) = 1 + 4t\sin\left(\dfrac{1}{t}\right) - 2\cos\left(\dfrac{1}{t}\right)$ for $t \neq 0$, it follows that $|f'(t)| \leq 7$ for all $t \in (-1,1)$. To show that f is not one-to-one in any neighborhood of 0, we observe that $f'(\frac{1}{k\pi}) = 1 + 2(-1)^k$, so that $f(t)$ is decreasing at $t = \frac{1}{k\pi}$ if k is odd and increasing if k is even. It follows that the minimum value of $f(t)$ on the interval $\left[\frac{1}{(2k+1)\pi}, \frac{1}{2k\pi}\right]$ is assumed at an interior point, so that $f(t)$ cannot be one-to-one on this interval.

Exercise 9.17 Let $\mathbf{f} = (f_1, f_2)$ be the mapping of R^2 into R^2 given by

$$f_1(x,y) = e^x \cos y, \quad f_2(x,y) = e^x \sin y.$$

(*a*) What is the range of $\mathbf{f}$?
(*b*) Show that the Jacobian of $\mathbf{f}$ is not zero at any point of R^2. Thus every point of R^2 has a neighborhood in which $\mathbf{f}$ is one-to-one. Nevertheless, $\mathbf{f}$ is not one-to-one on R^2.
(*c*) Put $\mathbf{a} = (0, \pi/3)$, $\mathbf{b} = f(\mathbf{a})$, let $\mathbf{g}$ be the continuous inverse of $\mathbf{f}$, defined in a neighborhood of $\mathbf{b}$, such that $\mathbf{g}(\mathbf{b}) = \mathbf{a}$. Find an explicit formula for $\mathbf{g}$, compute $\mathbf{f}'(\mathbf{a})$ and $\mathbf{g}'(\mathbf{b})$, and verify the formula (52).

(*d*) What are the images under $\mathbf{f}$ of lines parallel to the coordinate axes?

Solution. (*a*) The range of $\mathbf{f}$ is all of R^2 except the point $(0,0)$. Indeed if $(u,v) \neq (0,0)$, choose y so that

$$\cos y = \frac{u}{\sqrt{u^2+v^2}}, \quad \sin y = \frac{v}{\sqrt{u^2+v^2}},$$

and let $x = \ln\sqrt{u^2+v^2}$, so that $e^x = \sqrt{u^2+v^2}$. It is then obvious from the equations defining y and x that $u = e^x \cos y$ and $v = e^x \sin y$. Hence every point except $(0,0)$ is in the range of $\mathbf{f}$. The point $(0,0)$ is not in the range, since $u^2+v^2 = e^{2x} > 0$ for any point $(u,v) = \mathbf{f}(x,y)$.

(*b*) The Jacobian of $\mathbf{f}(x,y)$ is e^{2x}, which is never zero. However, since $\mathbf{f}(x,y+2\pi) = \mathbf{f}(x,y)$, it follows that $\mathbf{f}$ is not one-to-one.

(*c*) By our definition $\mathbf{b} = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$. We can therefore take $y = \arctan\left(\frac{v}{u}\right)$ for (u,v) near $\mathbf{b}$, the arctangent being between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. Thus we have $\mathbf{g}(u,v) = \left(\ln\sqrt{u^2+v^2}, \arctan\left(\frac{v}{u}\right)\right)$. We then have

$$\mathbf{f}'(x,y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}, \quad \mathbf{g}'(u,v) = \begin{pmatrix} \frac{u}{u^2+v^2} & \frac{v}{u^2+v^2} \\ \frac{-v}{u^2+v^2} & \frac{u}{u^2+v^2} \end{pmatrix}.$$

When we take $u = e^x \cos y$ and $v = e^x \sin y$, we find that

$$\mathbf{g}'(\mathbf{f}(x,y)) = \begin{pmatrix} e^{-x}\cos y & e^{-x}\sin y \\ -e^{-x}\sin y & e^{-x}\cos y \end{pmatrix}.$$

It is then a routine computation to verify that $\mathbf{g}'(\mathbf{f}(x,y))\mathbf{f}'(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Likewise we find

$$\mathbf{f}'(\mathbf{g}(u,v)) = \begin{pmatrix} u & -v \\ v & u \end{pmatrix},$$

and a routine computation shows that $\mathbf{f}'(\mathbf{g}(u,v))\mathbf{g}'(u,v) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

(*d*) The family of lines $x = c$ maps to the family of concentric circles $u^2+v^2 = e^{2c}$. The lines $y = c$ map to half-lines $v = Ku$, $u \geq 0$, where $K = \tan y$. (If y is an odd multiple of $\frac{\pi}{2}$, the half-line is either the positive or negative u-axis.

Exercise 9.18 Answer analogous questions for the mapping defined by

$$u = x^2 - y^2, \quad v = 2xy.$$

Solution. (*a*) the range of the mapping $\mathbf{f}(x,y) = (x^2-y^2, 2xy)$ is the entire plane R^2. Indeed, every point (u,v) except $(0,0)$ has two distinct preimages, one of which is

$$x = \sqrt{\frac{\sqrt{u^2+v^2}+u}{2}}, \quad y = (\operatorname{sgn} v)\sqrt{\frac{\sqrt{u^2+v^2}-u}{2}}.$$

(The other preimage is $-x$, $-y$, with this x and this y.)
(b) The Jacobian of $\mathbf{f}$ vanishes only at $x = y = 0$. Indeed,

$$\mathbf{f}'(x,y) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}.$$

Hence the Jacobian is $4(x^2+y^2)$.
(c) Taking $\mathbf{a} = (3,4)$, so that $\mathbf{b} = (-7,24)$, we can take, locally

$$g(u,v) = \left(\sqrt{\frac{\sqrt{u^2+v^2}+u}{2}}, \sqrt{\frac{\sqrt{u^2+v^2}-u}{2}}\right).$$

We then have

$$\mathbf{g}'(u,v) = \begin{pmatrix} \frac{1}{4}\sqrt{\frac{2}{\sqrt{u^2+v^2}+u}}\left(1+\frac{u}{\sqrt{u^2+v^2}}\right) & \frac{1}{4}\sqrt{\frac{2}{\sqrt{u^2+v^2}+u}}\left(\frac{v}{\sqrt{u^2+v^2}}\right) \\ \frac{1}{4}\sqrt{\frac{2}{\sqrt{u^2+v^2}-u}}\left(-1+\frac{u}{\sqrt{u^2+v^2}}\right) & \frac{1}{4}\sqrt{\frac{2}{\sqrt{u^2+v^2}-u}}\left(\frac{v}{\sqrt{u^2+v^2}}\right) \end{pmatrix}$$

Noting that the defining relations imply $u^2+v^2 = (x^2+y^2)^2$, we see that

$$\mathbf{g}'(\mathbf{f}(x,y)) = \begin{pmatrix} \frac{x}{2(x^2+y^2)} & \frac{y}{2(x^2+y^2)} \\ \frac{-y}{2(x^2+y^2)} & \frac{x}{2(x^2+y^2)} \end{pmatrix},$$

from which we see easily that $\mathbf{g}'(\mathbf{f}(x,y))\mathbf{f}'(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The corresponding equality with $\mathbf{g}$ and $\mathbf{f}$ interchanged is likewise simple, though more cumbersome to write out.

Exercise 9.19 Show that the system of equations

$$\begin{aligned} 3x+y-z+u^2 &= 0 \\ x-y+2z+u &= 0 \\ 2x+2y-3z+2u &= 0 \end{aligned}$$

can be solved for x, y, u in terms of z; for x, z, u in terms of y; for y, z, u in terms of x; but not for x, y, z in terms of u.

Solution. Adding the last two equations and subtracting the first yields $3u - u^2 = 0$, whence either $u = 0$ or $u = 3$. Hence unless u has one of these two values, there are no solutions at all. Therefore the system cannot generally be solved for x, y, z in terms of u. If one of these two equations holds, we can solve just the last two equations for any two of the variables x, y, z in terms of the third. The remaining equation will then automatically be satisfied. For example,

$$x = -\frac{z}{4}, \quad y = \frac{7z}{4}, \quad u = 0; \quad x = -\frac{9+z}{4}, \quad y = \frac{3+7z}{4}, \quad u = 3.$$

We could also have

$$x = -\frac{y}{7}, \quad z = \frac{4y}{7}, \quad u = 0; \quad x = \frac{60+4y}{7}, \quad z = \frac{4y-3}{7}, \quad u = 3.$$

Finally, we could also have

$$y = -7x, \quad z = -4x, \quad u = 0; \quad y = \frac{7x-60}{4}, \quad z = 9-4x, \quad u = 3.$$

Note that the matrix of the derivative of the transformation $\mathbf{f}(x, y, z, u) = (3x + y - z + u^2, x - y + 2z + u, 2x + 2y - 3z + 2u)$ is

$$\mathbf{f}'(x, y, z, u) = \begin{pmatrix} 3 & 1 & -1 & 2u \\ 1 & -1 & 2 & 1 \\ 2 & 2 & -3 & 2 \end{pmatrix}$$

and any 3×3 submatrix containing the last column is invertible when $u = 0$ or $u = 3$. However, the first three columns of this matrix does not form an invertible matrix.

Exercise 9.20 Take $n = m = 1$ in the implicit function theorem, and interpret the theorem (as well as its proof) graphically.

Solution. The theorem asserts that if $f(x, y)$ is continuously differentiable in a neighborhood of (x_0, y_0), $f(x_0, y_0) = 0$, and $D_2 f(x_0, y_0) \neq 0$, then there exist 1) an interval $I = (x_0 - \delta, x_0 + \delta)$, 2) an interval $J = (y_0 - \eta, y_0 + \eta)$, and 3) a continuously differentiable function $\varphi : I \to J$ such that for all $(x, y) \in I \times J$ the equation $f(x, y) = 0$ holds if and only if $y = \varphi(x)$.

The proof amounts to the argument that, since $D_2 f(x_0, y_0) \neq 0$ and f is continuously differentiable, it must be that $D_2 f(x, y) \neq 0$ for all (x, y) near (x_0, y_0). Hence the function $g(y) = f(x_0, y)$ is strictly monotonic near $y = y_0$. Therefore, since $g(y_0) = 0$, there is a small interval $[y_0 - \eta, y_0 + \eta]$ such that $g(y_0 - \eta)$ and $g(y_0 + \eta)$ have opposite signs. By the continuity of $f(x, y)$, it follows that $f(x, y_0 - \eta)$ has the same sign as $f(x_0, y_0 - \eta)$ if x is near x_0, and similarly $f(x, y_0 + \eta)$ has the same sign as $f(x_0, y_0 + \eta)$ for x near x_0. That is, $f(x, y_0 - \eta)$ and $f(x, y_0 + \eta)$ have opposite signs if x is near x_0. It follows that there is a point $\varphi(x) \in (y_0 - \eta, y_0 + \eta)$ such that $f(x, \varphi(x)) = 0$. By restricting the neighborhood so that $D_2 f(x, y)$ is of constant sign, we assure that $g_x(y) = f(x, y)$ is monotonic on $[y_0 - \eta, y_0 + \eta]$ for each x near x_0. It then follows that there can be at most one value of y in $(y_0 - \eta, y_0 + \eta)$ satisfying the equation $f(x, y) = 0$. That is, the function $\varphi(x)$ is unique. This proves all but the differentiability of φ.

The graphical interpretation is that, near a point on a smooth curve $f(x, y) = 0$ where the tangent is not vertical ($D_2 f(x_0, y_0) \neq 0$) the curve intersects each vertical line exactly once.

Exercise 9.21 Define f in R^2 by

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

(*a*) Find the four points in R^2 at which the gradient of f is zero. Show that f has exactly one local maximum and one local minimum in R^2.
(*b*) Let X be the set of all $(x,y) \in R^2$ at which $f(x,y) = 0$. Find those points of S that have no neighborhoods in which the equation $f(x,y) = 0$ can be solved for y in terms of x (or for x in terms of y). Describe S as precisely as you can.

Solution. (*a*) We have $\nabla f(x,y) = 6(x^2 - x)\mathbf{i} + 6(y^2 + y)\mathbf{j}$. Hence $\nabla f(x,y) = \mathbf{0}$ precisely at the four points $(0,0)$, $(1,0)$, $(0,-1)$, $(1,-1)$. Since the Hessian matrix of f is

$$\begin{pmatrix} 12x - 2 & 0 \\ 0 & 12y + 2 \end{pmatrix}$$

this matrix has a positive determinant when $x > \frac{1}{6}$ and $y > \frac{-1}{6}$ or when $x < \frac{1}{6}$ and $y < \frac{-1}{6}$. Thus $(1,0)$ and $(0,-1)$ are possible extrema. Since $12x - 2 > 0$ at $(1,0)$, that point is a minimum. Likewise $(0,-1)$ is a maximum.
(*b*) Since $f(x,y) = (x+y)[2x^2 - 2xy + 2y^2 - 3x + 3y]$, the equation $f(x,y) = 0$ has the real solution $y = -x$ for every real value of x. In addition, if $-\frac{1}{2} \le x \le \frac{3}{2}$, it has the real solutions

$$y = \frac{2x - 3 + \sqrt{9 + 12x - 12x^2}}{4}, \quad y = \frac{2x - 3 - \sqrt{9 + 12x - 12x^2}}{4}.$$

According to the implicit function theorem, the only possible points near which there might not be a unique solution are for y in terms of x are those where $y = 0$ or $y = -1$. The corresponding values of x are $x = 0$ and $x = \frac{3}{2}$ for $y = 0$ and $x = 1$ and $x = -\frac{1}{2}$ for $y = -1$.

We observe that both solutions $y = -x$ and $y = \dfrac{2x - 3 + \sqrt{9 + 12x - 12x^2}}{4}$ tend to 0 as $x \to 0$. Hence there is no unique solution for y near $(0,0)$. As $x \uparrow \frac{3}{2}$, the quantity under the radical sign tends to zero, and hence these two solutions converge toward the common value $y = 0$. Hence the point $\left(\frac{3}{2}, 0\right)$, is another point around which the solution for y is not unique. The two radicals also tend to zero as $x \downarrow -\frac{1}{2}$, causing the two values of y both to tend toward -1, so that $\left(-\frac{1}{2}, -1\right)$ is not a point of unique solvability. Finally, as $x \to 1$, the three y values tend toward -1, $\frac{1}{2}$, and -1. Since two of these values are identical, there is no unique solution around the point $(1,-1)$.

Finally, the three x-values corresponding to any y are

$$x = -y, \; x = \frac{2y + 3 \pm \sqrt{9 - 12y - 12y^2}}{4},$$

where the quantity under the radical is nonnegative in the range $-\frac{3}{2} \le y \le \frac{1}{2}$. The values where $D_1 f(x,y) = 0$ are $x = 0$ and $x = 1$, and the four points near which a solution for x might not be unique are $(0,0)$, $\left(0, -\frac{3}{2}\right)$, $(1,-1)$, and

$(1, \frac{1}{2})$. As y tends to zero, two of these tend to zero. Hence $(0,0)$ is not a point of unique solvability for x in terms of y. As y tends to -1, two of the x-values tend to 1, so that $(1,-1)$ is not a point of unique solvability for x. Finally, as y tends to $-\frac{3}{2}$ or $\frac{1}{2}$, the radical disappears, and so once again two of the x values tend to the same value, namely 1 as $y \to \frac{1}{2}$ and 0 as $y \to -\frac{3}{2}$. Thus these four points are not points of unique solvability for x.

In sum, the points near which the equation $f(x,y) = 0$ does not define either y as a function of x or x as a function of y are $(0,0)$ and $(1,-1)$.

Exercise 9.22 Give a similar discussion for

$$f(x,y) = 2x^3 + 6xy^2 - 3x^2 + 3y^2.$$

Solution. The gradient is

$$\nabla 6(x^2 + y^2 - x)\mathbf{i} + 6(2xy + y)\mathbf{j}$$

As we see from solving the appropriate equations, this gradient vanishes at the points $(0,0)$ and $(1,0)$. The point $(0,0)$ is a saddle point, since $f(x,0)$ is negative for $x < 0$ and $f(0,y)$ is positive for y near zero. The Hessian determinant is positive at $(1,0)$, and the upper left-hand entry is also; hence $(1,0)$ is a minimum.

Because the equation $f(x,y) = 0$ can be written as

$$(6x + 3)y^2 = (3 - 2x)x^2,$$

there will be real solutions y if and only if $-\frac{1}{2} < x \le \frac{3}{2}$. (When $x = -\frac{1}{2}$, the equation does not contain y.) In this range there are two distinct values of y except for $x = 0$ and $x = \frac{3}{2}$. Hence the two points on the locus of $f(x,y) = 0$ at which the equation cannot be solved for y are $(0,0)$ and $(\frac{3}{2}, 0)$.

Since the equation is cubic in x, its solvability is more complicated from this point of view. Every value of y gives at least one value of x (but those x-values always lie between $-\frac{1}{2}$ and $\frac{3}{2}$). To find the points where two of the three (complex) x-roots coincide, we observe that at such points $D_1 f(x,y) = 0$, and hence also $3f(x,y) - xD_1 f(x,y) = 0$. This last equation says $x^2 - 4xy^2 + 3y^2 = 0$, i.e., $y^2 = \dfrac{x^2}{4x+3}$. Substituting this value of y^2 into $f(x,y) = 0$, we get either $x = 0$ and $y = 0$ or

$$x^2 = \frac{3}{4}.$$

Since we have to have $-\frac{1}{2} < x$, we must have $x = \frac{\sqrt{3}}{2}$, and this gives $y^2 = \frac{2\sqrt{3}-3}{4}$. Hence the points near which $f(x,y) = 0$ cannot be solved uniquely for x are $(0,0)$ and $(\frac{\sqrt{3}}{2}, \pm\frac{\sqrt{2\sqrt{3}-3}}{4})$.

Exercise 9.23 Define f in R^3 by

$$f(x, y_1, y_2) = x^2 y_1 + e^x + y_2.$$

Show that $f(0,1,-1) = 0$, $(D_1 f)(0,1,-1) \neq 0$, and that there exists therefore a differentiable function g in some neighborhood of $(1,-1)$ in R^2 such that $g(1,-1) = 0$ and

$$f(g(y_1, y_2), y_1, y_2) = 0.$$

Find $(D_1 g)(1,-1)$ and $(D_2 g)(1,-1)$.

Solution. The proof that $f(0,1,-1) = 0$ is a routine computation. We have $(D_1 f)(x, y_1, y_2) = 2xy_1 + e^x$, so that $(D_1 f)(0,1,-1) = 1 \neq 0$. To find the partial derivatives of g we use the chain rule. Let $\psi(y_1, y_2) = f(g(y_1, y_2), y_1, y_2) \equiv 0$. Then

$$0 = D_1\psi(y_1, y_2) = D_1 f(g(y_1, y_2), y_1, y_2) D_1 g(y_1, y_2) + D_2 f(g(y_1, y_2), y_1, y_2),$$

so that

$$0 = \left(2y_1 g(y_1, y_2) + e^{g(y_1, y_2)}\right) D_1 g(y_1, y_2) + \left(g(y_1, y_2)\right)^2.$$

Similarly, setting

$$0 = D_2\psi(y_1, y_2) = D_1 f(g(y_1, y_2), y_1, y_2) D_2 g(y_1, y_2) + D_3 f(g(y_1, y_2), y_1, y_2),$$

we find

$$0 = \left(2y_1 g(y_1, y_2) + e^{g(y_1, y_2)}\right) D_2 g(y_1, y_2) + 1.$$

Taking $y_1 = 1$, $y_2 = -1$, $g(y_1, y_2) = 0$, we get

$$D_1 g(1,-1) = 0, \quad D_2 g(1,-1) = -1.$$

Exercise 9.24 For $(x, y) \neq (0,0)$, define $\mathbf{f} = (f_1, f_2)$ by

$$f_1(x, y) = \frac{x^2 - y^2}{x^2 + y^2}, \quad f_2(x, y) = \frac{xy}{x^2 + y^2}.$$

Compute the rank of $\mathbf{f}'(x, y)$, and find the range of $\mathbf{f}$.

Solution. The matrix of $\mathbf{f}'(x, y)$ is

$$\begin{pmatrix} \dfrac{4xy^2}{(x^2+y^2)^2} & \dfrac{-4x^2 y}{(x^2+y^2)^2} \\ \dfrac{y(y^2-x^2)}{(x^2+y^2)^2} & \dfrac{x(x^2-y^2)}{(x^2+y^2)^2} \end{pmatrix}.$$

Its determinant is 0 at every point. Hence its rank is either 0 or 1 at every point. Since the point $(0,0)$ is excluded from the domain, the rank is 1 at every point. The range must therefore be 1-dimensional, i.e., there is some non-trivial relation connecting f_1 and f_2. Indeed, it is easy to verify that if $u = f_1(x,y)$ and $v = f_2(x,y)$, then

$$u^2 + 4v^2 = 1.$$

Thus the range of $\mathbf{f}$ is a subset of this ellipse. In fact, it is all of this ellipse. The point $(1,0)$ is its own image, and the point $(-1,0)$ is the image of $(0,1)$. For any other point (u,v) on this ellipse we have $-1 < u < 1$ and $v = \pm\frac{1}{2}\sqrt{1-u^2}$. The point (u,v) is the image of the point $\left(1, \pm\sqrt{\frac{1-u}{1+u}}\right)$ (and, of course, many other points as well).

Exercise 9.25 Suppose $A \in L(R^n, R^m)$, let r be the rank of A.
(a) Define S as in the proof of Theorem 9.32. Show that SA is a projection in R^n whose nullspace is $\mathcal{N}(A)$ and whose range is $\mathcal{R}(S)$. *Hint:* By (68), $SASA = SA$.
(b) Use (a) to show that

$$\dim \mathcal{N}(A) + \dim \mathcal{R}(A) = n.$$

Solution. We recall that S is defined by first choosing a basis for the range of A, say $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$, then choosing vectors $\{\mathbf{z}_1, \ldots, \mathbf{z}_r\}$ such that $A\mathbf{z}_i = \mathbf{y}_i$ for $i = 1, 2, \ldots, r$. We then define $S\mathbf{y}_i = \mathbf{z}_i$ on the vectors $\mathbf{y}_i$ (and S arbitrary on any set of vectors $\mathbf{y}_{r+1}, \ldots, \mathbf{y}_m$ that can be adjoined to $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$ so as to make a basis of R^m). Thus S is a left inverse of the restriction of A to the subspace spanned by $\mathbf{z}_1, \ldots, \mathbf{z}_r$. Since $A\mathbf{x}$ belongs to the range of A, it follows, as in (68), that $ASA\mathbf{x} = A\mathbf{x}$, from which we conclude that $SASA\mathbf{x} = SA\mathbf{x}$, i.e., SA is a projection. Then every vector $\mathbf{x}$ has the unique decomposition $\mathbf{x} = SA\mathbf{x} + (\mathbf{x} - SA\mathbf{x})$, where the first vector on the right belongs to the range of SA and the second to the nullspace of this projection. The two subspaces have only the zero vector in common. Since S is an isomorphism of the range of A, the range of SA has the same dimension as the range of A. Since $A = ASA$, the nullspace of SA is the same as the nullspace of A. Thus $n = \dim \mathcal{N}(SA) + \dim \mathcal{R}(SA) = \dim \mathcal{N}(A) + \dim \mathcal{R}(A)$.

Exercise 9.26 Show that the existence (and even the continuity) of $D_{12}f$ does not imply the existence of $D_1 f$. For example, let $f(x,y) = g(x)$, where g is nowhere differentiable.

Solution. The second sentence in the exercise *is* its solution. Since $D_2 f$ is identically zero, $D_{12}f$ is also identically zero, hence certainly continuous.

Exercise 9.27 Put $f(0,0) = 0$, and

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

if $(x,y) \neq (0,0)$. Prove that
(a) f, $D_1 f$, and $D_2 f$ are continuous in R^2;
(b) $D_{12} f$ and $D_{21} f$ exist at every point of R^2, and are continuous except at $(0,0)$;
(c) $(D_{12} f)(0,0) = 1$, and $(D_{21} f)(0,0) = -1$.

Solution. (a) The continuity of f is obvious at every point except $(0,0)$; at $(0,0)$ it follows from the inequality $|f(x,y)| \le \frac{1}{2}(x^2+y^2)$. It is also clear that $D_1 f(0,0) = 0 = D_2 f(0,0)$. For $(x,y) \neq (0,0)$ we have $D_1 f(x,y) = \dfrac{x^4 y + 4x^2 y^3 - y^5}{(x^2+y^2)^2}$ and $D_2 f(x,y) = \dfrac{x^5 - 4x^3 y^2 - xy^4}{(x^2+y^2)^2}$. The continuity of the partial derivatives at every point except $(0,0)$ is obvious. It is easy to see that these derivatives satisfy the inequalities $|D_1 f(x,y)| \le 2|y|$ and $|D_2 f(x,y)| \le 2|x|$, so that $D_1 f$ and $D_2 f$ are also continuous at $(0,0)$.
(b) Since $f(x,y)$ is a rational function with non-zero denominator for $(x,y) \neq (0,0)$, it has continuous partial derivatives of all orders on this set.

(c) Since $D_1 f(0,y) = -y$ and $D_2 f(x,0) = x$, it follows that $D_{21} f(0,y) = -1$ for all y and $D_{12} f(x,0) = 1$ for all x.

Exercise 9.28 For $t \ge 0$ put

$$\varphi(x,t) = \begin{cases} x & (0 \le x \le \sqrt{t}) \\ -x + 2\sqrt{t} & (\sqrt{t} \le x \le 2\sqrt{t}) \\ 0 & (\text{otherwise}), \end{cases}$$

and put $\varphi(x,t) = -\varphi(x,|t|)$ if $t < 0$.

Show that φ is continuous on R^2, and

$$(D_2\varphi)(x,0) = 0$$

for all x. Define

$$f(t) = \int_{-1}^{1} \varphi(x,t)\,dx.$$

Show that $f(t) = t$ if $|t| < \frac{1}{4}$. Hence

$$f'(0) \neq \int_{-1}^{1} (D_2\varphi)(x,0)\,dx.$$

Solution. This function is zero in the (closed) left half-plane of the xt-plane and on the positive x-axis. Since the functions by which it is defined are continuous, we need only verify that they agree on the boundary curves $x = \sqrt{t}$ and $x = 2\sqrt{t}$ in the first quadrant that separate the three different regions of definition. This is a routine computation.

Likewise the computation showing that $(D_2\varphi)(x,0) = 0$ is routine, since for each $x > 0$ $\varphi(x,t) = 0$ for $0 \le t \le \frac{1}{4}x^2$, while $\varphi(x,t) = 0$ for all t if $x \le 0$.

If $0 < t < \frac{1}{4}$, then

$$\begin{aligned} f(t) &= \int_0^{\sqrt{t}} x\,dx + \int_{\sqrt{t}}^{2\sqrt{t}} -x + 2\sqrt{t}\,dt \\ &= \frac{1}{2}t - \frac{1}{2}(4t - t) + 2\sqrt{t}(2\sqrt{t} - \sqrt{t}) \\ &= \frac{t}{2} - \frac{3t}{2} + 4t - 2t = t. \end{aligned}$$

Obviously $f(0) = 0$, and if $t < 0$, then $f(t) = -f(-t) = t$. Therefore $f'(0) = 1$. However

$$\int_{-1}^{1} (D_2\varphi)(x, 0)\,dx = 0.$$

Note: This result is possible only because $D_2\varphi(x,t)$ is not bounded on $[-1,1] \times [-a, a]$ for any $a > 0$. Also note that having -1 as the lower limit of the integral was a needless complication. The problem would have been more effective it the lower limit had been 0.

Exercise 9.29 Let E be an open set in R^n. The classes $\mathcal{C}'(E)$ and $\mathcal{C}''(E)$ are defined in the text. By induction $\mathcal{C}^{(k)}(E)$ can be defined as follows for all positive integers k: To say that $f \in \mathcal{C}^{(k)}(E)$ means that the partial derivatives $D_1 f, \ldots, D_n f$ belong to $\mathcal{C}^{(k-1)}(E)$.

Assume $f \in \mathcal{C}^{(k)}(E)$, and show (by repeated application of Theorem 9.41) that the kth-order derivative

$$D_{i_1 i_2 \ldots i_k} f = D_{i_1} D_{i_2} \ldots D_{i_k} f$$

is unchanged if the subscripts $i_1, \ldots, i_k$ are permuted.

For instance, if $n \geq 3$, then

$$D_{1213} f = D_{3112} f$$

for every $f \in \mathcal{C}^{(4)}$.

Solution. If the permutation leaves i_k fixed, this follows from the result for $k-1$ applied to $D_{i_k} f$. To get the general result, we observe that by the case $k = 2$ we have $D_{i_{k-1} i_k} f = D_{i_k i_{k-1}} f$. Hence the result holds for any permutation that maps i_{k-1} to i_k. But any permutation that maps i_j to i_k ($j \neq k, k-1$) can be written as the composition of a permutation that maps i_j to i_{k-1}, leaving i_k fixed, followed by the interchange of i_{k-1} and i_k, followed by a second permutation that leaves i_k fixed. Therefore the result applies to all permutations whatsoever.

Exercise 9.30 Let $f \in \mathcal{C}^{(m)}(E)$, where E is an open subset of R^n. Fix $\mathbf{a} \in E$, and suppose $\mathbf{x} \in R^n$ is so close to $\mathbf{0}$ that the points

$$\mathbf{p}(t) = \mathbf{a} + t\mathbf{x}$$

lie in E whenever $0 \leq t \leq 1$. Define

$$h(t) = f(\mathbf{p}(t))$$

for all $t \in R^1$ for which $\mathbf{p}(t) \in E$.
(*a*) For $1 \leq k \leq m$, show (by repeated application of the chain rule) that

$$h^{(k)}(t) = \sum (D_{i_1 \ldots, i_k} f)(\mathbf{p}(t)) x_{i_1} \ldots x_{i_k}.$$

The sum extends over all ordered k-tuples $(i_1, \ldots, i_k)$ in which each i_j is one of the integers $1, \ldots, n$.
(*b*) By Taylor's theorem (5.15)

$$h(1) = \sum_{k=0}^{m-1} \frac{h^{(k)}(0)}{k!} + \frac{h^{(m)}(t)}{m!}$$

for some $t \in (0, 1)$. Use this to prove Taylor's theorem in n variables by showing that the formula

$$f(\mathbf{a} + \mathbf{x}) = \sum_{k=0}^{m-1} \frac{1}{k!} \sum (D_{i_1 \ldots i_k} f)(\mathbf{a}) x_{i_1} \ldots x_{i_k} + r(\mathbf{x})$$

represents $f(\mathbf{a} + \mathbf{x})$ as the sum of its so-called "Taylor polynomial of degree $m - 1$," plus a remainder that satisfies

$$\lim_{\mathbf{x} \to \mathbf{0}} \frac{r(\mathbf{x})}{|\mathbf{x}|^{m-1}} = 0.$$

Each of the inner sums extends over all ordered k-tuples $(i_1, \ldots, i_k)$, as in part (*a*); as usual, the zero-order derivative of f is simply f, so that the constant term of the Taylor polynomial of f at $\mathbf{a}$ is $f(\mathbf{a})$.
(*c*) Exercise 29 shows that repetition occurs in the Taylor polynomial as written in part (*b*). For instance D_{113} occurs three times, as D_{113}, D_{131}, D_{311}. The sum of the corresponding three terms can be written in the form

$$3(D_1^2 D_3 f)(\mathbf{a}) x_1^2 x_3.$$

Prove (by calculating how often each derivative occurs) that the Taylor polynomial in (*b*) can be written in the form

$$\sum \frac{(D_1^{s_1} \cdots D_n^{s_n} f)(\mathbf{a})}{s_1! \cdots s_n!} x_1^{s_1} \cdots x_n^{s_n}.$$

Here the summation extends over all ordered n-tuples $(s_1, \ldots, s_n)$ such that each s_i is a nonnegative integer, and $s_1 + \cdots + s_n \le m - 1$.

Solution. (*a*) This formula is a simple application of the chain rule together with the fact that $D\mathbf{p}^{(i)}(t) = x_i$. The proof proceeds by induction on k.

(*b*) The formula is an immediate application of the fact that $\mathbf{p}(1) = \mathbf{a} + \mathbf{x}$, so that $h(1) = f(\mathbf{a} + \mathbf{x})$. The right-hand side is then an immediate application of the fact that $\mathbf{p}(0) = \mathbf{a}$. The only assertion that requires verification is that on the order of the remainder. The one-variable Taylor's theorem gives $r(\mathbf{x}) = \sum (D_{i_1,\ldots,i_m} f(\mathbf{p}(t)) x_{i_1} \cdots x_{i_m}$ for some $t \in (0,1)$, so that $|r(\mathbf{x})| \le K|\mathbf{x}|^m$ for some constant K. The assertion as to the order of r follows from this fact.

(*c*) If $s_1 + \cdots + s_n \le m-1$, the number of terms having the derivative combination $D_1^{s_1} \cdots D_n^{s_n} f$ is $\binom{s_1 + \cdots + s_n}{s_1, \ldots, s_n} = \dfrac{(s_1 + \cdots + s_n)!}{s_1! \cdots s_n!}$. Thus the $k!$ that occurs in the one-variable Taylor's theorem is $(s_1 + \cdots + s_n)!$; and when the terms are consolidated, this factor cancels the numerator of the multinomial symbol, effectively being replaced by $s_1! \cdots s_n!$.

Exercise 9.31 Suppose $f \in C^{(3)}$ in some neighborhood of a point $\mathbf{a} \in R^2$, the gradient of f is $\mathbf{0}$ at $\mathbf{a}$, but not all second-order derivatives of f are 0 at $\mathbf{a}$. Show how one can then determine from the Taylor polynomial of f at $\mathbf{a}$ (of degree 2) whether f has a local maximum or a local minimum, or neither, at the point $\mathbf{a}$.

Extend this to R^n in place of R^2.

Solution. Let us simply do R^n in the first place and save the trouble of doing R^2. According to Taylor's theorem

$$f(\mathbf{a}+\mathbf{x}) - f(\mathbf{a}) = \frac{1}{2} \sum_{i_1, i_2} (D_{i_1 i_2} f)(\mathbf{a}) x_{i_1} x_{i_2} + r(\mathbf{x}),$$

where $|\mathbf{x}|^{-2} r(\mathbf{x}) \to 0$ as $\mathbf{x} \to \mathbf{0}$. Note that the Taylor polynomial can be concisely written as $\frac{1}{2}\langle A\mathbf{x}, \mathbf{x}\rangle$, where A is the $n \times n$ Hessian matrix whose i,j entry is $D_{ij} f(\mathbf{a})$ and the angle brackets denote the inner product. If A is positive-definite, i.e., if $\langle A\mathbf{x}, \mathbf{x}\rangle > 0$ when $\mathbf{x} \ne \mathbf{0}$, there is a positive constant c such that $\langle A\mathbf{x}, \mathbf{x}\rangle \ge c|\mathbf{x}|^2$. (The constant c is the minimum value of $\langle A\mathbf{x}, \mathbf{x}\rangle$ on the unit sphere $|\mathbf{x}| = 1$.) Hence if $\delta > 0$ is chosen so that $|r(\mathbf{x})| < c|\mathbf{x}|^2$ when $0 < |\mathbf{x}| < \delta$, we see that $f(\mathbf{a}+\mathbf{x}) - f(\mathbf{a}) > 0$ if $0 < |\mathbf{x}| < \delta$, i.e., $\mathbf{a}$ is a local minimum of f. Likewise if A is negative-definite, then $\mathbf{a}$ is a local maximum of f.

It is well-known from linear algebra that a necessary and sufficient condition for positive-definiteness of the matrix A is that the principal minors be positive, i.e., the $k \times k$-submatrix consisting of the elements in the first k rows and columns of A has a positive determinant. For negative-definiteness the corresponding criterion is that this minor have the same sign as $(-1)^k$.

There are no other resonably regular cases that guarantee a maximum or minimum. A nonnegative-definite or nonpositive-definite matrix may well fail to guarantee a maximum or minimum, even in R^1. If the quadratic form $\langle A\mathbf{x}, \mathbf{x}\rangle$ assumes both signs, then the point $\mathbf{a}$ is definitely not either a maximum or a minimum. (If $\langle A\mathbf{x}, \mathbf{x}\rangle > 0$, then $f(\mathbf{a}+t\mathbf{x}) - f(\mathbf{a}) > 0$ for small values of t, while if $\langle A\mathbf{x}, \mathbf{x}\rangle < 0$, then $f(\mathbf{a}+t\mathbf{x}) - f(\mathbf{a}) < 0$ for small values of t.)

Chapter 10

Integration of Differential Forms

Exercise 10.1 Let H be a compact convex set in R^k with nonempty interior. Let $f \in \mathcal{C}(H)$, put $f(\mathbf{x}) = 0$ in the complement of H and define $\int_H f$ as in Definition 10.3.

Prove that $\int_H f$ is independent of the order in which the integrations are carried out.

Hint: Approximate f by functions that are continuous on R^k and whose supports are in H, as was done in Example 10.4.

Solution. We first give the definition of $\int_H f$, namely $\int_I f$, where I is any k-cell containing H. This definition is unambiguous, since if I and J are both k-cells containing H, each of the single integrals carried out is an integral over the same line segment for both cells, namely the intersection of the path of integration with H.

There seems to be no way to avoid somehow proving that the boundary of H, denoted ∂H, has "measure zero." The definition of the integral as an iterated integral makes that problem slightly more difficult than it would be otherwise, although we can show how to avoid this approach in two dimensions. We shall reserve that discussion until after the proof, which is rather lengthy. The length of the proof is due to the fact that integrals are really defined only over parallelepipeds. The point of the exercise is to enlarge the class of sets over which one can integrate. Our challenge is to show that the boundary of H can be enclosed in a finite set of parallelepipeds whose total volume can be arbitrarily small.

Our first job is to show that the hypersphere in R^k has measure zero. As the proof of that fact involves some work with $(k-2)$-dimensional hyperspheres in k-dimensional space, we need to make several definitions in order to express these ideas properly.

First, for each real number z and each positive number r, $S_r^{k-2}(z)$ denotes the $(k-2)$-dimensional hypersphere in R^k having radius r and center at the point $(0,0,\ldots,0,z)$, that is,

$$S_r^{k-2}(z) = \{\mathbf{x} : x_1^2 + \cdots + x_{k-1}^2 = r^2,\ x_k = z\}.$$

When $z = 0$, we shall write simply S_r^{k-2} and identify this sphere with the same $(k-2)$-dimensional hypersphere in R^{k-1}. We observe that $S_r^0(z)$ consists of the two points (r, z) and $(-r, z)$ in R^2. Another way of defining the set $S_r^{k-2}(z)$ is as the intersection of the $(k-1)$-dimensional sphere of radius r centered at $(0,0,\ldots,0,z)$ in R^k with its equatorial hyperplane $P_z = \{(x_1,\ldots,x_k) : x_k = z\}$.

Second, for each $\mathbf{a} \in R^k$ and each $\delta > 0$, $I_{\mathbf{a}}^k(\delta)$ is the closed hypercube of side δ in R^k whose "lower left" corner is $\mathbf{a}$, that is

$$I_{\mathbf{a}}^k(\delta) = \{\mathbf{x} : a_j \le x_j \le a_j + \delta,\ j = 1,\ldots,k\}.$$

Third, the set of points $(m_1,\ldots,m_k) \in R^k$ having integer coordinates will be denoted $\mathcal{Z}^k$.

Fourth, for all real numbers r and δ such that $0 < \delta < r$, $C_{r,\delta}^k$ is the set of lattice points $\mathbf{m} \in \mathcal{Z}^k$ for which the closed hypercube of side δ with lower left corner $\delta\mathbf{m}$ intersects the hypersphere S_r^{k-1} in R^k. That is,

$$C_{r,\delta}^k = \{\mathbf{m} : I_{\delta\mathbf{m}}^k(\delta) \cap S_r^{k-1} \ne \emptyset\}.$$

Fifth, $N(r,\delta,k)$ is the number of points in $C_{r,\delta}^k(z)$, that is, the number of hypercubes of side δ with lower left hand corner at a point $\delta\mathbf{m}$, $\mathbf{m} \in \mathcal{Z}^k$, that intersect the hypersphere S_r^{k-1}. Our main goal in the first stage of the proof will be to prove the estimate $N(r,\delta,k) \le 6^{k^2}\left(\frac{r}{\delta}\right)^{k-1}$. (A smaller constant than 12^{k^2} could easily be attained, but we have no need of any improvement, and this constant seems to be the one that makes the argument simplest.)

Sixth, and finally, $A_{r,\delta}^k$ is the union of all the hypercubes $I_{\delta\mathbf{m})}^k(\delta)$ that intersect the hypersphere S_r^{k-1}, that is, for which $\mathbf{m} \in C_{r,\delta}^k$. This set is a finite union of compact sets, hence is compact. Obviously it contains the hypersphere S_r^{k-1}. What is slightly less obvious is that its interior contains this sphere. In fact no point of the sphere can be a limit point of points exterior to $A_{r,\delta}^k$, since if $\{\mathbf{x}_n\}$ is a sequence of points such that each $\mathbf{x}_n$ belongs to a hypercube $I_{\delta\mathbf{m}_n}^k(\delta)$ not contained in $A_{r,\delta}^k$, and $\mathbf{x}_n \to \mathbf{x}$, some set $I_{\delta\mathbf{m}_{n_0}}^k(\delta)$ must occur infinitely often. (Any bounded neighborhood of $\mathbf{x}$ intersects only finitely many of these hypercubes.) Since $I_{\delta\mathbf{m}_{n_0}}^k(\delta)$ is closed, this implies that $\mathbf{x}$ belongs to $I_{\delta\mathbf{m}_{n_0}}^k(\delta)$. Since $I_{\delta\mathbf{m}_{n_0}}^k(\delta)$ is not contained in $A_{r,\delta}^k$ it follows that $\mathbf{x} \notin S_r^{k-1}$. Thus no sequence of points exterior to $A_{r,\delta}^k$ can approach a point of S_r^{k-1}. It follows that S_r^{k-1} contains no points of the boundary of $A_{r,\delta}^k$ and is therefore contained in the interior of this set.

With these definitions out of the way we can proceed to the proof, which we break into several stages, each broken into several steps, in order to make navigating easier.

Stage 1. Establish that the sphere S_r^{k-1} in R^k has k-dimensional content 0.

Step 1. Establish that $A^k_{r,\delta} \cup A^k_{r+\delta}$ contains the closed k-dimensional annulus consisting of the region between S^{k-1}_r and $S^{k-1}_{r+\delta}$, that is, all the points $\mathbf{x} \in R^k$ such that $r \le |\mathbf{x}| \le r + \delta$.

To this end, let $\mathbf{x}$ belong to this annulus, so that $r \le |\mathbf{x}| \le r + \delta$. Since $S^{k-1}_s \subset A^k_{s,\delta}$ for each s, we can assume $r < |\mathbf{x}| < r + \delta$. Let $\mathbf{m} = (m_1, \ldots, m_k)$ be a lattice point in $\mathcal{Z}^k$ such that $\mathbf{x} \in I^k_{\delta\mathbf{m}}(\delta)$. Let $n_j = m_j$ if $m_j \ge 0$ and $n_j = m_j + 1$ if $m_j < 0$, so that $(\delta n_1, \ldots, \delta n_k)$ is a corner of $I_{\delta\mathbf{m}}(\delta)$. Since $|n_j| = \min(|m_j|, |m_j + 1|)$ for all j, $(\delta\mathbf{n})$ is the unique point of $I^k_{\delta\mathbf{m}}(\delta)$ closest to the origin. In particular $|\delta\mathbf{n}| \le |\mathbf{x}|$. The lattice point $\mathbf{n}' = (n_1 + \varepsilon_1(n_1), \ldots, n_k + \varepsilon_k(n_k))$, where $\varepsilon_j(t)$ is 1 if $t \ge 0$ and -1 if $t < 0$, is such that $\delta\mathbf{n}'$ is the corner of $I_{\delta\mathbf{m}}(\delta)$ opposite to $\delta\mathbf{n}$ and is the unique point of $K_{\delta\mathbf{m}}(\delta)$ farthest from the origin. In particular $|\delta\mathbf{n}'| \ge |\mathbf{x}|$.

We claim first that $|\mathbf{n}'| - |\mathbf{n}| \ge 1$. Indeed, we have

$$\begin{aligned}|\mathbf{n}'|^2 &= (n_1^2 + \cdots + n_k^2) + 2(n_1\varepsilon_1(n_1) + \cdots + n_k\varepsilon_k(n_k) + ((\varepsilon_1(n_1))^2 + \cdots + (\varepsilon_k(n_k))^2) \\ &= (n_1^2 + \cdots + n_k^2) + 2(|n_1| + \cdots + |n_k|) + k,\end{aligned}$$

and

$$\big(|\mathbf{n}| + 1\big)^2 = (n_1^2 + \cdots + n_k^2) + 2\sqrt{n_1^2 + \cdots + n_k^2} + 1,$$

so that the desired inequality follows from the two inequalities $k \ge 1$ and $\sqrt{n_1^2 + \cdots + n_k^2} \le |n_1| + \cdots + |n_k|$. This argument shows in general that, for any $\mathbf{m} \in \mathcal{Z}^k$, if $\mathbf{b}$ and $\mathbf{c}$ are the points in $I^k_{\delta\mathbf{m}}(\delta)$ of minimal and maximal absolute value respectively, then $|\mathbf{c}| - |\mathbf{b}| \ge \delta$.

From this we deduce a corollary: *Let r be any positive real number larger than δ. If $r \le s \le r + \delta$ and $\mathbf{m} \in C^k_{s,\delta}$, then either $\mathbf{m} \in C^k_{r,\delta}$ or $\mathbf{m} \in C^k_{r+\delta,\delta}$.* In plain words, if $I^k_{\delta\mathbf{m}}(\delta)$ meets S^{k-1}_s for some $s \in [r, r + \delta]$, it must meet either S^{k-1}_r or $S^{k-1}_{r+\delta}$.

To prove this corollary, we note that the assumption $\mathbf{m} \in C^k_{s,\delta}$ says that there exists $\mathbf{x} \in R^k$ such that $\mathbf{x} \in S^{k-1}_s \cap I^k_{\delta\mathbf{m}}(\delta)$. Now suppose $\mathbf{m}$ belongs to neither of the sets $C^k_{r,\delta}$ and $C^k_{r+\delta,\delta}$. Then the set $I^k_{\delta\mathbf{m}}(\delta)$ contains no points of S^{k-1}_r. If $\mathbf{b}$ is the point of $I^k_{\delta\mathbf{m}}(\delta)$ of smallest norm, it follows that $|\mathbf{b}| > r$. (For $I^k_{\delta\mathbf{m}}(\delta)$ contains the point $\mathbf{x}$ of norm $s \ge r$. Since $I^k_{\delta\mathbf{m}}(\delta)$ is a connected set, if it contained a point of norm less than or equal to r it would also contain a point of S^{k-1}_r.) Similarly, if the set $I^k_{\delta\mathbf{m}}(\delta)$ contains no points of $S^{k-1}_{r+\delta}$, then $|\mathbf{c}| < r + \delta$. But then it follows that $|\mathbf{c}| - |\mathbf{b}| < r + \delta - r = \delta$, contrary to what has been proved.

Another way of stating what was just proved is that if $\mathbf{x} \in R^k$ is such that $r \le |\mathbf{x}| \le r + \delta$, then $\mathbf{x} \in A^k_{r,\delta} \cup A^k_{r+\delta,\delta}$. That is, the union $A^k_{r,\delta} \cup A^k_{r+\delta,\delta}$ contains the entire annulus of points $\mathbf{x}$ such that $r \le |\mathbf{x}| \le r + \delta$. Step 1 of the proof is now complete.

Step 2. Assuming $k > 1$, estimate the number of k-dimensional hypercubes $I^k_{\delta\mathbf{m}}(\delta)$ that intersect various zones on the $(k-1)$-sphere S^{k-1}_r in R^k.

We divide the upper hemisphere, consisting of $\mathbf{x}$ such that $|\mathbf{x}| = r$ and $x_k \geq 0$ into half-open zones

$$Z_p = \{(x_1, \ldots, x_{k-1}, x_k) : x_1^2 + \cdots + x_k^2 = r, p\delta \leq x_k < (p+1)\delta\},$$

for $0 \leq p \leq \left[\frac{r}{\delta}\right] - 1$. Here $[a]$ denotes the integer part of a, that is, the integer q such that $q \leq a < q+1$, and we assume $0 < \delta < r$. Between the top zone $Z_{[\frac{r}{\delta}]-1}$ and the "north pole" (the point $(0, 0, \ldots, r)$, there is a closed "cap" of height η for some $\eta \in [0, \delta)$. The hypercubes $I^k_{\delta\mathbf{m}}(\delta)$ (where $m_k = \left[\frac{r}{\delta}\right] - 1$ or $m_k = \left[\frac{r}{\delta}\right]$) intersecting this cap must be handled separately from those intersecting the other zones.

We shall prove that *the lattice points* $\mathbf{m} \in \mathcal{Z}^k$ *for which* $m_k = p$ *and* $I^k_{\delta\mathbf{m}}(\delta)$ *intersects* Z_p *are precisely those whose bottom face* $x_k = p\delta$ *intersects the half-closed* $(k-1)$*-dimensional annulus between* $S^{k-2}_s(p\delta)$ *and* $S^{k-2}_t(p\delta)$ *in the hyperplane* $x_k = p\delta$. (This annulus is closed at t and open at s, where $s = \sqrt{r^2 - ((p+1)\delta)^2}$ and $t = \sqrt{r^2 - (p\delta)^2}$.)

Indeed, this fact is nearly obvious, as the zone Z_p is the union of the $(k-2)$-dimensional spheres $S^{k-2}_{\sqrt{r^2-u^2}}(u)$ for $p\delta \leq u < (p+1)\delta$. If $(x_1, \ldots, x_{k-1}, u) \in I^k_{\delta\mathbf{m}}(\delta) \cap Z_p$ and $m_k = p$, then $p\delta \leq u < (p+1)\delta$ and $x_1^2 + \cdots + x_{k-1}^2 = r^2 - u^2$, so that $s < \sqrt{x_1^2 + \cdots + x_{k-1}^2} \leq t$. Thus the point $(x_1, \ldots, x_{k-1}, \delta p)$ belongs to both $I^k_{\delta\mathbf{m}}(\delta)$ and to the annulus. Conversely, if $I^k_{\delta\mathbf{m}}(\delta)$ with $m_k = p$ intersects the annulus, then this hypercube contains a point $(x_1, \ldots, x_{k-1}, p\delta)$ with $s^2 < x_1^2 + \cdots + x_{k-1}^2 \leq t^2$. Setting $u = \sqrt{r^2 - (x_1^2 + \cdots + x_{k-1}^2)}$, we have $p\delta \leq u < (p+1)\delta$, and therefore $(x_1, \ldots, x_{k-1}, u) \in Z_p \cap I^k_{\delta\mathbf{m}}(\delta)$.

To estimate the total number of hypercubes $I^k_{\delta\mathbf{m}}(\delta)$ that intersect the hypersphere S^{k-1}_r, we need an estimate of the number that intersect each zone Z_p. If $I^k_{\delta\mathbf{m}}(\delta)$ intersects Z_p, then $m_k = p$ or $m_k = p - 1$. If $I^k_{(\delta m_1, \ldots, \delta m_{k-1}, \delta(p-1))}(\delta)$ intersects Z_p, the intersection must be in the hyperplane $x_k = p\delta$, and hence $I^k_{(\delta m_1, \ldots, \delta m_{k-1}, \delta p)}(\delta)$ also intersects Z_p. Hence we can get a (loose, but safe) upper bound on the number of hypercubes $I^k_{\delta\mathbf{m}}(\delta)$ that intersect Z_p by counting those for which $m_k = p$ and doubling. (The case of the bottom layer Z_0 is special, and the "northern cap" mentioned above will be handled separately.)

The fact that a hypercube $I^k_{\delta\mathbf{m}}(\delta)$ intersecting Z_p must intersect the annulus shows that we need only estimate of the width of the annulus, that is, the number $t - s = \sqrt{r^2 - (p\delta)^2} - \sqrt{r^2 - ((p+1)\delta)^2}$. For that width we have the following simple result:

$$\sqrt{r^2 - (p\delta)^2} - \sqrt{r^2 - ((p+1)\delta)^2} \leq \frac{(2p+1)\delta^2}{\sqrt{r^2 - (p\delta)^2}}.$$

The proof of this inequality is straightforward:

$$\begin{aligned}
\sqrt{r^2 - (p\delta)^2} - \sqrt{r^2 - ((p+1)\delta)^2} &= \frac{(r^2 - (p\delta)^2) - (r^2 - ((p+1)\delta)^2)}{\sqrt{r^2 - (p\delta)^2} + \sqrt{r^2 - ((p+1)\delta)^2}} \\
&\leq \frac{(2p+1)\delta^2}{\sqrt{r^2 - (p\delta)^2}}.
\end{aligned}$$

Now let j be the largest integer not larger than $\sqrt{\left(\frac{r}{\delta}\right)^2 - (p+1)^2}$ and l the smallest integer not smaller than $\sqrt{\left(\frac{r}{\delta}\right)^2 - p^2}$. We have just shown that

$$0 < l - j < 2 + \frac{(2p+1)\delta}{\sqrt{r^2 - (p\delta)^2}} = 2 + \frac{2p+1}{\sqrt{\left(\frac{r}{\delta}\right)^2 - p^2}}.$$

It is this inequality that provides the required estimate of the width of the annulus corresponding to the zone Z_p.

Step 3. Prove the estimate $N^k_{r,\delta} \le 6^{k^2}\left(\frac{r}{\delta}\right)^{k-1}$.

We first do the case $k = 1$. This case is very straightforward. If $I^1_{(\delta m,z)}(\delta) \cap S^0_r$, then either $m\delta \le r \le (m+1)\delta$ or $m\delta \le -r \le (m+1)\delta$. If $r = k\delta$ for some integer k, there are four values of m for which one of these two sets of inequalities hold, namely $-k-1$, $-k$, $k-1$, and k. Otherwise there are only two such integers m, namely $\left[\frac{r}{\delta}\right]$ and $\left[\frac{-r}{\delta}\right]$. Thus we actually have $N^0_{r,\delta} \le 4 < 6^1\left(\frac{r}{\delta}\right)^0$.

We now proceed by induction, supposing the theorem proved for dimensions less than k, and we assume $k \ge 2$. We consider all the lattice points $\mathbf{m} \in R^k$ such that $m_k = p$ and $I^k_{\delta\mathbf{m}}(\delta)$ intersects Z_p. From what we have shown above in Step 1 and Step 2, if $\mathbf{m}$ has this property, then the bottom face of $I^k_{\delta\mathbf{m}}(\delta)$ intersects one of the spheres $S^{k-2}_{s\delta}(p\delta)$, where s is an integer such that $j \le s \le l$. The number of such $\mathbf{m}$ is at most $N^{k-1}_{s\delta,\delta}$ and hence is at most $6^{(k-1)^2}s^{k-2}$, which is certainly no larger than

$$6^{(k-1)^2}\left(1 + \sqrt{\left(\frac{r}{\delta}\right)^2 - p^2}\right)^{k-2}.$$

Because of our estimate of $l - j$, we see that the total number of $\mathbf{m}$ for which $m_k = p$ and $I^k_{\delta\mathbf{m}}(\delta)$ intersects Z_p is at most

$$6^{(k-1)^2}\left(2 + \frac{2p+1}{\sqrt{\left(\frac{r}{\delta}\right)^2 - p^2}}\right)\left(2^{k-2} + 2^{k-2}\left(\left(\frac{r}{\delta}\right)^2 - p^2\right)^{\frac{k-2}{2}}\right),$$

where we have used the inequality $(1+t)^q \le 2^q + (2t)^q$ with $q = k-2$. We expand this last product into a sum of six terms:

$$6^{(k-1)^2}\left(2^{k-1} + 2^{k-1}\left(\left(\frac{r}{\delta}\right)^2 - p^2\right)^{\frac{k-2}{2}} + 2^{k-1}\frac{p}{\sqrt{\left(\frac{r}{\delta}\right)^2 - p^2}}\right.$$
$$\left. + \frac{2^{k-2}}{\sqrt{\left(\frac{r}{\delta}\right)^2 - p^2}} + 2^{k-1}p\left(\left(\frac{r}{\delta}\right)^2 - p^2\right)^{\frac{k-3}{2}} + 2^{k-2}\left(\left(\frac{r}{\delta}\right)^2 - p^2\right)^{\frac{k-3}{2}}\right)$$
$$= 6^{(k-1)^2}\left(I_1(p) + I_2(p) + I_3(p) + I_4(p) + I_5(p) + I_6(p)\right).$$

We need to estimate the sum of each of these terms over p from 0 to $\left[\frac{r}{\delta}\right] - 1$.

For $I_1(p)$ we have the simple estimate

$$\sum_{p=0}^{[\frac{r}{\delta}]-1} I_1(p) = 2^{k-1}[\frac{r}{\delta}] \leq 2^{k-1}(\frac{r}{\delta}) < 2^{k-1}(\frac{r}{\delta})^{k-1},$$

since $k > 1$ and $r > \delta$.

For $I_2(p)$ we have

$$\sum_{p=0}^{[\frac{r}{\delta}]-1} I_2(p) = 2^{k-1}(\frac{r}{\delta})^{k-2} \sum_{p=0}^{[\frac{r}{\delta}]-1} \left(1-\frac{p\delta}{r}\right)^{\frac{k-2}{2}}$$
$$< 2^{k-1}(\frac{r}{\delta})^{k-2}(\frac{r}{\delta}) = 2^{k-1}(\frac{r}{\delta})^{k-1}.$$

Here we have used the fact that $0 < 1 - (\frac{p\delta}{r})^2 \leq 1$ for the values of p in the range of summation, as we shall do twice more below.

For $I_3(p)$ we have

$$\begin{aligned}
\sum_{p=0}^{[\frac{r}{\delta}]-1} I_3(p) &= 2^{k-1}(\frac{\delta}{r}) \sum_{p=0}^{[\frac{r}{\delta}]-1} \frac{p}{\sqrt{1-(\frac{p\delta}{r})^2}} \\
&\leq 2^{k-1}(\frac{\delta}{r}) \int_1^{\frac{r}{\delta}} \frac{x}{\sqrt{1-(\frac{\delta x}{r})^2}}\,dx \\
&= 2^{k-1}(\frac{r}{\delta}) \int_{\frac{\delta}{r}}^1 \frac{y}{\sqrt{1-y^2}}\,dy \\
&< 2^{k-1}(\frac{r}{\delta}) \int_0^1 \frac{y}{\sqrt{1-y^2}}\,dy \\
&= 2^{k-1}\frac{r}{\delta} \leq 2^{k-1}(\frac{r}{\delta})^{k-1}.
\end{aligned}$$

Since $I_4(p) \leq \frac{1}{2}I_3(p)$ for $p = 1, 2, \ldots, [\frac{r}{\delta}] - 1$, it is clear that

$$\sum_{p=0}^{[\frac{r}{\delta}]-1} I_4(p) < 2^{k-2} + 2^{k-2}(\frac{r}{\delta})^{k-1} < 2^{k-1}(\frac{r}{\delta})^{k-1}.$$

For $I_5(p)$ we have

$$\sum_{p=0}^{[\frac{r}{\delta}]-1} I_5(p) = 2^{k-1}(\frac{r}{\delta})^{k-3} \sum_{p=0}^{[\frac{r}{\delta}]-1} p\left(1-\frac{p\delta}{r}\right)^{\frac{k-3}{2}}$$
$$< 2^{k-2}(\frac{r}{\delta})^{k-3}(\frac{r}{\delta})^2 = 2^{k-2}(\frac{r}{\delta})^{k-1}.$$

For $I_6(p)$ we have

$$\sum_{p=0}^{[\frac{r}{\delta}]-1} I_6(p) = 2^{k-2}\Big(\frac{r}{\delta}\Big)^{k-3} \sum_{p=0}^{[\frac{r}{\delta}]-1} \Big(1 - \frac{p\delta}{r}\Big)^{\frac{k-3}{2}}$$

$$< 2^{k-2}\Big(\frac{r}{\delta}\Big)^{k-3}\Big(\frac{r}{\delta}\Big) < 2^{k-2}\Big(\frac{r}{\delta}\Big)^{k-1}.$$

Adding all these estimates, we find a sum that is at most

$$6^{(k-1)^2}\big(2^{k-1}+2^{k-1}+2^{k-1}+2^{k-2}+2^{k-2}+2^{k-2}\big)\Big(\frac{r}{\delta}\Big)^{k-1} \le 6^{(k-1)^2+1}2^{k-1}\Big(\frac{r}{\delta}\Big)^{k-1}.$$

If we wish to count the total number of hypercubes $I^k_{\delta\mathbf{m}}(\delta)$ that intersect the zones $Z_1, \ldots, Z_{[\frac{r}{\delta}]-1}$, we recall our previous observation that such a hypercube can intersect Z_p only if $m_k = p$ or $m_k = p-1$, and if a hypercube $I^k_{\delta\mathbf{m}}(\delta)$ with $m_k = p-1$ intersects Z_p, then so does the hypercube $I^k_{\delta(\mathbf{m}+\mathbf{e}_k)}(\delta)$, and the latter has already been counted. Hence in estimating the total number of hypercubes that intersect one of these zones we are more than safe in simply doubling the estimate we have already obtained.

As for the zone Z_0, any hypercube $I^k_{\delta\mathbf{m}}(\delta)$ with $m_k = -1$ that intersects its bottom edge (the hypersphere S^{k-2}_r) also meets the reflection of Z_0 through the plane $x_k = 0$, and hence the reflection of that hypercube has already been counted among those that meet Z_0. When we double our count to include the hypercubes meeting the "southern" hemisphere, all these hypercubes will automatically be counted. Thus it remains only to estimate the hypercubes that meet the "northern arctic zone," then double the count.

Hence we now consider the cap at the top of the hemisphere, whose boundary is the $(k-2)$-dimensional hypersphere

$$S^{k-2}_s\Big(\Big[\frac{r}{\delta}\Big]\delta\Big),$$

where

$$s = \delta\sqrt{\Big(\frac{r}{\delta}\Big)^2 - \Big[\frac{r}{\delta}\Big]^2}.$$

If $\mathbf{m}$ is such that $I^k_{\delta\mathbf{m}}(\delta)$ meets this set, then we must have $[\frac{r}{\delta}] - 1 \le m_k \le [\frac{r}{\delta}]$, so that there are only two possible values for m_k. As for m_j, $j < k$, we certainly have $-\frac{s}{\delta} - 1 \le m_j \le \frac{s}{\delta}$, so that there are at most $2^k\Big(\dfrac{s}{\delta}+1\Big)^{k-1}$ such hypercubes not already counted.

Since $\big[\frac{r}{\delta}\big] \le \frac{r}{\delta} < \big[\frac{r}{\delta}\big] + 1$, we easily find that

$$s \le 2\sqrt{r\delta},$$

so that $\frac{s}{\delta}+1 \le 2\sqrt{\frac{r}{\delta}}+1$. Once more using the inequality $(1+t)^q \le 2^q(1+t^q)$ for positive t, with $q = k-1$, we find that the number of hypercubes meeting the northern polar cap is at most $2^{2k-1}\big(1 + (\frac{r}{\delta})^{\frac{k-1}{2}}\big)$, which is less than $2^{2k}(\frac{r}{\delta})^{k-1}$

Adding the numbers up and then doubling to count the hypercubes that meet the southern hemisphere, we find that the hypersphere S_r^{k-1} meets at most

$$\left[6^{(k-1)^2+1}2^k + 2^{2k+1}\right)\left(\frac{r}{\delta}\right)^{k-1}.$$

The constant coefficient here is less than 6^{k^2}. This is directly computable for $k = 2$ and $k = 3$. Indeed, the quantity 2^{2k+1} is less than 6^k, while $2^k < 6^{\frac{k}{2}}$. The extremely weak inequality $a + b < ab$, which is valid for positive integers a and b both larger than 1, then implies that for $k \geq 4$ the coefficient is at most

$$6^{(k-1)^2+1+(k/2)+k} = 6^{k^2-k/2+2} \leq 6^{k^2}.$$

Stage 1 in the proof is now complete. We have shown that the $(k-1)$-sphere S_r^{k-1} intersects at most $6^{k^2}\left(\frac{r}{\delta}\right)^{k-1}$ hypercubes from the family $I^k_{\delta\mathbf{m}}(\delta)$. As the volume of each hypercube is δ^k, it follows that the $(k-1)$-sphere is contained in a finite union of cubes of total k-dimensional volume $6^{k^2}r^{k-1}\delta$. Since δ is an arbitrary positive number, the k-dimensional volume of S_r^{k-1} is zero.

We now move on to the second stage of the proof.

Stage 2. Given any convex set H in R^k with non-empty interior, construct a homeomorphism $\mathbf{T}$ of R^k onto itself that maps $S^{k-1} = S_1^{k-1}$ to ∂H, the inside of the unit ball to the interior of H, and the outside to the exterior of H, and satisfies a Lipschitz condition on a neighborhood of S^{k-1}. To get this result we need some more background work on general convex sets in R^k.

Let C be a bounded convex set in R^k, and let $\mathbf{z}$ be an interior point of C. For each point $\mathbf{x}$ on the unit sphere S^{k-1} in R^k, let $\psi(\mathbf{x})$ be the distance from $\mathbf{z}$ to the complement of C in the direction of $\mathbf{x}$, that is,

$$\psi(\mathbf{x}) = \inf\{t > 0 : \mathbf{z} + t\mathbf{x} \notin C\} = \sup\{t > 0 : \mathbf{z} + t\mathbf{x} \in C\}.$$

Step 1. Prove that the function $\psi : S^{k-1} \to (0, +\infty)$ is continuous, in fact, that it satisfies a Lipschitz condition: for some constant K, $|\psi(\mathbf{x}) - \psi(\mathbf{y})| \leq K|\mathbf{x} - \mathbf{y}|$.

There exist positive numbers a and b such that $a \leq \psi(\mathbf{x}) \leq b$ for all $\mathbf{x} \in S^{k-1}$. Indeed we can let a be the radius of the largest open ball about $\mathbf{z}$ that is contained in C and b the radius of the smallest closed ball about $\mathbf{z}$ containing C.

The result now follows from a lemma.

Let $0 \leq s \leq \psi(\mathbf{x})$. Then C contains the open ball of radius $\left(1 - \dfrac{s}{\psi(\mathbf{x})}\right)a$ about $\mathbf{z} + s\mathbf{x}$.

Proof: If $s = \psi(\mathbf{x})$, this ball is empty, and if $s = 0$ the assertion is merely the definition of a. Hence assume $0 < s < \psi(\mathbf{x})$. Now suppose $|\mathbf{y} - (\mathbf{z} + s\mathbf{x})| < \left(1 - \dfrac{s}{\psi(\mathbf{x})}\right)a$. Let $t = 1 - \dfrac{s}{\psi(\mathbf{x})}$, so that $0 < t < 1$, and let $r = \dfrac{1}{2}\min\left(a - \dfrac{|\mathbf{y} - (\mathbf{z} + s\mathbf{x})|}{t}, \dfrac{s}{t}\right)$, so that $r > 0$. Let $\mathbf{w} = \dfrac{\mathbf{y} - (\mathbf{z} + s\mathbf{x})}{t} + r\mathbf{x}$, and let $u = \dfrac{s - tr}{1 - t}$. We claim that $\mathbf{z} + \mathbf{w} \in C$ and that $\mathbf{z} + u\mathbf{x} \in C$, so that $\mathbf{y} = t(\mathbf{z} + \mathbf{w}) + (1 - t)(\mathbf{z} + u\mathbf{x}) \in C$.

The first claim will follow if we show that $|\mathbf{w}| < a$. In fact

$$\begin{aligned} |\mathbf{w}| &\leq \frac{|\mathbf{y} - (\mathbf{z} + s\mathbf{x})|}{t} + r \ \ (\text{since } |\mathbf{x}| = 1), \\ &\leq \frac{|\mathbf{y} - (\mathbf{z} + s\mathbf{x})|}{t} + \frac{1}{2}\Big(a - \frac{|\mathbf{y} - (\mathbf{z} + s\mathbf{x})||}{t}\Big) \\ &< \frac{|\mathbf{y} - (\mathbf{z} + s\mathbf{x})|}{t} + a - \frac{|\mathbf{y} - (\mathbf{z} + s\mathbf{x})|}{t} = a. \end{aligned}$$

The second claim will follow if we prove $0 < u < \psi(\mathbf{x})$. In fact $u = (s - tr) \cdot \dfrac{1}{1 - t} = (s - tr) \cdot \dfrac{\psi(\mathbf{x})}{s} = \Big(1 - \dfrac{tr}{s}\Big)\psi(\mathbf{x}) < \psi(\mathbf{x})$. Since $r \leq \dfrac{s}{2t}$, we have $u \geq \dfrac{1}{2}\psi(\mathbf{x}) > 0$.

Finally, the last claim is a routine computation:

$$\begin{aligned} t(\mathbf{z} + \mathbf{w}) + (1 - t)(\mathbf{z} + u\mathbf{x}) &= \mathbf{z} + t\mathbf{w} + (1 - t)u\mathbf{x} \\ &= \mathbf{z} + \mathbf{y} - (\mathbf{z} + s\mathbf{x}) + tr\mathbf{x} + (1 - t)u\mathbf{x} \\ &= \mathbf{y} + \big(tr - s + (1 - t)u\big)\mathbf{x} \\ &= \mathbf{y} \ \ (\text{since } (1 - t)u = s - tr). \end{aligned}$$

The lemma is now proved.

Taking $\mathbf{y} = \mathbf{z} + s\mathbf{v}$ in this lemma (where $|\mathbf{v}| = 1$), we see that $\mathbf{y} \in C$ (and hence $\psi(\mathbf{v}) \geq s$) if

$$|\mathbf{v} - \mathbf{x}| < \Big(\frac{1}{s} - \frac{1}{\psi(\mathbf{x})}\Big)a.$$

Now let $t > \psi(\mathbf{x})$, $|\mathbf{v}| = 1$, and $|\mathbf{v} - \mathbf{x}| < \Big(\dfrac{1}{\psi(\mathbf{x})} - \dfrac{1}{t}\Big)a$. Choose $t' \in (\psi(\mathbf{x}), t)$ such that

$$|\mathbf{v} - \mathbf{x}| < \Big(\frac{1}{t'} - \frac{1}{t}\Big)a.$$

If $\psi(\mathbf{v}) \geq t$, we have *a fortiori* $\psi(\mathbf{v}) > t'$ and

$$|\mathbf{x} - \mathbf{v}| < \Big(\frac{1}{t'} - \frac{1}{\psi(\mathbf{v})}\Big)a,$$

which, as already shown, implies $\psi(\mathbf{x}) \geq t'$, contradicting the choice of t'. Therefore $\psi(\mathbf{v}) < t$.

To summarize, if $s < \psi(\mathbf{x}) < t$, then $s \leq \psi(\mathbf{v}) < t$ provided

$$|\mathbf{x} - \mathbf{v}| < \min\bigg(\Big(\frac{1}{s} - \frac{1}{\psi(\mathbf{x})}\Big)a, \Big(\frac{1}{\psi(\mathbf{x})} - \frac{1}{t}\Big)a\bigg).$$

This proves that ψ is continuous. Specializing to the case where $s = \psi(\mathbf{x}) - \varepsilon$ and $t = \psi(\mathbf{x}) + \varepsilon$, we see that $|\psi(\mathbf{x}) - \psi(\mathbf{v})| \leq \varepsilon$ provided $|\mathbf{x} - \mathbf{v}| < \dfrac{a\varepsilon}{\psi(\mathbf{x})(\psi(\mathbf{x}) + \varepsilon)}$.

Again, *a fortiori*,

$$|\mathbf{x}-\mathbf{v}| < \frac{a\varepsilon}{b(b+\varepsilon)} \Rightarrow |\psi(\mathbf{x})-\psi(\mathbf{v})| \le \varepsilon.$$

We now claim that

$$|\mathbf{x}-\mathbf{v}| \le \frac{a\varepsilon}{2b^2} \Rightarrow |\psi(\mathbf{x})-\psi(\mathbf{v})| \le \varepsilon.$$

This follows from the previous statement and the continuity of ψ (together with the fact that a closed ball on the sphere S is the closure of the open ball with the same center and radius) if $\varepsilon \le b-a$. If $\varepsilon > b-a$, the second inequality automatically holds because $\psi(\mathbf{x})$ and $\psi(\mathbf{v})$ differ by at most $b-a$. Specializing to equality in the hypothesis, we deduce the Lipschitz inequality

$$|\psi(\mathbf{x})-\psi(\mathbf{v})| \le \frac{2b^2}{a}|\mathbf{x}-\mathbf{v}|.$$

We remark that the statement that $\mathbf{y}$ belongs to the interior, boundary, and exterior of C is equivalent to $|\mathbf{y}-\mathbf{z}| < \psi\Big(\frac{\mathbf{y}-\mathbf{z}}{|\mathbf{y}-\mathbf{z}|}\Big)$, $|\mathbf{y}-\mathbf{z}| = \psi\Big(\frac{\mathbf{y}-\mathbf{z}}{|\mathbf{y}-\mathbf{z}|}\Big)$, or $|\mathbf{y}-\mathbf{z}| > \psi\Big(\frac{\mathbf{y}-\mathbf{z}}{|\mathbf{y}-\mathbf{z}|}\Big)$.

Step 2. Use the function $\psi(\mathbf{x})$ to define a homeomorphism of R^k onto itself that maps S^{k-1} to ∂H and is Lipschitz in a neighborhood of S^{k-1}. Such a homeomorphism $\mathbf{T}(\mathbf{x})$ is defined for all $\mathbf{x} \in R^k$ as follows. We set $\mathbf{T}(\mathbf{0}) = \mathbf{z}$ and

$$\mathbf{T}(\mathbf{x}) = \mathbf{z} + \psi\Big(\frac{\mathbf{x}}{|\mathbf{x}|}\Big)\mathbf{x}$$

if $\mathbf{x} \ne \mathbf{0}$. Since $|\mathbf{T}(\mathbf{x}) - \mathbf{T}(\mathbf{0})| \le M|\mathbf{x}|$, where $M = \sup\{\psi(\mathbf{y}) : |\mathbf{y}| = 1\}$, it is clear that $\mathbf{T}$ is continuous at $\mathbf{0}$. At all other points it is a composition of continuous functions, hence continuous. Since $\frac{\mathbf{T}(\mathbf{x})-\mathbf{z}}{|\mathbf{T}(\mathbf{x})-\mathbf{z}|} = \frac{\mathbf{x}}{|\mathbf{x}|}$, we have the continuous inverse function

$$\mathbf{x} = \frac{\mathbf{T}(\mathbf{x})-\mathbf{z}}{\psi\big((\mathbf{T}(\mathbf{x})-\mathbf{z})/|\mathbf{T}(\mathbf{x})-\mathbf{z}|\big)},$$

That is, for $\mathbf{y} \ne \mathbf{z}$,

$$\mathbf{T}^{-1}(\mathbf{y}) = \frac{\mathbf{y}-\mathbf{z}}{\psi\big((\mathbf{y}-\mathbf{z})/|\mathbf{y}-\mathbf{z}|\big)},$$

which is not $\mathbf{0}$. Thus the mapping is one-to-one and onto.

The mapping also satisfies a Lipschitz condition on the exterior of each ball about $\mathbf{0}$, that is, on the set $E_\eta = \{\mathbf{x} : |\mathbf{x}| \ge \eta\}$ for each $\eta > 0$. To see this we observe that for any $\mathbf{x}$ and $\mathbf{y}$ in this set,

$$|\mathbf{T}(\mathbf{x}) - \mathbf{T}(\mathbf{y})| = \Big|\psi\Big(\frac{\mathbf{x}}{|\mathbf{x}|}\Big)\mathbf{x} - \psi\Big(\frac{\mathbf{y}}{|\mathbf{y}|}\Big)\mathbf{y}\Big|$$

$$
\begin{aligned}
&\le \left|\psi\Big(\frac{\mathbf{x}}{|\mathbf{x}|}\Big) - \psi\Big(\frac{\mathbf{y}}{|\mathbf{y}|}\Big)\right||\mathbf{y}| + \psi\Big(\frac{\mathbf{x}}{|\mathbf{x}|}\Big)\mathbf{x} - \mathbf{y}| \\
&\le K\left|\frac{\mathbf{x}}{|\mathbf{x}|} - \frac{\mathbf{y}}{|\mathbf{y}|}\right| + M|\mathbf{x} - \mathbf{y}| \\
&\le \Big(\frac{2K}{\eta} + M\Big)|\mathbf{x} - \mathbf{y}|.
\end{aligned}
$$

Here we have used the fact that

$$
\begin{aligned}
\left|\frac{\mathbf{x}}{|\mathbf{x}|} - \frac{\mathbf{y}}{|\mathbf{y}|}\right| &= \frac{1}{|\mathbf{x}|\,|\mathbf{y}|}\big||\mathbf{y}|\mathbf{x} - |\mathbf{x}|\mathbf{y}\big| \\
&= \frac{1}{|\mathbf{x}|\,|\mathbf{y}|}\big|(|\mathbf{y}| - |\mathbf{x}|)\mathbf{x} + |\mathbf{x}|(\mathbf{x} - \mathbf{y})\big| \\
&\le \frac{1}{|\mathbf{y}|}\big||\mathbf{y}| - |\mathbf{x}|\big| + \frac{1}{|\mathbf{y}|}|\mathbf{x} - \mathbf{y}| \\
&= \frac{2}{|\mathbf{y}|}|\mathbf{y} - \mathbf{x}|.
\end{aligned}
$$

The statements about the images of the inside of the unit ball, the unit sphere, and the outside are now obvious. For example, as remarked above, $\mathbf{y} \in \partial H$ if and only if $|\mathbf{y} - \mathbf{z}| = \psi\Big(\dfrac{\mathbf{y} - \mathbf{z}}{|\mathbf{y} - \mathbf{z}|}\Big)$. But this is equivalent to the statement that $\mathbf{T}^{-1}(\mathbf{y}) = \frac{\mathbf{y}-\mathbf{z}}{|\mathbf{y}-\mathbf{z}|}$, which says precisely that $\mathbf{T}^{-1}(\mathbf{y})$ belongs to the unit sphere.

We have now finished Stage 2 of the proof and are ready for the third and final stage.

Stage 3. For each $\delta > 0$, approximate a function $f(\mathbf{x})$ that is continuous on H by a function $f_\delta(\mathbf{x})$ that is continuous on all of R^k and such that the iterated integrals of f and f_δ differ by at most a fixed multiple of δ no matter what order they are taken in.

To that end, we first let $\delta \in (0, 1/\sqrt{k})$ be given. According to what was proved in Stage 1, the hypersphere S^{k-1} is contained in the interior of the set of hypercubes $I^k_{\delta\mathbf{m}}(\delta)$ that intersect it, and there are at most $6^{k^2}\delta^{1-k}$ of these hypercubes. In each hypercube $I^k_{\delta\mathbf{m}}(\delta)$ from this family we choose and keep fixed one point $\mathbf{x_m}$ belonging to S^{k-1}. The image of these hypercubes under $\mathbf{T}$ is a compact set containing ∂H in its interior, and each of them is contained in a hypercube of side at most $2L\sqrt{k}\delta$ centered at $\mathbf{T}(\mathbf{x_m}) \in \partial H$, where L is the Lipschitz constant for the mapping $\mathbf{T}$ on the set $E_{1-\delta}$, so that the total volume of these hypercubes is at most $6^{k^2}(2L\sqrt{k})^k\delta$. Let $c > 0$ be the distance from H to the complement of the union of these hypercubes.

We define $f_\delta(\mathbf{x})$ as a continuous function that equals $f(\mathbf{x})$ for $\mathbf{x} \in H$, while for $\mathbf{x}$ not in the interior of H we set $f_\delta(\mathbf{x}) = \max\big(0, 1 - \frac{d(\mathbf{x},H)}{c}\big)f\big(\theta(\mathbf{x})\big)$. Here $\theta(\mathbf{x})$ is the unique point of H closest to $\mathbf{x}$ and $d(\mathbf{x}, H)$ is the distance from $\mathbf{x}$ to H. On the boundary of H, where we have apparently given two definitions of f_δ we have $d(\mathbf{x}, H) = 0$, so that the two definitions are consistent. Hence the piecewise-defined function will be continuous if each of the pieces is. The piece defined on H is continuous by assumption, so that we need only concern ourselves with the second definition. It is well-known that $d(\mathbf{x}, H)$ is a continuous function of $\mathbf{x}$. It is somewhat less obvious that $\theta(\mathbf{x})$ is continuous, so that we must prove that fact.

First we show that there *is* a unique point $\theta(\mathbf{x})$ in H closest to $\mathbf{x}$. This is obvious if $\mathbf{x} \in H$, so we assume $\mathbf{x} \notin H$. Let $c = \min\{|\mathbf{x} - \mathbf{z}| : \mathbf{z} \in H\}$, and suppose $\mathbf{z}$ and $\mathbf{w}$ are two points of H such that $|\mathbf{x} - \mathbf{z}| = c = |\mathbf{x} - \mathbf{w}|$. Then the point $\mathbf{w} + t(\mathbf{z} - \mathbf{w})$ belongs to H for $0 \le t \le 1$, and so the quadratic function $|\mathbf{x} - \mathbf{w} - t(\mathbf{z} - \mathbf{w})|^2 = |\mathbf{x} - \mathbf{w}|^2 - 2t(\mathbf{x} - \mathbf{w}) \cdot (\mathbf{z} - \mathbf{w})) + t^2|\mathbf{z} - \mathbf{w}|^2$ has its minimum value c on $[0, 1]$ at *both* endpoints. But this is impossible for a non-constant quadratic function whose leading coefficient is positive. Hence the function is constant, that is, $\mathbf{z} = \mathbf{w}$. Now suppose $\mathbf{x}_n \to \mathbf{x}$. We claim $\theta(\mathbf{x}_n) \to \theta(\mathbf{x})$.

Since H is compact, we can pass to a subsequence if necessary and assume that $\theta(\mathbf{x}_n) \to \mathbf{z}$ for some point $\mathbf{z} \in H$. Certainly $|\mathbf{x}_n - \theta(\mathbf{x}_n)| \to |\mathbf{x} - \mathbf{z}|$. But $|\mathbf{x}_n - \theta(\mathbf{x}_n)| = d(\mathbf{x}_n, H) \to d(\mathbf{x}, H)$, so that $|\mathbf{x} - \mathbf{z}| = d(\mathbf{x}, H) = |\mathbf{x} - \theta(\mathbf{x})|$. As H contains only one point satisfying this equality, we must have $\mathbf{z} = \theta(\mathbf{x})$. Thus $\theta(\mathbf{x})$ is a continuous function, and therefore $f_\delta(\mathbf{x})$ is continuous on all of R^k.

It is now clear that $|f(\mathbf{x})|$ and $|f_\delta(\mathbf{x})|$ have the same maximum value, say J, and that f and f_δ differ only on the finite set of hypercubes covering ∂H. The iterated integrals of the two functions, taken in any order, over this finite set of hypercubes differ by at most $6^{k^2} JL(2\sqrt{k})^k \delta$. Thus the iterated integral of f differs from the iterated integral of f_δ by at most this amount, and since all the iterated integrals of f_δ are equal, it follows that any two iterated integrals of f differ by arbitrarily small amounts, hence are equal.

The proof is, at long last, complete.

Because this proof is so long and involved, it may be worthwhile to look at an alternative proof that works only for the case $k = 2$ and does not generalize to higher dimensions. To this end, let $k = 2$. we define two functions $m(x)$ and $M(x)$, as follows: The domain of both functions is the projection of H on the x-axis, that is, the set $\prod(H)$ consisting of x such that there exists y for which $(x, y) \in H$. By definition $m(x)$ is the minimal y for which $(x, y) \in H$, and $M(x)$ is the maximal y for which $(x, y) \in H$. We claim that these functions are continuous on $\prod(H)$. Indeed, suppose $(x^{(n)}, y^{(n)}) \in H$ and $x^{(n)} \to x$. Without loss of generality we can assume that $x^{(n)} > x$ for all x. (By passing to a subsequence if necessary, we can have either $x^{(n)} < x$ for all n or $x^{(n)} > x$ or $x^{(n)} = x$ for all n. The last case is trivial, and the other two cases are handled by identical arguments.) Some subsequence of $M(x^{(n)})$ converges to a value z. Since $(x^{(n)}, M(x^{(n)})) \in H$, and H is closed, it follows that (x, z) belongs to H. It is clear then that the assumption $z > M(x)$ contradicts the definition of $M(x)$ as the maximal number y for which $(x, y) \in H$. Hence it suffices to prove that $z \ge M(x)$. This will certainly be the case if $M(x^{(n)} \ge M(x)$ for all n. Hence assume that n_0 is an index for which $M(x^{(n_0)}) < M(x)$. Now $x^{(n_0)} > x$, since if the two were equal, $M(x^{(n_0)})$ would equal $M(x)$. We observe that if $t \in [0, 1]$, then the point $(tx^{(n_0)} + (1 - t)x, tM(x^{(n_0)}) + (1 - t)M(x))$ belongs to H. In particular, taking $t = \frac{x^{(n)}-x}{x^{(n_0)}-x}$, we find that $tx^{(n_0)} + (1-t)x = x^{(n)}$. It therefore follows that $M(x^{(n)}) \geq \frac{x^{(n)}-x}{x^{(n_0)}-x} M(x^{(n_0)} + \frac{x^{(n_0)}-x^{(n)}}{x^{(n_0)}-x} M(x) \to M(x)$. Therefore $z \geq M(x)$.

(It is this part of the argument that does not generalize to R^3, as shown by the the convex set

$$H = \{(1-t, ty, tz) : 0 \leq t \leq 1, -1 \leq y \leq 1, y^2 \leq z \leq 1\}.$$

On this set, if we define $M(y,z) = \sup\{x : (x,y,z) \in H\}$, we have $M(s,s^2) = 0$ for $s \neq 0$, but $M(0,0) = 1$.)

It now follows that $M(x)$ is continuous on H, and the proof that $m(x)$ is continuous is similar.

Now let H be a convex closed set in R^2 containing an interior point. For each $\delta > 0$, we let H_δ be the δ-neighborhood of H, that is, the set of points whose distance from H is at most δ. It is clear that H_δ is a convex set containing H in its interior. If f is a continuous function on H, we extend f to a function f_δ defined on all of R^2, as above.

By our definition

$$\int_H f(x,y)\,dy\,dx = \int_a^b \int_{m(x)}^{M(x)} f(x,y)\,dy\,dx,$$

where $[a,b]$ is the projection of H on the x-axis and for each $x \in [a,b]$

$$m(x) = \min\{t : (x,t) \in H\}$$

and

$$M(x,y) = \max\{t : (x,t) \in H\}.$$

We intend to show that the when these integrals are evaluated, the resulting value is the limit of the same integrals evaluated for f_δ, and of course the same for the integrals in reverse order. Hence these two iterated integrals are equal.

To that end, let A be the maximal value of $|f(x,y)|$, which is also the maximal value of $|f_\delta(x,y)|$. As we have set $f(x,y) = 0$ on the complement of H, the two functions $f(x,y)$ and $f_\delta(x,y)$ differ only on the set $H_\delta \setminus H$, and by no more than A at any point.

Let $P_\delta = [a - \lambda(\delta), b + \mu(\delta)]$ be the projection of H_δ on the x-axis. We claim that $\lambda(\delta)$ and $\mu(\delta)$ both tend to zero as δ tends to zero. For certainly $\lambda(\delta)$ decreases as δ decreases. Let its limit be c. There is a point $(a - \lambda(\delta), y_\delta) \in H_\delta$ for each $\delta > 0$. If y is a limit point of y_δ as $\delta \to 0$, then, since $(a - \lambda(\eta), y_\eta) \in H_\eta \subset H_\delta$ for $\eta < \delta$ and H_δ is closed, it follows that $(a - c, y) \in H_\delta$ for all $\delta > 0$, and therefore, since $\bigcap_{\delta > 0} H_\delta = H$, that $(a - c, y) \in H$. By definition of a, it then follows that $a \leq a - c \leq a$, and so $c = 0$. The proof that $\mu \to 0$ is similar.

Let $m_\delta(x)$ and $M_\delta(x)$ be the functions corresponding to $m(x)$ and $M(x)$ for H_δ. For $x \in [a,b]$ we have $m_\delta(x) < m(x) \leq M(x) < M_\delta(x)$. An argument similar to the one just given shows that $m(x) - m_\delta(x)$ and $M_\delta(x) - M(x)$ tend monotonically to zero for each $x \in [a,b]$. Since these are continuous functions, this convergence is *uniform*. let $\varphi(\delta) = \max_{x\in[a,b]}\{m(x) - m_\delta(x), M_\delta(x) - M(x)\}$, so that $\varphi(\delta) \to 0$ as $\delta \to 0$.

Now the double integrals that we wish to evaluate are

$$\int_a^b \int_{m(x)}^{M(x)} f(x,y)\,dy\,dx$$

and

$$\int_{a-\lambda(\delta)}^{b+\mu(\delta)} \int_{m_\delta(x)}^{M_\delta(x)} f_\delta(x,y)\,dy\,dx.$$

Now fix a number N larger than twice the absolute value of any coordinate of any point in H_1, and assume $\delta < \min(1,N)$. We observe that the difference between the two integrals is

$$\int_{a-\lambda(\delta)}^{a} \int_{-N}^{N} f_\delta(x,y)\,dy\,dx + \int_a^b \int_{m_\delta(x)}^{m(x)} f_\delta(x,y)\,dy\,dx$$
$$+ \int_a^b \int_{M(x)}^{M_\delta(x)} f_\delta(x,y)\,dy\,dx + \int_b^{b+\mu(\delta)} \int_{-N}^{N} f_\delta(x,y)\,dy\,dx.$$

This expression is assuredly not larger than

$$2AN\big(\lambda(\delta) + \varphi(\delta) + \mu(\delta)\big),$$

and hence it tends to zero as $\delta \to 0$. The same is true of the integral in the reverse order, and for the same reasons. Since the integral of f_δ is the same in either order, it follows that the integral of f is also the same in either order.

Exercise 10.2 For $i = 1, 2, 3, \ldots$, let $\varphi_i \in \mathcal{C}(R^1)$ have support in $(2^{-i}, 2^{1-i})$, such that $\int \varphi_i = 1$. Put

$$f(x,y) = \sum_{i=1}^{\infty} [\varphi_i(x) - \varphi_{i+1}(x)]\varphi_i(y).$$

The f has compact support in R^2, f is continuous except at $(0,0)$, and

$$\int dy \int f(x,y)\,dx = 0 \text{ but } \int dx \int f(x,y)\,dy = 1.$$

Observe that f is unbounded in every neighborhood of $(0,0)$.

Solution: The computation is straightforward:

$$\int f(x,y)\,dx = \sum_{i=1}^{\infty} \varphi_i(y)[1-1] = 0;$$

$$\int f(x,y)\,dy = \sum_{i=1}^{\infty} [\varphi_i(x) - \varphi_{i+1}(x)] = \varphi_1(x).$$

To justify the first of these, we observe that the sum is finite for each fixed y, since $\varphi_i(y) = 0$ for $i > -\log_2(y)$ if $y > 0$. Likewise the second sum is finite for each fixed x. The result now follows. The function must be unbounded, since the integral of φ_i must be 1, even though the support of that function has length 2^{-i}.

Exercise 10.3 (*a*) If $\mathbf{F}$ is as in Theorem 10.7, put $\mathbf{A} = \mathbf{F}'(0)$, $\mathbf{F}_1(\mathbf{x}) = \mathbf{A}^{-1}\mathbf{F}(\mathbf{x})$. Then $\mathbf{F}_1'(\mathbf{0}) = I$. Show that

$$\mathbf{F}_1(\mathbf{x}) = \mathbf{G}_n \circ \mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_1(\mathbf{x})$$

in some neighborhood of $\mathbf{0}$, for certain primitive mappings $\mathbf{G}_1, \ldots, \mathbf{G}_n$. This gives another version of Theorem 10.7:

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}'(\mathbf{0})\mathbf{G}_n \circ \mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_1(\mathbf{x}).$$

(*b*) Prove that the mapping $(x, y) \to (y, x)$ of R^2 onto R^2 is not the composition of any two primitive mappings, in any neighborhood of the origin. (This shows that the flips B_1 cannot be omitted from the statement of Theorem 10.7.)

Solution: (*a*) According to the proof of Theorem 10.7, the flips are needed only to interchange m and k, where k is the first index not less than m for which $D_m\alpha_k(\mathbf{0}) \neq 0$. Here

$$\mathbf{F}_m'(\mathbf{0})\mathbf{e}_m = \sum_{i=m}^{n} (D_m\alpha_i)(\mathbf{0})\mathbf{e}_i.$$

But in that proof $\mathbf{F}_1 = \mathbf{F}$, and since in the present case $\mathbf{F}'(\mathbf{0})$ is the identity, B_1 is the identity. But then the definition of $\mathbf{G}_1(\mathbf{x})$ as

$$\mathbf{G}_1(\mathbf{x}) = \mathbf{x} + \big[\alpha_1(\mathbf{x}) - x_1\big]\mathbf{e}_1$$

implies that $\mathbf{G}_1'(\mathbf{0})$ is also the identity. Suppose we know that B_j, $\mathbf{F}_m'(\mathbf{0})$, and $\mathbf{G}_j'(\mathbf{0})$ are all equal to the identity for $j \le m$. Then the inductive definition of $\mathbf{F}_{m+1}$ as $\mathbf{F}_{m+1}(\mathbf{y}) = \mathbf{F}_m \circ \mathbf{G}_m^{-1}(\mathbf{y})$ implies that $\mathbf{F}_{m+1}'(\mathbf{0})$ is also the identity, from which it then follows that $\mathbf{F}_{m+1}'(\mathbf{0})$, B_{m+1}, and $\mathbf{G}_{m+1}'(\mathbf{0})$ are all equal to the identity. Thus the decomposition of $\mathbf{F}_1$ involves no flips, as asserted.

(*b*) If this map were a composition of two primitive maps, its derivative at $(0, 0)$ would be the product of two matrices of the form

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ c & d \end{pmatrix} = \begin{pmatrix} a + bc & bd \\ c & d \end{pmatrix}.$$

Since this matrix must be $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, it follows that $c = 1$, $d = 0$. But then the second column of the product of the two matrices is zero, which is a contradiction.

Exercise 10.4 For $(x, y) \in R^2$, define

$$\mathbf{F}(x, y) = (e^x \cos y - 1, e^x \sin y).$$

Prove that $\mathbf{F} = \mathbf{G}_2 \circ \mathbf{G}_1$, where

$$\begin{aligned} \mathbf{G}_1(x, y) &= (e^x \cos y - 1, y) \\ \mathbf{G}_2(x, y) &= (u, (1 + u) \tan v) \end{aligned}$$

are primitive in some neighborhood of $(0, 0)$.

Compute the Jacobians of $\mathbf{G}_1$, $\mathbf{G}_2$, and $\mathbf{F}$ at $(0, 0)$. Define

$$\mathbf{H}_2(x, y) = (x, e^x \sin y)$$

and find

$$\mathbf{H}_1(u, v) = (h(u, v), v)$$

so that $\mathbf{F} = \mathbf{H}_1 \circ \mathbf{H}_2$ in some neighborhood of $(0, 0)$.

Solution: The equation $\mathbf{F} = \mathbf{G}_2 \circ \mathbf{G}_1$ is a routine computation, and the fact that $\mathbf{G}_1$ and $\mathbf{G}_2$ are primitive is immediate.

The Jacobians of $\mathbf{G}_1$ and $\mathbf{G}_2$ are

$$\mathbf{G}_1'(x, y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ 0 & 1 \end{pmatrix}, \quad \mathbf{G}_2'(u, v) = \begin{pmatrix} 1 & 0 \\ \tan v & (1 + u) \sec^2 v \end{pmatrix},$$

so that each of them equals the identity at $(0, 0)$. It therefore follows that $\mathbf{F}'(0, 0) = I$ also.

If we take $h(u, v) = (\sqrt{e^{2u} - v^2} - 1, v)$, the primitive mapping $\mathbf{H}_1(u, v) = (h(u, v), v)$ will yield $\mathbf{H}_1 \circ \mathbf{H}_2 = \mathbf{F}$.

Exercise 10.5 Formulate and prove an analogue of Theorem 10.8, in which K is a compact subset of an arbitrary metric space. (Replace the functions φ_i that occur in the proof of Theorem 10.8 by functions of the type constructed in Exercise 22 of Chap. 4.)

Solution: We are given a compact set K in a metric space X (say with metric d) and a cover of K by open sets V_i, $i = 1, 2, \ldots, n$. (We may as well assume a finite number of sets, since we can find a finite subcover of any infinite cover.)

We need to construct continuous functions ψ_i, $i = 1, 2, \ldots, n$ such that $0 \leq \psi_i(x) \leq 1$ for all i and all $x \in X$, the support of $\psi_i(x)$ is contained in V_i, and $\sum_{i=1}^{n} \psi_i(x) = 1$ for all $x \in K$.

To do this, let $\eta > 0$ be a Lebesgue number for the covering of K by the sets V_i, that is such that the η-neighborhood of every point $x \in K$ is contained in some V_i. Let $\varepsilon \in (0, \eta)$, and let U_i be the set of points whose distance from the complement of V_i is larger than ε and W_i the set of points whose distance from the complement of V_i is larger than $\frac{\epsilon}{2}$. Since the distance from x to the complement of V_i is a continuous function of x, it follows that U_i and W_i are open sets. It is obvious that the closure of U_i is contained in W_i and the closure of W_i is contained in V_i. We note that $K \subset \bigcup_{i=1}^{n} U_i$. For if $x \in K$ there exists V_i such that the η-neighborhood of x is contained in V_i, and hence the distance from x to the complement of that V_i is at least η.

Now let A_i be the closure of U_i, and B_i the complement of W_i. Define

$$\varphi_i(x) = \frac{d(x, B_i)}{d(x, A_i) + d(x, B_i)}.$$

Then $\varphi_i(x)$ is 1 on A_i (and hence certainly on U_i) and 0 on B_i, $\varphi_i(x)$ is continuous, and $0 \le \varphi_i(x) \le 1$ for all x. Since the support of $\varphi_i(x)$ is the closure of W_i, it is contained in V_i. Since $\varphi_i(x) > 0$ for x in W_i, the sum $\varphi(x) = \sum_{i=1}^{n} \varphi_i(x)$ is positive on the open set $W = \bigcup_{i=1}^{n} W_i$, which contains K. Now let L be the complement of W, and define a continuous function $\psi(x)$ by

$$\psi(x) = \frac{d(x, L)}{d(x, K) + d(x, L)},$$

so that $0 \le \psi(x) \le 1$ for all x, $\psi(x) = 1$ if $x \in K$, and $\psi(x) = 0$ if $x \in L$. If we now define $\psi_i(x) = 0$ for $x \notin W$ and

$$\psi_i(x) = \frac{\varphi_i(x)\psi(x)}{\varphi(x)},$$

then $\psi_i(x)$ is continuous on the entire space. Its restriction to L is continuous. If we can show that its restriction to the closure of W is continuous, we shall be done. But it is obvious that it is continuous on W itself, and so we need only show that it is continuous at a point of ∂W. Hence let $x_n \to x \in \partial W$. Since $\varphi_i(x)/\varphi(x)$ is bounded, and $\psi(x_n) \to 0$, it follows that $\psi_i(x_n) \to 0 = \psi_i(x)$, and hence ψ_i is continuous at x.

The construction is now complete.

Exercise 10.6 Strengthen the conclusion of Theorem 10.8 by showing that the functions ψ_i can be made differentiable, and even infinitely differentiable. (Use Exercise 1 of Chap. 8 in the construction of the auxiliary functions φ_i.)

Solution: The function $\varphi_i(\mathbf{x})$ is required to have only three properties: 1) $\varphi_i(\mathbf{x}) = 1$ for $|\mathbf{x} - \mathbf{a}_i| \le r_i$; 2) $\varphi_i(\mathbf{x}) = 0$ for $|\mathbf{x} - \mathbf{a}_i| \ge s_i$; 3) $0 =\le \varphi_i(\mathbf{x}) \le 1$ for all $\mathbf{x}$. These properties can be achieved with an infinitely differentiable function $\varphi_i(\mathbf{x})$. To construct such a function, we go to the function $f(t)$ in Exercise 1 of Chapter 8, namely

$$f(t) = e^{-\frac{1}{t^2}}$$

for $t \neq 0$ and $f(0) = 0 = f^{(n)}(0)$ for all positive integers n, $f^{(n)}(t)$ being the nth derivative of $f(t)$. It was established in that exercise that $f(t)$ is infinitely differentiable, and it is obvious that $f(t)$ is strictly increasing for nonnegative values of t.

Let

$$g(t) = \frac{f\big(f(1) - f(t)\big)}{f\big(f(1)\big)}.$$

Then it is obvious that $g(t)$ is an infinitely differentiable function that decreases from 1 to 0 as x increases from 0 to 1. If we show that $g^{(n)}(0) = 0 = g^{(n)}(1)$ for all positive integers n, it will follow that the function

$$h(t) = \begin{cases} 1, & t \leq 0, \\ g(t), & 0 \leq t \leq 1, \\ 0, & 1 \leq t \end{cases}$$

is also a $\mathcal{C}^\infty$ function, and we can then take

$$\varphi(\mathbf{x}) = h\Big(\frac{|\mathbf{x}|^2 - r_i^2}{s_i^2 - r_i^2}\Big).$$

But it is easy to prove these properties by showing inductively that for all integers j and k with $0 \leq j \leq n-k$ and $1 \leq k \leq n$ there exist infinitely differentiable functions $\theta_{j,k,n}(t)$ such that

$$g^{(n)}(t) = \sum_{\substack{0 \leq j \leq n-k \\ 1 \leq k \leq n}} \theta_{j,k,n}(t) f^{(k)}\big(f(1) - f(t)\big) f^{(n-k-j+1)}(t).$$

In fact the chain rule shows that

$$\theta_{0,1,1}(t) = -\frac{1}{f\big(f(1)\big)}.$$

Then, assuming there exist such functions $\theta_{j,k,n}(t)$, we find

$$\begin{aligned} g^{(n+1)}(t) \quad = \quad & \sum_{\substack{0 \leq j \leq n-k \\ 1 \leq k \leq n}} \Big\{ \theta'_{j,k,n}(t) f^{(k)}\big(f(1) - f(t)\big) f^{(n-k-j+1)}(t) \\ & + \theta_{j,k,n}(t)\big(-f'(t)\big) f^{(k+1)}\big(f(1) - f(t)\big) f^{n-k-j+1}(t) \\ & + \theta_{j,k,n}(t) f^{(k)}\big(f(1) - f(t)\big) f^{(n-k-j+2)}. \Big\} \end{aligned}$$

Each term in this expression contains a factor $f^{(s)}\big(f(1) - f(t)\big) f^{(n+1-s-r+1)}(t)$ with $0 \leq r \leq n+1-s$, $1 \leq s \leq n+1$ and with a coefficient that is infinitely differentiable. Thus when suitably rearranged, this sum has the appropriate form

$$g^{(n+1)}(t) = \sum_{\substack{0 \leq j \leq n+1-k \\ 1 \leq k \leq n+1}} \theta_{j,k,n+1}(t) f^{(k)}\big(f(1) - f(t)\big) f^{(n-k-j+1)}(t)$$

with infinitely differentiable functions $\theta_{j,k,n+1}$. Since each term contains a factor $f^{(k)}(f(1)-f(t))f^{(l)}(t)$ with $k \geq 1$, it follows that each term vanishes when $t = 0$ or $t = 1$, and hence that $g^{(n)}(1) = 0 = g^{(n)}(0)$ for $n = 1, 2, \ldots$.

Exercise 10.7 (*a*) Show that the simplex Q^k is the smallest convex subset of R^k that contains $\mathbf{0}, \mathbf{e}_1, \ldots, \mathbf{e}_k$.

(*b*) Show that affine mappings take convex sets to convex sets.

Solution: (*a*) By definition $Q^k = \{\mathbf{x} : x_1 + \cdots x_k \leq 1, x_j \geq 0, j = 1, \ldots, k\}$. It is obvious that Q^k contains all the points $\mathbf{0}, \mathbf{e}_1, \ldots, \mathbf{e}_k$. It is nearly obvious that Q^k is convex. Indeed, if $\mathbf{x}$ and $\mathbf{y}$ are points of Q^k and $0 < t < 1$, then $t\mathbf{x} + (1-t)\mathbf{y} = \mathbf{z}$, where $z_j = tx_j + (1-t)y_j$. Since $x_j \geq 0$ and $y_j \geq 0$ and $0 < t < 1$, it is clear that $z_j \geq 0$. Simple algebra shows that $z_1 + \cdots + z_k \leq t + (1-t) = 1$, so that $\mathbf{z} \in Q^k$ also. Thus Q^k is convex.

Now let C be any convex set containing these points, and let $\mathbf{x} \in Q^k$. We need to show that $\mathbf{x} \in C$. We shall show by induction that the point $x_1\mathbf{e}_1 + \cdots + x_j\mathbf{e}_j$ is in C whenever $x_1 \geq 0, \ldots, x_j \geq 0$ and $x_1 + \cdots + x_j \leq 1$. If $j = 1$, this is obvious, since $x_1\mathbf{e}_1 = x_1\mathbf{e}_1 + (1 - x_1)\mathbf{0}$ and by assumption $0 \leq x_1 \leq 1$.

Suppose the theorem is true for j, and let $c = x_1 + \cdots + x_{j+1} \leq 1$, $x_1 \geq 0, \ldots, x_{j+1} \geq 0$. If $c = 0$, the point $x_1\mathbf{e}_1 + \cdots + x_{j+1}\mathbf{e}_{j+1}$ is $\mathbf{0}$, and hence belongs to C. Therefore we assume $c > 0$. Since $\mathbf{e}_{j+1} \in C$, we need only consider the case $x_{j+1} < 1$. By the induction assumption, taking $x_l' = \frac{x_l}{1-x_{j+1}}$ for $l = 1, \ldots, j$, we find that the point $\mathbf{y} = x_1'\mathbf{e}_1 + \cdots + x_j'\mathbf{e}_j$ belongs to C, and therefore the point $(1 - x_{j+1})\mathbf{y} + x_{j+1}\mathbf{e}_{j+1} = x_1\mathbf{e}_1 + \cdots + x_j\mathbf{e}_j + x_{j+1}\mathbf{e}_{j+1}$ does also.

(*b*) Let $\mathbf{A}(\mathbf{x})$ be an affine mapping, that is, $\mathbf{A}(\mathbf{x}) = \mathbf{x}_0 + \mathbf{T}(\mathbf{x})$, where $\mathbf{T}(\mathbf{x})$ is a linear transformation, let C be any convex set, and let $\mathbf{u} \in \mathbf{A}(C)$, $\mathbf{v} \in \mathbf{A}(C)$. We need to show that $t\mathbf{u} + (1-t)\mathbf{v} \in \mathbf{A}(C)$ for all $t \in (0,1)$. But this is trivial, since if $\mathbf{u} = \mathbf{A}(\mathbf{x})$ and $\mathbf{v} = \mathbf{A}(\mathbf{y})$, then $t\mathbf{u} + (1-t)\mathbf{v} = \mathbf{A}(t\mathbf{x} + (1-t)\mathbf{y})$ and $t\mathbf{x} + (1-t)\mathbf{y} \in C$.

Exercise 10.8 Let H be the parallelogram in R^2 whose vertices are $(1,1)$, $(3,2)$, $(4,5)$, $(2,4)$. Find the affine map T which sends $(0,0)$ to $(1,1)$, $(1,0)$ to $(3,2)$, $(0,1)$ to $(2,4)$. Show that $J_T = 5$. Use T to convert the integral

$$\alpha = \int_H e^{x-y}\,dx\,dy$$

to an integral over I^2 and thus compute α.

Solution: Clearly the constant term in an affine mapping is the image of $(0,0)$, which in the present case is to be $(1,1)$. Thus we are looking for a linear transformation L such that $(3,2) = (1,1) + L(1,0)$ and $(2,4) = (1,1) + L(0,1)$, which is to say $L(1,0) = (2,1)$ and $L(0,1) = (1,3)$. Obviously $L(x,y) = (2x+y, x+3y)$. Then $J_T = 2 \cdot 3 - 1 \cdot 1 = 5$. The inverse of T is given by $T^{-1}(u,v) = L^{-1}((u,v)-(1,1))$. Simple algebra then reveals that

$$T^{-1}(u,v) = \Big(\frac{-2+3u-v}{5}, \frac{-1-u+2v}{5}\Big).$$

The parallelogram H is the image of the unit square S under T, and so

$$\alpha = \int_{T(S)} e^{x-y}\, dx\, dy = \int_S e^{T^{-1}(u,v)} |J_{T^{-1}}|\, du\, dv.$$

Thus

$$\begin{aligned}\alpha &= \int_0^1 \int_0^1 e^{\frac{-1+4u-3v}{5}} \frac{1}{5}\, du\, dv \\ &= \frac{e^{-\frac{1}{5}}}{5} \int_0^1 e^{\frac{4u}{5}}\, du \int_0^1 e^{\frac{-3v}{5}}\, dv \\ &= e^{-\frac{1}{5}} \cdot \frac{5}{4} \cdot (e^{\frac{4}{5}} - 1) \cdot \Big(\frac{-5}{3}\Big) \cdot (e^{\frac{-3}{5}} - 1).\end{aligned}$$

Exercise 10.9 Define $(x,y) = T(r,\theta)$ on the rectangle

$$0 \le r \le a, \quad 0 \le \theta \le 2\pi$$

by the equations

$$x = r\cos\theta, \quad y = r\sin\theta.$$

Show that T maps this rectangle onto the closed disc D with center at $(0,0)$ and radius a, that T is one-to-one on the interior of the rectangle, and that $J_T(r,\theta) = r$. If $f \in \mathcal{C}(D)$, prove the formula for integration in polar coordinates:

$$\int_D f(x,y)\, dx\, dy = \int_0^a \int_0^{2\pi} f(T(r,\theta)) r\, dr\, d\theta.$$

Hint: Let D_0 be the interior of D, minus the interval from $(0,0)$ to $(0,a)$. As it stands, Theorem 10.9 applies to continuous functions whose support lies in D_0. To remove this restriction, proceed as in Example 10.4.

Solution: The simple geometry of this transformation allows a fairly straightforward proof. Let $\varepsilon \in (0, \min(\pi, a/2))$. Let $H_\varepsilon = \{(r,\theta) : \varepsilon \le r \le a-\varepsilon, \varepsilon \le \theta \le 2\pi - \varepsilon\}$. The transformation T is one-to-one on H_ε. Let $\varphi_\varepsilon(x,y)$ be a continuous function on all of R^2 such that $\varphi_\varepsilon(x,y) = 1$ for $(x,y) \in T(H_\varepsilon)$, $\varphi_\varepsilon(x,y) = 0$ for $(x,y) \notin T(H_{\varepsilon/2})$ and $0 \le \varphi_\varepsilon(x,y) \le 1$. Define $f_\varepsilon(x,y) = f(x,y)\varphi_\varepsilon(x,y)$ for $(x,y) \in D$ and $f_\varepsilon(x,y) = 0$ for $(x,y) \notin D$. Then $f_\varepsilon(x,y) = f(x,y)$ except for $(x,y) \in D \setminus T(H_\varepsilon)$. Hence $f_\varepsilon(T(x,y)) = f(x,y)$ on $[0,a] \times [0,2\pi] \setminus H_\varepsilon$. Let

M be the maximum of $|f(x,y)|$ on D. Since the support of f_ε is contained in $T(H_{\varepsilon/2})$, which in turn is contained in D_0, we certainly have

$$\int_{R^2} f_\varepsilon(x,y)\,dx\,dy = \int_{R^2} f_\varepsilon(r\cos\theta, r\sin\theta)r\,dr\,d\theta.$$

We need to see how much each of these integrals differs from the corresponding integral of f. We first look at $f_\varepsilon(x,y)$. In evaluating its integral we can confine ourselves to the square $-a \le x \le a$, $-a \le y \le a$, since D is contained in that square. We first exclude the three intervals $-a \le y \le -a+\varepsilon$, $\min(-\varepsilon, -a\sin\varepsilon) \le y \le \max(\varepsilon, a\sin\varepsilon)$, and $a-\varepsilon \le y \le a$. When y is not in these intervals, we have $\varepsilon^2 \le y^2 \le (a-\varepsilon)^2$, and $f(x,y)$ and $f_\varepsilon(x,y)$ can differ only on the two intervals where $\sqrt{(a-\varepsilon)^2 - y^2} \le |x| \le \sqrt{a^2-y^2}$, each of which has length

$$\begin{aligned}\sqrt{a^2-y^2} - \sqrt{(a-\varepsilon)^2-y^2} &= \frac{(a^2-y^2) - ((a-\varepsilon)^2 - y^2)}{\sqrt{a^2-y^2}+\sqrt{(a-\varepsilon)^2-y^2}}\\ &\le \frac{2a\varepsilon}{\sqrt{a^2-y^2}} \le \frac{2a\varepsilon}{\sqrt{2a\varepsilon+\varepsilon^2}} \le \frac{2a}{\sqrt{2a+\varepsilon}}\sqrt{\varepsilon} \le \sqrt{2a\varepsilon}.\end{aligned}$$

Since the maximum possible difference between $f(x,y)$ and $f_\varepsilon(x,y)$ is M, we see that

$$\left|\int f(x,y)\,dx - \int f_\varepsilon(x,y)\,dx\right| \le 2M\sqrt{2a\varepsilon}$$

if y is not in one of the three excluded intervals.

If y is in one of the three excluded intervals, since f and f_ε can differ by at most M, we have

$$\left|\int f(x,y)\,dx - \int f_\varepsilon(x,y)\,dx\right| \le 2Ma.$$

Since the total length of the excluded y-intervals is at most $(2a+3)\varepsilon$, and the total length of the interval over which y varies is at most $2a$, we see that

$$\left|\int\int f(x,y)\,dx\,dy - \int\int f_\varepsilon(x,y)\,dx\,dy\right| \le 4Ma\sqrt{2a\varepsilon} + 2Ma(2a+3)\varepsilon.$$

Thus this approximation can be made arbitrarily good by taking ε sufficiently small.

As for the integral with respect to r, θ, we observe that we can confine ourselves to the rectangle $0 \le r \le a$, $0 \le \theta \le 2\pi$, and that $f_\varepsilon(r\cos\theta, r\sin\theta) = f(r\cos\theta, r\sin\theta)$ for $\varepsilon \le r \le a-\varepsilon$ and $\varepsilon \le \theta \le 2\pi - \varepsilon$. Thus, excluding the intervals $0 \le \theta \le \varepsilon$, and $2\pi - \varepsilon \le \theta \le 2\pi$, we find that for θ not in these intervals $f(r\cos\theta, r\sin\theta)$ and $f_\varepsilon(r\cos\theta, r\sin\theta)$ can differ (by at most M) only on the two intervals $0 \le r \le \varepsilon$ and $a - \varepsilon \le r \le a$. Hence as before, if θ is not in one of these two intervals, then

$$\left|\int f(r\cos\theta, r\sin\theta)r\,dr - \int f_\varepsilon(r\cos\theta, r\sin\theta)r\,dr\right| \le 2Ma\varepsilon.$$

On the other hand, if θ is in one of these two intervals, we have

$$\left| \int f(r\cos\theta, r\sin\theta) r\,dr - \int f_\varepsilon(r\cos\theta, r\sin\theta) r\,dr \right| \le Ma^2.$$

Since the exceptional intervals have total length 2ε and the total length of the θ interval is 2π, we see that

$$\left| \iint f(r\cos\theta, r\sin\theta) r\,dr\,d\theta - \iint f_\varepsilon(r\cos\theta, r\sin\theta) r\,dr\,d\theta \right| \le 4\pi Ma\varepsilon + 2Ma^2\varepsilon.$$

Hence these two integrals also can be made arbitrarily close together by choosing ε sufficiently small. Since the two integrals of f_ε are equal for each $\varepsilon > 0$, it follows that the other two are also equal.

Exercise 10.10 Let $a \to \infty$ in Exercise 9, and prove that

$$\int_{R^2} f(x,y)\,dx\,dy = \int_0^\infty \int_0^{2\pi} f\big(T(r,\theta)\big) r\,d\theta\,dr,$$

for continuous functions f that decrease sufficiently rapidly as $|x| + |y| \to \infty$. (Find a more precise formulation.) Apply this to

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

to derive formula (101) of Chap. 8.

Solution: Without striving for ultimate generality, we shall assume that there are positive numbers K and δ such that $|f(x,y)| \le K(x^2+y^2)^{-1-\delta}$ for all $(x,y) \ne (0,0)$. (Such an estimate holds for (x,y) ranging over any *bounded* set merely because $f(x,y)$ is continuous.) Let $D_a = \{(x,y) : 0 \le x^2 + y^2 \le a^2\}$ and $S_a = \{(x,y) : |x| \le a, |y| \le a\}$. Since both D_a and S_a are convex sets, the functions $g_a(x,y) = \chi_{D_a}(x,y) f(x,y)$ and $h_a(x,y) = \chi_{S_a}(x,y) f(x,y)$ are both integrable over R^2. We shall show that

$$\lim_{a\to\infty} \int_{R^2} g_a(x,y)\,dx\,dy = \int_{R^2} f(x,y)\,dx\,dy = \lim_{a\to\infty} \int_{R^2} h_a(x,y)\,dx\,dy.$$

Our job is simpler if we first show that

$$\lim_{a\to\infty} \left(\int_{R^2} g_a(x,y)\,dx\,dy - \int_{R^2} h_a(x,y)\,dx\,dy \right) = 0.$$

As before, we let $M = \sup\{|f(x,y)|\}$. Since $g_a(x,y) = h_a(x,y)$ except for $(x,y) \in S_a \setminus D_a$, and on this set $g_a(x,y) = 0$ and $|h_a(x,y)| \le Ka^{-2-2\delta}$, the maximum possible difference in these two integrals is $4Ka^{-2\delta}$, which does indeed tend to zero as $a \to \infty$.

It now suffices to show only the second of the two equalities given above, i.e., that

$$\int_{R^2} f(x,y)\,dx\,dy = \lim_{a\to\infty} \int_{R^2} h_a(x,y)\,dx\,dy.$$

To that end, we fix y. We then have, if $|y| \geq a$, so that $h_a(x,y) = 0$,

$$\begin{aligned}\int_{-\infty}^{\infty} f(x,y) - h_a(x,y)\,dx &\leq K \int_{-\infty}^{\infty} \frac{1}{(x^2+y^2)^{1+\delta}}\,dx \\ &\leq \int_{-\infty}^{-|y|} \frac{K}{(x^2)^{1+\delta}}\,dx + \int_{-|y|}^{|y|} \frac{K}{(y^2)^{1+\delta}}\,dx + \int_{|y|}^{\infty} \frac{K}{(x^2)^{1+\delta}}\,dx \\ &\leq \frac{2K|y|^{-1-2\delta}}{1+2\delta} + 2K|y|^{-1-2\delta} \leq 4K|y|^{-1-2\delta}.\end{aligned}$$

If $|y| \leq a$, we note that $f(x,y) = h_a(x,y)$ for $-a \leq x \leq a$, and so

$$\begin{aligned}\int_{-\infty}^{\infty} f(x,y) - h_a(x,y)\,dx &\leq \int_{-\infty}^{-a} \frac{K}{(x^2)^{1+\delta}}\,dx + \int_{a}^{\infty} \frac{K}{(x^2)^{1+\delta}}\,dx \\ &\leq \frac{2Ka^{-1-2\delta}}{1+2\delta} \leq 2Ka^{-1-2\delta}.\end{aligned}$$

Applying these two inequalities we find that

$$\begin{aligned}\left|\int_{R^2} f(x,y) - h_a(x,y)\,dx\,dy\right| &\leq 4K \int_{-\infty}^{-a} |y|^{-1-2\delta}\,dy + 4Ka^{-2\delta} + \\ &+ 4K \int_{a}^{\infty} y^{-1-2\delta}\,dy \leq 4K\Big(1+\frac{1}{\delta}\Big)a^{-2\delta}.\end{aligned}$$

The desired formula is now proved by merely remarking that

$$\int_{R^2} h_a(x,y)\,dx\,dy = \int_0^a \int_0^{2\pi} f(r\cos\theta, r\sin\theta) r\,d\theta\,dr.$$

The fact that the limit on the right-hand side exists as $a \to \infty$ follows from the fact that the limit on the left-hand side does, but can also be proved directly, since $|f(r\cos\theta, r\sin\theta)r| \leq Kr^{-1-2\delta}$.

Applying this formula with $f(x,y) = e^{-x^2-y^2}$, we find that

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-t^2}\,dt\right)^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2} e^{-y^2}\,dx\,dy = \int_0^{\infty} \int_0^{2\pi} e^{-r^2} r\,d\theta\,dr = \\ &= 2\pi \int_0^{\infty} e^{-r^2} r\,dr = \pi \int_0^{\infty} e^{-u}\,du = \pi.\end{aligned}$$

In other words,

$$\int_{-\infty}^{\infty} e^{-t^2}\,dt = \sqrt{\pi}.$$

Exercise 10.11 Define $(u, v) = T(s, t)$ on the strip

$$0 < s < \infty, \quad 0 < t < 1$$

by setting $u = s - st$, $v = st$. Show that T is a 1-1 mapping of the strip onto the positive quadrant Q in R^2. Show that $J_T(s, t) = s$.

For $x > 0$, $y > 0$ integrate

$$u^{x-1}e^{-u}v^{y-1}e^{-v}$$

over Q, use Theorem 10.9 to convert the integral to one over the strip, and derive formula (96) of Chap. 8 in this way.

(For this application, Theorem 10.9 has to be extended so as to cover certain improper integrals. Provide this extension.)

Solution: It is easy to compute the inverse of T, namely

$$s = u + v, \quad t = \frac{v}{u+v},$$

and this inverse is defined on the entire (u, v)-plane with the line $v = -u$ removed. It is obvious that v is positive if and only if s and t have the same sign, and that u is positive if and only if s and $1 - t$ have the same sign.

Thus if u and v are both positive, then t and $1-t$ have the same sign, which happens if and only if $0 < t < 1$. In this case s must also be positive. Conversely, the equations that give s and t show that if u and v are both positive, then $s > 0$ and $0 < t < 1$. The Jacobian matrix of T is

$$\begin{pmatrix} 1-t & -s \\ t & s \end{pmatrix},$$

so that $J_T(s, t) = s$.

The integral of $u^{x-1}e^{-u}v^{y-1}e^{-v}$ over the quadrant is

$$\int_0^\infty u^{x-1}e^{-u}\,du \int_0^\infty v^{y-1}e^{-v}\,dv = \Gamma(x)\Gamma(y).$$

According to Theorem 10.9

$$\int_0^\infty \int_0^1 f(s - st, st)s\,dt\,ds = \int_0^\infty \int_0^\infty f(u, v)\,du\,dv$$

for any function $f(u, v)$ having compact support contained in the open quadrant. Assuming this theorem remains valid for the particular function we have in mind, we get

$$\Gamma(x)\Gamma(y) = \int_0^\infty s^{x+y-1}e^{-s}\,ds \int_0^1 t^{y-1}(1-t)^{x-1}\,dt =$$

$$= \Gamma(x+y) \int_0^1 t^{x-1}(1-t)^{y-1}\,dt,$$

which is indeed formula (96) of Chapter 8.

Thus we need only justify the use of Theorem 10.9 with the function f in the unbounded regions. To do this, we first show that Theorem 10.9 applies to the function $f(u,v)\varphi_\delta(u,v)$, where $\varphi_\delta(u,v)$ is the characteristic function of the set

$$E_\delta = \{(u,v) : \delta \le u \le \delta^{-1},\, \delta \le v \le \delta^{-1}\}.$$

Since this function is positive on E_δ, it is easy to modify it and make it into a continuous nonnegative function f_η that vanishes outside the set $E_{\delta-\eta}$, for $\eta < \delta$ and this can be done without increasing its maximal value. Theorem 10.9 applies to f_η, and it is easy to see that the integral of f_η on both sides of the formula tends to the integral of $f\varphi_\delta$ as $\eta \to \delta$. (Indeed, there is a constant ε such that $T^{-1}(u,v) = (s,t)$ lies in the strip $\varepsilon \le t \le 1-\varepsilon$ whenever $(u,v) \in E_{\delta-\eta}$ and $\eta \le \frac{\delta}{2}$. In that case, for each fixed t, the distance between the rightmost points (s,t) in $T^{-1}(E_{\delta-\eta})$ and in $T^{-1}(E_\delta)$ is at most $\frac{\delta-\eta}{\varepsilon}$. A similar statement applies to the leftmost points in the two regions, showing that the usual argument applies: The integrals of f and $f\varphi_\delta$ over each horizontal line differ by at most $\frac{2M(\delta-\eta)}{\varepsilon}$, except for a small range of t whose length tends to zero with $\delta-\eta$, on which the difference is bounded. It then follows that both of the integrals of f_η tend toward the corresponding integrals of $f\varphi_\delta$

It then remains only to prove that the integral of $f\varphi_\delta$ tends to the integral of f on both sides of the formula. Since these integrals increase as δ decreases, there is no question that the limit exists, and we need only show that in both cases the limit is the integral in the formula. This is nearly immediate in the case of the integral over the quadrant. As for the integral over the strip, the set $T^{-1}(E_\delta)$ contains the region $\delta^{1/2} \le s \le \frac{1}{\delta-\delta^{3/2}}$, $\delta^{1/2} \le t \le 1-\delta^{1/2}$. For these inequalities imply that $\delta \le st \le \frac{1}{\delta}$, and since $1-t$ satisfies the same inequalities as t, we also have $\delta \le s(1-t) \le \frac{1}{\delta}$. The integral of $f(s-st,st)s$ over the two strips $0 \le t \le \sqrt{\delta}$ and $1-\sqrt{\delta} \le t \le 1$ tends to zero with δ, and for each t with $\sqrt{\delta} \le t \le 1-\sqrt{\delta}$ the integral

$$\int_{\sqrt{\delta}}^{1/(\delta-\delta^{3/2})} f(s-st,st)s\,ds$$

differs from the integral from 0 to ∞ by less than

$$t^{1-x}(1-t)^{1-y}\Big(\int_0^{\sqrt{\delta}} s^{x+y-1}\,ds + \int_{1/(\delta-\delta^{3/2})}^{\infty} s^{x+y-1}e^{-s}\,ds\Big).$$

The first of these integrals is explicitly calculable and tends to zero as $\delta \to 0$. In the second we use the fact that $e^{-s} < \frac{n!}{s^n}$ for all $s > 0$ and take $n \ge x+y+1$. It then follows that the integral of $f(s-st,st)\varphi_\delta(s-st,st)s$ over each of these horizontal line differs from the integral of $f(s-st,st)s$ by an amount that tends to zero uniformly for $\sqrt{\delta} \le t \le 1-\sqrt{\delta}$.

The proof is now complete.

Exercise 10.12 Let I^k be the set of all $\mathbf{u} = (u_1, \ldots, u_k) \in R^k$ with $0 \le u_i \le 1$ for all i; let Q^k be the set of all $\mathbf{x} = (x_1, \ldots, x_k) \in R^k$ with $x_i \ge 0$, $\sum x_k \le 1$. (I^k is the unit cube; Q^k is the standard simplex in R^k.) Define $\mathbf{x} = T(\mathbf{u})$ by

$$\begin{aligned} x_1 &= u_1 \\ x_2 &= (1-u_1)u_2 \\ &\cdots\cdots\cdots\cdots \\ x_k &= (1-u_1)\cdots(1-u_{k-1})u_k. \end{aligned}$$

Show that

$$\sum_{i=1}^{k} x_i = 1 - \prod_{i=1}^{k}(1-u_i).$$

Show that T maps I^k onto Q^k, that T is 1-1 in the interior of I^k, and that its inverse S is defined in the interior of Q^k by $u_1 = x_1$ and

$$u_i = \frac{x_i}{1 - x_1 - \cdots - x_{i-1}}$$

for $i = 2, \ldots, k$. Show that

$$J_T(\mathbf{u}) = (1-u_1)^{k-1}(1-u_2)^{k-2}\cdots(1-u_{k-1}),$$

and

$$J_S(\mathbf{x}) = \left[(1-x_1)(1-x_1-x_2)\cdots(1-x_1-\cdots-x_{k-1})\right]^{-1}.$$

Solution: The first identity is easily proved by induction on k. It is obvious for $k = 1$, and

$$\begin{aligned} \sum_{i=1}^{k+1} x_i &= x_{k+1} + \sum_{i=1}^{k} x_k \\ &= (1-u_1)\cdots(1-u_k)u_{k+1} + 1 - (1-u_1)\cdots(1-u_k) \\ &= 1 - (1-u_1)\cdots(1-u_k)(1-u_{k+1}). \end{aligned}$$

The defining formulas and the formula just proved show that $\mathbf{x} \in Q^k$ whenever $\mathbf{u} \in I^k$. In the process of showing that T is onto, we shall prove the inverse formula. Let $\mathbf{x} \in Q^k$, and assume for the moment that $\sum\limits_{i=1}^{k-1} x_i < 1$. Then all of the equations given as inverse equations are defined. We need only show that the defining equations yield $\mathbf{x}$ when applied to the left-hand sides of these equations. Certainly we do have $x_1 = u_1$. Suppose that $x_r = (1-u_1)\cdots(1-u_{r-1})u_r$ for $r < j$. For the moment assume $u_r \ne 0$.

$$\begin{aligned} (1-u_1)\cdots(1-u_{r-1})(1-u_r)u_{r+1} &= x_r\left(1 - \frac{1}{u_r}\right)u_{r+1} = \\ &= x_r \cdot \frac{1 - x_1 - \cdots - x_r}{x_r} \cdot \frac{x_{r+1}}{1 - x_1 - \cdots - x_r} = x_{r+1}. \end{aligned}$$

If $u_l \neq 0$, but $u_j = 0$ for $l < j \leq r$, then $x_j = 0$ also for these values, and $u_{r+1} = \frac{x_{r+1}}{1-x_1-\cdots-x_l}$. We then have

$$\begin{aligned} x_{r+1} &= (1-u_1)\cdots(1-u_l)u_{r+1} \\ &= x_l\Big(1-\frac{1}{u_l}\Big)u_{r+1} \\ &= x_l \cdot \frac{1-x_1-\cdots-x_l}{x_l} \cdot \frac{x_{r+1}}{1-x_1-\cdots-x_l} \\ &= x_{r+1}. \end{aligned}$$

Finally, if $u_1 = u_2 = \cdots = u_r = 0$, we have simply $u_{r+1} = x_{r+1}$ in both sets of equations. Thus in all cases the point $\mathbf{u} \in I^k$ is a preimage of the point $\mathbf{x} \in Q^k$.

It remains only to consider the case when $\sum_{i=1}^{r} x_i = 1$ for some $r < k$. For these points $x_{r+1} = \cdots = x_k = 0$.

To find preimages of these points, let r be the first index for which $\sum\limits_{i=1}^{r} x_i = 1$. If $r = 1$, we have $x_2 = \cdots = x_k = 0$, and this point is its own preimage. In general the preïmage of the point $\mathbf{x}$ for which $x_{r+1} = \cdots = x_k = 0$ is $\mathbf{u}$, where $u_1, \ldots, u_r$ are given by the formulas for S. The formulas imply $u_r = 1$. The values of $u_{r+1}, \cdots, u_k$ are then arbitrary, since the formulas that define T will automatically make the remaining x_i equal to zero.

The Jacobian matrix is a triangular matrix whose diagonal consists of the entries 1, $(1-u_1)$, $(1-u_1)(1-u_2)$,..., $(1-u_1)\cdots(1-u_{k-1})$, and this fact yields the formula for $J_T(\mathbf{u})$ immediately.

Likewise, the Jacobian of S is triangular and has diagonal entries 1, $\frac{1}{1-x_1}$, $\frac{1}{1-x_1-x_2}, \ldots, \frac{1}{1-x_1-x_2\ldots-x_{k-1}}$, from which again the formula for $J_S(\mathbf{x})$ is immediate.

Exercise 10.13 Let $r_1, \ldots, r_k$ be nonnegative integers, and prove that

$$\int_{Q^k} x_1^{r_1} \cdots x_k^{r_k}\, dx = \frac{r_1! \cdots r_k!}{(k + r_1 + \cdots + r_k)!}.$$

Hint: Use Exercise 12, Theorems 10.9 and 8.20.

Note that the special case $r_1 = \cdots = r_k = 0$ shows that the volume of Q^k is $1/k!$.

Solution: Following the hint, we rewrite the integral in terms of $\mathbf{u}$, getting

$$\int_{I^k} u_1^{r_1} \cdots u_k^{r_k} (1-u_1)^{r_2+\cdots+r_k}(1-u_2)^{r_3+\cdots+r_k} \cdots$$
$$(1-u_{k-1})^{r_k}(1-u_1)^{k-1}(1-u_2)^{k-2}\cdots(1-u_{k-1})\, du_1 \cdots du_k.$$

This integral is the product

$$\prod_{i=1}^{k} \int_0^1 u_i^{r_i}(1-u_i)^{k-i+r_{i+1}\cdots+r_k}\, du_i,$$

which by formula (96) of Chapter 8 (just proved in Exercise 11 above) equals the product

$$\prod_{i=1}^{k} \frac{\Gamma(r_i+1)\Gamma(k+1-i+r_{i+1}\cdots+r_k)}{\Gamma(k+2-i+r_i+r_{i+1}\cdots+r_k)}.$$

When this product is evaluated, the numerator $\Gamma(k+1-i+r_{i+1}\cdots+r_k)$ in each factor cancels the denominator $\Gamma(k+2-(i+1)+r_{i+1}\cdots+r_k)$ in the next factor. Thus the product "telescopes" to the product of the factors $\Gamma(r_i+1)$ in the numerators divided by the first denominator $\Gamma(k+1+r_1+\cdots+r_k)$. Considering that $\Gamma(n+1)=n!$ for integers n, we therefore get the required formula.

Theoretically we ought to be worried about the fact that T is not 1-1 on the entire cube I^k. This problem, however, is handled by the same reasoning used in Exercises 9, 10, and 11, and need not be repeated.

Exercise 10.14 Prove formula (46).

Solution: Formula (46) asserts that $\prod\limits_{p<q} \operatorname{sgn}(j_q - j_p)$ is -1 if the permutation $j_1,\ldots,j_k$ is odd and 1 if the permutation is even. We observe that this product is $(-1)^k$, where k is the number of pairs (j_p, j_q) for which $j_p > j_q$. Since $\operatorname{sgn}(j_q - j_p) = 1$ if $j_p < j_q$ and $\operatorname{sgn}(j_q - j_p) = -1$ if $j_p > j_q$, we need to show that the parity of k is the same as the parity of the number of interchanges that will be used in converting this permutation to the identity. (As a corollary, that parity will be the same, no matter what particular sequence of interchanges is used to get to the identity.) This equality is obvious if the permutation is the identity to begin with. Suppose then that $j_m > j_n$ and $m < n$. The elements j_i, $m < i < n$ are of three kinds: Set A, those for which $j_i < j_n$; set B, those for which $j_n < j_i < j_m$; and set C, those for which $j_m < j_i$. Before j_m and j_n are interchanged, there is one out-of-order pair (j_m, j_i) for each $j_i \in A$, one out-of-order pair (j_i, j_n) for each $j_i \in C$, and two out-of-order pairs (j_m, j_i) and (j_i, j_n) for each $j_i \in B$. After the switch there is one out-of-order pair (j_n, j_i) for each $j_i \in A$, and one pair (j_i, j_m) for each $j_i \in C$. There are no pairs involving any $j_i \in B$. Hence, *when an out-of-order pair (j_m, j_n) is put in the right order by interchanging its elements, the number of out-of-order pairs decreases by $2|B|+1$, where B is the set of elements j_i between j_m and j_n that are in the wrong order relative to both j_m and j_n and $|B|$ is the number of elements in B.*

Of course the number would *increase* by an odd number if we foolishly interchanged a pair that were not out-of-order relative to each other. (The number would increase by $2|B|+1$, where $|B|$ is the number of elements between them that were in the correct order relative to both elements of the interchanged pair.) In any case, each interchange of two elements changes the number of inversions (out-of-order pairs) by an odd number, so that an odd number of interchanges, starting from the identity, will result in an odd number of inversions, and an even number of interchanges will result in an even number of inversions.

Exercise 10.15 If ω and λ are k- and m-forms, respectively, prove that

$$\omega \wedge \lambda = (-1)^{km} \lambda \wedge \omega$$

Solution: Because of the associative and distributive laws, it suffices to prove this in the case when $\omega = f\, dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ and $\lambda = g\, dx_{i_{k+1}} \wedge \cdots \wedge dx_{i_{k+m}}$. In that case

$$\omega \wedge \lambda = fg\, dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{i_{k+1}} \wedge \cdots \wedge dx_{i_{k+m}}.$$

For each $j = 1, 2, \ldots, k$ exactly m interchanges of adjacent basic one-forms will move $dx_{i_{k+1-j}}$ to the position just right of $dx_{i_{k+m}}$, if these moves are made in increasing order of j. Thus a total of km interchanges will exactly reverse λ and ω. The result now follows from the alternating property of the wedge product on basis elements.

Exercise 10.16 If $k \geq 2$ and $\sigma = [\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_k]$ is an oriented affine k-simplex, prove that $\partial^2 \sigma = 0$, directly from the definition of the boundary operator ∂. Deduce from this that $\partial^2 \Psi = 0$ for every chain Ψ.

Hint: For orientation, do it first for $k = 2$, $k = 3$. In general, if $i < j$, let σ_{ij} be the $(k-2)$-simplex obtained by deleting $\mathbf{p}_i$ and $\mathbf{p}_j$ from σ. Show that each σ_{ij} occurs twice in $\partial^2 \sigma$ with opposite sign.

Solution. For $k = 2$ we have

$$\partial \sigma = [\mathbf{p}_1, \mathbf{p}_2] - [\mathbf{p}_0, \mathbf{p}_2] + [\mathbf{p}_0, \mathbf{p}_1],$$

so that

$$\partial^2 \sigma = (\mathbf{p}_2 - \mathbf{p}_1) - (\mathbf{p}_2 - \mathbf{p}_0) + (\mathbf{p}_1 - \mathbf{p}_0) = 0.$$

For $k = 3$ we have

$$\partial \sigma = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3] - [\mathbf{p}_0, \mathbf{p}_2, \mathbf{p}_3] + [\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_3] - [\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2],$$

so that

$$\begin{aligned}
\partial^2 \sigma &= \big([\mathbf{p}_2, \mathbf{p}_3] - [\mathbf{p}_1, \mathbf{p}_3] + [\mathbf{p}_1, \mathbf{p}_2]\big) \\
&\quad -\big([\mathbf{p}_2, \mathbf{p}_3] - [\mathbf{p}_0, \mathbf{p}_3] + [\mathbf{p}_0, \mathbf{p}_2]\big) \\
&\quad +\big([\mathbf{p}_1, \mathbf{p}_3] - [\mathbf{p}_0, \mathbf{p}_3] + [\mathbf{p}_0, \mathbf{p}_1]\big) \\
&\quad -\big([\mathbf{p}_1, \mathbf{p}_2] - [\mathbf{p}_0, \mathbf{p}_2] + [\mathbf{p}_0, \mathbf{p}_1]\big) \\
&= 0.
\end{aligned}$$

In general the order in which $\mathbf{p}_i$ and $\mathbf{p}_j$ are omitted from σ determines the sign that σ_{ij} will have. If $\mathbf{p}_j$ is omitted first, the resulting $(k-1)$-simplex $\sigma_j = [\mathbf{p}_0, \ldots, \mathbf{p}_{j-1}, \mathbf{p}_{j+1}, \ldots, \mathbf{p}_k]$ will acquire the sign $(-1)^j$. If $\mathbf{p}_i$ is then omitted, the resulting $(k-2)$-simplex will acquire a factor of $(-1)^i$, resulting in $(-1)^{i+j}\sigma_{ij}$.

However, if $\mathbf{p}_i$ is omitted first, $\mathbf{p}_j$ will move forward one position in the resulting $(k-1)$-simplex σ_i, and when it is subsequently omitted, a factor of

$(-1)^{j-1}$ will be affixed, resulting in $(-1)^{i+j-1}\sigma_{ij}$. Hence the two occurrences of σ_{ij} in the second boundary will cancel each other.

The linearity of the boundary operator, operating on a base of simplexes, then shows that ∂^2 is the zero operator on all chains.

Exercise 10.17 Put $J^2 = \tau_1 + \tau_2$, where

$$\tau_1 = [\mathbf{0}, \mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2], \quad \tau_2 = -[\mathbf{0}, \mathbf{e}_2, \mathbf{e}_2 + \mathbf{e}_1].$$

Explain why it is reasonable to call J^2 the positively oriented unit square in R^2. Show that ∂J^2 is the sum of 4 oriented affine simplexes. Find these. What is $\partial(\tau_1 - \tau_2)$?

Solution: Although J^2 is really a collection of two affine mappings, the *ranges* of these mappings cover the unit square, the diagonal from $(0,0)$ to $(1,1)$ being covered twice with opposite orientations in the two mappings. In both cases, the sense of orientation is such that the cross product of the last two vertices of the simplex is $\mathbf{e}_3$, which is a reasonable definition of the positive orientation on the unit square.

By routine computation,

$$\begin{aligned}\partial J^2 &= ([\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2] \\ &\qquad -[\mathbf{0}, \mathbf{e}_1 + \mathbf{e}_2] + [\mathbf{0}, \mathbf{e}_1]) - ([\mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2] - [\mathbf{0}, \mathbf{e}_1 + \mathbf{e}_2] + [\mathbf{0}, \mathbf{e}_2]) \\ &= [\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2] + [\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_2] + [\mathbf{e}_2, \mathbf{0}] + [\mathbf{0}, \mathbf{e}_1].\end{aligned}$$

Again, by routine computation,

$$\begin{aligned}\partial(\tau_1 - \tau_2) &= ([\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2] - [\mathbf{0}, \mathbf{e}_1 + \mathbf{e}_2] + [\mathbf{0}, \mathbf{e}_1]) \\ &\qquad +([\mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2] - [\mathbf{0}, \mathbf{e}_1 + \mathbf{e}_2] + [\mathbf{0}, \mathbf{e}_2]) \\ &= [\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2] - [\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_2] - [\mathbf{e}_2, \mathbf{0}] + [\mathbf{0}, \mathbf{e}_1] - 2[\mathbf{0}, \mathbf{e}_1 + \mathbf{e}_2].\end{aligned}$$

Exercise 10.18 Consider the oriented affine 3-simplex

$$\sigma_1 = [\mathbf{0}, \mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3]$$

in R^3. Show that σ_1 (regarded as a linear transformation) has determinant 1. Thus σ_1 is positively oriented.

Let $\sigma_2, \ldots, \sigma_6$ be five other oriented simplexes, obtained as follows: There are five permutations (i_1, i_2, i_3) of $(1,2,3)$ distinct from $(1,2,3)$. Associate with each (i_1, i_2, i_3) the simplex

$$s(i_1, i_2, i_3)[\mathbf{0}, \mathbf{e}_{i_1}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \mathbf{e}_{i_3}]$$

where s is the sign that occurs in the definition of the determinant. (This is how τ_2 was obtained from τ_1 in Exercise 17.)

Show that $\sigma_2, \ldots, \sigma_6$ are positively oriented.

Put $J^3 = \sigma_1 + \cdots + \sigma_6$. Then J^3 may be called the positively oriented unit cube in R^3.

Show that ∂J^3 is the sum of 12 oriented affine 2-simplexes. (These 12 triangles cover the surface of the unit cube I_3.)

Show that $\mathbf{x} = (x_1, x_2, x_3)$ is in the range of σ_1 if and only if $0 \le x_3 \le x_2 \le x_1 \le 1$.

Show that the ranges of $\sigma_1, \ldots, \sigma_6$ have disjoint interiors, and that their union covers I^3. (Compare with Exercise 13; note that $3! = 6$.)

Solution. We first show that each of these simplexes is positively oriented. To that end, it is convenient to refer to the simplex $[\mathbf{0}, \mathbf{e}_{i_1}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \mathbf{e}_{i_3}]$ corresponding to the permutation (i_1, i_2, i_3) as $\sigma^{(i_1,i_2,i_3)}$.

The simplex $\sigma^{(i_1,i_2,i_3)}$, regarded as a linear transformation, maps (x, y, z) to $(x+y+z)\mathbf{e}_{i_1} + (y+z)\mathbf{e}_{i_2} + z\mathbf{e}_{i_3}$. Its matrix therefore has $(1\ \ 1\ \ 1)$ as row i_1, $(0\ \ 1\ \ 1)$ as row i_2, and $(0\ \ 0\ \ 1)$ as row i_3. By interchanging rows in correspondence with the interchanges needed to convert the permutation (i_1, i_2, i_3) to the identity, we can convert this matrix to an upper-triangular matrix with 1's on the main diagonal. The determinant of the matrix is therefore $s(i_1, i_2, i_3)$, so that the simplex $s(i_1, i_2, i_3)[\mathbf{0}, \mathbf{e}_{i_1}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \mathbf{e}_{i_3}]$ is positively oriented.

The boundary of $\sigma^{(i_1,i_2,i_3)}$ consists of four terms, two of which ($[\mathbf{e}_{i_1}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \mathbf{e}_{i_3}]$ and $-[\mathbf{0}, \mathbf{e}_{i_1}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2}]$) are not shared with any other $\sigma^{(i)}$. The other two terms ($-[\mathbf{0}, \mathbf{e}_{i_1}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \mathbf{e}_{i_3}]$ and $[\mathbf{0}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \mathbf{e}_{i_3}]$ are shared with $\sigma^{(i_1,i_3,i_2)}$ and $\sigma^{(i_2,i_1,i_3)}$ respectively. As these two permutations each differ from (i_1, i_2, i_3) by a single interchange, the sign of each of these terms will be opposite in its two occurrences, and hence they will cancel out. Thus the boundary of J^3 will consist of a total of 12 oriented affine 2-simplexes.

A point $\mathbf{x} = (x_1, x_2, x_3)$ is in the range of σ_1 if and only if there are numbers $r, s, t \in [0, 1]$ such that $r + s + t \le 1$ and $\mathbf{x} = r\mathbf{e}_1 + s(\mathbf{e}_1 + \mathbf{e}_2) + t(\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3)$, that is, $x_1 = r + s + t$, $x_2 = s + t$, and $x_3 = t$. If such numbers r, s, t exist, obviously $0 \le x_3 \le x_2 \le x_1 \le 1$. Conversely, if these conditions hold, there will be such numbers r, s, t, namely $t = x_3$, $s = x_2 - x_3$, and $r = x_1 - x_2$.

The interior of the range of $\sigma^{(i_1,i_2,i_3)}$ is the set of all $\mathbf{x} = (x_1, x_2, x_3)$ such that $0 < x_{i_3} < x_{i_2} < x_{i_1} < 1$. For the range of this simplex is the set of $\mathbf{x}$ for which each of these inequalities or the corresponding equality holds. If equality holds in any of them, the point can be approached by points outside the range, as one can easily see. That the union covers I^3 is also obvious. Indeed, the characterization of the range of σ_1 applies to all $\sigma^{(i_1,i_2,i_3)}$ and shows that this range is contained in I^3. Thus we need only show the reverse inclusion.

If $\mathbf{x} = (x_1, x_2, x_3) \in I^3$, let i_1 be the smallest subscript i for which $x_i = \max\{x_1, x_2, x_3\}$. Let i_2 be the first subscript for which $x_i = \max\left(\{x_1, x_2, x_3\} \setminus \{x_{i_1}\}\right)$. Finally, let i_3 be such that $\{x_{i_3}\} = \{x_1, x_2, x_3\} \setminus \{x_{i_1}, x_{i_2}\}$. By construction $0 \le x_{i_3} \le x_{i_2} \le x_{i_1} \le 1$, and so, by the argument given above, $\mathbf{x}$ belongs to the range of $\sigma^{(i_1,i_2,i_3)}$. Symmetry shows that all of these simplexes have the same volume, which must therefore be $1/6$. (Remember that we showed back in Exercise 1 that the boundary of a convex set in R^k has k-dimensional volume 0, so that the volume of each of these sets equals the volume of its interior. As the interiors are disjoint, the sum of their volumes is at most 1. Since the simplexes together cover I^3, the sum of their volumes is at least 1. Therefore it is exactly 1.)

Exercise 10.19 Let J^2 and J^3 be as in Exercise 17 and 18. Define

$$\begin{aligned} B_{01}(u,v) &= (0,u,v), & B_{11}(u,v) &= (1,u,v), \\ B_{02}(u,v) &= (u,0,v), & B_{12}(u,v) &= (u,1,v), \\ B_{03}(u,v) &= (u,v,0), & B_{13}(u,v) &= (u,v,1). \end{aligned}$$

These are affine and map R^2 into R^3.

Put $\beta_{ri} = B_{ri}(J^2)$, for $r = 0, 1$, $i = 1, 2, 3$. Each β_{ri} is an affine-oriented 2-chain. (See Sec. 10.30.) Verify that

$$\partial J^3 = \sum_{i=1}^{3} (-1)^i (\beta_{0i} - \beta_{1i}),$$

in agreement with Exercise 18.

Solution. Although we did not spell it out in our solution of Exercise 18, the boundary of J^3 is the 2-chain

$$\sum_{i_1,i_2,i_3} s(i_1,i_2,i_3)\big([\mathbf{e}_{i_1}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \mathbf{e}_{i_3}] - [\mathbf{0}, \mathbf{e}_{i_1}, \mathbf{e}_{i_1} + \mathbf{e}_{i_2}]\big).$$

This sum can be rearranged as a sum of three terms, each of which consists of four terms. For example, the terms in the sum for which $i_=1$ can be written as

$$\begin{aligned} &\big([\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3] \\ &\qquad - [\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_3, \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3]\big) - \big([\mathbf{0}, \mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2] - [\mathbf{0}, \mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_3]\big). \end{aligned}$$

For $i_1 = 2$ we get a similar set of four terms, namely,

$$\begin{aligned} &\big(-[\mathbf{e}_2, \mathbf{e}_2 + \mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3] + [\mathbf{e}_2, \mathbf{e}_2 + \mathbf{e}_3, \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3]\big) \\ &\qquad\qquad + \big([\mathbf{0}, \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2] - [\mathbf{0}, \mathbf{e}_2, \mathbf{e}_2 + \mathbf{e}_3]\big). \end{aligned}$$

Finally, for $i_1 = 3$ we have

$$\begin{aligned} &\big([\mathbf{e}_3, \mathbf{e}_3 + \mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3] - [\mathbf{e}_3, \mathbf{e}_3 + \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3]\big) \\ &\qquad\qquad - \big([\mathbf{0}, \mathbf{e}_3, \mathbf{e}_3 + \mathbf{e}_1] - [\mathbf{0}, \mathbf{e}_3, \mathbf{e}_3 + \mathbf{e}_2]\big). \end{aligned}$$

Now consider the 2-chain β_{01}. According to the notation of Eq. (88), it is $B_{01}(\tau_1) + B_{01}(\tau_2)$. Letting $(u,v) = \tau_1(x,y) = (x+y)\mathbf{e}_1 + y\mathbf{e}_2$, and then $(u,v) = \tau_2(x,y) = (x+y)\mathbf{e}_2 + y\mathbf{e}_1$ (and keeping in mind the orientation assigned to τ_2), we see that $\beta_{01}(x,y) = B_{01}(x+y,y) - B_{01}(y,x+y) = (0,x+y,y) -$

$(0, y, x+y) = [\mathbf{0}, \mathbf{e}_2, \mathbf{e}_2+\mathbf{e}_3] - [\mathbf{0}, \mathbf{e}_3, \mathbf{e}_3+\mathbf{e}_2]$. Notice that these two terms occur in the expression for ∂J^3, in the groupings for $i_1 = 2$ and $i_1 = 3$ respectively, but each occurs with the opposite sign. Hence these terms can be accounted for in ∂J^3 by being grouped together and written as $-\beta_{01}$. Similarly when we look at β_{11}, we find that it is the 2-chain whose points are $(1, x+y, y) - (1, y, x+y)$, which is $[\mathbf{e}_1, \mathbf{e}_1+\mathbf{e}_2, \mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3] - [\mathbf{e}_1, \mathbf{e}_1+\mathbf{e}_3, \mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3]$. Again these terms occur in the expression for $\mathbf{e}_1$, this time with exactly the same signs, so that they can be accounted for by grouping them and writing them as the term β_{11}. Thus four of the twelve simplexes in ∂J^3 are accounted for by the expression $(-1)^1(\beta_{01} - \beta_{11})$. The other 8 simplexes are accounted for similarly.

Exercise 10.20 State conditions under which the formula

$$\int_\Phi f\, d\omega = \int_{\partial\Phi} f\omega - \int_\Phi (df) \wedge \omega$$

is valid, and show that it generalizes the formula for integration by parts.
Hint: $d(f\omega) = (df) \wedge \omega + f\, d\omega$.

Solution. Given the formula in the hint, we need only invoke Stokes' Theorem. For any chain Φ satisfying the hypotheses of that theorem we shall have

$$\int_\Phi d(f\omega) = \int_{\partial\Phi} f\omega,$$

which is precisely the given theorem. The ordinary formula for integration by parts follows by considering a 0-form fg.

Exercise 10.21 As in Example 10.36, consider the 1-form

$$\eta = \frac{x\, dy - y\, dx}{x^2 + y^2}$$

in $R^2 \setminus \{\mathbf{0}\}$.
(*a*) Carry out the computation that leads to formula (113), and prove that $d\eta = 0$.
(*b*) Let $\gamma(t) = (r\cos t, r\sin t)$, for some $r > 0$, and let Γ be a $\mathcal{C}''$-curve in $R^2 \setminus \{\mathbf{0}\}$, with parameter interval $[0, 2\pi]$, with $\Gamma(0) = \Gamma(2\pi)$, such that the intervals $[\gamma(t), \Gamma(t)]$ do not contain $\mathbf{0}$ for any $t \in [0, 2\pi]$. Prove that

$$\int_\Gamma \eta = 2\pi.$$

Hint: For $0 \le t \le 2\pi$, $0 \le u \le 1$, define

$$\Phi(t, u) = (1-u)\Gamma(t) + u\gamma(t).$$

then Φ is a 2-surface in $R^2 \setminus \{\mathbf{0}\}$ whose parameter domain is the indicated rectangle. Because of cancellations (as in Example 10.32),

$$\partial\Phi = \Gamma - \gamma.$$

Use Stokes' theorem to deduce that

$$\int_\Gamma \eta = \int_\gamma \eta$$

because $d\eta = 0$.
(*c*) Take $\Gamma(t) = (a\cos t, b\sin t)$ where $a > 0$, $b > 0$ are fixed. Use part (*b*) to show that

$$\int_0^{2\pi} \frac{ab}{a^2\cos^2 t + b^2\sin^2 t}\, dt = 2\pi.$$

(*d*) Show that

$$\eta = d\Big(\arctan\frac{y}{x}\Big)$$

in any convex open set in which $x \neq 0$, and that

$$\eta = d\Big(-\arctan\frac{x}{y}\Big)$$

in any convex open set in which $y \neq 0$.

Explain why this justifies the notation $\eta = d\theta$, in spite of the fact that η is not exact in $R^2 \setminus \{\mathbf{0}\}$.
(*e*) Show that (*b*) can be derived from (*d*).
(*f*) If Γ is any closed $\mathcal{C}'$-curve in $R^2 \setminus \{\mathbf{0}\}$, prove that

$$\frac{1}{2\pi}\int_\Gamma \eta = \operatorname{Ind}(\Gamma).$$

(See Exercise 23 of Chap. 8 for the definition of the index of a curve.)

Solution. (*a*) By the rules for computing line integrals, given that $x = r\cos t$ and $y = r\sin t$,

$$\int_\gamma \eta = \int_0^{2\pi} \frac{(r\cos t)(r\cos t)\,dt - (r\sin t)(-r\sin t)\,dt}{r^2\cos^2 t + r^2\sin^2 t} = \int_0^{2\pi} dt = 2\pi.$$

(*b*) Let $\Gamma(t) = \big(X(t), Y(t)\big)$ and $\gamma(t) = \big(x(t), y(t)\big)$. Following the hint, observing that the hypothesis that the interval from $\Gamma(t)$ to $\gamma(t)$ does not pass through $\mathbf{0}$, we find that $\Phi(t,u)$ is indeed a 2-surface in $R^2 \setminus \{\mathbf{0}\}$, and making it into a singular 2-chain by regarding the domain as an affine 2-chain, as in Exercise 17, we find by Stokes' theorem that

$$\begin{aligned} 0 &= \int_\Phi d\eta = \int_{\partial\Phi} \eta \\ &= -\int_\Gamma \eta + \int_\gamma \eta + \int_\delta \eta - \int_\varepsilon \eta, \end{aligned}$$

where δ is the curve $\delta(u) = \Phi(2\pi, u) = (1-u)\Gamma(2\pi) + u\gamma(2\pi)$ and ε is the curve $\varepsilon(u) = \Phi(0,u) = (1-u)\Gamma(0) + u\gamma(0)$. Since δ and ε are the same curve, the last two terms in this expression cancel each other, yielding the required result.

(*c*) We need only verify that $\Phi(t,u) \neq \mathbf{0} = (0,0)$. But this is clear: If $((1-u)a + ur)\cos t = 0$, then $t = \frac{\pi}{2}$ or $t = \frac{3\pi}{2}$, since $(1-u)a + ur \geq \min(a,r) > 0$. But this means that $((1-u)b + ur)\sin t \neq 0$, since t is not a multiple of π. The result now follows.

(*d*) It is a routine computation that the differential of $\arctan \frac{y}{x}$ is η in the entire right or left half-plane, and similarly for $\pi - \arctan \frac{x}{y}$, which is after all $\operatorname{arccot} \frac{x}{y}$, which in turn is $\arctan \frac{y}{x}$ wherever both functions are defined. Thus *locally* we have $\eta = d\theta$, even though θ is not defined *globally* in $R^2 \setminus \{\mathbf{0}\}$.

(*e*) Break the integral over γ into five parts: $0 \leq t \leq \frac{\pi}{4}$, $\frac{\pi}{4} \leq t \leq \frac{3\pi}{4}$, $\frac{3\pi}{4} \leq t \leq \frac{5\pi}{4}$, $\frac{5\pi}{4} \leq t \leq \frac{7\pi}{4}$, $\frac{7\pi}{4} \leq t \leq 2\pi$. In the first, third, and fifth parts we have $\eta = d\big(\arctan \frac{y}{x}\big)$, and in the second and fourth we have $\eta = d\big(-\arctan \frac{x}{y}\big)$. Now in the first, third, and fifth parts, $\frac{y}{x} = \frac{\sin t}{\cos t} = \tan t$, so that either $t = \arctan \frac{y}{x}$ or $t = \pi + \arctan \frac{y}{x}$ on these arcs. In either case the integral over the these parts is just the difference in t at the endpoints. Hence these three integrals contribute $\frac{\pi}{4} + \frac{\pi}{2} + \frac{\pi}{4} = \pi$ to the integral. On the other parts $\frac{x}{y} = \cot t = \tan\big(\frac{\pi}{2} - t\big)$. Hence, once again, $\arctan \frac{x}{y}$ is either $\frac{\pi}{2} - t$ or $\frac{3\pi}{2} - t$. In either case, these two integrals contribute $\frac{\pi}{2} + \frac{\pi}{2} = \pi$ to the integral, and provides the result of (*b*).

(*f*) The definition of $\operatorname{Ind}(\Gamma)$ is defined by regarding $\Gamma(t)$ as a curve $X(t) + Y(t)i$ in the complex plane, in which case

$$\operatorname{Ind}(\Gamma) = \frac{1}{2\pi i}\int_0^{2\pi} \frac{\Gamma'(t)}{\Gamma(t)}\,dt = \frac{1}{2\pi i}\int_0^{2\pi} \frac{(X(t) - Y(t)i)(X'(t) + Y'(t)i)}{(X(t))^2 + Y(t))^2}\,dt.$$

Since we know the imaginary part is zero, we consider only the real part, which is

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{X(t)Y'(t) - Y(t)X'(t)}{(X(t))^2 + Y(t))^2}\,dt = \frac{1}{2\pi}\int_\Gamma \eta.$$

(Incidentally, it follows from Stokes' theorem that the imaginary part of this complex integral is zero, since it is

$$-\frac{1}{2\pi}\int_0^{2\pi} \frac{X(t)X'(t) + Y(t)Y'(t)}{(X(t))^2 + (Y(t))^2}\,dt = -\frac{1}{4\pi}\int_\Gamma d\zeta = -\frac{1}{4\pi}\int_{\partial\Gamma} \zeta$$

where $\zeta(x,y) = \ln\left(x^2 + y^2\right)$. This last integral is zero, since Γ is a closed curve.)

Exercise 10.22 As in Example 10.37, define ζ in $R^3 - \mathbf{0}$ by

$$\zeta = \frac{x\,dy \wedge dz + y\,dz \wedge dx + z\,dx \wedge dy}{r^3},$$

where $r = (x^2 + y^2 + z^2)^{1/2}$, let D be the rectangle given by $0 \le u \le \pi$, $0 \le v \le 2\pi$, and let Σ be the 2-surface in R^3, with parameter domain D, given by

$$x = \sin u \cos v, \quad y = \sin u \sin v, \quad z = \cos u.$$

(a) Prove that $d\zeta = 0$ in $R^3 \setminus \mathbf{0}$.

(b) Let S denote the restriction of Σ to a parameter domain $E \subset D$. Prove that

$$\int_S \zeta = \int_E \sin u \, du \, dv = A(S),$$

where A denotes area, as in Sec. 10.43. Note that this contains (115) as a special case.

(c) Suppose g, h_1, h_2, h_3, are $\mathcal{C}''$-functions on $[0,1]$, $g > 0$. Let $(x, y, z) = \Phi(s, t)$ define a 2-surface Φ, with parameter domain I^2, by

$$x = g(t)h_1(s), \quad y = g(t)h_2(s), \quad z = g(t)h_3(s).$$

Prove that

$$\int_\Phi \zeta = 0,$$

directly from (35).

Note the shape of the range of Φ: For fixed s, $\Phi(s,t)$ runs over an interval on a line through $\mathbf{0}$. The range of Φ thus lies in a "cone" with vertex at the origin.

(d) Let E be a closed rectangle in D, with edges parallel to those of D. Suppose $f \in \mathcal{C}''(D)$, $f > 0$. Let Ω be the 2-surface with parameter domain E, defined by

$$\Omega(u, v) = f(u, v)\Sigma(u, v).$$

Define S as in (b) and prove that

$$\int_\Omega \zeta = \int_S \zeta = A(S).$$

(Since S is the "radial projection" of Ω onto the unit sphere, this result makes it reasonable to call $\int_\Omega \zeta$ the "solid angle" subtended by the range of Ω at the origin.)

Hint: Consider the 3-surface Ψ given by

$$\Psi(t, u, v) = [1 - t + tf(u, v)]\Sigma(u, v),$$

where $(u, v) \in E$, $0 \le t \le 1$. For fixed v, the mapping $(t, u) \to \Psi(t, u, v)$ is a 2-surface Φ to which (c) can be applied to show that $\int_\Phi \zeta = 0$. The same thing holds when u is fixed. By (a) and Stokes' theorem,

$$\int_{\partial\Psi} \zeta = \int_\Psi d\zeta = 0.$$

(e) Put $\lambda = -(z/r)\eta$, where

$$\eta = \frac{x\,dy - y\,dx}{x^2 + y^2},$$

as in Exercise 21. Then λ is a 1-form in the open set $V \subset R^3$ in which $x^2+y^2 > 0$. Show that ζ is *exact in* V by showing that

$$\zeta = d\lambda.$$

(f) Derive (d) from (e), without using (c).
Hint: To begin with, assume $0 < u < \pi$ on E. By (e),

$$\int_\Omega \zeta = \int_{\partial\Omega} \lambda \quad \text{and} \quad \int_S \zeta = \int_{\partial S} \lambda.$$

Show that the two integrals of λ are equal, by using part (d) of Exercise 21, and by noting that z/r is the same at $\Sigma(u,v)$ as at $\Omega(u,v)$.

(g) Is ζ exact in the complement of every line through the origin?

Solution. (a) We note that, since $\frac{\partial r}{\partial x} = xr^{-1}$, we have

$$\frac{\partial}{\partial x}\frac{x}{r^3} = r^{-3} - 3x^2r^{-5} = r^{-5}(r^2 - 3x^2).$$

By symmetry we have analogous relations for the partial derivatives of yr^{-3} and zr^{-3} with respect to y and z respectively. Since $dx \wedge dy \wedge dz = dy \wedge dz \wedge dx = dz \wedge dx \wedge dy$, we find that

$$d\zeta = r^{-5}(r^2 - 3x^2 + r^2 - 3y^2 + r^2 - 3z^2)\,dx \wedge dy \wedge dz = 0.$$

(b) Since $r\big(\Sigma(u,v)\big) = 1$, we have only to note that the differentials pull back to D as $dy \wedge dz = \frac{\partial(y,z)}{\partial(u,v)}\,du \wedge dv = \sin^2 u \cos v\,du \wedge dv$, $dz \wedge dx = \sin^2 u \sin v\,du \wedge dv$ and $dx \wedge dy = \sin u \cos u\,du \wedge dv$. The integrand then pulls back as $(\sin^3 u + \sin u \cos^2 u)\,du \wedge dv = \sin u\,du \wedge dv$. The reference to Sec. 10.43 must be a misprint for Sec. 10.46.

(c) For the application to be made in part (d) below we actually need to allow the function $g(t)$ to depend on s also. Thus we consider $g(s,t)$ instead of $g(t)$. Using only the definition (35) for the integral, we need to get the pullbacks of the wedge products to the parameter domain $[0,1] \times [0,1]$. Since $dx = \frac{\partial g}{\partial t} h_1(s)\,dt + (g(s,t)h_1'(s) + h_1(s)\frac{\partial g}{\partial s}\,ds$, with similar expressions for dy and dz, we find that $dy \wedge dz = g(s,t)\frac{\partial g}{\partial t}\big(h_3(s)h_2'(s) - h_3'(s)h_2(s)\big)\,ds \wedge dt$, $dz \wedge dx = g(s,t)\frac{\partial g}{\partial t}\big(h_3'(s)h_1(s) - h_3(s)h_1'(s)\big)\,ds \wedge dt$, and $dx \wedge dy = g(s,t)\frac{\partial g}{\partial t}\big(h_1'(s)h_2(s) - h_1(s)h_2'(s)\big)\,ds \wedge dt$. Thus, *assuming $h_1(t)$, $h_2(t)$, and $h_3(t)$ do not vanish simultaneously*, we have

$$\int_\Phi \zeta = \int_0^1 \int_0^1 \frac{\frac{\partial g}{\partial t}}{g(s,t)} \frac{(h_1(s)h_2(s)h_3(s))' - (h_1(s)h_2(s)h_3(s))'}{(h_1(s))^1 + (h_2(s))^2 + (h_3(s))^2}\,ds\,dt = 0.$$

(d) Using part (c), as amended, we note that $\partial\Psi$ consists of six mappings $\Psi(1,u,v) = \Omega(u,v)$, $\Psi(0,u,) = S(u,v)$, $\Psi(t,b,v)$, $\Psi(t,a,v)$, $\Psi(t,u,d)$, and $\Psi(t,u,c)$, where $E = [a,b]\times[c,d]$. By part (c) the integrals over each of the last 4 surfaces are all zero. Since $d\zeta = 0$, Stokes' theorem implies that

$$\int_\Omega \zeta - \int_S \zeta = 0.$$

(e) By straightforward computation,

$$\begin{aligned} d\lambda &= -d(z/r)\wedge\eta - (z/r)\,d\eta \\ &= \frac{xz\,dx + yz\,dy + (z^2-r^2)\,dz}{r^3}\wedge\eta \\ &= \frac{(x^2z+y^2z)\,dx\wedge dy}{r^3(x^2+y^2)} - \frac{x\,dz\wedge dy - y\,dz\wedge dx}{r^3} \\ &= \zeta. \end{aligned}$$

(f) Again by Stokes' theorem we must have

$$\int_\Omega \zeta = \int_\Omega d\lambda = \int_{\partial\Omega} \lambda.$$

But η is independent of z, and z/r is the same for both $S(u,v)$ and $\Omega(u,v)$. Therefore

$$\int_{\partial\Omega} \lambda = \int_{\partial S} \lambda.$$

(g) Yes, ζ is exact on the complement of every line through the origin. Indeed, for every line through the origin there is a rotation T that maps that line to the z-axis. By Theorem 10.22, part (c) we have $d(\lambda_T) = (d\lambda)_T = \zeta_T$. However, $\zeta_T = \zeta$, as one can easily compute. Indeed, since r is invariant under T, we need only show that $x\,dy\wedge dz + y\,dz\wedge dx + z\,dx\wedge dy$ is rotation-invariant. To that end, suppose $(u,v,w) = T(x,y,z)$, say $u = t_{11}x + t_{12}y + t_{13}z$, so that $du = t_{11}\,dx + t_{12}\,dy + t_{13}\,dz$, etc. We then have $dv\wedge dw = (t_{22}t_{33} - t_{32}t_{23})\,dy\wedge dz + (t_{23}t_{31} - t_{33}t_{21})\,dz\wedge dx + (t_{21}t_{32} - t_{31}t_{22})\,dx\wedge dy$, etc. and so $u\,dv\wedge dw + v\,dw\wedge du + w\,du\wedge dv$ works out (after tedious computation) to precisely $x\,dy\wedge dz + y\,dz\wedge dx + z\,dx\wedge dy$.

Exercise 10.23 Fix n. Define $r_k = (x_1^2 + \cdots + x_k^2)^{1/2}$ for $1 \le k \le n$, let E_k be the set of all $\mathbf{x} \in R^n$ at which $r_k > 0$, and let ω_k be the $(k-1)$-form defined in E_k by

$$\omega_k = (r_k)^{-k}\sum_{i=1}^{k}(-1)^{i-1}x_i\,dx_1\wedge\cdots\wedge dx_{i-1}\wedge dx_{i+1}\wedge\cdots\wedge dx_k.$$

Note that $\omega_2 = \eta$, $\omega_3 = \zeta$ in the notation of Exercises 21 and 22. Note also that

$$E_1 \subset E_2 \subset \cdots \subset E_n = R^n \setminus \{\mathbf{0}\}.$$

(*a*) Prove that $d\omega_k = 0$ in E_k.

(*b*) For $k = 2, \ldots, n$, prove that ω_k is exact in E_{k-1}, by showing that

$$\omega_k = d(f_k \omega_{k-1}) = (df_k) \wedge \omega_{k-1},$$

where $f_k(\mathbf{x}) = (-1)^k g_k(x_k/r_k)$ and

$$g_k(t) = \int_{-1}^{t} (1-s)^{(k-3)/2}\, ds \quad (-1 < t < 1).$$

Hint: f_k satisfies the differential equations

$$\mathbf{x} \cdot (\nabla f_k)(\mathbf{x}) = 0$$

and

$$(D_k f_k)(\mathbf{x}) = \frac{(-1)^k (r_{k-1})^{k-1}}{(r_k)^k}.$$

(*c*) Is ω_n exact in E_n?

(*d*) Note that (*b*) is a generalization of part (*e*) of Exercise 22. Try to extend some of the other assertions of Exercises 21 and 22 to ω_n, for arbitrary n.

Solution. (*a*) Computation shows that $d\Big(\sum_{i=1}^{k} (-1)^{i-1} x_i\, dx_1 \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots \wedge dx_k\Big) = k dx_1 \wedge \cdots \wedge dx_k$, and $\frac{\partial r_k}{\partial x_j} = \frac{x_j}{r_k}$ for $j \le k$, so that $d(r_k) = -k(r_k)^{-k-2} \sum_{j=1}^{k} x_j\, dx_j$, we find that

$$d\omega_k = k(r_k)^{-k}\, dx_1 \wedge \cdots \wedge dx_k - k(r_k)^{-k-2} \sum_{j=1}^{k} x_j^2 dx_1 \wedge \cdots \wedge dx_k =$$

$$= k(r_k)^{-k-2}\Big(r_k^2 - \sum_{j=1}^{k} x_j^2\Big)\, dx_1 \wedge \cdots \wedge dx_k = 0.$$

This argument shows, incidentally, that $d\omega_k = 0$ in $E_n = R^n \setminus \{\mathbf{0}\}$.

(*b*) We compute that

$$\begin{aligned}
df_k &= (-1)^k (1 - x_k^2/r_k^2)^{(k-3)/2}\Big(\big(r_k^{-1} - x_k^2 r_k^{-3}\big) dx_k - \sum_{i=1}^{k-1} x_k x_i r_k^{-3}\, dx_i\Big) \\
&= (-1)^k (r_{k-1}/r_k)^{k-3}\Big(\big(r_k^{-3} r_{k-1}^2\big) dx_k - r_k^{-3} \sum_{i=1}^{k-1} x_i x_k\Big)\, dx_i \\
&= (-1)^k (r_k)^{-k}\Big(r_{k-1}^{k-1} dx_k - r_{k-1}^{k-3} \sum_{i=1}^{k-1} x_i x_k\, dx_i\Big).
\end{aligned}$$

Hence, since $(df_k) \wedge \omega_{k-1} = (-1)^{k-2}\omega_{k-1} \wedge (df_k)$, the first term in this last expression contributes

$$(r_k)^{-k}\sum_{i=1}^{k-1}(-1)^{i-1}x_i\,dx_i \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots \wedge dx_{k-1} \wedge dx_k$$

to the wedge product. As this contribution is all of ω_k except the last term $r_k^{-k}(-1)^{k-1}x_k\,dx_1 \wedge \cdots \wedge dx_{k-1}$, we must endeavor to show that the contribution of the remaining terms amounts to this expression. Since any term containing a repeated factor dx_j is zero, we see that the rest of the expression is

$$(-1)^{k-1}x_k(r_k)^{-k}(r_{k-1})^{k-3}\Big(\sum_{i=1}^{k-1}x_i\,dx_i\Big) \wedge (r_{k-1})^{k-1} \times$$
$$\times \sum_{i=1}^{k-1}(-1)^{i-1}x_i\,dx_1 \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots \wedge dx_{k-1},$$

which is easily seen to be the same as

$$(-1)^{k-1}r_k^{-k}x_kr_{k-1}^{-2}\sum_{i=1}^{k-1}x_i^2dx_1 \wedge \cdots \wedge dx_{k-1} = (-1)^{k-1}r_k^{-k}x_k\,dx_1 \wedge \cdots \wedge dx_{k-1},$$

exactly as required. Thus we have computed this result by "brute force," arrogantly ignoring the hint.

For the benefit of those who wish to *use* the hint, here is an alternative approach. The wedge product $(df_k) \wedge \omega_{k-1}$ is the sum of $D_kf_k(\mathbf{x})\,dx_k \wedge \omega_{k-1}$ and

$$r_{k-1}^{-k-1}\Big(\sum_{i=1}^{k-1}x_iD_if(\mathbf{x})\Big)\,dx_1 \wedge \cdots \wedge dx_k =$$
$$= r_{k-1}^{-k-1}(\mathbf{x}\cdot(\nabla f_k)(\mathbf{x}) - x_kD_kf_k(\mathbf{x}))\,dx_1 \wedge \cdots \wedge dx_k,$$

and hence, by the first differential equation, equals

$$D_kf_k(\mathbf{x})\,dx_k \wedge \omega_{k-1} - r_{k-1}^{-k-1}x_kD_kf_k(\mathbf{x})\,dx_1 \wedge \cdots dx_{k-1},$$

so that the second equation yields the result immediately. The two differential equations themselves are routine computations.

(*c*) No, ω_n is not exact in E_n for any n, since its integral over the $(n-1)$-sphere equals $\frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}$, as will be shown below in the answer to part (*d*). (If it were exact, say the differential of λ, this integral would equal thte integral of λ over the boundary of the $(n-1)$-sphere, which is the 0 $(n-2)$-chain.)

(*d*) We can parameterize the $(n-1)$-sphere Σ^{n-1} by the mapping T_n defined by

$$x_1 \;=\; \sin t_1 \sin t_2 \cdots \sin t_{n-1},$$

$$\begin{aligned}
x_2 &= \cos t_1 \sin t_2 \cdots \sin t_{n-1},\\
x_3 &= \cos t_2 \sin t_3 \cdots \sin t_{n-1},\\
\cdots & \cdots \cdots\\
x_{n-1} &= \cos t_{n-2} \sin t_{n-1}\\
x_n &= \cos t_{n-1},
\end{aligned}$$

where $0 \le t_1 \le 2\pi$ and $0 \le t_j \le \pi$ for $2 \le j \le n-1$ That is, the domain of T_n is the parallelepiped $D = [0, 2\pi] \times [0, \pi]^{n-2}$. This is known to be true for $n = 2$ and $n = 3$, and follows easily by induction on n. Suppose, for example, we know it is true for $n-1$, and suppose $x_1^2 + \cdots + x_n^2 = 1$. If $x_n = \pm 1$, we can take $t_{n-1} = 0$ or π, and the values of the other angles can be anything. If $-1 < x_n < 1$, there is precisely one angle $t_{n-1} \in (0, \pi)$ such that $x_n = \cos t_{n-1}$. But then the point $(x_1/\sin t_{n-1}, \ldots, x_{n-1}/\sin t_{n-1})$ belongs to Σ^{n-2}, and hence, by induction, can be written as

$$\begin{aligned}
x_1/\sin t_{n-1} &= \sin t_1 \cdots \sin t_{n-2},\\
x_2/\sin t_{n-1} &= \cos t_1 \cdots \sin t_{n-2},\\
\cdots & \cdots \cdots\\
x_{n-2}/\sin t_{n-1} &= \cos t_{n-3} \sin t_{n-2}\\
x_{n-1}/\sin t_{n-1} &= \cos t_{n-2}.
\end{aligned}$$

This completes the induction. Observe that the angle t_1 requires the entire range $[0, 2\pi]$. That is, all points on the unit circle in R^2 can be written as $(\cos t, \sin t)$ only if t is allowed to range from 0 to 2π. Otherwise put, the $(n-1)$-sphere is parameterized by $n-2$ latitude angles and one longitude angle.

We can easily show by induction that the pullback of ω_n is

$$(\omega_n)_{T_n} = (-1)^{n-1} \sin t_2 \sin^2 t_3 \cdots \sin^{n-2} t_{n-1}\, dt_1 \wedge \cdots \wedge dt_{n-1}.$$

To make the induction work, we need to distinguish the x_i's in various numbers of dimensions; hence let the transformation T_n be defined by giving its components $x_i^{(n)}$, $i \le n$, by the equations

$$\begin{aligned}
x_1^{(n)} &= \sin t_1 \sin t_2 \cdots \sin t_{n-1},\\
x_2^{(n)} &= \cos t_1 \sin t_2 \cdots \sin t_{n-1},\\
x_3^{(n)} &= \cos t_2 \sin t_3 \cdots \sin t_{n-1},\\
\cdots & \cdots \cdots\\
x_{n-1}^{(n)} &= \cos t_{n-2} \sin t_{n-1},\\
x_n^{(n)} &= \cos t_{n-1},
\end{aligned}$$

Thus we have $x_n^n = \cos t_{n-1}$ and $x_j^{(n)} = x_j^{(n-1)} \sin t_{n-1}$ for $j < n$. Suppose we have proved that

$$\left(\sum_{i=1}^{n-1} (-1)^{i-1} x_i^{(n-1)} dx_1^{(n-1)} \wedge \cdots \wedge dx_{i-1}^{(n-1)} \wedge dx_{i+1}^{(n-1)} \wedge \cdots \wedge dx_{n-1}^{(n-1)}\right)_{T_{n-1}} =$$

$$= (-1)^{n-2}(\sin t_1 \sin^2 t_2 \cdots \sin^{n-3} t_{n-2}\, dt_1 \wedge \cdots \wedge dt_{n-2}.$$

We observe that the Jacobian matrix of the transformation T_n is the $n \times (n-1)$ matrix

$$\frac{\partial(x_1^{(n)}, \ldots, x_n^{(n)})}{\partial(t_1, \ldots, t_{n-1})} = \begin{pmatrix} \frac{\partial x_1^{(n-1)}}{\partial t_1} \sin t_{n-1} & \cdots & \frac{\partial x_1^{(n-1)}}{\partial t_{n-2}} \sin t_{n-1} & x_1^{(n-1)} \cos t_{n-1} \\ \frac{\partial x_2^{(n-1)}}{\partial t_1} \sin t_{n-1} & \cdots & \frac{\partial x_2^{(n-1)}}{\partial t_{n-2}} \sin t_{n-1} & x_2^{(n-1)} \cos t_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_{n-1}^{(n-1)}}{\partial t_1} \sin t_{n-1} & \cdots & \frac{\partial x_{n-1}^{(n-1)}}{\partial t_{n-2}} \sin t_{n-1} & x_{n-1}^{(n-1)} \cos t_{n-1} \\ 0 & \cdots & 0 & -\sin t_{n-1} \end{pmatrix}$$

It follows immediately, when we expand the determinant of the first $n-1$ rows along the last column, that

$$\begin{aligned} &(dx_1^{(n)} \wedge \cdots \wedge dx_{n-1}^{(n)})_{T_n} = \sin^{n-2} t_{n-1} \cos t_{n-1} \times \\ &\quad \times \sum_{i=1}^{n-1} (-1)^{n-1+i} x_i^{(n-1)} \frac{\partial(x_1^{(n-1)}, \ldots, x_{n-1}^{(n-1)})}{\partial(t_1, \ldots, t_{n-2})} dt_1 \wedge \cdots \wedge dt_{n-1} r = \\ &\quad = (-1)^n \sin^{n-2} t_{n-1} \cos t_{n-1} \Big(\sum_{i=1}^{n-1} (-1)^{i-1} x_i^{(n-1)}\, dx_1^{(n-1)} \wedge \cdots \\ &\qquad \cdots \wedge dx_{i-1}^{(n-1)} \wedge dx_{i+1}^{(n-1)} \wedge \cdots \wedge dx_{n-1}^{(n-1)} \Big)_{T_{n-1}} \wedge dt_{n-1} \\ &\quad = \sin^{n-2} t_{n-1} \cos t_{n-1} \big(\sin t_2 \sin^2 t_3 \cdots \sin^{n-3} t_{n-2} \big) dt_1 \wedge \cdots \wedge dt_{n-1}. \end{aligned}$$

Hence

$$\begin{aligned} &(-1)^{n-1} \big(x_n^{(n)}\, dx_1^{(n)} \wedge \cdots \wedge dx_{n-1}^{(n)} \big)_{T_n} = \\ &\quad = (-1)^{n-1} \cos^2 t_{n-1} \sin t_2 \sin^2 t_3 \cdots \sin^{n-2} t_{n-1} \big)\, dt_1 \wedge \cdots \wedge dt_{n-1}. \end{aligned}$$

Next, omitting row i ($i < n$) and expanding the resulting determinant along the last row, we find that

$$\begin{aligned} &(dx_1^{(n)} \wedge \cdots \wedge dx_{i-1}^{(n)} \wedge dx_{i+1}^{(n)} \wedge \cdots \wedge dx_n^{(n)})_{T_n} = \\ &= -\sin^{n-1} t_{n-1} (dx_1^{(n-1)} \wedge \cdots \wedge dx_{i-1}^{(n-1)} \wedge dx_{i+1}^{(n-1)} \wedge \cdots \wedge dx_{n-1}^{(n-1)})_{T_{n-1}} \wedge dt_{n-1}, \end{aligned}$$

so that

$$\begin{aligned} &\Big(\sum_{i=1}^{n-1} (-1)^{i-1} x_i^{(n)} dx_1^{(n)} \wedge \cdots \wedge dx_{i-1}^{n} \wedge dx_{i+1}^{(n)} \wedge \cdots \wedge dx_n^{(n)} \Big)_{T_n} = \\ &\quad = -\sin^n t_{n-1} \Big(\sum_{i=1}^{n-1} (-1)^{i-1} x_i^{(n-1)} dx_1^{(n-1)} \wedge \cdots \wedge dx_{i-1}^{(n-1)} \wedge dx_{i+1}^{(n-1)} \wedge \cdots \\ &\qquad \cdots \wedge dx_{n-1}^{(n-1)} \Big)_{T_{n-1}} \wedge dt_{n-1}, \end{aligned}$$

and again by induction this is

$$(-1)^{n-1}\sin^2 t_{n-1}\big(\sin t_2 \sin^2 t_3 \cdots \sin^{n-2} t_{n-1}\big)dt_1 \cdots dt_{n-1}.$$

Combining these results we find that

$$\Big(\sum_{i=1}^{n}(-1)^{i-1}x_i^{(n)}dx_1^{(n)}\wedge\cdots\wedge dx_{i-1}^{(n)}\wedge dx_{i+1}^{(n)}\wedge\cdots\wedge dx_n^{n)}\Big)_{T_n} =$$

$$= (-1)^{n-1}\sin t_2 \sin^2 t_3 \cdots \sin^{n-2} t_{n-1}dt_1\wedge\cdots\wedge dt_{n-1}.$$

The induction is now complete.

Except for the unimportant factor of -1, this formula gives results consistent with the known results for the area of the $(n-1)$-sphere, namely a total area of

$$A_{n-1} = \frac{2\pi^{\frac{n}{2}}}{\Gamma\big(\frac{n}{2}\big)}.$$

This is easily verified for $n = 2$ and $n = 3$. In general

$$A_{n-1} = A_{n-2}\int_0^{\pi}\sin^{n-2} t_{n-1}\,dt_{n-1} =$$

$$2A_{n-2}\int_0^{\frac{\pi}{2}}\sin^{n-2} s\,ds = 2A_{n-2}\frac{\Gamma\big(\frac{1}{2}\big)\Gamma\big(\frac{n-1}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}.$$

It easily follows, since $\Gamma\big(\frac{1}{2}\big) = \sqrt{\pi}$, that the formula for the surface area of $\Sigma^{(n-1)}$ is valid for all n.

Similarly we can show the analog of part (c) of Exercise 22, namely that

$$\int_{\Phi}\omega_n = 0$$

for any $(n-1)$-dimensional surface given by a mapping of the form

$$\Phi(s_1,\ldots,s_{n-2},t) = \big(g(s_1,\ldots,s_{n-2},t)h_1(s_1,\ldots,s_{n-2}),\ldots$$
$$\ldots,g(s_1,\ldots,s_{n-2},t)h_n(s_1,\ldots,s_{n-2})\big).$$

Indeed, the pullback of ω_n is

$$(\omega_n)_\Phi =$$
$$= \begin{vmatrix} gh_1 & \frac{\partial g}{\partial s_1}h_1 + g\frac{\partial h_1}{\partial s_1} & \cdots & \frac{\partial g}{\partial s_{n-2}}h_1 + g\frac{\partial h_1}{\partial s_{n-2}} & \frac{\partial g}{\partial t}h_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ gh_n & \frac{\partial g}{\partial s_1}h_n + g\frac{\partial h_n}{\partial s_1} & \cdots & \frac{\partial g}{\partial s_{n-2}}h_n + g\frac{\partial h_n}{\partial s_{n-2}} & \frac{\partial g}{\partial t}h_n \end{vmatrix}\, ds_1\wedge\cdots\wedge ds_{n-2}\wedge dt.$$

But this determinant is zero, since the first and last columns are proportional.

We can now prove that if $f(t_1,\ldots,t_{n-1}) > 0$ and

$$\Omega(t_1,\ldots,t_{n-1}) = f(t_1,\ldots,t_{n-1})\Sigma^{(n-1)}(t_1,\ldots,t_{n-1}),$$

then

$$\int_{\Omega} \omega_n = \int_S \omega_n = \pm A_{n-1}(S).$$

To do so, we consider the n-surface in R^n given by

$$\Psi(t_1,\ldots,t_{n-1},t) = [1-t+tf(t_1,\ldots,t_{n-1})]\Sigma^{(n-1)}(t_1,\ldots,t_{n-1}),$$

for $0 \le t \le 1$ and $t_1,\ldots,t_n$ ranging over a parallelepiped contained in the interior of D with boundary faces parallel to those of D. For each fixed t_j, this Ψ is an $(n-1)$-surface of the form just considered, and hence the integral of ω_n over it is zero. This applies in particular to the faces of the closed parallelepiped E. Since $\int_{\partial\Psi} \omega_n = \int_{\Psi} d\omega_n = 0$, it then follows that, up to a factor of ± 1,

$$\int_{\Omega} \omega_n = \int_S \omega_n = A_{n-1}(S).$$

Finally, as in Exercise 22, ω_n is exact in the complement of every $(n-2)$-hyperplane through the origin, since there is a rotation that maps the complement of that hyperplane to E_{n-1}, while ω_n is rotation-invariant.

Exercise 10.24 Let $\omega = \sigma a_i(\mathbf{x})\,dx_i$ be a 1-form of class C'' in a convex open set $E \subset R^n$. Assume $d\omega = 0$ and prove that ω is exact in E by completing the following outline:

Fix $\mathbf{p} \in E$. Define

$$f(\mathbf{x}) = \int_{[\mathbf{p},\mathbf{x}]} \omega \quad (\mathbf{x} \in E).$$

Apply Stokes' theorem to affine-oriented 2-simplexes $[\mathbf{p},\mathbf{x},\mathbf{y}]$ in E. Deduce that

$$f(\mathbf{y}) - f(\mathbf{x}) = \sum_{i=1}^{n} (y_i - x_i) \int_0^1 a_i((1-t)\mathbf{x} + t\mathbf{y})\,dt$$

for $\mathbf{x} \in E$, $\mathbf{y} \in E$. Hence $D_i f(\mathbf{x}) = a_i(\mathbf{x})$.

Solution. Because $d\omega = 0$, the integral of ω over the boundary of the oriented 2-simplex $[\mathbf{p},\mathbf{x},\mathbf{y}]$ is zero. That is

$$\int_{[\mathbf{x},\mathbf{y}]} \omega - \int_{[\mathbf{p},\mathbf{y}]} \omega + \int_{[\mathbf{p},\mathbf{x}]} \omega = 0,$$

which can be rewritten as

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_{[\mathbf{x},\mathbf{y}]} \omega = \sum_{i=1}^{n} (y_i - x_i) \int_0^1 a_i((1-t)\mathbf{x} + t\mathbf{y})\,dt.$$

Differentiating with respect to y_i, we find

$$D_i f(\mathbf{y}) = \int_0^1 a_i((1-t)\mathbf{x}+t\mathbf{y})\,dt + \sum_{j=1}^{n}(y_j - x_j)\int_0^1 tD_i a_j((1-t)\mathbf{x}+t\mathbf{y})\,dt.$$

The fact that $d\omega = 0$ says that $D_i a_j = D_j a_i$, so that we have

$$\begin{aligned}
D_i f(\mathbf{y}) &= \int_0^1 a_i((1-t)\mathbf{x}+t\mathbf{y})dt + \int_0^1 \sum_{j=1}^{n} t(y_j - x_j)D_j a_i((1-t)\mathbf{x}+t\mathbf{y})dt \\
&= \int_0^1 a_i((1-t)\mathbf{x}+t\mathbf{y})\,dt + \int_0^1 t\frac{d}{dt}a_i((1-t)\mathbf{x}+t\mathbf{y})\,dt \\
&= \int_0^1 a_i((1-t)\mathbf{x}+t\mathbf{y})\,dt + ta_i((1-t)\mathbf{x}+t\mathbf{y})\Big|_0^1 \\
&\qquad - \int_0^1 a_i((1-t)\mathbf{x}+t\mathbf{y})\,dt \\
&= a_i(\mathbf{y}).
\end{aligned}$$

Thus $\omega = df$.

Exercise 10.25 Assume that ω is a 1-form in an open set $E \subset R^n$ such that

$$\int_\gamma \omega = 0$$

for every closed curve γ in E of class C'. Prove that ω is exact in E, by imitating part of the argument sketched in Exercise 24.

Solution. We first observe that Stokes' theorem and the argument of Theorem 10.15 show that $d\omega = 0$ in E. (Theorem 10.15 actually shows that if some component of $d\omega$ is nonzero at some point of E, then there is a 2-surface Φ in E whose domain is a 2-cell in R^2 for which $\int_\Phi d\omega \neq 0$. Then by Stokes' theorem, $\int_{\partial\Phi}\omega \neq 0$ also, contradicting the assumption of the problem.

In each connected component E_α of E, we choose a fixed point $\mathbf{x}_\alpha$. There is a ball of some positive radius r_α centered at $\mathbf{x}_\alpha$ and contained in E. Let this ball be B_α. Exercise 24 shows that there is a function $f(\mathbf{x})$ such that $\omega(\mathbf{x}) = df(\mathbf{x})$ inside B_α. By subtracting a constant from f we can assume that $f(\mathbf{x}_\alpha) = 0$.

Now consider the set S of all points $\mathbf{x} \in E_\alpha$ having the property that there exist a connected open set $F_\mathbf{x}$ containing $\mathbf{x}$ and $\mathbf{x}_\alpha$ and a function $f_\mathbf{x}$ defined on $F_\mathbf{x}$ such that $df_\mathbf{x} = \omega$ on $F_\mathbf{x}$ and $f_{\mathbf{x}_\alpha} = 0$. It is clear that S is an open connected subset of E_α, being the union of all the connected open sets $F_\mathbf{x}$, which have the common point $\mathbf{x}_\alpha$. It is also clear that there is a function f defined on S such that $df = \omega$ on S. In fact we can define $f(\mathbf{x}) = f_\mathbf{x}(\mathbf{x})$, and this definition is unambiguous, since if $f_\mathbf{x}$ and $f_\mathbf{y}$ are both defined at $\mathbf{z}$, then

$$f_\mathbf{x}(\mathbf{z}) = \int_\gamma df_\mathbf{x} = \int_\gamma \omega = \int_\delta \omega = \int_\delta df_\mathbf{y} = f_\mathbf{y}(\mathbf{z}).$$

Here γ is a path in $E_{\mathbf{x}}$ from $\mathbf{x}_\alpha$ to $\mathbf{z}$, and δ is a path in $E_{\mathbf{y}}$ from $\mathbf{x}_\alpha$ to $\mathbf{z}$. The path $\gamma - \delta$ lies in E and is a closed loop, so that

$$\int_{\gamma-\delta} \omega = 0.$$

We need only show that $S = E_\alpha$. But if not, then E_α contains a boundary point $\mathbf{x} \in S$. Some open ball B about $\mathbf{x}$ is contained in E, and this open ball contains a point $\mathbf{y} \in S$. But then there exists a function g such that $dg = \omega$ in B, and subtracting a constant makes it possible to ensure that $g(\mathbf{y}) = f_{\mathbf{y}}(\mathbf{y}) = f(\mathbf{y})$. We claim that $g(\mathbf{z}) = f(\mathbf{z})$ on the entire set $S \cap B$. In fact this argument merely repeats the argument just given to show that f is unambiguously defined. It then follows that $\mathbf{y}$ is contained in the connected open set $S \cap B$ and that the function h defined to be f on S and g on B has the property that $dh = \omega$ on $S \cap B$. By definition, this means $\mathbf{y} \in S$, which contradicts the assumption that $\mathbf{y}$ is a boundary point of S. Therefore $S = E_\alpha$.

Thus we can find a primitive for ω on each connected component of E. These primitives can be pieced together to provide a single primitive for ω on E.

Exercise 10.26 Assume ω is a 1-form in $R^3 \setminus \{\mathbf{0}\}$, of class C' and $d\omega = 0$. Prove that ω is exact in $R^3 \setminus \{\mathbf{0}\}$.

Hint: Every closed continuously differentiable curve in $R^3 \setminus \{\mathbf{0}\}$ is the boundary of a 2-surface in $R^3 \setminus \{\mathbf{0}\}$. Apply Stokes' theorem and Exercise 25.

Solution. Given the assumption in the hint, the solution is easy. By Exercise 25 we need only show that the integral of ω over every closed curve is zero. By the assertion in the hint, this closed curve is the boundary of a two-surface. By Stokes' theorem, the integral of ω over the curve equals the integral of $d\omega$ over the 2-surface.

To prove the claim that every continuously differentiable curve in $R^3 \setminus \{\mathbf{0}\}$ is the boundary of a two-surface, we may assume that the curve is of the form $\mathbf{x}(t)$, $0 \le t \le 1$ and $\mathbf{x}(0) = \mathbf{x}(1)$. Let $\mathbf{x}(t) = \big(x(t), y(t), z(t)\big)$. We shall show first of all that there is some line through the origin in R^3 that does not intersect the curve.

To that end, we observe that the intersection of a sphere of radius ρ in R^3 with a ball of radius r ($r \le 2\rho$) about a point of the sphere is a spherical cap whose area is πr^2. (Note that this result is independent of ρ. It is a remarkable fact, whose proof is a routine computation.) Since the area of the whole sphere is $4\pi\rho^2$, it follows that half of any given hemisphere cannot be covered by fewer than ρ^2/r^2 such spherical caps. Now, since $\mathbf{x}(t) \ne \mathbf{0}$ and $\mathbf{x}'(t)$ is continuous, it follows that $\mathbf{v}(t) = \mathbf{x}(t)/|\mathbf{x}(t)|$ is a Lipschitz function, that is, there exists a constant M such that $|\mathbf{v}(s) - \mathbf{v}(t)| \le M|t - s|$ for all s and t. In particular the image of each interval $[k/n, (k+1)/n]$ is contained in a spherical cap of radius M/n. Thus the complete curve is contained in a set of n spherical caps of radius at most M/n. But to cover the half of any given hemisphere of the unit sphere requires at least $\frac{n^2}{M^2}$ such caps. Hence, if $n > M^2$, the projection of the curve $\mathbf{x}(t)$ on the unit sphere is contained in a set of spherical caps covering less than half of the upper hemisphere and less than half of the lower hemisphere. Hence there are two antipodal points $\mathbf{x}_0$ and $-\mathbf{x}_0$ on the unit sphere not in its image. That means there is at least one line through the origin that the curve does not intersect.

This line through the origin gives us a sense of positive rotation from $\mathbf{x}(t)$ to $\mathbf{x}(t+\frac{1}{2})$ for each $t \in [0, \frac{1}{2}]$. We can then construct a $\mathcal{C}'$-curve $\gamma_t(s)$ in $R^3\backslash\{\mathbf{0}\}$ that goes from $\mathbf{x}(t)$ to $\mathbf{x}(t+\frac{1}{2})$ by letting cylindrical coordinates vary linearly with respect to s. To be specific, we can assume without loss of generality that the line is the z-axis. In that case, the radial coordinate $r(t) = \sqrt{x^2(t) + y^2(t)}$ is never zero and is a continuously differentiable function of position. We choose $\theta(t)$ as the cylindrical polar coordinate of $\mathbf{x}(t)$ in a continuously differentiable manner for $0 \le t \le 1$. (This is possible by piecing together sections of this function over sufficiently small intervals.) We then define $\gamma(s,t) = \big(x(s,t), y(s,t), z(s,t)\big)$ for $0 \le s \le 1$, $0 \le t \le 1/2$ by

$$\begin{aligned} x(t,u) &= (1-u)r(t)\cos\big((1-u)\theta(t)\big) + ur(1-t)\cos\big(u\theta(1-t)\big)\big), \\ y(t,u) &= (1-u)r(t)\sin\big((1-u)\theta(t)\big) + ur(1-t)\sin\big(u\theta(1-t)\big)\big), \\ z(t,u) &= (1-u)z(t) + uz(1-t). \end{aligned}$$

We let the boundary of this cell be $\delta_1+\delta_2+\delta_3+\delta_4$. Here δ_1 is $\gamma(t,0)$, $0 \le t \le 1/2$, which is just $\mathbf{x}(t)$ over the same interval; δ_2 is $\gamma(1/2,u)$, which is the "line segment" from $\mathbf{x}(1/2)$ to $\mathbf{x}(1/2)$, whose range is just a point, and hence counts as 0 when regarded as a 1-chain; δ_3 is $\gamma(1/2-t,1)$ which is just $\mathbf{x}(t+\frac{1}{2})$, so that $\delta_1+\delta_3$ represents $\mathbf{x}(t)$ as t goes from 0 to 1. Finally δ_4 is $\gamma(0,u)$, which is the line segment from $\mathbf{x}(1)$ to $\mathbf{x}(0)$, and since $\mathbf{x}$ is a closed curve, these two points are the same. Hence once again δ_4 counts as 0 when regarded as a 1-chain. Thus the boundary of γ is indeed the curve $\mathbf{x}$.

Exercise 10.27 Let E be an open 3-cell in R^3, with edges parallel to the coordinate axes. Suppose $(a,b,c) \in E$, $f_t \in \mathcal{C}'(E)$ for $i = 1, 2, 3$,

$$\omega = f_1\,dy \wedge dz + f_2\,dz \wedge dx + f_3\,dx \wedge dy,$$

and assume that $d\omega = 0$ in E. Define

$$\lambda = g_1\,dx + g_2\,dy$$

where

$$\begin{aligned} g_1(x,y,z) &= \int_c^z f_2(x,y,s)\,ds - \int_b^y f_3(x,t,c)\,dt \\ g_2(x,y,z) &= -\int_c^z f_1(x,y,s)\,ds, \end{aligned}$$

for $(x, y, z) \in E$. Prove that $d\lambda = \omega$ in E.

Evaluate these integrals when $\omega = \zeta$ and thus find the form λ that occurs in part (e) of Exercise 22.

Solution. Since

$$d\lambda = -\frac{\partial g_2}{\partial z}\, dy \wedge dz + \frac{\partial g_1}{\partial z}\, dz \wedge dx + \left(\frac{\partial g_2}{\partial x} - \frac{\partial g_1}{\partial y}\right) dx \wedge dy,$$

we need only show that

$$\begin{aligned} \frac{\partial g_2}{\partial z} &= -f_1, \\ \frac{\partial g_1}{\partial z} &= f_2, \\ \frac{\partial g_2}{\partial x} - \frac{\partial g_1}{\partial y} &= f_3. \end{aligned}$$

The first two equations are immediate. As for the third, direct computation shows that

$$D_1 g_2(x, y, z) - D_2 g_1(x, y, z) = \int_c^z -(D_1 f_1(x, y, s) + D_2 f_2(x, y, s))\, ds + f_3(x, y, c).$$

Now the assumption that $d\omega = 0$ says that

$$D_1 f_1(x, y, s) + D_2 f_2(x, y, s) = -D_3 f_3(x, y, s),$$

Substituting this value into the last expression and evaluating the integral using the fundamental theorem of calculus yields the result $d\lambda = \omega$.

Taking

$$\begin{aligned} f_1(x, y, z) &= \frac{x}{(x^2 + y^2 + z^2)^{3/2}}, \\ f_2(x, y, z) &= \frac{y}{(x^2 + y^2 + z^2)^{3/2}}, \\ f_3(x, y, z) &= \frac{z}{(x^2 + y^2 + z^2)^{3/2}}, \end{aligned}$$

we get

$$\begin{aligned} g_2(x, y, z) = -\int_c^z \frac{x}{(x^2 + y^2 + s^2)^{3/2}}\, ds &= \\ &= \frac{1}{x^2 + y^2}\left(\frac{cx}{\sqrt{x^2 + y^2 + c^2}} - \frac{zx}{\sqrt{x^2 + y^2 + z^2}}\right), \end{aligned}$$

$$\begin{aligned} g_1(x, y, z) &= \int_c^z \frac{y}{(x^2 + y^2 + s^2)^{3/2}}\, ds - \int_b^y \frac{c}{(x^2 + t^2 + c^2)^{3/2}}\, dt \\ &= \frac{1}{x^2 + y^2}\left(\frac{yz}{\sqrt{x^2 + y^2 + z^2}} - \frac{yc}{\sqrt{x^2 + y^2 + c^2}}\right) \\ &\quad - \frac{1}{x^2 + c^2}\left(\frac{cy}{\sqrt{x^2 + y^2 + c^2}} - \frac{bc}{\sqrt{x^2 + b^2 + c^2}}\right). \end{aligned}$$

It is a routine computation to verify that these functions do indeed provide a primitive for ω.

Exercise 10.28 Fix $b > a > 0$, define

$$\Phi(r, \theta) = (r\cos\theta, r\sin\theta)$$

for $a \le r \le b$, $0 \le \theta \le 2\pi$. (The range of Φ is an annulus in R^2.) Put $\omega = x^3\,dy$, and compute both

$$\int_\Phi d\omega \text{ and } \int_{\partial\Phi} \omega$$

to verify that they are equal.

Solution. Since $d\omega = 3x^2\,dx \wedge dy$, we have $(d\omega)_\Phi = -r\,dr \wedge d\theta$, and

$$\int_\Phi d\omega = -\int_a^b \int_0^{2\pi} 3r^3\cos^2\theta\,d\theta\,dr = \frac{3\pi}{4}(a^4 - b^4).$$

For the integral over the boundary we have $dy = r\cos\theta\,d\theta + \sin\theta\,dr$, and we get

$$\int_0^{2\pi} (a^4 - b^4)\cos^4\theta\,d\theta = \frac{3\pi}{4}(a^4 - b^4).$$

Exercise 10.29 Prove the existence of a function α with the properties needed in the proof of Theorem 10.38, and prove that the resulting function F is of class C'. (Both assertions become trivial if E is an open cell or an open ball, since α can then be taken to be a constant. Refer to Theorem 9.42.)

Solution. We are given a convex open set $V \subseteq R^p$ whose projection on R^{p-1} is the convex open set U. We need to show that there is a continuously differentiable function $\alpha : U \to R$ whose graph is contained in V. If V is a cell or an open ball, there exists a section $x_p = c$ of it (many sections, if it is a cell), whose projection is U, and we can simply define $\alpha(\mathbf{y}) = c$ for all $\mathbf{y} \in U$.

Now write V as a countable union of open balls $V = \bigcup\limits_{i=1}^{\infty} B_i$. Also write V as the union of an increasing sequence of compact sets K_n such that $K_n \subseteq \operatorname{int}(K_{n+1})$.

We claim that, as in Theorem 10.8, there exist continuous functions ψ_i such that the support of ψ_i is contained in the projection of B_i on R^{p-1}, $0 \le \psi_i(\mathbf{y}) \le 1$ for all $\mathbf{y}$, and $\sum\limits_{i=1} \psi_i(\mathbf{y}) = 1$ for all $\mathbf{y} \in U$. Moreover, this sum is *locally finite,* that is, each point $\mathbf{y}$ has a neighborhood $U_\mathbf{y}$ such that the set of indices i for which $\psi_i(\mathbf{z}) \ne 0$ for some $\mathbf{z} \in U_\mathbf{y}$ is finite.

To construct such functions, for each $\mathbf{x} \in V$, let $i(\mathbf{x})$ be the smallest index r such that $\mathbf{x} \in B_i$. Then, as in the proof of Theorem 10.8, for each $\mathbf{x} \in K_1$, choose open balls $B(\mathbf{x})$ and $W(\mathbf{x})$ centered at $\mathbf{x}$ such that

$$\overline{B(\mathbf{x})} \subset W(\mathbf{x}) \subset \overline{W(\mathbf{x})} \subset B_{i(\mathbf{x})}.$$

Since K_1 is compact, there are points $\mathbf{x}_{11}, \ldots, \mathbf{x}_{1N_1}$ such that

$$K_1 \subseteq B(\mathbf{x}_{11}) \cup \cdots \cup B(\mathbf{x}_{1N_1}).$$

For later convenience we define $L_1 = K_1$.

Now let $L_2 = K_2 \setminus \bigcup_{j=1}^{N_1} B(\mathbf{x}_{1j})$. For each $\mathbf{x} \in L_2$ there are open balls $B(\mathbf{x})$ and $W(\mathbf{x})$ centered at $\mathbf{x}$ such that

$$\overline{B(\mathbf{x})} \subset W(\mathbf{x}) \subset \overline{W(\mathbf{x})} \subset B_{i(\mathbf{x})} \setminus K_1.$$

Since L_2 is compact, we choose a finite set of points $\mathbf{x}_{21}, \ldots, \mathbf{x}_{2N_2}$ such that

$$L_2 \subseteq B(\mathbf{x}_{21}) \cup \cdots \cup B(\mathbf{x}_{2N_2}).$$

Notice that $K_2 \subset \bigcup_{k=1}^{2} \bigcup_{j=1}^{N_k} B(\mathbf{x}_{kj})$.

Now suppose we have chosen a (possibly empty) collection of open balls $B(\mathbf{x}_{kj})$ and $W(\mathbf{x}_{kj})$, $1 \leq j \leq N_k$, $1 \leq k \leq r$, centered at $\mathbf{x}_{kj} \in L_k = K_k \setminus \bigcup_{i=1}^{k-1} \bigcup_{j=1}^{N_i} B(\mathbf{x}_{ij})$, and such that

$$\overline{B(\mathbf{x}_{kj})} \subset W(\mathbf{x}_{kj}) \subset \overline{W(\mathbf{x}_{kj})} \subset B_{i(\mathbf{x}_{kj})} \setminus K_{k-1},$$

and

$$K_r \setminus \bigcup_{k=1}^{r-1} \bigcup_{j=1}^{N_k} B(\mathbf{x}_{kj}) \subset \bigcup_{j=1}^{N_r} B(\mathbf{x}_{rj}),$$

and

$$K_s \subset \bigcup_{k=1}^{s} \bigcup_{j=1}^{N_k} B(\mathbf{x}_{kj}),$$

for $1 \leq s \leq r-1$. It then follows from the last two relationships that the last one also holds with $s = r$. By then considering the compact set $L_{r+1} = K_{r+1} \setminus \bigcup_{k=1}^{r} \bigcup_{j=1}^{N_k} B(\mathbf{x}_{kj})$ and repeating the argument, we can assume that the sets $B(\mathbf{x}_{kj})$ and $W(\mathbf{x}_{kj})$ with these properties have been chosen for all k and all j, $1 \leq k < \infty$, $1 \leq j \leq N_k$. It follows in particular that

$$V = \bigcup_{n=1}^{\infty} K_n \subset \bigcup_{k=1}^{\infty} \bigcup_{j=1}^{N_k} B(\mathbf{x}_{kj}).$$

Now let $\widetilde{K}_k$, $\widetilde{B}(\mathbf{x}_{kj})$, and $\widetilde{W}(\mathbf{x}_{kj})$ be respectively the projections on R^{p-1} of K_k, $B(\mathbf{x}_{kj})$, and $W(\mathbf{x}_{kj})$, and let

$$\widetilde{L}_r = \widetilde{K}_k \setminus \bigcup_{k=1}^{r-1} \bigcup_{j=1}^{N_k} \widetilde{B}(\mathbf{x}_{kj}).$$

We then choose functions φ_{jk} as smooth as we like such that $\varphi_{kj}(\mathbf{y}) = 1$ on $\widetilde{B}(\mathbf{x}_{kj})$ (and hence also on $\overline{\widetilde{B}(\mathbf{x}_{kj})}$), $\varphi_{kj}(\mathbf{y}) = 0$ outside $\widetilde{W}(\mathbf{x}_{kj})$, and $0 \leq \varphi_{kj}(\mathbf{y}) \leq 1$ on R^{p-1}. Let $\varphi_j(\mathbf{y}) = \varphi_{1j}(\mathbf{y})$ for $1 \leq j \leq N_1$ and $\varphi_j(\mathbf{y}) =$

$\varphi_{k,j-(N_1+\cdots+N_{k-1})}(\mathbf{y})$ for $N_1+\cdots+N_{k-1} < j \leq N_1+\cdots+N_k$, $2 \leq k < \infty$. We define $\mathbf{x}_j$ analogously. Let $\mathbf{x}_j = (\mathbf{y}_j, c_j)$.

We then proceed to define $\psi_1(\mathbf{y}) = \varphi_1(\mathbf{y})$ and

$$\psi_{j+1}(\mathbf{y}) = \left(1-\varphi_1(\mathbf{y})\right)\cdots\left(1-\varphi_j(\mathbf{y})\right)\varphi_{j+1}(\mathbf{y})$$

for $j = 1, 2, \ldots$, as in Theorem 10.8. It is obvious that the support of ψ_j is contained in the closure of $\widetilde{W}(\mathbf{x}_j)$ and hence in $\widetilde{B}_{i(\mathbf{x}_j)} \setminus \widetilde{K}_{k-1} \subseteq U \setminus \widetilde{K}_{k-1}$ when $N_1+\cdots+N_{k-1} < j \leq N_1+\cdots+N_k$.

Now by the choice of the sets K_n, if $\mathbf{y} \in U$, there is some n such that $\mathbf{y} \in \widetilde{K}_n \subset \operatorname{int}\left(\widetilde{K}_{n+1}\right)$, and hence $\psi_j(\mathbf{y}) = 0$ on the open neighborhood $\operatorname{int} \widetilde{K}_{n+1}$ of $\mathbf{y}$ if $j > N_1+\cdots+N_{n+1}$. Therefore the sum and product

$$\sum_{j=1}^{\infty} \psi_j(\mathbf{y}) = 1 - \prod_{i=1}^{\infty}[1-\varphi_i(\mathbf{y})]$$

are both locally finite at each point. (Local finiteness of the product means all but a finite number of factors equal 1 on a neighborhood of each point.) However, if $y \in U$, then $\mathbf{y} \in \widetilde{B}(\mathbf{x}_j)$ for some j, and so $\varphi_j(\mathbf{y}) = 1$, from which it then follows that

$$\sum_{j=1}^{\infty} \psi_j(\mathbf{y}) = 1$$

for all $\mathbf{y} \in U$.

Since we have defined c_j so that $\mathbf{x}_j = (\mathbf{y}_j, c_j)$, it follows that the projection of the c_j-section of $B(\mathbf{x}_j)$ on R^{p-1}, which we denote C_j, is the same as the projection of $B(\mathbf{x}_j)$ on this subspace. That is, it is $\widetilde{B}(\mathbf{x}_j)$. We can now let $\alpha(\mathbf{y}) = \sum\limits_{j=1}^{\infty} c_j\psi_1(\mathbf{y})$. For then at each $\mathbf{y} \in U$ there is a finite integer n such that

$$(\mathbf{y}, \alpha(\mathbf{y})) = \psi_1(\mathbf{y})(\mathbf{y}, c_1) + \cdots + \psi_n(\mathbf{y})(\mathbf{y}, c_n).$$

Since $\psi_k(\mathbf{y}) = 0$ if $\mathbf{y} \notin C_k$ and $(\mathbf{y}, c_k) \in B_k \subset V$ if $\mathbf{y} \in C_k$, it follows that $(\mathbf{y}, \alpha(\mathbf{y}))$ is a weighted average of points in V, hence belongs to V for all $\mathbf{y} \in U$.

Exercise 10.30 If $\mathbf{N}$ is the vector given by (135), prove that

$$\det\begin{bmatrix} \alpha_1 & \beta_1 & \alpha_2\beta_3 - \alpha_3\beta_2 \\ \alpha_2 & \beta_2 & \alpha_3\beta_1 - \alpha_1\beta_3 \\ \alpha_3 & \beta_3 & \alpha_2\beta_2 - \alpha_2\beta_1 \end{bmatrix} = |\mathbf{N}|^2.$$

Also, verify Eq. (137).

Solution. The equation in the problem is a straightforward computation, and amounts merely to expanding the determinant along the last column. Likewise Eq. (137), which merely asserts that a cross product is perpendicular to each of the factors, is routine. The two inner products in the equation can be obtained by replacing the last column of this determinant by either $(\alpha_1, \alpha_2, \alpha_3)$ or $(\beta_1, \beta_2, \beta_3)$. In each case, the result is a determinant with two equal columns, which is therefore zero.

Exercise 10.31 Let $E \subset R^3$ be open, suppose $g \in C''(E)$, $h \in C''(E)$, and consider the vector field

$$\mathbf{F} = g\nabla h.$$

(a) Prove that

$$\nabla \cdot \mathbf{F} = g\nabla^2 h + (\nabla g) \cdot (\nabla h)$$

where $\nabla^2 h = \nabla \cdot (\nabla h) = \sum \partial^2 h/\partial x_i^2$ is the so-called "Laplacian" of h. (b) If Ω is a closed subset of E with positively oriented boundary $\partial\Omega$ (as in Theorem 10.51), prove that

$$\int_\Omega [g\nabla^2 h + (\nabla g) \cdot (\nabla h)]\, dV = \int_{\partial\Omega} g\frac{\partial h}{\partial n}\, dA$$

where (as is customary) we have written $\partial h/\partial n$ in place of $(\nabla h) \cdot \mathbf{n}$. (Thus $\partial h/\partial n$ is the directional derivative of h in the direction of the outward normal to $\partial\Omega$, the so-called *normal derivative* of h.) Interchange g and h, subtract the resulting formula from the first one, to obtain

$$\int_\Omega (g\nabla^2 h - h\nabla^2 g)\, dV = \int_{\partial\Omega} \left(g\frac{\partial h}{\partial n} - h\frac{\partial g}{\partial n}\right) dA.$$

These two formulas are usually called *Green's identities.*

(c) Assume that h is *harmonic* in E; this means that $\nabla^2 h = 0$. Take $g = 1$ and conclude that

$$\int_{\partial\Omega} \frac{\partial h}{\partial n}\, dA = 0.$$

Take $g = h$, and conclude that $h = 0$ in Ω if $h = 0$ on $\partial\Omega$.

(d) Show that Green's identities are also valid in R^2.

Solution. Part (a) is simply the product rule for derivatives.

The main equation in part (b) is simply the divergence theorem applied to $\mathbf{F}$. Green's identities then follow by completely routine computation.

(c) Taking $g = 1$ forces $\partial g/\partial n = 0$ and $\nabla^2 g = 0$. Since $\nabla^2 h = 0$ by the assumption that h is harmonic, the result follows. For the other assertion of this part we have to go back to the main equation before taking $g = h$. When we do, we actually get a slightly stronger assertion: $\nabla h = 0$ in Ω, and so h is constant on each component of Ω, if either $h = 0$ or $\partial h/\partial n = 0$ on all of $\partial\Omega$. When $h = 0$ on $\partial\Omega$, obviously the constant value of h must be 0.

(d) The "two-dimensional" divergence theorem is simply Green's theorem. That is, the assertion that

$$\int_{\Omega} \nabla \cdot \mathbf{F} = \int_{\partial\Omega} \mathbf{k} \times \mathbf{F}$$

follows upon applying Green's theorem to the one-form $\omega = -F_2\,dx + F_1\,dy$ corresponding to the vector field $\mathbf{k}\times\mathbf{F} = -F_2\mathbf{i} + F_1\mathbf{j}$. Because the dot and cross operations can be interchanged in the scalar triple product, integrating $\mathbf{k} \times \mathbf{F}$ along a curve, that is, taking the product $\mathbf{k} \times \mathbf{F} \cdot \mathbf{r}$, where $\mathbf{r}$ is the tangent to the curve, and then integrating, is the same as integrating $\mathbf{F} \cdot \mathbf{k} \times \mathbf{r}$, which is the normal component of $\mathbf{F}$. All the same identities now follow.

Exercise 10.32 Fix δ, $0 < \delta < 1$. Let D be the set of all $(\theta, t) \in R^2$ such that $0 \le \theta \le \pi$, $-\delta \le t \le \delta$. Let Φ be the 2-surface in R^3 with parameter domain D given by

$$\begin{aligned} x &= (1 - t\sin\theta)\cos 2\theta \\ y &= (1 - t\sin\theta)\sin 2\theta \\ z &= t\cos\theta \end{aligned}$$

where $(x, y, z) = \Phi(\theta, t)$. note that $\Phi(\pi, t) = \Phi(0, -t)$ and that Φ is one-to-one on the rest of D.

The range $M = \Phi(D)$ is known as a *Möbius band.* It is the simplest example of a nonorientable surface.

Prove the various assertions made in the following description: Put $\mathbf{p}_1 = (0, -\delta)$, $\mathbf{p}_2 = (\pi, -\delta)$, $\mathbf{p}_3 = (\pi, \delta)$, $\mathbf{p}_4 = (0, \delta)$, $\mathbf{p}_5 - \mathbf{p}_1$. Put $\gamma_i = [\mathbf{p}_i, \mathbf{p}_{i-1}]$, $i = 1, 2 \ldots, 4$, and put $\Gamma_i = \Phi \circ \gamma_i$. Then

$$\partial\Phi = \Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4.$$

Put $\mathbf{z} = (1, 0, -\delta)$, $\mathbf{b} = (1, 0, \delta)$. Then

$$\Phi(\mathbf{p}_1) = \Phi(\mathbf{p}_3) = \mathbf{a}, \quad \Phi(\mathbf{p}_2) = \Phi(\mathbf{p}_4) = \mathbf{b},$$

and $\partial\Phi$ can be described as follows.

Γ_1 spirals up from $\mathbf{a}$ to $\mathbf{b}$; its projection into the (x, y)-plane has winding number $+1$ around the origin. (See Exercise 23, Chap. 8).

$\Gamma_2 = [\mathbf{b}, \mathbf{a}]$.

Γ_3 spirals up from $\mathbf{a}$ to $\mathbf{b}$; its projection into the (x, y)-plane has winding number -1 around the origin.

$\Gamma_4 = [\mathbf{b}, \mathbf{a}]$.

Thus $\partial\Phi = \Gamma_1 + \Gamma_3 + 2\Gamma_2$.

If we go from $\mathbf{a}$ to $\mathbf{b}$ along Γ_1 and continue along the "edge" of M until we return to $\mathbf{a}$, the curve traced out is

$$\Gamma = \Gamma_1 - \Gamma_3,$$

which may also be represented on the parameter interval $[0, 2\pi]$ by the equations

$$\begin{aligned} x &= (1+\delta\sin\theta)\cos 2\theta, \\ y &= (1+\delta\sin\theta)\sin 2\theta, \\ z &= -\delta\cos\theta. \end{aligned}$$

It should be emphasized that $\Gamma \neq \partial\Phi$: Let η be the 1-form discussed in Exercises 21 and 22. Since $d\eta = 0$, Stokes' theorem shows that

$$\int_{\partial\Phi} \eta = 0,$$

But although Γ is the "geometric" boundary of M, we have

$$\int_{\Gamma} \eta = 4\pi.$$

In order to avoid this possible source of confusion, Stokes' formula (Theorem 10.50) is frequently stated only for orientable surfaces Φ.

Solution. The claim about the boundary $\partial\Phi$ follows immediately from the definition of a boundary. The domain D is a cell whose boundary is $\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$, so that by definition $\partial\Phi = \Phi(\gamma_1) + \Phi(\gamma_2) + \Phi(\gamma_3) + \Phi(\gamma_4)$.

The claims that $\Phi(\mathbf{p}_1) = \Phi(\mathbf{p}_3) = \mathbf{a}$ and $\Phi(\mathbf{p}_2) = \Phi(\mathbf{p}_4) = \mathbf{b}$ are routine computations.

The description of Γ_1 follows from the fact that γ_1 can be described as the set $(\theta, -\delta)$, $0 \le \theta \le \pi$, so that the projection of Γ_1 in the (x, y)-plane is the set of all points $\big(x(\theta), y(\theta)\big)$, where

$$\begin{aligned} x(\theta) &= (1+\delta\sin\theta)\cos 2\theta \\ y(\theta) &= (1+\delta\sin\theta)\sin 2\theta. \end{aligned}$$

Regarding the pair $\big(x(\theta), y(\theta)\big)$ as the complex number $z(\theta) = x(\theta) + iy(\theta) = (1+\delta\sin\theta)(\cos 2\theta + i\sin 2\theta)$, and using the definition of the winding number, we find this winding number to be

$$n = \frac{1}{2\pi i}\int_0^{\pi} \frac{z'(\theta)}{z(\theta)}\,d\theta.$$

Now, $z'(\theta) = 2(1+\delta\sin\theta)(-\sin 2\theta + i\cos 2\theta) + \delta\cos\theta(\cos 2\theta + i\sin 2\theta)$, so that we get

$$n = \frac{1}{\pi i}\left(\int_0^{\pi} \frac{-\sin 2\theta + i\cos 2\theta}{\cos 2\theta + i\sin 2\theta}\,d\theta + \delta\int_0^{\pi} \frac{\cos\theta}{1+\delta\sin\theta}\,d\theta\right).$$

But $-\sin 2\theta + i\cos 2\theta = i(\cos 2\theta + i\sin 2\theta)$, so that the first integral is just πi, and that term contributes $+1$ to the winding number. The second integral is just $\ln(1+\delta\sin\theta)$, and since this function has the value 0 at both $\theta = 0$ and $\theta = \pi$, it contributes nothing.

As for Γ_2, since $\theta = \pi$, it is given by $\big(x(t), y(t), z(t)\big)$, $-\delta \le t \le \delta$, where $x(t) = 1$, $y(t) = 0$, $z(t) = -t$. It therefore describes the line segment from **b** to **a** as t goes from $-\delta$ to δ.

The descriptions of Γ_3 and Γ_4 are justified exactly as was just done for Γ_1 and Γ_2.

As both Γ_1 and Γ_3 spiral upward from **a** to **b**, it is manifest that $\Gamma_1 - \Gamma_3$ represents a spiral that goes from **a** to **b** and back again. It is also easy to see that this spiral does not intersect itself, as the ranges of Γ_1 and Γ_3 meet only in **a** and **b**. For suppose θ and φ are such that $\Gamma_1(\theta) = \Gamma_3(\varphi)$. This means in particular that $-\delta\cos\theta = \delta\cos\varphi$, and so $\theta = \pi - \varphi$. It then follows that $(1 + \delta\sin\theta)\cos 2\theta = (1 - \delta\sin\theta)\cos 2\theta$, so that either $\cos 2\theta = 0$ or $\sin\theta = 0$. Since we also have $(1 + \delta\sin\theta)\sin 2\theta = -(1 - \delta\sin\theta)\sin 2\theta$, the possibility that $\cos 2\theta = 0$ is ruled out, and so $\sin\theta = 0$, i.e., $\theta = 0$ or $\theta = \pi$, meaning the point in common is either **a** or **b**, as asserted.

As for the description of $\Gamma_1 - \Gamma_3$, it is clear that the mapping $T(\theta)$ given by the equations

$$\begin{aligned} x &= (1 + \delta\sin\theta)\cos 2\theta \\ y &= (1 + \delta\sin\theta)\sin 2\theta \\ z &= -\delta\cos\theta \end{aligned}$$

has the property that $T(\theta + \pi)$ is given by the equations

$$\begin{aligned} x &= (1 - \delta\sin\theta)\cos 2\theta \\ y &= (1 - \delta\sin\theta)\sin 2\theta \\ z &= \delta\cos\theta. \end{aligned}$$

Hence it equals describes $\Gamma_1(-\delta, \theta)$ on the interval $[0, \pi]$ and $-\Gamma_3(\delta, \theta)$ (since Γ_3 is given by the latter formulas, but is traversed with θ decreasing from π to 0).

Since $x^2 + y^2 = (1 - t\sin\theta)^2 \ge (1 - \delta)^2 > 0$ on all of M, it follows that η is defined on M. On Γ_1 and Γ_3 we have $\eta = 2\,d\theta$, so that

$$\int_\Gamma \eta = 4\pi.$$

Chapter 11

The Lebesgue Theory

Exercise 11.1 If $f \geq 0$ and $\int_E f\,d\mu = 0$, prove that $f(x) = 0$ almost everywhere on E. *Hint:* Let E_n be the subset of E on which $f(x) > 1/n$. Write $A = \cup E_n$. Then $\mu(A) = 0$ if and only if $\mu(E_n) = 0$ for every n.

Solution. The assertion in the hint is immediate. If $\mu(A) = 0$, then $\mu(E_n) = 0$ also, since $E_n \subseteq A$. Conversely, letting $F_n = E_n \setminus \bigcup_{k=1}^{n-1} E_k$, we have $F_n \subset E_n$, $F_m \cap F_n = \emptyset$ if $m \neq n$, and $\cup F_n = \cup E_n = A$. Hence if $\mu(E_n) = 0$, then $\mu(F_n) = 0$ also, and therefore $\mu(A) = 0$ by the countable additivity of μ.

Given the hint, the solution is immediate, since A is the subset of E on which $f(x) > 0$. If $\mu(F_n) > 0$ for any n, then $\int_E f\,d\mu \geq \int_{F_n} f\,d\mu \geq \mu(F_n)/n > 0$.

Exercise 11.2 If $\int_A f\,d\mu = 0$ for every measurable subset A of a measurable set E, then $f(x) = 0$ almost everywhere on E.

Solution. The hypothesis applies in particular if A is the set on which $f(x) > 0$. Since $\chi_A f \geq 0$, the preceding exercise shows that $\mu(A) = 0$. Likewise, taking B as the set on which $-f(x) > 0$, we find that $\mu(B) = 0$. Hence $f(x) = 0$ for almost every x.

Exercise 11.3 If $\{f_n\}$ is a sequence of measurable functions, prove that the set of points x at which $\{f_n(x)\}$ converges is measurable.

Solution. This set can be written as

$$\bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} \bigcap_{l=m}^{\infty} \{x : |f_k(x) - f_l(x)| < \frac{1}{n}\}.$$

For this set is the set of x such that for every n there exists m such that $|f_k(x) - f_l(x)| < 1/n$ for all $k \geq m$, $l \geq m$. That is precisely the Cauchy criterion for convergence.

Exercise 11.4 If $f \in \mathcal{L}(\mu)$ on E and g is bounded and measurable on E, then $fg \in \mathcal{L}(\mu)$ on E.

Solution. This follows immediately from the dominated convergence theorem and the fact that $|g(x)| \leq M$ for some constant M. (Take $f_n(x) = g_n(x)f(x)$ for all n, where $g_n(x)$ is a sequence of simple functions converging to $g(x)$ almost everywhere. We can assume $|g_n(x)| \leq M$ and let the dominating function be $M|f(x)|$.)

Exercise 11.5 Put

$$g(x) = \begin{cases} 0 & (0 \leq x \leq \frac{1}{2}), \\ 1 & (\frac{1}{2} < x \leq 1), \end{cases}$$
$$f_{2k}(x) = g(x) \quad (0 \leq x \leq 1),$$
$$f_{2k+1}(x) = g(1-x) \quad (0 \leq x \leq 1).$$

Show that

$$\liminf_{n\to\infty} f_n(x) = 0 \quad (0 \leq x \leq 1),$$

but

$$\int_0^1 f_n(x)\,dx = \frac{1}{2}.$$

[Compare with (77).]

Solution. Since for each $x \in [0, \frac{1}{2}]$ we have $f_{2k}(x) = 0$ for all k, it follows that the inferior limit at such an x is zero. The same is true for $x \in [\frac{1}{2}, 1]$, since $f_{2k+1}(x) = 0$ for all these x. The value of the integral is immediate, since each $f_n(x)$ is a step function.

The point of this exercise is that strict inequality can easily occur in Fatou's Lemma.

Exercise 11.6 Let

$$f_n(x) = \begin{cases} \dfrac{1}{n} & (|x| \leq n), \\ 0 & (|x| > n). \end{cases}$$

Then $f_n(x) \to 0$ uniformly on R^1, but

$$\int_{-\infty}^{\infty} f_n\,dx = 2 \quad (n = 1, 2, 3, \ldots).$$

(We write $\int_{-\infty}^{\infty}$ in place of $\int_{R^1}$.) Thus uniform convergence does not imply dominated convergence in the sense of Theorem 11.32. However, on sets of finite measure, uniformly convergent sequences of bounded functions do satisfy Theorem 11.32.

Solution. The uniform convergence to zero is obvious, since $0 \le f_n(x) \le 1/n$ for all x and all n.

Again, since $f_n(x)$ is a step function, the value of the integral is immediate.

Exercise 11.7 Find a necessary and sufficient condition that $f \in \mathcal{R}(\alpha)$ on $[a, b]$. *Hint:* Consider Example 11.6(*b*) and Theorem 11.33.

A bounded function f belongs to $\mathcal{R}(\alpha)$ on $[a, b]$ if and only if the following two conditions hold:

(*i*) f is right-continuous wherever α is not right-continuous and left-continuous wherever α is not left-continuous (that is, one of f and α is right-continuous at each point and one is left-continuous);

(*ii*) the set of points where α is continuous and f is not continuous is a set of zero α-variation. That is, this set has μ_α-measure zero, where μ_α is the regular Borel measure generated by the function α, as in Example 11.6(*b*).

To prove this fact, all we have to do is copy the proof of Theorem 11.33, *mutatis mutandis*, specifically, replacing dx by $d\alpha$ and Δx by $\Delta\alpha$ at every stage. It will follow as a corollary of the proof that if $f \in \mathcal{R}(\alpha)$, then $f \in \mathcal{L}(\mu_\alpha)$ and

$$\int_{[a,b]} f\, d\mu_\alpha = \mathcal{R} \int_a^b f(x)\, d\alpha(x).$$

In modifying the proof we need to clear out just one case in order to make the changes run smoothly. To that end, we note that if f and α both have a one-sided discontinuity from the same side and at the same point, it is impossible for f to belong to $\mathcal{R}(\alpha)$. Indeed, suppose p is a common right-sided discontinuity of both f and α. For any partition P we have $x_k \le p < x_{k+1}$ for some index k, and then

$$U(P, f, \alpha) - L(P, f, \alpha) \ge o \cdot (\alpha(p+) - \alpha(p)) > 0,$$

where o is the limit of $\sup\limits_{p \le x < p+h} f(x) - \inf\limits_{p \le x < p+h} f(x)$ as $h \downarrow 0$. (The function $f(x)$ is right-continuous at p if and only if $o = 0$.)

Note that if f and α are discontinuous from opposite sides at a point, it is quite possible that $f \in \mathcal{R}(\alpha)$. For example, let $f(x) = \chi_{[0,1/2]}(x)$ and $\alpha(x) = \chi_{[1/2,\infty)}(x)$. Then for any partition P of $[0, 1]$ containing $1/2$, we have $x_k = \frac{1}{2}$ for some k, and

$$U(P, f, \alpha) - L(P, f, \alpha) = 1 - 1 = 0.$$

(It is for this reason that I define $\mathcal{R}(\alpha)$ differently in my courses. I require that for each $\varepsilon > 0$ there must exist $\delta > 0$ such that $U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$ for *all* partitions P such that $\max\limits_{1 \le k \le n} (x_k - x_{k-1}) < \delta$. When this is done, conditions (*i*) and (*ii*) are no longer sufficient for f to belong to $\mathcal{R}(\alpha)$, and the theory is somewhat simpler. Except for special considerations at discontinuities of α, however, the results are the same in both theories.)

We now suppose that condition (i) holds and prove the necessity of condition (ii). To avoid having to single out the endpoints in what follows, we simply extend α outside the interval $[a, b]$ by specifying $\alpha(x) = \alpha(b)$ for $x > b$ and $\alpha(x) = \alpha(a)$ for $x < a$.

Suppose that f is in $\mathcal{R}(\alpha)$. Let $\{P_k\}$ be a sequence of partitions such that P_{k+1} is a refinement of P_k, the distance between adjacent points of P_k is less than $\frac{1}{k}$, P_k contains all points x at which $\alpha(x+) - \alpha(x-) > \frac{1}{k}$, and

$$U(P_k, f, \alpha) \to \mathcal{R}\overline{\int} f\, d\alpha, \quad L(P_k, f, \alpha) \to \mathcal{R}\underline{\int} f\, d\alpha.$$

We note that every discontinuity of α belongs to some partition P_k. Assume P_k consists of the points $a = x_{k,0} < x_{k,1} < \cdots < x_{k,n_k} = b$.

As in the proof of Theorem 11.33, we define $U_k(x) = M_i$ and $L_k(x) = m_i$ for $x_{k,i-1} < x \le x_{k,i}$, $1 \le i \le n_k$. For definiteness we define $U_k(a) = M_1$ and $L_k(a) = m_1$. Then by definition of the upper and lower sume, the definition of $\mu_\alpha((a, b])$, and the definition of the integral of a simple function,

$$\begin{aligned} \int_{[a,b]} U_k\, d\mu_\alpha &= U(P_k, f, \alpha) + M_1\big(\alpha(a+) - \alpha(a)\big), \\ \int_{[a,b]} L_k\, d\mu_\alpha &= L(P_k, f, \alpha) + m_1\big(\alpha(a+) - \alpha(a)\big). \end{aligned}$$

By condition (i), either $M_1 - m_1 \to 0$ as $k \to \infty$ or $\alpha(a+) = \alpha(a)$. It then follows that the monotonic sequences L_k and U_k have limits L and U that are measurable, and either

$$\int_{[a,b]} L\, d\mu_\alpha = \mathcal{R}\underline{\int} f\, d\alpha, \quad \int_{[a,b]} U\, d\mu_\alpha = \mathcal{R}\overline{\int} f\, d\alpha,$$

(when $\alpha(a+) = \alpha(a)$) or

$$\begin{aligned} \int_{[a,b]} L\, d\mu_\alpha &= \mathcal{R}\underline{\int} f\, d\alpha + f(a)\big(\alpha(a+) - \alpha(a)\big), \\ \int_{[a,b]} U\, d\mu_\alpha &= \mathcal{R}\overline{\int} f\, d\alpha + f(a)\big(\alpha(a+) - \alpha(a)\big), \end{aligned}$$

(when $\alpha(a+) > \alpha(a)$).

If these two integrals are the same, it follows that $U(x) = L(x)$ almost everywhere with respect to the measure μ_α. If x is not a point of any partition P_k and $U(x) = L(x)$, then f is continuous at x. As for points of the partition, they are either points of discontinuity of α or points x such that $\mu_\alpha(\{x\}) = 0$. Since there are only countably many points in all the partitions, the partition points x for which $\mu_\alpha(\{x\}) = 0$ form a set of measure zero. Thus the set of discontinuities of $f(x)$ can be written as the union $A \cup B$, where A consists of points where f is continuous from only one side and α is discontinuous from that side (these points are all among the points of partition), and B consists of the points of discontinuity of f where α is continuous. We have just shown that B is of zero α-variation, as claimed.

Conversely, if f satisfies these two conditions, we note that $U(x) = L(x)$ at all points where $f(x)$ is continuous. Hence if the discontinuities of $f(x)$ other than one-sided discontinuities at points where α is continuous from the side on which f is discontinuous form a set of zero α-variation, then $U(P_k, f, \alpha) - L(P_k, f, \alpha) \to 0$.

Exercise 11.8 If $f \in \mathcal{R}$ on $[a, b]$ and if $F(x) = \int_a^x f(t)\,dt$, prove that $F'(x) = f(x)$ almost everywhere on $[a, b]$.

Solution. We know by Theorem 6.20 that $F'(x) = f(x)$ at every point where $f(x)$ is continuous. Theorem 11.33 shows that if f is Riemann-integrable on $[a, b]$, then it is continuous almost everywhere. Hence the result follows.

Exercise 11.9 Prove that the function F given by (96) is continuous on $[a, b]$.

Solution. The function F is the one in the preceding exercise. Its continuity follows from the dominated convergence theorem, taking $|f|$ as the "dominating" function and letting $f_n = \chi_{[a,x_n]} f$ where $\{x_n\}$ is any sequence of numbers converging to x, so that f_n converges pointwise to $\chi_{[a,x]} f$ except possibly at the point x, which is a set of measure 0. The dominated convergence theorem then guarantees that $F(x_n) \to F(x)$. Since the sequence $\{x_n\}$ is arbitrary, it follows that F is continuous at x, as in Theorem 4.2.

Exercise 11.10 If $\mu(X) < +\infty$ and $f \in \mathcal{L}^2(\mu)$ on X, prove that $f \in \mathcal{L}(\mu)$ on X. If

$$\mu(X) = +\infty,$$

this is false. For instance, if

$$f(x) = \frac{1}{1 + |x|},$$

then $f \in \mathcal{L}^2$ on R^1, but $f \notin \mathcal{L}$ on R^1.

Solution. This follows from Theorems 11.27 and 11.29, if we let $A = \{x : |f(x)| \le 1\}$ and $B = \{x : |f(x)| > 1\}$. We can then write

$$|f| \le \chi_A + \chi_B \cdot |f|^2.$$

and χ_A is integrable by Theorem 11.23(a).

As for the counterexample, we have

$$|f|^2 \leq \chi_{[-1,1]} + \chi_{[1,\infty)} \cdot \frac{1}{x^2},$$

which implies that $f \in \mathcal{L}^2$, and

$$f(x) \geq \chi_{[0,n]}(x)\frac{1}{1+x},$$

so that

$$\int f\,dx \geq \int_0^n \frac{1}{1+x}\,dx = \ln(1+n) \to \infty.$$

Hence $f \notin \mathcal{L}$.

Exercise 11.11 If $f, g \in \mathcal{L}(\mu)$ on X, define the distance between f and g by

$$\int_X |f-g|\,d\mu.$$

Prove that $\mathcal{L}(\mu)$ is a complete metric space.

Solution. We have to regard functions equal almost everywhere as the same function. Given that, it does follow that if $d(f,g) = 0$, then $f = g$. The fact that $d(f,g) = d(g,f)$ is immediate from the definition and the triangle inequality follows from simply integrating the triangle inequality for the values of the functions. Hence $\mathcal{L}$ is a metric space.

To prove that it is complete, we merely repeat the reasoning of Theorem 11.42, replacing $\mathcal{L}^2$ by $\mathcal{L}$ and taking the function $g(x)$ to be identically equal to 1. When this is done, every step in the proof of Theorem 11.42 follows for $\mathcal{L}$.

Exercise 11.12 Suppose

(a) $|f(x,y) \leq 1$ if $0 \leq x \leq 1$, $0 \leq y \leq 1$,

(b) for fixed x, $f(x,y)$ is a continuous function of y,

(c) for fixed y, $f(x,y)$ is a continuous function of x.

Put

$$g(x) = \int_0^1 f(x,y)\,dy \quad (0 \leq x \leq 1).$$

Is g continuous?

Solution. Yes, $g(x)$ is continuous. Let $x_n \to x$. Then by (c), $f(x_n,y) \to f(x,y)$ for each $y \in [0,1]$, in particular for almost every y. Since $|f(x_n,y)| \leq 1$ for all x_n and y by assumption (a), and the set $[0,1]$ has finite measure, it follows from the dominated convergence theorem that $g(x_n) \to g(x)$.

Note that property (b) was used only to guarantee that $g(x)$ is actually defined. Thus the word *continuous* could be replaced by *integrable* in this condition.

Exercise 11.13 Consider the functions

$$f_n(x) = \sin nx \quad (n = 1, 2, 3, \ldots, -\pi \le x \le \pi)$$

as points of $\mathcal{L}^2$. Prove that the set of these points is closed and bounded, but not compact.

Solution. We compute by brute force that

$$\|f_m - f_n\|^2 = \begin{cases} 0, & \text{if } m = n, \\ 2\pi, & \text{if } m \ne n. \end{cases}$$

Further, it is easy to see that $\|f_n\|^2 = \pi$. Hence the set $\{f_n\}$ is bounded and has no limit points. (The $\sqrt{\frac{\pi}{2}}$-neighborhood of any point contains at most one point of this set.) Having no limit points, it contains all of its limit points and is therefore closed. Being infinite, if it were compact, it *would* have a limit point. Therefore it is not compact.

Exercise 11.14 Prove that a complex function f is measurable if and only if $f^{-1}(V)$ is measurable for every open set V in the plane.

Solution. By definition $f = u + iv$, where u and v are real-valued, is measurable if and only if u and v are.

Suppose f is measurable (that is, u and v are measurable). Let V be any open set in the plane and $(x, y) \in V$. Then there exists $\delta > 0$ such that the square $S(x, y) = (x - \delta, x + \delta) \times (y - \delta, y + \delta)$ is contained in V. The union of these open squares is all of V, and there is a countable set of points $(x_n, y_n) \in V$ such that $\bigcup\limits_{n=1}^{\infty} S(x_n, y_n) = V$. (This is proved by appealing to Exercise 23 of Chapter 2.) But then

$$f^{-1}(V) = \bigcup_{n=1}^{\infty} f^{-1}(S(x_n, y_n)) = \bigcup_{n=1}^{\infty} u^{-1}(x_n - \delta, x_n + \delta) \cap v^{-1}(y_n - \delta, y_n + \delta).$$

It follows that $f^{-1}(V)$ is measurable.

Conversely if $f^{-1}(V)$ is measurable for every open set in the plane , then in particular this set is measurable if $V = (a, b) \times R^1$ (where $f^{-1}(V) = u^{-1}((a, b))$) or $V = R^1 \times (a, b)$ (where $f^{-1}(V) = v^{-1}((a, b))$), and hence both u and v are measurable. By definition, that means that f is measurable.

Exercise 11.15 Let $\mathcal{R}$ be the ring of all elementary subsets of $(0, 1]$. If $0 < a \le b \le 1$, define

$$\phi([a, b]) = \phi([a, b)) = \phi((a, b]) = \phi((a, b)) = b - a,$$

but define

$$\phi((0, b)) = \phi((0, b]) = 1 + b$$

if $0 < b \leq 1$. Show that this gives an additive set function ϕ on $\mathcal{R}$, which is not regular and which cannot be extended to a countably additive set function on a σ-ring.

Solution. In brief, since an elementary set A is a finite disjoint union of intervals, $\phi(A)$ is the sum of the lengths of those intervals if 0 is not the endpoint of any interval in A and 1 larger than the sum of the lengths of the intervals if 0 is one of the endpoints. In particular $\phi(A) < 1$ if A is a closed set, since 0 cannot be the endpoint of any closed set that is a finite union of intervals in $(0, 1]$.

(This alternate definition is independent of the particular way in which the set A is represented as a finite disjoint union of intervals, since if $A = \bigcup_{i=1}^{m} I_i = \bigcup_{j=1}^{n} J_j$, where each of the collections $\{I_i\}$ and $\{J_j\}$ is a set of pairwise disjoint intervals, one can easily verify that

$$|I_i| = \sum_{j=1}^{n} |I_i \cap J_j|, \quad |J_j| = \sum_{i=1}^{m} |I_i \cap J_j|,$$

so that $\sum_{i=1}^{m} |I_i| = \sum_{j=1}^{n} |J_j| = \sum_{i,j} |I_i \cap J_j|$. HEre $|I|$ is the length of the interval I.)

If two elementary sets A and B are disjoint, at most one of them can have the point 0 as the endpoint of one of its intervals. Then $\phi(A \cup B)$ is the sum of the lengths of the intervals in $A \cup B$ if neither set contains an interval having 0 as the endpoint, and 1 larger than this sum if one of them does contain an interval with 0 as endpoint. In either case $\phi(A \cup B) = \phi(A) + \phi(B)$ when $A \cap B = \emptyset$. Thus the function ϕ is additive.

The function ϕ is not regular, however, since there is no closed subset of $(0, c]$ that can approximate $(0, c]$ if $c < 1$. For $\phi((0, c]) = 1 + c$, but $\phi(A) \leq 1$ if A is closed.

The function ϕ also cannot be extended to a countably additive set function on a σ-ring, since

$$(0, \frac{1}{2}] = \bigcup_{n=1}^{\infty} (\frac{1}{2^{n+1}}, \frac{1}{2^n}],$$

and

$$\phi((0, \frac{1}{2}]) = \frac{3}{2}, \quad \sum_{n=1}^{\infty} \phi((\frac{1}{2^{n+1}}, \frac{1}{2^n}]) = \frac{1}{2}.$$

Exercise 11.16 Suppose $\{n_k\}$ is an increasing sequence of positive integers and E is the set of all $x \in (-\pi, \pi)$ at which $\{\sin n_k x\}$ converges. Prove that $m(E) = 0$. *Hint:* For every $A \subset E$,

$$\int_A \sin n_k x \, dx = 0,$$

and

$$2\int_A (\sin n_k x)^2\,dx = \int_A (1-\cos 2n_k x)\,dx \to m(A) \text{ as } k\to\infty.$$

Solution. The two statements in the hint follow from the Riemann–Lebesgue lemma (or from Bessel's inequality applied to the Fourier series of χ_A, if you wish). Let $f(x)$ be the limit of $\sin n_k x$ on the set E. Then, since termwise integration is justified by the dominated convergence theorem, we have

$$\int_A [(f(x))^2 - \frac{1}{2}]\,dx = 0,$$

for all A. Hence, by Exercise 2 above, $f(x) = \pm\frac{1}{\sqrt{2}}$ almost everywhere on E. If we let A be the set of points of E at which $f(x) = \frac{1}{\sqrt{2}}$, we find that $\int_A f(x)\,dx = 0$, and so by Exercise 1, $f(x) = 0$ almost everywhere on A. Since in fact $f(x) \neq 0$ on A, it follows that A has measure 0. Similarly the set where $f(x) = -\frac{1}{\sqrt{2}}$ has measure 0.

Exercise 11.17 Suppose $E \subset (-\pi,\pi)$, $m(E) > 0$, $\delta > 0$. Use the Bessel inequality to prove that there are at most finitely many integers n such that $\sin nx \geq \delta$ for all $x \in E$.

Solution. For any integer with this property we have

$$\int_E \sin nx\,dx \geq \delta\mu(E),$$

and the Bessel inequality implies that this inequality can hold for only a finite number of n. (The integral is the imaginary part of the Fourier coefficient of the $\mathcal{L}^2$-function χ_E.)

Exercise 11.18 Suppose $f \in \mathcal{L}^2(\mu)$, $g \in L^2(\mu)$. Prove that

$$\left|\int f\bar{g}\,d\mu\right|^2 = \int |f|^2\,d\mu \int |g|^2\,d\mu$$

if and only if there is a constant c such that $g(x) = cf(x)$ almost everywhere. (Compare Theorem 11.35.)

Solution. There is a slight mistake in the statement of the problem, since equality certainly holds if $f(x)$ is identically zero, whether $g(x)$ equals zero or not. We must either assume that $f(x)$ is not identically zero, or allow the possibility that $f(x) = cg(x)$.

Equality can hold if $g(x) = 0$ almost everywhere, and in that case $c = 0$ in the relation $g(x) = cf(x)$. Hence assume now that $\int |g|^2\,d\mu > 0$. The inequality

$$0 \le \int (|f| + \lambda|g|)^2\,d\mu,$$

which holds for real values of λ, is equivalent to the inequality

$$-2\lambda \int |fg|\,d\mu \le \int |f|^2\,d\mu + \lambda^2 \int |g|^2\,d\mu.$$

In this inequality take $\lambda = -\sqrt{\dfrac{\int |f|^2\,d\mu}{\int |g|^2\,d\mu}}$. The result is

$$2\frac{\int |fg|\,d\mu \sqrt{\int |f|^2\,d\mu}}{\sqrt{\int |g|^2\,d\mu}} \le 2\int |f|^2\,d\mu,$$

which is equivalent to

$$\left[\int |fg|\,d\mu\right]^2 \le \int |f|^2\,d\mu \int |g|^2\,d\mu.$$

Hence the equality in the problem can hold only if equality holds in this last equality, which, since it implies that

$$\int \bigl(|f| + \lambda|g|\bigr)^2\,d\mu = 0,$$

implies that $|f| = -\lambda|g|$ almost everywhere. In particular f vanishes almost everywhere that g vanishes. In addition, the equality in the hypothesis of the problem requires that

$$\left|\int f\overline{g}\,d\mu\right| = \int |fg|\,d\mu.$$

If both sides of this last equality are zero, then at almost every point either $f(x) = 0$ or $g(x) = 0$. Since $|f| = -\lambda|g|$, it then follows that in fact either both functions vanish identically, a case we have already discussed, or $\lambda = 0$, in which case only f vanishes identically. In either case we do have the kind of linear dependence specified in the amended statement of the problem.

Hence assume that neither side of this equality is zero. Let ω be the complex number

$$\omega = \frac{\overline{\int f\overline{g}\,d\mu}}{|\int f\overline{g}\,d\mu|},$$

so that $|\omega| = 1$. We note that

$$\int \omega f\overline{g}\,d\mu = \omega \int f\overline{g}\,d\mu = \left|\int f\overline{g}\,c\mu\right| \le \int |f\overline{g}|\,d\mu = \int |\omega f\overline{g}|\,d\mu.$$

This means that the real parts of the two integrals on the extremes here are equal, and the imaginary parts of both are zero. Taking just the real parts, since $\operatorname{Re}(\omega f \overline{g}) \leq |\omega f \overline{g}|$, this implies that the real part of $\omega f \overline{g}$ is equal to $|fg| = -\lambda g \overline{g}$ almost everywhere, and therefore that the imaginary part is zero almost everywhere. But then, almost everywhere where g does not vanish, we can cancel $\overline{g}$ from the equality, getting $f = -\lambda \overline{\omega} g$ wherever g does not vanish. Since this equality also holds almost everywhere where g does vanish, we are done.