PART D

Complex Analysis

Chap. 13 Complex Numbers and Functions. Complex Differentiation

Complex numbers appeared in the textbook before in different topics. Solving linear homogeneous ODEs led to characteristic equations, (3), p. 54 in Sec. 2.2, with complex numbers in Example 5, p. 57, and Case III of the table on p. 58. Solving algebraic eigenvalue problems in Chap. 8 led to characteristic equations of matrices whose roots, the eigenvalues, could also be complex as shown in Example 4, p. 328. Whereas, in these type of problems, complex numbers appear almost naturally as complex roots of polynomials (the simplest being $x^2 + 1 = 0$), *it is much less immediate to consider* ***complex analysis****—the systematic study of complex numbers, complex functions, and "complex" calculus.* Indeed, complex analysis will be the direction of study in Part D. The area has important engineering applications in electrostatics, heat flow, and fluid flow. Further motivation for the study of complex analysis is given on p. 607 of the textbook.

We start with the basics in Chap. 13 by reviewing complex numbers $z = x + yi$ in Sec. 13.1 and introducing complex integration in Sec.13.3. Those functions that are differentiable in the complex, on some domain, are called **analytic** and will form the basis of complex analysis. Not all functions are analytic. This leads to the most important topic of this chapter, the **Cauchy–Riemann equations** (1), p. 625 in Sec. 13.4, which allow us to test whether a function is analytic. They are very short but you have to remember them! The rest of the chapter (Secs. 13.5–13.7) is devoted to elementary complex functions (exponential, trigonometric, hyperbolic, and logarithmic functions).

Your knowledge and understanding of real calculus will be useful. Concepts that you learned in real calculus carry over to complex calculus; however, be aware that *there are* ***distinct differences between real calculus and complex analysis*** that we clearly mark. For example, whereas the real equation $e^x = 1$ has only one solution, its complex counterpart $e^z = 1$ has infinitely many solutions.

Sec. 13.1 Complex Numbers and Their Geometric Representation

Much of the material may be familiar to you, but we start from scratch to assure everyone starts at the same level. This section begins with the four basic algebraic operations of complex numbers (addition, subtraction, multiplication, and division). Of these, the one that perhaps differs most from real numbers is **division** (or **forming a quotient**). *Thus make sure that you remember how to calculate the quotient of two complex numbers as given in equation* (7), **Example 2**, p. 610, and **Prob. 3**. In (7) we take the number z_2 from the denominator and form its complex conjugate $\bar{z}_2$ and a new quotient $\bar{z}_2/\bar{z}_2$. We multiply the given quotient by this new quotient $\bar{z}_2/\bar{z}_2$ (which is equal to 1 and thus allowed):

$$z = \frac{z_1}{z_2} = \frac{z_1}{z_2} \cdot 1 = \frac{z_1}{z_2} \cdot \frac{\bar{z}_2}{\bar{z}_2},$$

which we multiply out, recalling that $i^2 = -1$ [see (5), p. 609]. The final result is a complex number in a form that allows us to separate its real ($\operatorname{Re} z$) and imaginary ($\operatorname{Im} z$) parts. Also remember that $1/i = -i$ (see **Prob. 1**), as it occurs frequently. We continue by defining the **complex plane** and use it to graph complex numbers (note Fig. 318, p. 611, and Fig. 322, p. 612). We use equation (8), p. 612, to go from complex to real.

Problem Set. 13.1. Page 612

1. Powers of *i*. We compute the various powers of i by the rules of addition, subtraction, multiplication, and division given on pp. 609–610 of the textbook. We have formally that

$$\text{(I1)} \qquad \begin{aligned} i^2 &= i\,i \\ &= (0,1)(0,1) && \text{[by (1), p. 609]} \\ &= (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) && \text{[by (3), p. 609]} \\ &= (0-1, 0+0) && \text{(arithmetic)} \\ &= (-1, 0) \\ &= -1 && \text{[by (1)],} \end{aligned}$$

where in (3), that is, *multiplication of complex numbers*, we used $x_1 = 0$, $x_2 = 0$, $y_1 = 1$, $y_2 = 1$.

$$\text{(I2)} \qquad i^3 = i^2 i = (-1) \cdot i = -i.$$

Here we used (I1) in the second equality. To get (I3), we apply (I2) twice:

$$\text{(I3)} \qquad i^4 = i^2 i^2 = (-1) \cdot (-1) = 1.$$

$$\text{(I4)} \qquad i^5 = i^4 i = 1 \cdot i = i,$$

and the pattern repeats itself as summarized in the table below.
We use (7), p. 610, in the following calculation:

$$\text{(I5)} \qquad \frac{1}{i} = \frac{1\,\bar{i}}{i\,\bar{i}} = \frac{1\,(-i)}{i\,(-i)} = \frac{(1+0i)(0-i)}{(0+i)(0-i)} = \frac{1 \cdot 0 + 0 \cdot 1}{0^2 + 1^2} + i\,\frac{0 \cdot 0 - 1 \cdot 1}{0^2 + 1^2} = 0 - i = -i.$$

By (I5) and (I1) we get

$$\text{(I6)} \qquad \frac{1}{i^2} = \frac{1}{i} \cdot \frac{1}{i} = (-i)(-i) = (-1)i \cdot (-1)i = 1 \cdot i^2 = -1,$$

$$\frac{1}{i^3} = \left(\frac{1}{i}\right)^2 \left(\frac{1}{i}\right) = (-1)(-i) = i \qquad \text{[from (I6) and (I5)]},$$

$$\frac{1}{i^4} = \left(\frac{1}{i}\right)^2 \left(\frac{1}{i}\right)^2 = (-1)(-1) = 1,$$

and the pattern repeats itself. Memorize that $i^2 = -1$ and $1/i = -i$ as they will appear quite frequently.

	i^8	i^9	.	.	
	i^4	i^5	i^6	i^7	
Start ⟶	i^0	i	i^2	i^3	
	1	i	-1	$-i$	
	$1/i^4$	$1/i^3$	$1/i^2$	$1/i$	⟵ Start
	$1/i^8$	$1/i^7$	$1/i^6$	$1/i^5$	
	.	.	$1/i^{10}$	$1/i^9$	

Sec. 13.1. Prob. 1. Table of powers of i

3. Division of complex numbers

a. The calculations of (7), p. 610, in detail are

$$\begin{aligned}
z &= \frac{z_1}{z_2} = \frac{x_1 + iy_1}{x_2 + iy_2} && \text{(by definition of } z_1 \text{ and } z_2\text{)}\\
&= \frac{x_1 + iy_1}{x_2 + iy_2} \cdot \frac{x_2 - iy_2}{x_2 - iy_2} && \text{(N.B. corresponds to multiplication by 1)}\\
&= \frac{(x_1 + iy_1)(x_2 - iy_2)}{(x_2 + iy_2)(x_2 - iy_2)}\\
&= \frac{x_1x_2 - x_1iy_2 + iy_1x_2 - iy_1iy_2}{x_2x_2 - x_2iy_2 + iy_2x_2 - iy_2iy_2} && \text{(multiplying it out: (3) in notation (4), p. 609)}\\
&= \frac{x_1x_2 - ix_1y_2 + ix_2y_1 - i^2y_1y_2}{x_2^2 - ix_2y_2 + ix_2y_2 - i^2y_2^2} && \text{(grouping terms, using commutativity)}\\
&= \frac{x_1x_2 - ix_1y_2 + ix_2y_1 + y_1y_2}{x_2^2 + y_2^2} && \text{(using } i^2 = -1 \text{ and simplifying)}\\
&= \frac{x_1x_2 + +y_1y_2}{x_2^2 + y_2^2} + i\,\frac{x_2y_1 - x_1y_2}{x_2^2 + y_2^2} && \text{(breaking into real part and imaginary part).}
\end{aligned}$$

b. A practical example using (7) is

$$\begin{aligned}
\frac{26 - 18i}{6 - 2i} &= \frac{(26 - 18i)}{(6 - 2i)}\frac{(6 + 2i)}{(6 + 2i)} = \frac{26 \cdot 6 + 26 \cdot 2i - 18 \cdot 6i - 18 \cdot 2i^2}{6^2 + 2^2}\\
&= \frac{156 + 52i - 108i + 36}{36 + 4} = \frac{192 - 56i}{40} = 4.8 - 1.4i.
\end{aligned}$$

5. Pure imaginary number a. If $z = x + iy$ is pure imaginary, then $\bar{z} = -z$.
Proof. Let $z = x + iy$ be pure imaginary. Then $x = 0$, by definition on the bottom of p. 609. Hence

(A) $z = iy$ and (B) $\bar{z} = -iy$ (by definition. of complex conjugate, p. 612).

If we multipy both sides of (A) by -1, we get

$$-z = -iy,$$

which is equal to $\bar{z}$, hence

$$-z = \bar{z}.$$

b. If $\bar{z} = -z$ then $z = x + iy$ is pure imaginary.
Proof. Let $z = x + iy$ so that $\bar{z} = x - iy$. We are given that $\bar{z} = -z$, so

$$\bar{z} = x - iy = -z = -(x + iy) = -x - iy.$$

By the definition of equality (p. 609) we know that the real parts must be equal and that the imaginary parts must be equal. Thus

$$\begin{aligned} \operatorname{Re}\bar{z} &= \operatorname{Re}(-z), \\ x &= -x, \\ 2x &= 0, \\ x &= 0, \end{aligned}$$

and

$$\begin{aligned} \operatorname{Im}\bar{z} &= \operatorname{Im}(-z), \\ -y &= -\, y, \end{aligned}$$

which is true for any y. Thus

$$z = x + iy = iy.$$

But this means, by definition, that z is pure imaginary, as had to be shown.

11. Complex arithmetic

$$\begin{aligned} z_1 - z_2 &= (-2 + 11i) - (2 - i) \\ &= -2 + 11i - 2 + i = (-2 - 2) + (11 + 1)i = -4 + 12i \\ (z_1 - z_2)^2 &= (-4 + 12i)(-4 + 12i) = 16 - 48i - 48i - 144 = -128 - 96i \\ \frac{(z_1 - z_2)^2}{16} &= -\frac{128}{16} - \frac{96}{16}i = -\frac{8 \cdot 16}{16} - \frac{2^5 \cdot 3}{2^4} = -8 - 6i. \end{aligned}$$

Next consider

$$\left(\frac{z_1}{4} - \frac{z_2}{4}\right)^2.$$

We have

$$\frac{z_1}{4} = \frac{1}{4}(-2 + 11i) = -\frac{2}{4} + \frac{11}{4}i, \qquad \frac{z_2}{4} = \frac{2}{4} - \frac{1}{4}i.$$

Their difference is

$$\frac{z_1}{4} - \frac{z_2}{4} = -\frac{2}{4} - \frac{2}{4} + \left(\frac{11}{4} + \frac{1}{4}\right)i \ = -1 + 3i.$$

Hence

$$\left(\frac{z_1}{4} - \frac{z_2}{4}\right)^2 = (-1+3i)(-1+3i) = 1 - 3i - 3i + 9i^2 = 1 - 6i - 9 = -8 - 6i,$$

which is the same result as before.

19. Real part and imaginary part of $z/\bar{z}$. For $z = x + iy$, we have by (7), p. 610,

$$\frac{z}{\bar{z}} = \frac{z\,\bar{\bar{z}}}{\bar{z}\,\bar{\bar{z}}} = \frac{z\,z}{\bar{z}\,z}$$

since the conjugate of the conjugate of a complex number is the complex number itself (which you may want to prove!). Then

$$\frac{z}{\bar{z}} = \frac{z^2}{\bar{z}z} = \frac{(x+iy)^2}{x^2+y^2} = \frac{x^2 + 2ixy - y^2}{x^2+y^2} = \frac{x^2-y^2}{x^2+y^2} + i\,\frac{2xy}{x^2+y^2}.$$

Hence we get the result as shown on p. A34 of the textbook:

$$\operatorname{Re}\left(\frac{z}{\bar{z}}\right) = \frac{x^2-y^2}{x^2+y^2}; \qquad \operatorname{Im}\left(\frac{z}{\bar{z}}\right) = \frac{2xy}{x^2+y^2}.$$

Sec. 13.2 Polar Form of Complex Numbers. Powers and Roots

Polar coordinates, defined by (1) and (2) on p. 613, play a more important role in complex analysis than in calculus. Their study gives a deeper understanding of multiplication and division of complex numbers (pp. 615–616) and absolute values. More details are as follows.

The polar angle θ (taken counterclockwise, see Fig. 323, p. 614) of a complex number is determined only up to integer multiples of 2π. While often this is not essential, there are situations where it matters. For this purpose, we introduce the concept of the **principal value** Arg z in (5), p. 614, and illustrate it in **Example 1**, **Probs. 9** and **13**.

The triangle inequality defined in (6), p. 614, and illustrated in Example 2, p. 615, is very important since it will be used frequently in establishing bounds such as in Chap. 15.

Often it will be used in its generalized form (6*), p. 615, which can be understood by the following geometric reasoning. Draw several complex numbers as little arrows and let each tail coincide with the preceding head. This gives you a zigzaging line of n parts, and the left side of (6*) equals the distance from the tail of z_1 to the head of z_n. Can you "see" it? Now take your zigzag line and pull it taut; then you have the right side as the length of the zigzag line straightened out.

In almost all cases when we use (6*) in establishing bounds, it will not matter whether or not the right side of (6*) is much larger than the left. However, it will be essential that we have such an upper bound for the absolute value of the sum on the left, so that in a limit process, the latter cannot go to infinity.

The last topic is roots of complex numbers, illustrated in Figs. 327–329, p. 617, and **Prob. 21**. Look at these figures and see how, for different n, the roots of unity (16), p. 617, lie symmetrically on the unit circle.

Problem Set 13.2. Page 618

1. Polar form. Sketch $z = 1 + i$ to understand what is going on. Point z is the point $(1, 1)$ in the complex plane. From this we see that the distance of z from the origin is $|z| = \sqrt{2}$. This is the absolute value of z. Furthermore, z lies on the bisecting line of the first quadrant, so that its argument (the angle between the positive ray of the x-axis and the segment from 0 to z) is 45° or $\pi/4$.

Now we show how the results follow from (3) and (4), p. 613. In the notation of (3) and (4) we have $z = x + iy = 1 + i$. Hence the real part of z is $x = 1$ and the imaginary part of z is $y = 1$. From (3) we obtain

$$|\,z\,| = \sqrt{1^2 + 1^2} = \sqrt{2},$$

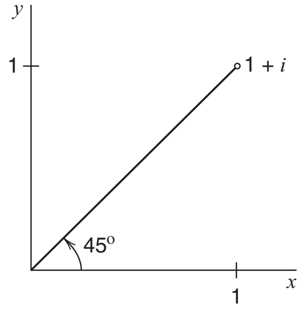
as before. From (4) we obtain

$$\tan\theta = \frac{y}{x} = 1, \qquad \theta = 45^\circ \text{ or } \frac{\pi}{4}.$$

Hence the polar form (2), p. 613, is

$$z = \sqrt{2}\left(\cos\frac{\pi}{4} + i\,\sin\frac{\pi}{4}\right).$$

Note that here we have explained the first part of **Example 1**, p. 614, in great detail.



Sec. 13.2 Prob. 1. Graph of $z = 1 + i$ in the complex plane

5. Polar form. We use (7), p. 610, in Sec. 13.1, to obtain

$$\text{(A)} \qquad \frac{\sqrt{2} + \frac{1}{3}i}{-\sqrt{8} - \frac{2}{3}i} = \frac{\sqrt{2} + \frac{1}{3}i}{-\sqrt{8} - \frac{2}{3}i} \cdot \frac{-\sqrt{8} + \frac{2}{3}i}{-\sqrt{8} + \frac{2}{3}i}.$$

The numerator of (A) simplifies to

$$\left(\sqrt{2} + \tfrac{1}{3}i\right)\left(-\sqrt{8} + \tfrac{2}{3}i\right) = -\sqrt{16} + \left(\tfrac{2}{3}\sqrt{2} - \tfrac{1}{3}\sqrt{8}\right)i - \tfrac{2}{9} = \tfrac{38}{9} + \left(\tfrac{2}{3}\sqrt{2} - \tfrac{1}{3}2\sqrt{2}\right)i = -\tfrac{38}{9}.$$

The denominator of (A) is

$$\left(-\sqrt{8}\right)^2 + \left(\tfrac{2}{3}\right)^2 = 8 + \tfrac{4}{9} = \tfrac{72}{9} + \tfrac{4}{9} = \tfrac{76}{9}.$$

Putting them together gives the simplification of (A), that is,

$$\frac{\sqrt{2} + \frac{1}{3}i}{-\sqrt{8} - \frac{2}{3}i} = \frac{-\frac{38}{9}}{\frac{76}{9}} = \left(-\frac{38}{9}\right)\left(\frac{9}{76}\right) = -\frac{38}{76} = -\frac{1}{2}.$$

Hence $z = -\frac{1}{2}$ corresponds to $\left(-\frac{1}{2}, 0\right)$ in the complex plane. Furthermore, by (3), p. 613,

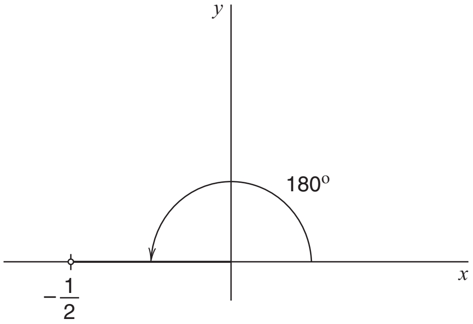
$$|z| = r = \sqrt{x^2 + y^2} = \sqrt{\left(-\tfrac{1}{2}\right)^2 + 0^2} = \tfrac{1}{2}$$

and by (4), p. 613,

$$\tan\theta = \frac{y}{x} = \frac{0}{-\frac{1}{2}} = 0; \qquad \theta = 180^\circ = \pi.$$

Hence by (2), p. 613, the desired polar form is

$$z = r(\cos\theta + i\sin\theta) = \tfrac{1}{2}\left(\cos\pi + i\sin\pi\right).$$



Sec. 13.2 Prob. 5. Graph of $z = -\frac{1}{2}$ in the complex plane

7. **Polar form.** For the given z we have

$$|z| = \sqrt{1^2 + \left(\tfrac{1}{2}\pi\right)^2} = \sqrt{1 + \tfrac{1}{4}\pi^2},$$

$$\tan\theta = \frac{y}{x} = \frac{\frac{1}{2}\pi}{1} = \frac{1}{2}\pi; \qquad \theta = \arctan\left(\frac{1}{2}\pi\right).$$

The desired polar form of z is

$$z = |z|\left(\cos\theta + i\sin\theta\right) = \sqrt{1 + \tfrac{1}{4}\pi^2}\left[\cos\left(\arctan\,\tfrac{1}{2}\pi\right) + i\sin\left(\arctan\,\tfrac{1}{2}\pi\right)\right].$$

9. **Principal argument.** The first and second quadrants correspond to $0 \leq \operatorname{Arg} z \leq \pi$. The third and fourth quadrants correspond to $-\pi < \operatorname{Arg} z \leq 0$. Note that Arg z is continuous on the positive real semiaxis and has a jump of 2π on the negative real semiaxis. This is a convenient convention. Points on the negative real semiaxis, e.g., -4.7, have the principal argument $\operatorname{Arg} z = \pi$.

To find the principal argument of $z = -1 + i$, we convert z to polar form:

$$|z| = \sqrt{(-1)^2 + 1^2} = \sqrt{2},$$

$$\tan\theta = \frac{y}{x} = \frac{1}{-1} = -1.$$

Hence

$$\theta = \tfrac{3}{4}\pi = 135^\circ.$$

Hence z, in polar form, is

$$z = \sqrt{2}\left(\cos\tfrac{3}{4}\pi + i\sin\tfrac{3}{4}\pi\right).$$

As explained near the end of p. 613, θ is called the argument of z and denoted by $\arg z$. Thus θ is

$$\theta = \arg z = \tfrac{3}{4}\pi \pm 2n\pi, \qquad n = 0, 1, 2, \cdots.$$

The reason is that sine and cosine are periodic with 2π, so 135° looks the same as $135° + 360°$, etc. To avoid this concern, we define the principal argument Arg z [see (5), p. 614]. We have

$$\operatorname{Arg} z = \tfrac{3}{4}\pi.$$

You should sketch the principal argument.

13. **Principal argument.** The complex number $1 + i$ in polar form is

$$1 + i = \sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right) \qquad \text{by } \textbf{Prob. 1.}$$

Then, using DeMoivre's formula (13), p. 616, with $r = \sqrt{2}$ and $n = 20$,

$$\begin{aligned}(1+i)^{20} &= \left(\sqrt{2}\right)^{20}\left[\cos\left(20\cdot\frac{\pi}{4}\right) + i\sin\left(20\cdot\frac{\pi}{4}\right)\right] \qquad \text{by } \textbf{Prob. 1}.\\ &= 2^{10}\left(\cos 5\pi + i\sin 5\pi\right)\\ &= 2^{10}\left(\cos\pi + i\sin\pi\right).\end{aligned}$$

Hence

$$\arg z = \pi \pm 2n\pi, \qquad n = 0, 1, 2, \cdots; \qquad \operatorname{Arg} z = \pi.$$

Furthermore, note that

$$(1+i)^{20} = 2^{10}\left(\cos\pi + i\sin\pi\right) = 2^{10}(-1 + i\cdot 0) = -2^{10} = -1024.$$

Graph the prinicipal argument.

17. **Conversion to $x + iy$.** To convert from polar form to the form $x + iy$, we have to evaluate $\sin\theta$ and $\cos\theta$ for the given θ. Here

$$\sqrt{8}\left(\cos\frac{1}{4}\pi + i\ \sin\frac{1}{4}\pi\right) = \sqrt{8}\left(\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}\right) = \frac{\sqrt{16}}{2} + \frac{\sqrt{16}}{2}i = 2 + 2i.$$

21. **Roots.** From Prob. 1 and Example 1, p. 614 in this section, we know that $1 + i$ in polar form is

$$1 + i = \sqrt{2}\left(\cos\tfrac{1}{4}\pi + i\cos\tfrac{1}{4}\pi\right).$$

Hence by (15), p. 617,

$$\sqrt[3]{1+i} = (1+i)^{1/3} = \left(\sqrt{2}\right)^{1/3}\left(\cos\frac{\frac{1}{4}\pi + 2k\pi}{3} + i\cos\frac{\frac{1}{4}\pi + 2k\pi}{3}\right).$$

Now we can simplify

$$\left(\sqrt{2}\right)^{1/3} = \left(2^{1/2}\right)^{1/3} = 2^{1/6}$$

and

$$\frac{\frac{1}{4}\pi + 2k\pi}{3} = \frac{\pi/4}{3} + \frac{2k\pi}{3} = \frac{\pi}{12} + \frac{8k\pi}{12} = \frac{\pi\,(1+8k)}{12}.$$

Hence

$$\sqrt[3]{1+i} = 2^{1/6}\left[\cos\frac{\pi\,(1+8k)}{12} + i\,\sin\frac{\pi\,(1+8k)}{12}\right],$$

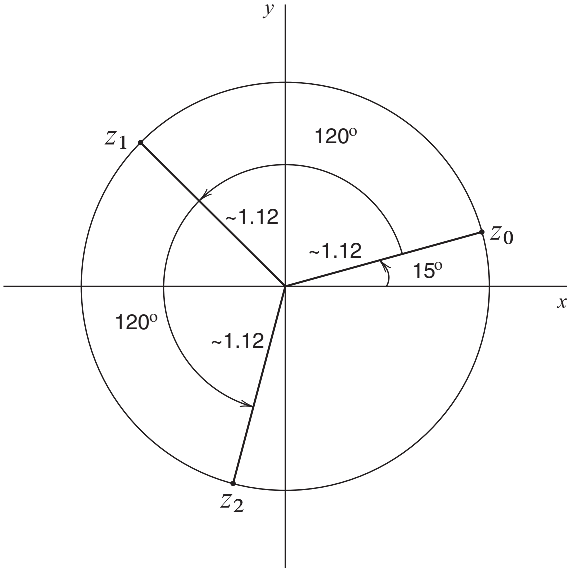
where $\quad k = 0, 1, 2$ (3 roots; thus 3 values of k). Written out we get

$$\text{For } k = 0 \qquad z_0 = 2^{1/6}\left(\cos\frac{\pi}{12} + i\,\sin\frac{\pi}{12}\right).$$

$$\text{For } k = 1 \qquad z_1 = 2^{1/6}\left(\cos\frac{9\pi}{12} + i\,\sin\frac{9\pi}{12}\right).$$

$$\text{For } k = 2 \qquad z_2 = 2^{1/6}\left(\cos\frac{17\pi}{12} + i\,\sin\frac{17\pi}{12}\right).$$

The three roots are regularly spaced around a circle of radius $2^{1/6} = 1.122\,5$ with center 0.



Sec. 13.2. Prob. 21. The three roots z_0, z_1, z_2 of $z = \sqrt[3]{1+i}$ in the complex plane

29. Equations involving roots of complex numbers. Applying the usual formula for the solutions of a quadratic equation

$$z = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

to

$$\text{(Eq)} \qquad z^2 - z + 1 - i = 0,$$

we first have

$$\text{(A)} \qquad z = \frac{-1 \pm \sqrt{1^2 - 4\cdot 1\cdot(1-i)}}{2\cdot 1} = \frac{-1 \pm \sqrt{-3+4i}}{2}.$$

Now, in (A), we have to simplify $\sqrt{-3+4i}$. Let $z = p + qi$ be a complex number where p, q are real. Then

$$z^2 = (p+qi)^2 = p^2 - q^2 + 2pqi = -3 + 4i.$$

We know that for two complex numbers to be equal, their real parts and imaginary parts must be equal, respectively. Hence, from the imaginary part

$$\begin{aligned} 2pq &= 4, \\ pq &= 2, \qquad \text{(B)} \\ p &= \frac{2}{q}. \end{aligned}$$

This can then be used, in the real part,

$$\begin{aligned} p^2 - q^2 &= -3, \\ p^2 + \frac{4}{p^2} &= -3, \\ p^4 + 4 &= -3p^2, \\ p^4 - 3p^2 + 4 &= 0. \end{aligned}$$

To solve this quartic equation, we set $h = p^2$ and get the quadratic equation

$$h^2 + 3h - 4 = 0,$$

which factors into

$$(h - 1)(h + 4) = 0 \qquad \text{so that} \quad h = 1 \quad \text{and} \quad h = -4.$$

Hence

$$p^2 = 1 \qquad \text{and} \qquad p^2 = -4.$$

Since p must be real, $p^2 = -4$ is of no interest. We are left with $p^2 = 1$ so

(C) (a) $p = 1$, (b) $p = -1$.

Substituting [C(a)] into (B) gives

$$pq = 1 \cdot q = 2 \qquad \text{so} \qquad q = 2.$$

Similarly, substituting [C(b)] into (B) gives

$$pq = (-1) \cdot q = 2 \qquad \text{so} \qquad q = -2.$$

We have $p = 1, q = 2$ and $p = -1, q = -2$. Thus, for $z = p + qi$ (see above), we get

$$1 + 2i \qquad \text{and} \qquad -1 - 2i = -(1 + 2i).$$

Hence (A) simplifies to

$$z = \frac{-1 \pm \sqrt{-3 + 4i}}{2} = \frac{-1 \pm \sqrt{(1 + 2i)^2}}{2} = \frac{-1 \pm (1 + 2i)}{2}.$$

This gives us the desired solutions to (Eq), that is,

$$z_1 = \frac{-1 + (1 + 2i)}{2} = \frac{2i}{2} = i$$

and

$$z_2 = \frac{-1 - (1 + 2i)}{2} = \frac{-2 - 2i}{2} = -1 - i.$$

Verify the result by plugging the two values into equation (Eq) and see that you get zero.

Sec. 13.3 Derivative. Analytic Function

The material follows the calculus you are used to with *certain differences* due to working in the complex plane with complex functions $f(z)$. In particular, *the concept of* **limit** *is different* as z may approach z_0 from any direction (see pp. 621–622 and **Example 4**). This also means that the **derivative**, which looks the same as in calculus, is different in complex analysis. Open the textbook on p. 623 and take a look at Example 4. We show from the definition of **derivative** (4), p. 622, which uses the concept of limit, that $f(z) = \bar{z}$ is not differentiable. The essence of the example is that approaching z along path I in **Fig. 334**, p. 623, gives a value different from that along path II. This is not allowed with limits (see pp. 621–622).

We call those functions that are differentiable in some domain **analytic** (p. 623). You can think of them as the "good functions," and they will form the preferred functions of complex analysis and its applications. Note that $f(z) = \bar{z}$ **is not analytic**. (You may want to build a small list of nonanalytic functions, as you encounter them. In Sec. 13.4 we shall learn a famous method for testing analyticity.)

The differentiation rules are the same as in real calculus (see Example 3, pp. 622–623 and Prob. 19). Here are two examples

$$f(z) = (1-z)^{16},$$

$$f'(z) = 16(1-z)^{15}(-1) = -16(1-z)^{15},$$

where the factor (-1) comes from the chain rule;

$$f(z) = i, \qquad f'(z) = 0$$

since i is a constant.

Go over the material to see that many concepts from calculus carry over to complex analysis. Use this section as a reference section for many of the concepts needed for Part D.

Problem Set 13.3. Page 624

1. Regions of practical interest. Closed circular disk. We want to write

$$|z + 1 - 5i| \le \tfrac{3}{2}$$

in the form

$$|z - a| \le p$$

as suggested on p. 619. We can write

$$\begin{aligned} |z + 1 - 5i| &= |z + (1 - 5i)| \\ &= |z - (-(1-5i))| \\ &= |z - (-1 + 5i)|. \end{aligned}$$

Hence the desired region

$$|z - (-1 + 5i)| \le \tfrac{3}{2}$$

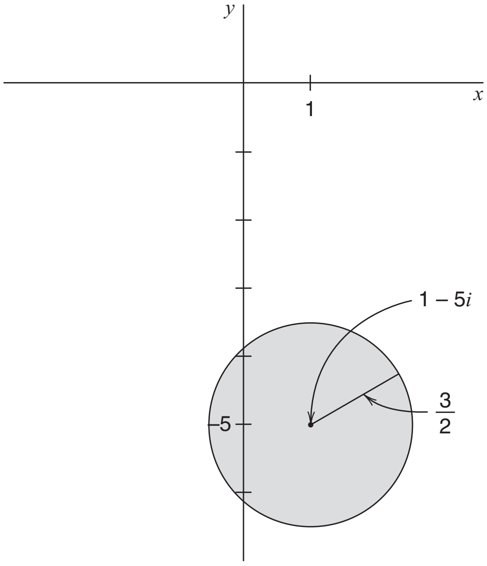
is a closed circular disk with center $-1 + 5i$ (**not** $1 - 5i$!) and radius $\frac{3}{2}$.

7. Regions. Half-plane. Let $z = x + yi$. Then $\operatorname{Re} z = x$ as defined on p. 609. We are required to determine what

$$\operatorname{Re} z \ge -1$$

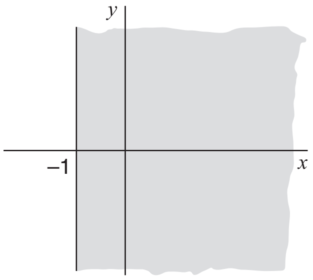
means. By our reasoning we have $\operatorname{Re} z = x \ge -1$ so that the region of interest is

$$x \ge -1.$$



Sec. 13.3. Prob. 1. Sketch of closed circular disk $|z + 1 - 5i| \leq \frac{3}{2}$

This is a closed right half-plane bounded by $x = -1$, that is, a half-plane to the right of $x = -1$ that includes the boundary.



Sec. 13.3. Prob. 7. Sketch of half-plane $\text{Re}\, z \geq -1$

11. Function values are obtained, as in calculus, by substitution of the given value into the given function. Perhaps a quicker solution than the one shown on p. A35 of the textbook, and following the approach of p. 621, is as follows. The function

$$f(z) = \frac{1}{1-z} \qquad \text{evaluated at } z = 1 - i$$

is

$$f(1-i) = \frac{1}{1-(1-i)} = \frac{1}{1-1+i} = \frac{1}{i} = -i,$$

with the last equality by (I5) in Prob. 1 of Sec. 13.1 on p. 258 of this Manual. Hence $\text{Re}\, f = \text{Re}(i) = 0$, $\text{Im}\, f = \text{Im}(i) = -1$.

17. Continuity. Let us use polar coordinates (Sec. 13.2) to see whether the function defined by

$$f(z) = \begin{cases} \dfrac{\text{Re}(z)}{1-|z|} & \text{for} \quad z \neq 0 \\ 0 & \text{for} \quad z = 0 \end{cases}$$

is continuous at $z = 0$. Then $x = r\cos\theta$, $y = r\sin\theta$ by (1), p. 613, and, using the material on p. 613, we get

$$f(z) = \frac{\operatorname{Re}(z)}{1-|z|} = \frac{x}{1-|z|} = \frac{r\cos\theta}{1-r}.$$

We note that as $r \to 0$,

$$1 - r \to 1 \quad \text{and} \quad r\cos\theta \to 0$$

so that

$$\frac{r\cos\theta}{1-r} \to 0 \quad \text{as} \quad r \to 0, \quad \text{for any value of } \theta.$$

By (3), p. 622, we can conclude that f is continuous at $z = 0$.

19. Differentiation. Note that differentiation in complex analysis is as in calculus. We have

$$f(z) = (z - 4i)^8,$$

$$f'(z) = 8(z - 4i)^7,$$

$$f'(3 + 4i) = 8(3 + 4i - 4i)^7 = 8 \cdot 3^7 = 8 \cdot 2187 = 17{,}496.$$

Remark. Be aware of the **chain rule**. Thus if, for example, we want to differentiate

$$g(z) = (-2z - 4i)^{10}, \qquad \text{then}$$

$$g'(z) = 10(-2z - 4i)^9 \cdot (-2) = -20(-2z - 4i)^9,$$

where the factor −2 comes in by the chain rule.

Sec. 13.4 Cauchy–Riemann Equations. Laplace's Equation

We discussed the concept of analytic functions in Sec. 13.4 and we learned that these are the functions that are differentiable in some domain and that operations of complex analysis can be applied to them. Unfortunately, not all functions are analytic as we saw in Example 4, p. 623. How can we tell whether a function is analytic? The Cauchy–Riemann equations (1), p. 625, allow us to test whether a complex function is analytic. Details are as follows.

If a complex function $f(z) = u(x, y) + iv(x, y)$ is analytic in D, then u and v satisfy the Cauchy–Riemann equations

$$(1) \qquad u_x = v_y, \qquad u_y = -v_x$$

(Theorem 1, p. 625) as well as Laplace's equations $\nabla^2 u = 0$, $\nabla^2 v = 0$ (Theorem 3, p. 628; see also Example 4, p. 629, and Prob. 15). The converse of Theorem 1 is also true (Theorem 2, p. 627), provided the derivatives in (1) are continuous. For these reasons the Cauchy–Riemann equations are most important in complex analysis, which is the study of *analytic* functions.

Examples 1 and **2**, p. 627, and **Probs. 3** and **5** use the Cauchy–Riemann equations to test whether the given functions are analytic. In particular, note that Prob. 5 gives complete details on how to use the Cauchy–Riemann equations (1), p. 625, in complex form (7), p. 628, and even how to conclude nonanlyticity by observing the given function. **You have to memorize the Cauchy–Riemann equations (1). Remember the minus sign in the second equation!**

Problem Set 13.4. Page 629

3. Check of analyticity. Cauchy–Riemann equations (1), p. 625. From the given function

$$\begin{aligned} f(z) &= e^{-2x}(\cos 2y - i \sin 2y) \\ &= e^{-2x}\cos 2y + i\left(-e^{-2x}\sin 2y\right) \end{aligned}$$

we see that the real part of f is

$$u = e^{-2x}\cos 2y$$

and the imaginary part is

$$v = -e^{-2x}\sin 2y.$$

To check whether f is analytic, we want to apply the important Cauchy–Riemann equations (1), p. 625. To this end, we compute the following four partial (real) derivatives:

$$\begin{aligned} u_x &= -2e^{-2x}\cos 2y, \\ v_y &= -e^{-2x}(2\cos 2y) = -2e^{-2x}\cos 2y, \\ u_y &= -e^{-2x}[(-\sin 2y)\cdot 2] = -2e^{-2x}\sin 2y, \\ v_x &= -e^{-2x}(-2)(\sin 2y) = 2e^{-2x}\sin 2y. \end{aligned}$$

Note that the factor -2 in u_x and the factor 2 in v_y result from the chain rule. Can you identify the use of the chain rule in the other two partial derivatives? We see that

$$u_x = \cos 2y = v_y$$

and

$$u_y = -2e^{-2x}\sin 2y = -v_x.$$

This shows that the Cauchy–Riemann equations are satisfied for all $z = x + iy$ and we conclude that f is indeed analytic.

In Sec. 13.5 we will learn that the given function f defines the complex exponential function e^z, with $z = -2x + i2y$ and that, in general, the complex exponential function is analytic.

5. Not analytic. We show that $f(z) = \mathrm{Re}\,(z^2) - i\,\mathrm{Im}\,(z^2)$ is not analytic in three different ways.

Solution 1. Standard solution in x,y coordinates. We have that, if $z = x + iy$, then

$$z^2 = (x+iy)(x+iy) = x^2 + 2ixy + i^2y^2 = (x^2 - y^2) + i(2xy).$$

Thus we see that

$$\mathrm{Re}\,(z^2) = x^2 - y^2 \quad \text{and} \quad \mathrm{Im}\,(z^2) = 2xy.$$

Thus the given function is

$$f(z) = x^2 - y^2 \; - i2y.$$

Hence

$$u = x^2 - y^2, \qquad v = -2xy.$$

To test whether $f(z)$ satisfies the Cauchy–Riemann equations (1), p. 625, we have to take four partial derivatives

$$u_x = 2x \quad \text{and} \quad v_y = -2x$$

so that

$$\text{(XCR1)} \qquad u_x \neq v_y.$$

(*We could stop here and have a complete answer that the given function is not analytic!* However, for demonstration purposes we continue.)

$$u_y = 2y \quad \text{and} \quad v_x = -2y$$

so that

$$\text{(CR2)} \qquad u_y = -v_x.$$

We see that although the given function $f(z)$ satisfies the second Cauchy–Riemann equation (1), p. 625, as seen by (CR2), it does not satisfy the first Cauchy–Riemann equation (1) as seen by (XCR1). We note that the functions $u(x, y)$, $v(x, y)$ are continuous and conclude by Theorems 1, p. 625, and 2, p. 627 that $f(z)$ **is not analytic.**

Solution 2. In polar coordinates. We have $z = r(\cos\theta + i\sin\theta)$ by (2), p. 613, so that $x = r\cos\theta$, $y = r\sin\theta$. Hence

$$x^2 - y^2 = r^2\cos^2\theta - r^2\sin^2\theta = r^2(\cos^2\theta - \sin^2\theta),$$

$$2xy = 2r^2\cos\theta\sin\theta.$$

Together, we get our given function $f(z)$ in polar coordinates

$$f(z) = r^2(\cos^2\theta - \cos^2\theta) - i2r^2\cos\theta\sin\theta = r^2(\cos^2\theta - \cos^2\theta - 2i\cos\theta\sin\theta).$$

We have

$$\begin{aligned} u &= r^2(\cos^2\theta - \sin^2\theta), \\ v &= -2r^2\cos\theta\sin\theta. \end{aligned}$$

Then the partial derivatives are

$$u_r = 2r(\cos^2\theta - \sin^2\theta)$$

and, by the product rule,

$$\begin{aligned} v_\theta &= (-2r^2)(-\sin\theta)(\sin\theta) + (-2r^2)\cos\theta\cos\theta \\ &= 2r^2(\sin^2\theta - \cos^2\theta), \end{aligned}$$

and

$$\frac{1}{r}v_\theta = 2r(\sin^2\theta - \cos^2\theta),$$

We see that $u_r = -(1/r)v_\theta$ so that

$$u_r \neq \frac{1}{r}v_\theta.$$

This means that f does not satisfy the first Cauchy–Riemann equation in polar coordinates (7), p. 628, and f is not analytic. (Again we could stop here. However, for pedagogical reasons we continue.)

$$\begin{aligned} v_r &= -4r\cos\theta\sin\theta, \\ u_\theta &= r^2(2\sin\theta\cos\theta) - 2\cos\theta(-\sin\theta) \\ &= 4r^2(\sin\theta\cos\theta) \end{aligned}$$

and

$$-\frac{1}{r}u_\theta = -4r(\sin\theta\cos\theta).$$

This shows that $f(z)$ satisfies the second Cauchy–Riemann equation in polar coordinates (7), that is,

$$v_r = -\frac{1}{r}u_\theta.$$

However, this does not help, since the first Cauchy–Riemann equation is not satisfied. We conclude that $f(z)$ is not analytic.

Solution 3. Observation about f(z). We note that

$$(\bar{z})^2 = (x - iy)(x - iy) = x^2 - 2ixy - y^2 = x^2 - y^2 - 2ixy.$$

We compare this with our given function and see that

$$f(z) = (\bar{z})^2 = \bar{z}\cdot\bar{z}.$$

Furthermore,

$$f(x) = (\bar{z})^2 = \overline{(z^2)}.$$

From Example 4, p. 623 in Sec. 13.3, we know that $\bar{z}$ is not differentiable so we conclude that the given $f(x) = \overline{(z^2)}$ is also not differentiable. Hence $f(z)$ is not analytic (by definition on p. 623).

Remark. Solution 3 is the most elegant one. Solution 1 is the standard one where we stop when the first Cauchy–Riemann equation is not satisfied. Solution 2 is included here to show how the Cauchy–Riemann equations are calculated in polar coordinates. (Here Solution 2 is more difficult than Solution 1 but sometimes conversion to polar makes calculating the partial derivatives simpler.)

15. Harmonic functions appear as real and imaginary parts of analytic functions.
First solution method. Identifying the function.
If you remember that the given function $u = x/(x^2 + y^2)$ is the real part of $1/z$, then you are done. Indeed,

$$\begin{aligned} \frac{1}{z} &= \frac{1}{x + iy} \\ &= \frac{1}{x + iy}\cdot\frac{x - iy}{x - iy} \\ &= \frac{x - iy}{x^2 + y^2} \\ &= \frac{x}{x^2 + y^2} + i\,\frac{-y}{x^2 + y^2} \end{aligned}$$

so that clearly

$$\operatorname{Re}\left(\frac{1}{z}\right) = \frac{x}{x^2 + y^2},$$

and hence the given function u is analytic. Moreover, our derivation also shows that a conjugate harmonic of u is $-y/(x^2 + y^2)$.

Second solution method. Direct calculation as in Example 4, p. 629.

If you don't remember that, you have to work systematically by differentiation, beginning with proving that u satisfies Laplace's equation (8), p. 628. Such somewhat lengthy differentiations, as well as other calculations, can often be simplified and made more reliable by introducing suitable shorter notations for certain expressions. In the present case we can write

$$u = \frac{x}{G}, \qquad \text{where} \qquad G = x^2 + y^2.$$

Then

$$\text{(A)} \qquad G_x = 2x, \qquad G_y = 2y.$$

By applying the product rule of differentiation (and the chain rule), not the quotient rule, we obtain the first partial derivative.

$$\text{(B)} \qquad u_x = \frac{1}{G} - \frac{x(2x)}{G^2}.$$

By differentiating this again, using the product and chain rules, we obtain the second partial derivative:

$$\text{(C)} \qquad u_{xx} = -\frac{2x}{G^2} - \frac{4x}{G^2} + \frac{8x^3}{G^3}.$$

Similarly, the partial derivative of u with respect to y is obtained from (A) in the form

$$\text{(D)} \qquad u_y = -\frac{2xy}{G^2}.$$

The partial derivative of this with respect to y is

$$\text{(E)} \qquad u_{yy} = -\frac{2x}{G^2} + \frac{8xy^2}{G^3}.$$

Adding (C) and (E) and remembering that $G = x^2 + y^2$ gives us

$$u_{xx} + u_{yy} = -\frac{8x}{G^2} + \frac{8x(x^2 + y^2)}{G^3} = -\frac{8x}{G^2} + \frac{8x}{G^2} = 0.$$

This shows that $u = x/G = x/(x^2 + y^2)$ satisfies Laplace's equation (8), p. 628, and thus is harmonic.

Next we want to determine a harmonic conjugate. From (D) and the second Cauchy–Riemann equation (1), p. 625, we obtain

$$u_y = -\frac{2xy}{G^2} = -v_x.$$

Integration of $2x/G^2 = G_x/G^2$, with respect to x, gives $-1/G$, so that integration of v_x, with respect to x, gives

$$\text{(F)} \qquad v = -\frac{y}{G} = -\frac{y}{x^2 + y^2} + h(y).$$

Now we show that $h(y)$ must be a constant. We obtain, by differentiating (F) with respect to y and taking the common denominator G^2, the following:

$$v_y = -\frac{1}{G} + \frac{2y^2}{G^2} = \frac{-x^2 + y^2}{G^2} + h'(y).$$

On the other hand, we have from (B) that

$$u_x = \frac{1}{G} - \frac{2x^2}{G^2} = \frac{y^2 - x^2}{G^2}.$$

By the first Cauchy–Riemann equation (1), p. 625, we have

$$v_y = u_x,$$

which means, written out, in our case

$$\frac{-x^2 + y^2}{G^2} + h'(y) = \frac{y^2 - x^2}{G^2}.$$

But this means that

$$h'(y) = 0 \quad \text{and hence} \quad h(y) = \text{const},$$

as we claimed. Since this constant is arbitrary, we can choose $h(y) = 0$ and obtain, from (F), the desired conjugate harmonic

$$v = -\frac{y}{x^2 + y^2} + h(y) = \frac{-y}{x^2 + y^2},$$

which is the same answer as in our first solution method.

Sec. 13.5 Exponential Function

Equation (1), p. 630, defines the complex exponential function. Equations (2) and (3) on that page are as in calculus. Note that equation (4), p. 631, is a special case of equation (3). The Euler formula (5), p. 631, is very important and gives the polar form (6) of

$$z = x + iy = r\,(\cos\theta + i\,\sin\theta) = re^{i\theta}.$$

It would be useful for you to remember equations (7), (8), and (9). The periodicity (12), p. 632, has no counterpart in real. It motivates the fundamental region (13), p. 632, of e^z.

Solving complex equations, such as Prob. 19, gives practice in the use of complex elementary functions and illustrates the difference between these functions and their *real* counterparts. In particular, Prob. 19 has infinitely many solutions in complex but only one solution in real!

Problem Set 13.5. Page 632

5. Function values. We note that

$$e^{2+3\pi i} = e^z = e^{x+iy}.$$

Thus we use (1), p. 630, with $x = 2$ and $y = 3\pi$ and obtain

$$\begin{aligned} e^{2+3\pi i} &= e^2(\cos 3\pi + i \sin 3\pi) \\ &= e^2(\cos(\pi + 2\pi) + i \sin(\pi + 2\pi)) && \big(\text{since } \cos(3\pi) = \cos(\pi + 2\pi), \text{ same for } \sin(3\pi)\big) \\ &= e^2(\cos \pi + i \sin \pi) && (\cos \text{ and } \sin \text{ both have periods of } 2\pi) \\ &= e^2(-1 + i \cdot 0) \\ &= -e^2 \approx -7.389. \end{aligned}$$

From (10), p. 631, we have the absolute value

$$\left|e^{2+3\pi i}\right| = |\, e^z \,| = e^x = e^2 \approx 7.389.$$

9. Polar form. We want to write $z = 4 + 3i$ in exponential form (6), p. 631. This means expressing it in the form

$$z = re^{i\theta}.$$

We have

$$r = |\, z \,| = \sqrt{x^2 + y^2} = \sqrt{4^2 + 3^2} = \sqrt{25} = 5.$$

We know, by Sec. 13.2, pp. 613–619, that the principal argument of the given z is

$$\operatorname{Arg} z = \arctan\left(\frac{y}{x}\right) = \arctan\left(\frac{3}{4}\right) = 0.643501.$$

Hence, by (6), p. 631, we get that z in polar form is

$$z = 5e^{i \arctan(3/4)} = 5e^{0.643501i}.$$

Checking the answer. By (2), p. 613, in Sec. 13.2, we know that any complex number $z = x + iy$ has polar form

$$z = r(\cos\theta + i \sin\theta).$$

Thus, for $z = 4 + 3i$, we have

$$\begin{aligned} z &= 5(\cos 0.643501 + i \sin 0.643501) \\ &= 5(0.8 + 0.6i) \\ &= 4 + 3i. \end{aligned}$$

15. Real and imaginary parts. We want to find the real and imaginary parts of $\exp(z^2)$. From the beginning of Sec. 13.5 of the textbook we know that the notation exp means

$$\exp(z^2) = e^{z^2}.$$

Now for $z = x + iy$,

$$z^2 = (x + iy)(x + iy) = x^2 - y^2 + i2xy.$$

Thus

$$e^{z^2} = e^{x^2-y^2+i2xy} = e^{x^2-y^2}e^{i2xy} \qquad \text{[by (3), p. 630].}$$

Now

$$e^{i2xy} = \cos(2xy) + i\,\sin(2xy) \qquad \text{[by (1), p. 630; (5), p. 631].}$$

Putting it together

$$\begin{aligned} e^{z^2} &= e^{x^2-y^2}[\cos(2xy) + i\,\sin(2xy)] \\ &= e^{x^2-y^2}\cos 2xy + i(e^{x^2-y^2}\sin 2xy). \end{aligned}$$

Hence

$$\operatorname{Re}\left[\exp(z^2)\right] = e^{x^2-y^2}\cos 2xy; \qquad \operatorname{Im}\left[\exp(z^2)\right] = e^{x^2-y^2}\sin 2xy,$$

as given on p. A36 of the textbook.

19. Equation. To solve

(A) $$e^z = 1$$

we set $z = x + iy$. Then

$$\begin{aligned} e^z &= e^{x+iy} = e^x e^{iy} = e^x(\cos y + i\,\sin y) && \text{[by (5), p. 631]} \\ &= e^x\cos y + ie^x\sin y \\ &= 1 && \text{[by (A)]} \\ &= 1 + i\cdot 0. \end{aligned}$$

Equate the real and imaginary parts on both sides to obtain

$$\text{(B)} \quad \operatorname{Re}(e^z) = e^x\cos\,y = 1, \qquad \text{(C)} \quad \operatorname{Im}(e^z) = e^x\sin\,y = 0.$$

Since $e^x > 0$ but the product in (C) must equal zero requires that

$$\sin y = 0 \quad \text{which means that} \quad \text{(D)} \quad y = 0, \pm\pi, \pm 2\pi, \pm 3\pi, \ldots.$$

Since the product in (B) is positive, $\cos y$ has to be positive. If we look at (D), we know that $\cos y$ is -1 for $y = \pm\pi, \pm 3\pi, \pm 5\pi, \ldots$ but $+1$ for $y = 0, \pm 2\pi, \pm 4\pi, \ldots$. Hence (B) and (D) give

(E) $$y = 0, \pm 2\pi, \pm 4\pi, \ldots.$$

Since (B) requires that the product be equal to 1 and the cosine for the values of y in (E) is 1, we have $e^x = 1$. Hence

(F) $$x = 0.$$

Then (E) and (F) together yield

$$x = 0 \quad y = 0, \pm 2\pi, \pm 4\pi, \ldots\ ,$$

and the desired solution to (A) is

$$z = x + yi = \pm 2n\pi i, \qquad n = 0, 1, 2, \ldots.$$

Note that (A), being complex, has infinitely many solutions in contrast to the same equation in real, which has only one solution.

Sec. 13.6 Trigonometric and Hyperbolic Functions. Euler's Formula

In complex, the exponential, trigonometric, and hyperbolic functions are related by the definitions (1), p. 633, and (11), p. 635, and by the Euler formula (5), p. 634, as well as by (14) and (15), p. 635. Thus we can convert them back and forth. Formulas (6) and (7) are needed for computing values. **Problem 9** uses such a formula to compute function values.

Problem Set 13.6. Page 636

1. Formulas for hyperbolic functions. To show that

$$\cosh z = \cosh x \cos y + i \sinh x \sin y$$

we do the following. We start with the definition of $\cosh z$. Since we want to avoid carrying a factor $\frac{1}{2}$ along, we multiply both sides of (11), p. 635, by 2 and get

$$\begin{aligned}
2\cosh z &= e^z + e^{-z} \\
&= e^{x+iy} + e^{-x-iy} && \text{(setting } z = x + iy) \\
&= e^x(\cos y + i \sin y) + e^{-x}(\cos y - i \sin y) && \text{(by (1), p. 630)} \\
&= \cos y(e^x + e^{-x}) + i \sin y(e^x - e^{-x}) \\
&= \cos y(2\cosh x) + i \sin y(2\sinh x) && \text{(by (17), p. A65 of Sec. A3.1 of App. 3)} \\
&= 2\cosh x \cos y + 2i \sinh x \sin y.
\end{aligned}$$

Division by 2 on both sides yields the desired result. Note that the formula just proven is useful because it expresses $\cosh z$ in terms of its real and imaginary parts.

The related formula for $\sinh z$ follows the same proof pattern, this time start with $2\sinh z = e^z - e^{-z}$. Fill in the details.

9. Function values. The strategy for **Probs. 6–12** is to find formulas in this section or in the problem set that allow us to get, as an answer, a real number or complex number. For example, the formulas in Prob. 1 are of the type we want for this kind of problem.

In the present case, by Prob. 1 (just proved before!), we denote the first given complex number by $z_1 = -1 + 2i$ so that $x_1 = -1$ and $y_1 = 2$ and use

$$\cosh z_1 = \cosh x_1 \cos y_1 + i \sinh x_1 \sin y_1.$$

Then

$$\cosh z_1 = \cosh(-1 + 2i) = \cosh(-1)\cos 2 + i \sinh(-1)\sin 2.$$

Now by (11), p. 635,

$$\cosh x_1 = \cosh(-1) = \frac{1}{2}(e^{-1} + e^1) = \frac{1 + e^2}{2e}; \qquad \sinh x_1 = \sinh(-1) = \frac{1 - e^2}{2e}.$$

Using a calculator (or CAS) to get the actual values we have

$$\begin{aligned}
\cosh(-1 + 2i) &= \frac{1 + e^2}{2e}\cos 2 + i\frac{1 - e^2}{2e}\sin 2 \\
&= 1.543081 \cdot (-0.4161468) + i(-1.752011) \cdot (0.9092974) \\
&= -0.642148 - 1.068607i,
\end{aligned}$$

which corresponds to the rounded answer on p. A36.

For the second function value $z_2 = -2 - i$ we notice that, by (1), p. 633,

$$\cos z = \tfrac{1}{2}\left(e^{iz} + e^{-iz}\right)$$

and, by (11), p. 635,

$$\cosh z = \tfrac{1}{2}\left(e^{z} + e^{-z}\right).$$

Now

$$\text{(A)} \qquad iz_2 = i(-2-i) = -2i - i^2 = 1 - 2i = z_1.$$

Hence

$$\begin{aligned}\cos z_2 &= \tfrac{1}{2}\left(e^{iz_2} + e^{-iz_2}\right)\\ &= \tfrac{1}{2}\left(e^{z_1} + e^{-z_1}\right) \qquad \text{[by (A)]}\\ &= \cosh z_1\\ &= \cosh(1-2i)\end{aligned}$$

so we get the same value as before!

13. Equations. We want to show that the complex cosine function is even.
First solution directly from definition (1), p. 633.
We start with

$$\cos(-z) = \tfrac{1}{2}\left(e^{i(-z)} + e^{-i(-z)}\right).$$

We see that for any complex number $z = x + iy$:

$$i(-z) = i[-(x+iy)] = i(-x-iy) = -ix - i^2y = y - ix.$$

Similarly,

$$-iz = -i(x+iy) = ix - i^2y = y - ix = i(-z).$$

So we have

$$\boxed{-iz = i(-z).}$$

Similarly,

$$-i(-z) = -i(-x-yi) = -y + xi$$

and

$$iz = i(x+iy) = ix + i^2y = -i(-z)$$

so that

$$\boxed{iz = -i(-z).}$$

Putting these two boxed equations to good use, we have

$$\cos(-z) = \tfrac{1}{2}\left(e^{i(-z)} + e^{-i(-z)}\right) = \tfrac{1}{2}\left(e^{-iz} + e^{iz}\right) = \tfrac{1}{2}\left(e^{iz} + e^{-iz}\right) = \cos z.$$

Thus $\cos(-z) = \cos z$, which means that the complex cosine function (like its real counterpart) is even.

Second solution by using (*6a*)*, p. 634.* From that formula we know

$$\cos z = \cos x \cosh y - i \sin x \sinh y.$$

We consider

$$\begin{aligned}\cos(-z) &= \cos(-x - iy)\\ &= \cos(-x + i(-y))\\ &= \cos(-x)\cosh(-y) - i\sin(-x)\sinh(-y)\\ &= \cos x \cosh y - i(-\sin x)(-\sinh y)\\ &= \cosh x \cosh y - i \sin x \sinh y\\ &= \cos z.\end{aligned}$$

The fourth equality used that, for real x and y, both cos and cosh are even and sin and sinh are odd, that is,

$$\begin{aligned}&\cos(-x) = \cos x; \qquad &&\cosh(-y) = \cosh y;\\ &\sin(-x) = -\sin x \qquad &&\sinh(-x) = -\sinh x.\end{aligned}$$

Similary, show that the complex sine function is odd, that is, $\sin(-z) = -\sin z$.

17. Equations. To solve the given complex equation, $\cosh z = 0$, we use that, by the first equality in Prob. 1, p. 636, of Sec. 13.6, the given equation is equivalent to a pair of real equations:

$$\begin{aligned}\text{Re}\,(\cosh z) &= \cosh x \cos y = 0,\\ \text{Im}\,(\cosh z) &= \sinh x \sin y = 0.\end{aligned}$$

Since $\cosh x \neq 0$ for all x, we must have $\cos y = 0$, hence $y = \pm(2n+1)\pi/2$ where $n = 0, 1, 2, \ldots$. For these y we have $\sin y \neq 0$, noting that the real cos and sin have no common zeros! Hence $\sinh x = 0$ so that $x = 0$. Thus our reasoning gives the solution

$$z = (x, y) = (0,\ \pm(2n+1)\pi/2), \qquad \text{that is,} \qquad z = \pm(2n+1)\pi i/2 \quad \text{where} \quad n = 0, 1, 2, \ldots.$$

Sec. 13.7 Logarithm. General Power. Principal Value

Work this section with extra care, so that you understand:

1. The meaning of formulas (1), (2), (3), p. 637.

2. The difference between the real logarithm $\ln x$, which is a function defined for $x > 0$, and the complex logarithm $\ln z$, which is an infinitely many-valued relation, which, by formula (3), p. 637, "decomposes" into infinitely many functions.

Example 1, p. 637, and **Probs. 5**, **15**, and **21** illustrate these formulas.
General powers z^c are defined by (7), p. 639, and illustrated in Example 3 at the bottom of that page.

Problem Set 13.7. Page 640

5. Principal value. Note that the real logarithm of a negative number is undefined. The principal value Ln z of $\ln z$ is defined by (2), p. 637, that is,

$$\operatorname{Ln} z = \ln|z| + i \operatorname{Arg} z$$

where Arg z is the principal value of $\arg z$. Now recall from (5), p. 614 of Sec. 13.2, that the principal value of the argument of z is defined by

$$-\pi < \operatorname{Arg} z \leq \pi.$$

In particular, for a negative real number we always have Arg $z = +\pi$, as you should keep in mind. From this, and (2), we obtain the answer

$$\operatorname{Ln}(-11) = \ln|-11| + i\pi = \ln 11 + i\pi.$$

15. All values of a complex logarithm. We need the absolute value and the argument of e^i because, by (1) and (2), p. 637,

$$\begin{aligned}\ln(e^i) &= \ln|e^i| + i \arg(e^i)\\ &= \ln|e^i| + i \operatorname{Arg}(e^i) \pm 2n\pi i, \qquad \text{where} \quad n = 0, 1, 2, \ldots.\end{aligned}$$

Now the absolute value of the exponential function e^z with a pure imaginary exponent always equals 1, as you should memorize; the derivation is

$$|e^{iy}| = |\cos y + i \sin y| = \sqrt{\cos^2 y + \sin^2 y} = 1.$$

(Can you see where this calculation would break down if y were not real?) In our case,

(A) $$|e^i| = 1, \qquad \text{hence} \qquad \ln|e^i| = \ln(1) = 0.$$

The argument of e^i is obtained from (10), p. 631 in Sec. 13.5, that is,

$$\arg(e^z) = \operatorname{Arg}(e^z) \pm 2n\pi = y \pm 2n\pi \quad \text{where} \quad n = 0, 1, 2, \ldots.$$

In our problem we have $z = i = x + iy$, hence $y = 1$. Thus

(B) $$\arg(e^i) = 1 \pm 2n\pi, \qquad \text{where} \quad n = 0, 1, 2, \ldots.$$

From (A) and (B) we obtain the answer

$$\begin{aligned}\ln(e^i) &= \ln|e^i| + i \ \arg(e^i)\\ &= 0 + i(1 \pm 2n\pi), \qquad \text{where} \quad n = 0, 1, 2, \cdots.\end{aligned}$$

21. Equation. We want to solve

$$\begin{aligned}\ln z &= 0.6 + 0.4i\\ &= \ln|z| + i \arg z \qquad \text{[by (1), p. 637].}\end{aligned}$$

We equate the real parts and the imaginary parts:

$$\begin{aligned}0.6 &= \ln|z|, \qquad \text{thus} \quad |z| = e^{0.6}.\\ 0.4 &= \arg z.\end{aligned}$$

Next we note that

$$z = e^{\ln z} = e^{\ln|z|+i \arg z} = e^{0.6}e^{0.4i}.$$

We consider

$$\begin{aligned} e^{0.4i} &= e^{0+0.4i} = e^{0}(\cos 0.4 + i \sin 0.4) \qquad \text{[by (1), p. 630, Sec. 13.5]} \\ &= \cos 0.4 + i \sin 0.4. \end{aligned}$$

Putting it together, we get

$$\begin{aligned} z &= e^{0.6}e^{0.4i} \\ &= e^{0.6}(\cos 0.4 + i \sin 0.4) \\ &= 1.822119 \cdot (0.921061 + 0.389418i) \\ &= 1.6783 + 0.70957i. \end{aligned}$$

23. General powers. Principal value. We start with the given equation and use (8), p. 640, and the definition of principal value to get

$$(1+i)^{1-1} = e^{(1-i)\,\mathrm{Ln}(1+i)}.$$

Now the principal value

$$\mathrm{Ln}\ (1+i) = \ln|1+i| + i\,\mathrm{Arg}(1+i) \qquad \text{[by (2), p. 637].}$$

Also

$$|1+i| = \sqrt{1^2+1^2} = \sqrt{2}$$

and

$$\mathrm{Arg}(1+i) = \frac{\pi}{4} \qquad \text{[see (5) and Example 1, both on p. 614].}$$

Hence

$$\mathrm{Ln}(1+i) = \ln\sqrt{2} + i\frac{\pi}{4}$$

so that

$$\begin{aligned} (1-i)\mathrm{Ln}(1+i) &= (1-i)\left(\ln\sqrt{2} + i\frac{\pi}{4}\right) \\ &= \ln\sqrt{2} + i\frac{\pi}{4} - i\ln\sqrt{2} - i^2\frac{\pi}{4} \\ &= \ln\sqrt{2} + \frac{\pi}{4} + i\left(\frac{\pi}{4} - \ln\sqrt{2}\right). \end{aligned}$$

Thus

$$\begin{aligned}(1+i)^{1-1} &= \exp\left[\ln\sqrt{2}+\frac{\pi}{4}+i\left(\frac{\pi}{4}-\ln\sqrt{2}\right)\right]\\ &= \exp\left(\ln\sqrt{2}+\frac{\pi}{4}\right)\cdot\exp\left[i\left(\frac{\pi}{4}-\ln\sqrt{2}\right)\right]\\ &= \exp\left(\ln\sqrt{2}\right)\cdot\exp\left(\frac{\pi}{4}\right)\cdot\left[\cos\left(\frac{\pi}{4}-\ln\sqrt{2}\right)+i\sin\left(\frac{\pi}{4}-\ln\sqrt{2}\right)\right] \qquad \text{[by (1), p. 630]}\\ &= \sqrt{2}e^{\pi/4}\left[\cos\left(\frac{\pi}{4}-\ln\sqrt{2}\right)+i\sin\left(\frac{\pi}{4}-\ln\sqrt{2}\right)\right].\end{aligned}$$

Numerical values are

$$\begin{aligned}\frac{\pi}{4}-\ln\sqrt{2} &= 0.4388246,\\ \cos\left(\frac{\pi}{4}-\ln\sqrt{2}\right) &= \cos(0.4388246)=0.9052517,\\ \sin\left(\frac{\pi}{4}-\ln\sqrt{2}\right) &= \sin(0.4388246)=0.4248757,\\ \sqrt{2}e^{\pi/4} &= 3.1017664.\end{aligned}$$

Hence $(1+i)^{1-1}$ evaluates to

$$(1+i)^{1-1}=2.8079+1.3179i.$$

Chap. 14 Complex Integration

The first method of integration ("indefinite integration and substitution of limits") is a direct analog of regular calculus and thus a good starting point for studying complex integration. The focal point of Chap. 14 is the very important **Cauchy integral theorem** (p. 653) in Sec. 14.2. This leads to **Cauchy's integral formula** (1), p. 660 in Sec. 14.3, allowing us to evaluate certain complex integrals whose integrand is of the form $f(z)/(z - z_0)$ with f being analytic. The chapter concludes with the surprising result that all analytic functions have derivatives of all orders. Complex integration has a very distinct flavor of its own and should therefore make an interesting study. The amount of theory in this chapter is very manageable but powerful in that it allows us to solve many different integrals.

General orientation. Chapter 13 provides the background material for Chap. 14. We can broadly classify the material in Chap. 14 as a ***first approach*** *to complex integration based on Cauchy's integral theorem and his related integral formula*. The groundwork to a *second approach* to complex integration is given in Chap. 15 with the actual method of integration ("residue integration") given in Chap. 16.

Prerequisite. You should remember the material of Chap. 13, including the concept of **analytic functions** (Sec. 13.3), the important **Cauchy–Riemann equations** of Sec. 13.4, and Euler's formula (5), p. 634 in Sec. 13.6. We make use of some of the properties of elementary complex functions when solving problems—so, if you forgot,—consult Chap. 13 in your textbook. You should recall how to solve basic real integrals (see inside cover of textbook if needed). You should also have some knowledge of roots of complex polynomials.

Sec. 14.1 Line Integral in the Complex Plane

The indefinite complex integrals are obtained from inverting, just as in regular calculus. Thus the starting point for the theory of complex integration is the consideration of definite complex integrals, which are defined as **complex line integrals** and explored on pp. 643–646. As an aside, the reader familiar with real line integrals (Sec. 10.1, pp. 413–419 in the text, pp. 169–172 in Vol. 1 of this Manual) will notice a similarity between the two. Indeed (8), p. 646, can be used to make the relationship between complex line integrals and real line integrals explicit, that is,

$$\begin{aligned}\int_C f(z)\,dz &= \int_C u\,dx - \int_C v\,dy + i\left[\int_C u\,dy + \int_C v\,dx\right]\\ &= \int_C (u\,dx - v\,dy) + i\left[\int_C u\,dy + v\,dx\right],\end{aligned}$$

where C is the curve of integration and the resulting integrals are real.

(However, having not studied real line integrals is not a hindrance to learning and enjoying complex analysis as we go in a systematic fashion with the only prerequisite for Part D being elementary calculus.)

The first practical method of complex integration involves *indefinite integration and substitution of limits* and is directly inspired from elementary calculus. It requires that the function be analytic. The details are given in Theorem 1, formula (9), p. 647, and illustrated below by **Examples 1–4** and **Probs. 23** and **27**.

A prerequisite to understanding the second practical method of integration (*use of a representation of a path*) is to understand **parametrization of complex curves** (**Examples 1–4**, p. 647, **Probs. 1, 7,** and **19**). Indeed, (10), p. 647, of **Theorem 2** is a more general approach than (9) of Theorem 1, because Theorem 2 applies to *any* continuous complex function not just analytic functions. However, the price of generality is a slight increase in difficulty.

Problem Set 14.1. Page 651

1. Path. We have to determine the path of

$$z(t) = (1 + \tfrac{1}{2}i)t \qquad (2 \le t \le 5).$$

Since the parametric representation

$$\begin{aligned} z(t) = x(t) + iy(t) &= \left(1 + \tfrac{1}{2}i\right)t \\ &= t + i \cdot \tfrac{1}{2}t \end{aligned}$$

is linear in the parameter t, the representation is that of a straight line in the complex z-plane. Its slope is positive, that is

$$\frac{y(t)}{x(t)} = \frac{\frac{1}{2}t}{t} = \frac{1}{2}.$$

The straight-line segment starts at $t = 2$, corresponding to

$$z_0 = z(2) = 2 + i \cdot \tfrac{1}{2} \cdot 2 = 2 + i$$

and ends at $t = 5$:

$$z_1 = z(5) = 5 + \tfrac{5}{2}i.$$

Sketch it.

7. Path. To identify what path is represented by

$$z(t) = 2 + 4e^{\pi it/2} \qquad \text{with} \qquad 0 \le t \le 2$$

it is best to derive the solution stepwise.

From Example 5, p. 648, we know that

$$z(t) = e^{it} \qquad \text{with} \qquad 0 \le t \le 2\pi$$

represents a unit circle (i.e., radius 1, center 0) traveled in the counterclockwise direction. Hence

$$z(t) = e^{\pi it/2} \qquad \text{with} \qquad 0 \le t \le 4$$

also represents that unit circle. Then

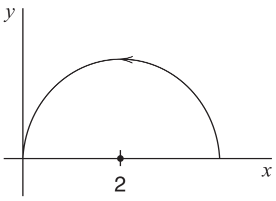
$$z(t) = 4e^{\pi it/2} \qquad \text{with} \qquad 0 \le t \le 2$$

represents a semicircle (half circle) of radius 4 with center 0 traversed in the counterclockwise direction.

Finally

$$z(t) = 2 + 4e^{\pi it/2} \qquad \text{with} \qquad 0 \le t \le 2$$

is a shift of that semicircle to center 2, corresponding to the answer on p. A36 in App. 2 of the textbook.



Sec. 14.1 Prob. 7. Semicircle

Remark. Our solution demonstrates a way of doing mathematics by going from a simpler problem, whose answer we know, to more difficult problems whose answers we infer from the simple problem.

19. Parametric representation. Parabola. We are given that

$$y = 1 - \tfrac{1}{4}x^2 \qquad \text{where} \qquad -2 \le x \le 2.$$

Hence we may set

$$x = t \qquad \text{so that} \qquad y = 1 - \tfrac{1}{4}x^2 = 1 - \tfrac{1}{4}t^2.$$

Now, for $x = t = -2$, we get

$$y = 1 - \tfrac{1}{4}t^2 = 1 - \tfrac{1}{4}(-2)^2 = 0$$

so that

$$z_0 = -2 + 0i.$$

Similarly, $z_1 = 2 + 0i$ and corresponds to $t = 2$. Hence

$$\begin{aligned} z(t) &= x(t) + iy(t) \\ &= t + i\left(1 - \tfrac{1}{4}t^2\right), \qquad (-2 \le t \le 2). \end{aligned}$$

21. Integration. Before we solve the problem we should use the Cauchy–Riemann equations to determine if the integrand $\operatorname{Re} z$ is analytic. The integrand

$$w = u + iv = f(z) = \operatorname{Re} z = x$$

is not analytic. Indeed, the first Cauchy–Riemann equation

$$u_v = v_y \qquad \text{[by (1), p. 625 in Sec. 13.4]}$$

is not satisfied because

$$u_x = 1 \qquad \text{but} \qquad v = 0 \qquad \text{so that } v_y = 0.$$

(The second Cauchy–Riemann equation is satisfied, but, of course, that is not enough for analyticity.) Hence we *cannot* apply the first method (9), p. 647, which would be more convenient, but we must use the second method (10), p. 647.

The shortest path from $z_0 = 1 + i$ to $z_1 = 3 + 3i$ is a straight-line segment with these points as endpoints. Sketch the path. The difference of these points is

$$\text{(A)} \qquad z_1 - z_0 = (3 + 3i) - (1 + i) = 2 + 2i.$$

We set

$$z(t) = z_0 + (z_1 - z_0)t. \tag{B}$$

Then, by taking the values $t = 0$ and $t = 1$, we have

$$z(0) = z_0 \qquad \text{and} \qquad z(1) = z_1$$

because z_0 cancels when $t = 1$. Hence (B) is a general representation of a segment with given endpoints z_0 and z_1, and t ranging from 0 to 1.

Now we start with Equation (B) and substitute (A) into (B) and, by use of $z_0 = z(0) = 1 + i$, we obtain

$$\begin{aligned} z(t) &= x(t) + iy(t) \\ &= z_0 + (z_1 - z_0)t \\ &= 1 + i + (2 + 2i)t \\ &= 1 + 2t + i(1 + 2t). \end{aligned} \tag{C}$$

We integrate by using (10), p. 647. In (10) we need

$$f(z(t)) = x(t) = 1 + 2t,$$

as well as the derivative of $z(t)$ with respect to t, that is,

$$\dot{z}(t) = \frac{dz}{dt} = 2 + 2i.$$

Both of these expressions are obtained from (C).

We are now ready to integrate. From (10), p. 647, we obtain

$$\begin{aligned} \int_C f(z)\, dz &= \int_a^b f[z(t)]\, \dot{z}(t)\, dt \\ &= \int_0^1 (1 + 2t)(2 + 2i)\, dt \\ &= (2 + 2i) \int_0^1 (1 + 2t)\, dt. \end{aligned}$$

Now

$$\int (1 + 2t)\, dt = \int dt + 2 \int t\, dt = t + 2\frac{t^2}{2} = t + t^2,$$

so that

$$\begin{aligned} \int_C f(z)\, dz &= (2 + 2i)\left(t + t^2\right)\Big|_0^1 \\ &= (2 + 2i)(1 + t^2) \\ &= 2(2 + 2i) \\ &= 4 + 4i, \end{aligned}$$

which is the final answer on p. A37 in App. 2 of the textbook (with a somewhat different parametrization).

23. Integration by the first method (Theorem 1, p. 647). From (3), p. 630 of Sec. 13.5 of the text, we know that e^z is analytic. Hence we use indefinite integration and substitution of upper and lower limits. We have

$$\int e^z \, dz = e^z + \text{const} \qquad \text{[by (2), p. 630].}$$

$$\text{(I1)} \qquad \int_{\pi i}^{2\pi i} e^z \, dz = \Big[e^z\Big]_{\pi i}^{2\pi i} = e^{2\pi i} - e^{\pi i}.$$

Euler's formula (5), p. 634, states that

$$e^{iz} = \cos z + i \sin z.$$

Hence

$$\begin{aligned} e^{2\pi i} &= \cos 2\pi + i \sin 2\pi = 1 + i \cdot 0 = 1, \\ e^{\pi i} &= \cos \pi + i \sin \pi = -1 + 0 = -1. \end{aligned}$$

Hence the integral (I1) evaluates to $1 - (-1) = 2$.

27. Integration by the first method (Theorem 1, p. 647). The integrand $\sec^2 z$ is analytic except at the points where $\cos z$ is 0 [see Example 2(b), pp. 634–635 of the textbook]. Since

$$(\tan z)' = \sec^2 z \qquad \text{[by (4), p. 634],}$$

$$\text{(I2)} \qquad \int_{\pi/4}^{\pi i/4} \sec^2 z \, dz = \Big[\tan z\Big]_{\pi/4}^{\pi i/4} = \tan \frac{1}{4}\pi i - \tan \frac{1}{4}\pi.$$

This can be simplified because

$$\tan \frac{1}{4}\pi = \frac{\sin \frac{1}{4}\pi}{\cos \frac{1}{4}\pi} = \frac{\frac{1}{\sqrt{2}}}{\frac{1}{\sqrt{2}}} = 1.$$

Also

$$\tan \frac{1}{4}\pi i = \frac{\sin \frac{1}{4}\pi i}{\cos \frac{1}{4}\pi i} = \frac{i \, \sinh \frac{1}{4}\pi}{\cosh \frac{1}{4}\pi} = i \, \tanh \frac{1}{4}\pi,$$

since, by (15), p. 635 of Sec. 13.6 of textbook,

$$\sin iz = i \sinh z$$

and

$$\cos iz = \cosh z$$

with $z = \frac{1}{4}\pi$.

A numeric value to six significant digits of the desired *real* hyperbolic tangent is 0.655794. Hence (I2) evaluates to

$$i\ \tanh \tfrac{1}{4}\pi - 1 = 0.655794i - 1.$$

Remember that the real hyperbolic tangent varies between -1 and 1, as can be inferred from the behavior of the curves of $\sinh x$ and $\cosh x$ in Fig. 551 and confirmed in Fig. 552, p. A65 (in Part A3.1 of App. 3 of the textbook).

Sec. 14.2 Cauchy's Integral Theorem

Cauchy's integral theorem, p. 653, is the most important theorem in the whole chapter. It states that the integral around a simple closed path (a *contour integral*) is zero, provided the integrand is an analytic function. Expressing this in a formula

$$(1) \qquad \oint_C f(z)\, dz = 0 \qquad \text{where } C \text{ is a simple closed path}$$

and C lives in a complex domain D that is simply connected. The little circle on the integral sign $\oint$ marks a contour integral.

Take a look at **Fig. 345**, p. 652, for the meaning of simple closed path and **Fig. 346**, p. 653, for a simply connected domain. In its basic form, Theorem 1 (Cauchy's integral theorem) requires that the path not touch itself (a circle, an ellipse, a rectangle, etc., but not a figure 8) and lies inside a domain D that has no holes (see Fig. 347, p. 653).

You have to memorize Cauchy's integral theorem. Not only is this theorem important by itself, as a main instrument of complex integration, it also has important implications explored further in this section as well as in Secs. 14.3 and 14.4.

Other highlights in Sec. 14.2 are path independence (Theorem 2, p. 655), **deformation of path** (p. 656, Example 6, Prob. 11), and extending Cauchy's theorem to multiply connected domains (pp. 658–659). We show where we *can* use Cauchy's integral theorem (**Examples 1** and **2**, p. 653, **Probs. 9** and **13**) and where we *cannot* use the theorem (**Examples 3** and **5**, pp. 653–654, **Probs. 11** and **23**). Often the decision hinges on the location of the points at which the integrand $f(z)$ is not analytic. If the points lie inside C (Prob. 23) then we cannot use Theorem 1 but use integration methods of Sec. 14.1. If the points lie outside C (Prob. 13) we can use Theorem 1.

Problem Set 14.2. Page 659

3. **Deformation of path.** In Example 4, p. 654, the integrand is not analytic at $z = 0$, but it is everywhere else. Hence we can deform the contour (the unit circle) into any contour that contains $z = 0$ in its interior. The contour (the square) in Prob. 1 is of this type. Hence the answer is yes.

9. **Cauchy's integral theorem is applicable** since $f(z) = e^{-z^2}$ is analytic for all z, and thus entire (see p. 630 in Sec. 13.5 of the textbook). Hence, by Cauchy's theorem (Theorem 1, p. 653),

$$\oint_C e^{-z^2} dz = 0 \qquad \text{with } C \text{ unit circle, counterclockwise.}$$

More generally, the integral is 0 around *any* closed path of integration.

11. **Cauchy's integral theorem (Theorem 1, p. 653) is not applicable. Deformation of path.** We see that $2z - 1 = 0$ at $z = \frac{1}{2}$. Hence, at this point, the function

$$f(z) = \frac{1}{2z - 1}$$

is not analytic. Since $z = \frac{1}{2}$ lies inside the contour of integration (the unit circle), Cauchy's theorem is not applicable. Hence we have to integrate by the use of path. However, we can choose a most convenient path by applying the principle of deformation of path, described on p. 656 of the textbook. This allows us to move the given unit circle e^{it} by $\frac{1}{2}$. We obtain the path C given by

$$z(t) = \tfrac{1}{2} + e^{it} \qquad \text{where} \qquad 0 \le t \le 2\pi.$$

Note that t is traversed counterclockwise as t increases from 0 to 2π, as required in the problem. Then

$$f(z) = f(z(t)) = \frac{1}{2z(t) - 1} = \frac{1}{2 \cdot \left(\frac{1}{2} + e^{it}\right) - 1} = \frac{1}{2e^{it}}.$$

Differentiation gives

$$\dot{z}(t) = ie^{it}, \qquad \text{(chain rule!)}.$$

Using the second evaluation method (Theorem 2, p. 647, of Sec. 14.1) we get

$$\begin{aligned}
\int_C f(x)\,dz &= \int_a^b f[z(t)]\dot{z}(t)\,dt \qquad \text{[by (10), p. 647]}\\
&= \int_0^{2\pi} \frac{1}{2e^{it}}\, ie^{it}\,dt\\
&= i\int_0^{2\pi} \frac{e^{it}}{2e^{it}}\,dt\\
&= i\int_0^{2\pi} \frac{1}{2}\,dt\\
&= i\left[\frac{t}{2}\right]_0^{2\pi}\\
&= \pi i.
\end{aligned}$$

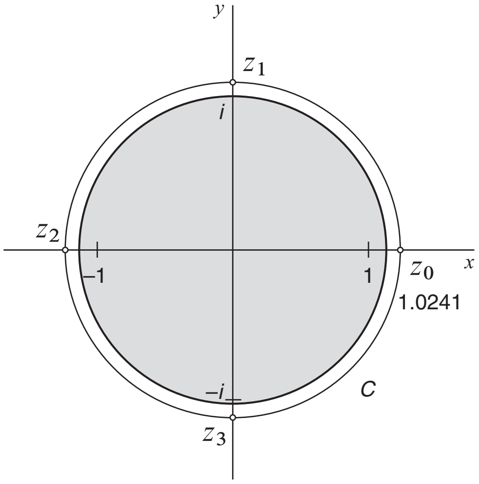
Note that the answer also follows directly from (3), p. 656, with $m = -1$ and $z_0 = \frac{1}{2}$.

13. Nonanalytic outside the contour. To solve the problem, we consider $z^4 - 1.1 = 0$, so that $z^4 = 1.1$. By (15), p. 617 of Sec. 13.2,

$$\sqrt[4]{z} = r\left(\cos\frac{\theta + 2k\pi}{4} + i\sin\frac{\theta + 2k\pi}{4}\right), \qquad k = 0, 1, 2, 3,$$

where $r = \sqrt[4]{1.1} = 1.0241$ and the four roots are

$$\begin{aligned}
z_0 &= \sqrt[4]{1.1}(\cos 0 + i\sin 0) &&= \sqrt[4]{1.1},\\
z_1 &= \sqrt[4]{1.1}\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right) &&= \sqrt[4]{1.1}\cdot i,\\
z_2 &= \sqrt[4]{1.1}\,(\cos\pi + i\sin\pi) &&= -\sqrt[4]{1.1},\\
z_3 &= \sqrt[4]{1.1}\left(\cos\frac{3\pi}{2} + i\sin\frac{3\pi}{2}\right) &&= -\sqrt[4]{1.1}\cdot i.
\end{aligned}$$



Sec. 14.2 Prob. 13. Area of integration C versus location of roots z_0, z_1, z_2, z_3 of denominator of integrand

Since z_0, z_1, z_2, z_3 all lie on the circle with center $(0, 0)$ and radius $r = \sqrt[4]{1.1} = 1.0241 > 1$, they are *outside* the given unit circle C. Hence $f(z)$ is analytic on and inside the unit circle C. Hence Cauchy's integral theorem applies and gives us

$$\oint_C f(z)\, dz = \oint_C \frac{1}{z^4 - 1.1}\, dz = 0.$$

23. Contour integration. We want to evaluate the contour integral

$$\oint_C \frac{2z-1}{z^2-z}\, dz \qquad \text{where } C \text{ as given in the accompanying figure on p. 659.}$$

We use partial fractions (given hint) on the integrand. We note that the denominator of the integrand factors into $z^2 - z = z(z-1)$ so that we write

$$\frac{2z-1}{z^2-z} = \frac{A}{z} + \frac{B}{z-1}.$$

Multiplying the expression by z and then substituting $z = 0$ gives the value for A:

$$\frac{2z-1}{z-1} = A + \frac{Bz}{z-1}, \qquad \frac{-1}{-1} = A + 0, \qquad \boxed{A = 1}.$$

Similarly, multiplying $z - 1$ and then substituting $z = 1$, gives the value for B:

$$\frac{2z-1}{z} = \frac{A(z-1)}{z} + B, \qquad \frac{1}{1} = 0 + B, \qquad \boxed{B = 1}.$$

Hence

$$\frac{2z-1}{z(z-1)} = \frac{1}{z} + \frac{1}{z-1}.$$

The integrand is not analytic at $z = 0$ and $z = 1$, which clearly lie inside C. Hence Cauchy's integral theorem, p. 653, does not apply. Instead we use (3), p. 656, with $m = -1$ for the two integrands obtained by partial fractions. Note that $z_0 = 0$, in the first integral, and then $z_0 = 1$ in the second. Hence we get

$$\oint_C \frac{2z-1}{z^2-z}\,dz = \oint_C \frac{1}{z}\,dz + \oint_C \frac{1}{z-1}\,dz = 2\pi i + 2\pi i = 4\pi i.$$

Sec. 14.3 Cauchy's Integral Formula

Cauchy's integral theorem leads to Cauchy's integral formula (p. 660):

$$\text{(1)} \qquad \oint_C \frac{f(z)}{z-z_0}dz = 2\pi i f(z_0).$$

Formula (1) evaluates contour integrals

$$\text{(A)} \qquad \oint_C g(z)\,dz$$

with an integrand

$$g(z) = \frac{f(z)}{z-z_0} \qquad \text{with } f(z) \text{ analytic.}$$

Hence one must first find

$$f(z) = (z-z_0)g(z).$$

For instance, in **Example 1**, p. 661 of the text,

$$g(z) = \frac{e^z}{z-2} \qquad \text{hence} \qquad f(z) = (z-2)g(z) = e^z.$$

The next task consists of identifying where the point z_0 lies with respect to the contour C of integration. If z_0 lies inside C (and the conditions of Theorem 1 are satisfied), then (1) is applied directly (Examples 1 and 2, p. 661). If z_0 lies outside C, then we use Cauchy's integral theorem of Sec. 14.3 (**Prob. 3**). We extend our discussion to several points at which $g(z)$ is not analytic.

Example 3, pp. 661–662, and Probs. **1** and **11** illustrate that the evaluation of (A) depends on the location of the points at which $g(z)$ is not analytic, relative to the contour of the integration. The section ends with multiply connected domains (3), p. 662 (Prob. 19).

Problem Set 14.3. Page 663

1. Contour integration by Cauchy's integral formula (1), p. 660. The contour $|z+1| = 1$ can be written as $|z-(-1)| = 1$. Thus, it is a circle of radius 1 with center -1. The given function to be integrated is

$$g(z) = \frac{z^2}{z^2-1}.$$

Our first task is to find out where $g(z)$ is not analytic. We consider

$$z^2 - 1 = 0 \qquad \text{so that} \qquad z^2 = 1.$$

Hence the points at which $g(z)$ is not analytic are

$$z = 1 \qquad \text{and} \qquad z = -1.$$

Our next task is to find out which of these two values lies inside the contour and make sure that neither of them lies on the contour (a case we would not yet be able to handle). The value $z = 1$ lies outside the circle (contour) and $z = -1$ lies inside the contour. We have

$$g(z) = \frac{z^2}{z^2 - 1} = \frac{z^2}{(z+1)(z-1)}.$$

Also

$$g(z) = \frac{z^2}{z^2 - 1} = \frac{f(z)}{z - z_0} = \frac{f(z)}{z - (-1)}.$$

Together

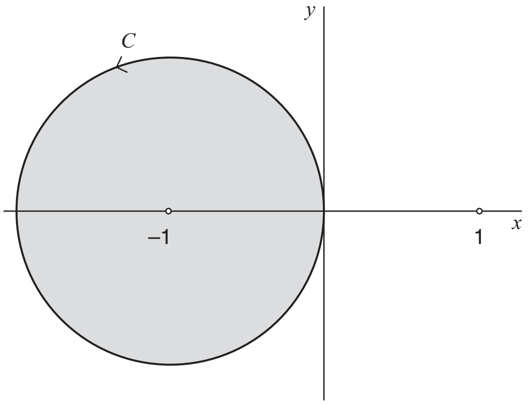
$$\frac{f(z)}{z+1} = \frac{z^2}{(z+1)(z-1)}.$$

Multiplying both sides by $z + 1$ gives

$$f(z) = \frac{z^2}{z-1},$$

which we use for (1), p. 660. Hence

$$\begin{aligned}
\oint_C \frac{z^2}{z^2-1}\,dz &= \oint_C \frac{f(z)}{z - z_0}\,dz && \text{[in the form (1), p. 660]} \\
&= \oint_C \frac{z^2/(z-1)}{z-(-1)}\,dz && \text{[Note } z_0 = -1\text{]} \\
&= 2\pi i\; f(z_0) \\
&= 2\pi i\; f(-1) \\
&= 2\pi i \cdot \frac{1}{-2} \\
&= -\pi i.
\end{aligned}$$



Sec. 14.3 Prob. 1. Contour C of integration

3. Contour integration. Cauchy's integral theorem, p. 653. The contour $C_3 : |\,z + i\,| = 1.4$ is a circle of radius 1.4 and center $z_0 = i$. Just as in **Prob. 1**, we have to see whether the points $z_1 = 1$ and $z_2 = -1$ lie inside the contour C_3. The distance between the points $z_0 = i$ and $z_1 = 1$ is, by (3) and Fig. 324, p. 614 in Sec. 13.2, as follows.

$$|z_0 - \ z_1| = |i - \ 1| = |-1 + i\,| = \sqrt{x^2 + y^2} = \sqrt{(-1)^2 + 1^2} = \sqrt{2} > 1.4.$$

Hence z_1 lies outside the circle C_3.

By symmetry $z_2 = 1$ also lies outside the contour.

Hence $g(z) = z^2/(z - 1)$ is analytic on and inside C_3. We apply Cauchy's integral theorem and get, by (1) on p. 653 in Sec. 14.2,

$$\oint_{C_3} \frac{z^2}{z^2 - 1} dz = 0 \qquad \left[\text{by setting } f(z) = \frac{z^2}{z^2 - 1} \text{ in (1)}\right].$$

11. Contour integral. The contour C is an ellipse with focal points 0 and $2i$. The given integrand is

$$g(z) = \frac{1}{z^2 + 4}.$$

We consider $z^2 + 4 = 0$ so that $z = \pm 2i$. Hence the points at which $g(z)$ is not analytic are $z = 2i$ and $z = -2i$.

To see whether these points lie inside the contour C we calculate for $z = 2i = x + yi$ so that $x = 0$ and $y = 2$ and

$$4x^2 + (y - 2)^2 = 4 \cdot 0^2 + (2 - 2)^2 = 0 < 4,$$

so that $z = 2i$ lies inside the contour. Similarly, $z = -2i$ corresponds to $x = 0$, $y = -2$ and

$$4x^2 + (y - 2)^2 = (-2 - 2)^2 = 16 > 4,$$

so that $z = -2i$ lies outside the ellipse.

We have

$$g(z) = \frac{1}{z^2 + 4} = \frac{f(z)}{z - z_0} = \frac{f(z)}{z - 2i}.$$

Together

$$\frac{f(z)}{z - 2i} = \frac{1}{z^2 + 4} = \frac{1}{(z + 2i)(z - 2i)}$$

where

$$f(z) = \frac{1}{z + 2i}.$$

Cauchy's integral formula gives us

$$\begin{aligned}\oint_C \frac{dz}{z^2+4} dz &= \oint_C \frac{f(z)}{z+z_0} dz \qquad \text{[by (1), p. 660]}\\ &= \oint_C \frac{1/(z+2i)}{z-2i}\, dz\\ &= 2\pi i f(z_0)\\ &= 2\pi i f(2i)\\ &= 2\pi i \frac{1}{2i+2i}\\ &= 2\pi i \frac{1}{4i}\\ &= \frac{1}{2}\pi.\end{aligned}$$

13. Contour integral. We use Cauchy's integral formula. The integral is of the form (1), p. 660, with $z - z_0 = z - 2$, hence $z_0 = 2$. Also, $f(z) = z + 2$ is analytic, so that we can use (1) and calculate

$$2\pi i f(2) = 8\pi i.$$

19. Annulus. We have to find the points in the annulus $1 < |\, z \,| < 3$ at which

$$g(z) = \frac{e^{z^2}}{z^2(z-1-i)} = \frac{e^{z^2}}{z^2[z-(1+i)]}$$

is not analytic. We see that $z = 1 + i$ is such a point in the annulus. Another point is $z = 0$, but this is not in the annulus, that is, not between the circles, but in the "hole." Hence we calculate

$$f(z) = [z-(1+i)]\, g(z) = \frac{e^{z^2}}{z^2}.$$

We evaluate it at $z = 1 + i$ and also note that

$$\text{(C)} \qquad z^2 = (1+i)^2 = 2i.$$

We obtain by Cauchy's integral formula, p. 660,

$$\begin{aligned}2\pi i f(1+i) &= 2\pi i \frac{e^{(1+i)^2}}{2i}\\ &= \pi e^{(1+i)^2}\\ &= \pi e^{2i} && \text{[by (C)]}\\ &= \pi(\cos 2 + i \sin 2) && \text{[by Euler's formula].}\end{aligned}$$

A numeric value is

$$\pi(-0.416147 + 0.909297i) = -1.30736 + 2.85664i.$$

Sec. 14.4 Derivatives of Analytic Functions

The main formula is (1), p. 664. It shows the surprising fact that complex analytic functions have derivatives of all orders. Be aware that, in the formula, the power in the denominator is one degree higher $(n+1)$ than the order of differentiation (n).

Problem Set 14.4. Page 667

1. Contour integration. Use of a third derivative. Using (1), p. 664, we see that the given function is

$$\frac{\sin z}{z^4} = \frac{f(z)}{(z-z_0)^{n+1}} \qquad \text{with} \qquad f(z) = \sin z; \qquad z_0 = 0 \qquad \text{and} \qquad n+1 = 4.$$

Thus $n = 3$. By Theorem 1, p. 664, we have

$$\text{(A)} \qquad \oint_C \frac{f(z)}{(z-z_0)^4}\,dz = \frac{2\pi i}{3!} f^{(3)}(z_0).$$

Since $f(z) = \sin z$, $f'(z) = \cos z$, $f''(z) = -\sin z$, so that

$$f^{(3)} = (-\sin z)' = -\cos z.$$

Furthermore $z_0 = 0$ and

$$f^{(3)}(z_0) = -\cos(0) = -1.$$

Hence, by (A), we get the answer that

$$\begin{aligned}\oint_C \frac{\sin z}{z^4}\,dz &= \frac{2\pi i}{3!}(-1)\\ &= -\frac{2}{3\cdot 2\cdot 1}\pi i\\ &= -\frac{1}{3}\pi i.\end{aligned}$$

5. Contour integration. This is similar to **Prob. 1**. Here the denominator of the function to be integrated is $\left(z-\frac{1}{2}\right)^4$; and $\left(z-\frac{1}{2}\right)^4 = 0$ gives $z_0 = \frac{1}{2}$ which lies inside the unit circle. To use Theorem 1, p. 664, we need the third derivative of $\cosh 2z$. We have, by the chain rule,

$$\begin{aligned}f(z) &= \cosh 2z\\ f'(z) &= 2\sinh 2z\\ f''(z) &= 4\sinh 2z\\ f^{(3)}(z) &= 8\sinh 2z.\end{aligned}$$

We evaluate the last equality at $z_0 = \frac{1}{2}$ and get

$$\begin{aligned} f^{(3)}\left(\frac{1}{2}\right) &= 8\sinh\left(2 \cdot \frac{1}{2}\right) \\ &= 8\sinh 1 \\ &= 8 \cdot \frac{1}{2}\left(e^1 - e^{-1}\right) \qquad \text{[by (17), p. A65 of Sec. A3.1 in App. 3]} \\ &= 4\left(e - \frac{1}{e}\right) \\ &= 9.40161. \end{aligned}$$

Thus

$$\begin{aligned} \oint_C \frac{\cosh 2z}{\left(z - \frac{1}{z}\right)^4}\, dz &= \frac{2\pi i}{3!} \cdot 9.40161 \\ &= \frac{1}{3}\pi i \cdot 9.40161 \\ &= 3.13387 \cdot \pi \cdot i \\ &= 9.84534 i. \end{aligned}$$

9. First derivative. We have to solve

$$\oint_C \frac{\tan \pi z}{z^2} dz, \qquad \text{with } C \text{ the ellipse } 16x^2 + y^2 = 1 \text{ traversed clockwise.}$$

The first derivative will occur because the given function is $(\tan \pi z)/z^2$. Now

$$\tan \pi z = \frac{\sin \pi z}{\cos \pi z} \qquad \text{is not analytic at the points} \qquad \pm\,(2n+1)\pi/2.$$

But all these infinitely many points lie outside the ellipse

$$\frac{x^2}{(\frac{1}{4})^2} + y^2 = 1$$

whose semiaxes are $\frac{1}{4}$ and 1. In addition,

$$\frac{\tan \pi z}{z^2} \qquad \text{is not analytic at} \qquad z = z_0 = 0,$$

where it is of the form of the integrand in (1′), p. 664. Accordingly, we calculate

$$f(z) = z^2 g(z) = \tan \pi z$$

and the derivative (chain rule)

$$f'(z) = \frac{\pi}{\cos^2 \pi z}.$$

Hence (1), p. 664, gives you the value of the integral in the *counterclockwise direction*, that is,

$$2\pi i f'(0) = \frac{2\pi i \cdot \pi}{1} = 2\pi^2 i. \tag{B}$$

Since the contour is to be traversed in the ***clockwise*** *direction*, we obtain a minus sign in result (B) and get the final answer $-2\pi^2 i$.

13. First derivative. Logarithm. The question asks us to evaluate

$$\oint_C \frac{\operatorname{Ln} z}{(z-2)^2} dz, \qquad C\ : |z-3| = 2 \text{ traversed counterclockwise.}$$

We see that the given integrand is $\operatorname{Ln}(z)/(z-2)^2$ and the contour of integration is a circle of radius 2 with center 3. At 0 and on the ray of the real axis, the function $\operatorname{Ln} z$ is not analytic, and it is essential that these points lie outside the contour. Otherwise, that is, if that ray intersected or touched the contour, we would not be able to integrate. Fortunately, in our problem, the circle is always to the right of these points.

In view of the fact that the integrand is not analytic at $z = z_0 = 2$, which lies inside the contour, then, according to (1), p. 664, with $n + 1 = 2$, hence $n = 1$, and $z_0 = 2$, the integral equals $2\pi i$ times the value of the first derivative of $\operatorname{Ln} z$ evaluated at at $z_0 = 2$. We have the derivative of $\operatorname{Ln} z$ is

$$(\operatorname{Ln} z)' = \frac{1}{z}$$

which, evaluated at $z = z_0 = 2$, is $\frac{1}{2}$. This gives a factor $\frac{1}{2}$ to the result, so that the final answer is

$$\tfrac{1}{2} \cdot 2\pi i = \pi i.$$

Chap. 15 Power Series, Taylor Series

We shift our studies from complex functions to *power series* of complex functions, marking the beginning of another **distinct** approach to complex integration. It is called "residue integration" and relies on *generalized* Taylor series—topics to be covered in Chap. 16. However, to properly understand these topics, we have to start with the topics of power series and Taylor series, which are the themes of Chap. 15.

The ***second approach*** *to complex integration based on residues* owes gratitude to Weierstrass (see footnote 5, p. 703 in the textbook), Riemann (see footnote 4, p. 625 in Sec. 13.4), and others. Weierstrass, in particular, championed the use of power series in complex analysis and left a distinct mark on the field through teaching it to his students (who took good lecture notes for posterity; indeed we own such a handwritten copy) and his relatively few but important publications during his lifetime. (His collected work is much larger as it also contains unpublished material.)

The two approaches of complex integration coexist and should not be a source of confusion. (For more on this topic turn to p. x of the Preface of the textbook and read the first paragraph.)

We start with convergence tests for complex series, which are quite similar to those for real series. Indeed, if you have a good understanding of real series, Sec. 15.1 may be a review and you could move on to the next section on power series and their **radius of convergence**. We learn that complex power series represent analytic functions (Sec. 15.3) and that, conversely, every analytic function can be represented by a power series in terms of a (complex) **Taylor series** (Sec. 15.4). Moreover, we can generate new power series from old power series (of analytic functions) by termwise differentiation and termwise integration. We conclude our study with uniform convergence.

From calculus, you want to review sequences and series and their convergence tests. You should remember **analytic functions** and **Cauchy's integral formula** (1), p. 660 in Sec. 14.3. A knowledge of how to calculate real Taylor series is helpful for Sec. 15.4. The material is quite hands-on in that you will construct power series and calculate their radii of convergence.

Sec. 15.1 Sequences, Series, Convergence Tests

This is similar to sequences and series in real calculus. Before you go on—*test your knowledge of real series and answer the following questions*: What is the harmonic series? Does it converge or diverge? Can you show that your answer is correct? Close the book and work on the problem. Compare your answer with the answer on p. 314 at the end of this chapter in this Manual. If you got a correct answer, great! If not, then you should definitely study Sec. 15.1 in the textbook.

Most important, from a practical point of view, is the **ratio test** (see Theorem 7, p. 676 and Theorem 8, p. 677).

The harmonic series is used in the proof of Theorem 8 (p. 677) and in the Caution after Theorem 3, p. 674. One difference between calculus and complex analysis is Theorem 2, p. 674, which treats the convergence of a complex series as the convergence of its real part and its complex part, respectively.

Problem Set 15.1. Page 679

3. Sequence. The sequence to be characterized is

$$z_n = \frac{n\pi}{4 + 2ni}.$$

First solution method:

$$\begin{aligned}
z_n &= \frac{n\pi}{4+2ni} \\
&= \frac{n\pi}{4+2ni} \cdot \frac{4-2ni}{4-2ni} \qquad \text{[by (7), p. 610 of Sec. 13.1]} \\
&= \frac{n\pi(4-2ni)}{4^2+(2n)^2} \\
&= \frac{4n\pi}{4^2+4n^2} + i\left(-\frac{n^2\pi}{8+2n^2}\right).
\end{aligned}$$

We have just written z_n in the form

$$z_n = x_n + iy_n.$$

By Theorem 1, p. 672, we treat each of the sequences $\{x_n\}$ and $\{y_n\}$ separately when characterizing the behavior of $\{z_n\}$. Thus

$$\begin{aligned}
\lim_{n\to\infty} x_n &= \lim_{n\to\infty} \frac{4n\pi}{4+4n^2} \\
&= \lim_{n\to\infty} \frac{\frac{n\pi}{n^2}}{\frac{1+n^2}{n^2}} \qquad \text{(divide numerator and denominator by } 4n^2\text{)} \\
&= \lim_{n\to\infty} \frac{\frac{n}{\pi}}{\frac{1}{n^2}+1} \\
&= \frac{\lim_{n\to\infty} \frac{\pi}{n}}{\lim_{n\to\infty}\left(\frac{1}{n^2}\right) + \lim_{n\to\infty} 1} \\
&= \frac{0}{0+1} = 0.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\lim_{n\to\infty} y_n &= \lim_{n\to\infty}\left(-\frac{n^2\pi}{8+2n^2}\right) \\
&= -\lim_{n\to\infty} \frac{\frac{n^2\pi}{n^2}}{\frac{8+2n^2}{n^2}} \\
&= -\lim_{n\to\infty} \frac{\pi}{\frac{8}{n^2}+2} \\
&= -\frac{\pi}{0+2} = -\frac{\pi}{2}.
\end{aligned}$$

Hence the sequence converges to

$$0 + i\left(-\frac{\pi}{2}\right) = -\frac{1}{2}\pi i.$$

Second solution method (as given on p. A38):

$$\begin{aligned} z_n &= \frac{n\pi}{4+2ni} \\ &= \frac{\frac{n\pi}{2ni}}{\frac{4+2ni}{2ni}} \qquad \text{(division of numerator and denominator by } 2ni\text{)} \\ &= \frac{\frac{\pi}{2i}}{\frac{2}{ni}+1} = \frac{\frac{\pi}{2}\cdot\frac{1}{i}}{\frac{2}{ni}+1} = \frac{\frac{\pi}{2}(-i)}{\frac{2}{ni}+1} = \frac{-\frac{1}{2}\pi i}{1+\frac{2}{ni}}. \end{aligned}$$

Now

$$\lim_{n\to\infty} z_n = \lim_{n\to\infty} \frac{-\frac{1}{2}\pi i}{1+\frac{2}{ni}} = \frac{\lim\limits_{n\to\infty}\left(-\frac{1}{2}\pi i\right)}{\lim\limits_{n\to\infty} 1 + \lim\limits_{n\to\infty}\frac{2}{ni}} = \frac{-\frac{1}{2}\pi i}{1+0} = -\frac{1}{2}\pi i.$$

Since the sequence converges it is also bounded.

5. **Sequence.** The terms $z_n = (-1)^n + 10i, \quad n = 1, 2, 3, \cdots$, are

$$z_1 = -1 + 10i, \qquad z_2 = 1 + 10i, \qquad z_3 = -1 + 10i, \qquad z_4 = 1 + 10i, \cdots .$$

The sequence is bounded because

$$\begin{aligned} |z_n| &= |(-1)^n + 10i| \\ &= \sqrt{[(-1)^n]^2 + 10^2} \\ &= \sqrt{1+100} \\ &= \sqrt{101} \\ &< 11. \end{aligned}$$

For odd subscripts the terms are $-1 + 10i$ and for even subscripts $1 + 10i$. The sequence has two limit points $-1 + 10i$ and $1 + 10i$, but, by definition of convergence (p. 672), it can only have one. Hence the sequence $\{z_n\}$ diverges.

9. **Sequence.** Calculate

$$\begin{aligned} |z_n| &= |0.9 + 0.1i|^{2n} \\ &= (|0.9 + 0.1i|^2)^n \\ &= (0.81 + 0.01)^n \\ &= 0.82^n \to 0 \qquad \text{as} \qquad n \to 0. \end{aligned}$$

Conclude that the sequence converges absolutely to 0.

13. **Bounded complex sequence.** To verify the claim of this problem, we first have to show that:

(i) If a complex sequence is bounded, then the two corresponding sequences of real parts and imaginary parts are also bounded.

Proof of (*i*). Let $\{z_n\}$ be an arbitrary complex sequence that is bounded. This means that there is a constant K such that

$$|z_n| < K \qquad \text{for all } n \text{ (i.e., all terms of the sequence).}$$

Set

$$z_n = x_n + i\, y_n$$

as on p. 672 of the text. Then

$$|z_n| = \sqrt{x_n^2 + y_n^2} \qquad \text{[by (3), p. 613 of Sec. 13.2]}$$

and

$$|z_n|^2 = x_n^2 + y_n^2.$$

Now

$$x_n^2 \le x_n^2 + y_n^2 = |z_n|^2 \qquad \text{since } x_n^2 \ge 0,\ y_n^2 \ge 0.$$

Furthermore,

$$x_n^2 = |x_n|^2 \qquad \text{since } x_n^2 \ge 0.$$

Thus

$$\begin{aligned} |x_n|^2 &\le |z_n|^2 \\ |x_n| &\le |z_n| \end{aligned}$$

so that

$$|x_n| < K.$$

Similarly,

$$|y_n|^2 \le y_n^2 \le |z_n|^2 < K^2$$

so that

$$|y_n| < K.$$

Since n was arbitrary, we have shown that $\{x_n\}$ and $\{y_n\}$ are bounded by some constant K.

Next we have to show that:

(ii) If the two sequences of real parts and imaginary parts are bounded, then the complex sequence is also bounded.

Proof of (*ii*). Let $\{x_n\}$ and $\{y_n\}$ be bounded sequences of the real parts and imaginary parts, respectively. This means that there is a constant L such that

$$|\, x_n \,| < \frac{L}{\sqrt{2}}, \qquad |y_n \,| < \frac{L}{\sqrt{2}}.$$

Then

$$|\, x_n \,|^2 < \frac{L^2}{2}, \qquad |\, y_n \,|^2 < \frac{L^2}{2},$$

so that

$$\begin{aligned} |\, z_n \,|^2 &= x_n^2 + y_n^2 \\ &< \frac{L^2}{2} + \frac{L^2}{2} \\ &< L^2. \end{aligned}$$

Hence $\{z_n\}$ is bounded.

19. Series convergent? Comparison test.

$$
\begin{aligned}
|z_n| &= \left|\frac{i^n}{n^2 - i}\right| \\
&= \frac{|i^n|}{|n^2 - i|} && \text{[by (10), p. 615 in Sec. 13.2]} \\
&= \frac{|i|^n}{|n^2 - i|} \\
&= \frac{1}{\sqrt{n^4 + 1}} && \text{[by (3), p. 613 in Sec. 13.2]} \\
&< \frac{1}{\sqrt{n^4}} = \frac{1}{n^2}.
\end{aligned}
$$

Since

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \ \text{converges} \qquad \text{[see p. 677 in the Proof of (c) of Theorem 8],}$$

we conclude, by the comparison test, p. 675, that the series given in this problem also converges.

23. Series convergent? Ratio test. We apply Theorem 8, p. 677. First we form the ratio z_{n+1}/z_n and simplify algebraically. Since

$$z_n = \frac{(-1)^n \, (1+i)^{2n+1}}{(2n)!},$$

the test ratio is

$$
\begin{aligned}
\frac{z_{n+1}}{z_n} &= \frac{(-1)^{n+1}(1+i)^{2(n+1)+1} \,/\, (2(n+1))!}{(-1)^n(1+i)^{2n+1} \,/\, (2n)!} \\
&= \frac{(-1)^{n+1}(1+i)^{2n+3}}{(2(n+1))!} \cdot \frac{(2n)!}{(-1)^n(1+i)^{2n+1}} \\
&= (-1)\frac{(1+i)^2}{(2n+2)!} \cdot \frac{(2n)!}{1} \\
&= (-1)\frac{(1+i)^2}{(2n+2)(2n+1)} \\
&= (-1)\frac{(2i)}{(2n+2)(2n+1)}\,.
\end{aligned}
$$

Then we take the absolute value of the ratio and simplify by (3), p. 613, of Sec. 13.2:

$$
\begin{aligned}
\left|\frac{z_{n+1}}{z_n}\right| &= \left|(-1)\frac{2i}{(2n+2)(2n+1)}\right| \\
&= \frac{1}{(n+1)(2n+1)}\,.
\end{aligned}
$$

Hence

$$\lim_{n\to\infty} \left|\frac{z_{n+1}}{z_n}\right| = \frac{1}{(n+1)(2n+1)} = L = 0.$$

because

$$\lim_{n\to\infty}\left(\frac{1}{(n+1)(2n+1)}\right) = \lim_{n\to\infty}\left(\frac{1}{n+2}\right)\cdot\lim_{n\to\infty}\left(\frac{1}{2n+1}\right) = 0\cdot 0 = 0$$

Thus, by the ratio test (Theorem 8), the series converges absolutely and hence converges.

Sec. 15.2 Power Series

Since analytic functions can be represented by infinite **power series** (1), p. 680,

$$a_0 + a_1(z-z_0) + a_2(z-z_0)^2 + \cdots, \tag{1}$$

such series are very important to complex analysis, much more so than in calculus. Here z_0, called the **center** of the series, can take on any complex number (once chosen, it is fixed). When $z_0 = 0$, then we get (2), p. 680. An example is

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \cdots. \tag{E}$$

More on this in Sec. 15.4. We want to know where (1) converges and use the **Cauchy–Hadamard formula** (6), p. 683, in Theorem 2 to determine the **radius of convergence** R, that is,

$$R = \lim_{n\to\infty}\left|\frac{a_n}{a_{n+1}}\right| \qquad \text{[remember that the } (n+1)\text{st term is in the denominator!].} \tag{6}$$

Formula (6) shows that the radius of convergence is the limit of the quotient $|\,a_n/a_{n+1}\,|$ (if it exists). This in turn is the reciprocal of the quotient $L^* = |\,a_{n+1}/a_n\,|$ in the ratio test (Theorem 8, p. 677). This is understandable; if the limit of L^* is small, then its reciprocal, the radius of convergence R, will be large. The following table characterizes (6).

Table. Area of convergence of power series (1)

Value of R	*Area of convergence of series* (1)	*Illustrative examples*
$R = c$ (c a constant: real, positive)	Convergence in disk $\lvert z - z_0\rvert < c$	Ex. 5, p. 683, Prob. 13
$R = \infty$	Convergence everywhere	Ex. 2, p. 680, series (E), Prob. 7
$R = 0$	Convergence only at the center $z = z_0$	Ex. 3, p. 681
Remarks: $R = \infty$ means the function is entire. $R = 0$ is the useless case.		

Problem Set 15.2. Page 684

7. Radius of convergence. The given series

$$\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}\left(z-\frac{1}{2}\pi\right)^{2n}$$

is in powers of $z - \frac{1}{2}\pi$, and its center is $\frac{1}{2}\pi$. We use the Cauchy–Hadamard formula (6), p. 683, to determine the radius of convergence R. We have

$$\frac{a_n}{a_{n+1}} = \frac{(-1)^n}{(2n)!}\cdot\frac{(2(n+1))!}{(-1)^{n+1}} = \frac{(-1)^n}{(-1)^{n+1}}\cdot\frac{(2(n+1))!}{(2n)!}.$$

We simplify the two fractions in the last equality:

$$\frac{(-1)^n}{(-1)^{n+1}} = \frac{(-1)^n}{(-1)^n(-1)} = -1$$

and

$$\frac{(2(n+1))!}{(2n)!} = \frac{(2n+2)!}{(2n)!} = \frac{(2n+2)(2n+1)2n\cdots 1}{2n\cdots 1} = (2n+2)(2n+1).$$

Together, the desired ratio simplifies to

$$\frac{a_n}{a_{n+1}} = -(2n+2)(2n+1),$$

and its absolute value is

$$\left|\frac{a_n}{a_{n+1}}\right| = (2n+2)(2n+1).$$

Now as $n \to \infty$

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

Hence

$$R = \infty.$$

This means that the series converges everywhere, see Example 2, p. 680, and the top of p. 683, of the textbook.

We were fortunate that the radius of convergence was ∞ because our series is of the form

$$\sum_{n=0}^{\infty} a_n z^{2n}.$$

Had R been finite, the radius of convergence would have been $\sqrt{R}$ (see the next problem).

Remark. Plausibility of result. From regular calculus you may recognize that the real series

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}\left(x - \frac{1}{2}\pi\right)^{2n} = \cos\left(x - \frac{1}{2}\pi\right)$$

is the Taylor series for $\cos\left(x - \frac{1}{2}\pi\right)$. The complex analog is $\cos\left(z - \frac{1}{2}\pi\right)$. Since the complex cosine function is an entire function, its has an infinite radius of convergence.

13. Radius of convergence. The given series is

$$\sum_{n=0}^{\infty} 16^n (z+i)^{4n}.$$

Since $z + i = z - (-i)$, the center of the series is $-i$. We can write the series as

$$\sum_{n=0}^{\infty} 16^n (z+i)^{4n} = \sum_{n=0}^{\infty} 16^n \left[(z+i)^4\right]^n = \sum_{n=0}^{\infty} 16^n t^n$$

where

(A) $$t = (z+i)^4.$$

We use the Cauchy–Hadamard formula (6), p. 683, to determine the radius of convergence R_t [where the subscript t refers to the substitution (A)]:

$$\frac{a_n}{a_{n+1}} = \frac{16^n}{16^{n+1}} = \frac{16^n}{16^n \cdot 16} = \frac{1}{16}.$$

Hence by (6), p. 683,

$$R_t = \lim_{n\to\infty} \frac{1}{16} = \frac{1}{16}.$$

This is the radius of convergence of the given series, regarded as a function of t. From (A) we have

$$z + i = t^{1/4}.$$

Hence the radius of convergence R_z, for the given series in z, is

$$R_z = (R_t)^{1/4} = \left(\tfrac{1}{16}\right)^{1/4} = \sqrt[4]{\tfrac{1}{16}} = \tfrac{1}{2}.$$

We denote R_z by R to signify that it is the wanted radius of convergence for the given series. Hence the series converges in the open disk

$$|z - (-i)| < \tfrac{1}{2} \qquad \text{with center } i \qquad \text{and} \qquad \text{radius } R = \tfrac{1}{2}.$$

15. Radius of convergence. Since the given series

$$\sum_{n=0}^{\infty} \frac{(2n)!}{4^n\,(n!)^2}\,(z-2i)^n$$

is in powers of $z - 2i$, its center is $2i$. We use (6), p. 683, to determine R

$$\frac{a_n}{a_{n+1}} = \frac{(2n)!}{4^n\,(n!)^2} \cdot \frac{4^{n+1}((n+1)!)^2}{(2(n+1))!}$$

which groups, conveniently, to

$$= \frac{(2n)!}{(2(n+1))!} \cdot \frac{4^{n+1}}{4^n} \cdot \frac{((n+1)!)^2}{(n!)^2}.$$

To avoid calculation errors, we simplify each fraction separately, that is,

$$\frac{(2n)!}{(2(n+1))!} = \frac{2n(2n-1)\cdots 1}{(2n+2)(2n+1)2n\cdots 1} = \frac{1}{(2n+2)(2n+1)},$$

$$\frac{4^{n+1}}{4^n} = 4,$$

and

$$\frac{((n+1)!)^2}{(n!)^2} = \left(\frac{(n+1)n\cdots 1}{n\cdots 1}\right)^2 = (n+1)^2.$$

Hence, putting the fractions together and further simplification gives us

$$\frac{a_n}{a_{n+1}} = \frac{4(n+1)^2}{(2n+2)(2n+1)} = \frac{4(n+1)(n+1)}{2(n+1)(2n+1)} = \frac{2(n+1)}{2n+1} = \frac{2n+2}{2n+1},$$

so that the final result is

$$\lim_{n\to\infty}\left|\frac{a_n}{a_{n+1}}\right| = \lim_{n\to\infty}\frac{2n+2}{2n+1} = \lim_{n\to\infty}\frac{\frac{2n+2}{n}}{\frac{2n+1}{n}} = \lim_{n\to\infty}\frac{2+\frac{2}{n}}{2+\frac{1}{n}}$$

$$= \frac{\lim\limits_{n\to\infty}\left(2+\frac{2}{n}\right)}{\lim\limits_{n\to\infty}\left(2+\frac{1}{n}\right)} = \frac{2+0}{2+0} = \frac{2}{2} = 1 = R$$

Thus the series converges in the open disk $|z-2i\ | < 1$ of radius $R = 1$ and center $2i$.

Sec. 15.3 Functions Given by Power Series

We now give some theoretical foundations for power series and show how we can develop a new power series from an existing one. This can be done in two ways. We can **differentiate** a power series term by term without changing the radius of convergence (Theorem 3, p. 687, Example 1, p. 688, Prob. 5). Similarly, we can **integrate** (Theorem 4, p. 688, Prob. 9). Most importantly, Theorem 5, p. 688, gives the reason why power series are of central importance in complex analysis since power series are analytic and so are “differentiated” power series (with the radius of convergence preserved).

Problem Set 15.3. Page 689

5. **Radius of convergence by differentiation: Theorem 3, p. 687**. We start with the geometric series

$$\text{(A)} \qquad g(z) = \sum_{n=0}^{\infty}\left(\frac{z-2i}{2}\right)^n = 1 + \frac{z-2i}{2} + \left(\frac{z-2i}{2}\right)^2 + \left(\frac{z-2i}{2}\right)^3 + \cdots .$$

Using Example 1, p. 680, of Sec. 15.2, we know that it converges for

$$\frac{|\,z-2i\,|}{2} < 1 \qquad \text{and thus for} \qquad |\,z-2i\,| < 2.$$

Theorem 3, p. 687, allows us to differentiate the series in (A), termwise, with the radius of convergence preserved. Hence we get

$$\text{(B)} \qquad g'(z) = 0 + \frac{1}{2} + 2\left(\frac{z-2i}{2}\right)\cdot\frac{1}{2} + 3\left(\frac{z-2i}{2}\right)^2\cdot\frac{1}{2} + \cdots$$

$$= \sum_{n=1}^{\infty}\frac{n(z-2i)^{n-1}}{2^n} \qquad \text{where} \qquad |\,z-2i\,| < 2.$$

Note that we sum from $n = 1$ because the term for $n = 0$ is 0.

Applying Theorem 3 to (B) yields

$$\text{(C)} \qquad g''(z) = \sum_{n=2}^{\infty}\frac{n(n-1)(z-2i)^{n-2}}{2^n} \qquad \text{where} \qquad |\,z-2i\,| < 2.$$

From (C) it follows that

$$\text{(D)} \qquad (z-2i)^2 g''(z) = \sum_{n=2}^{\infty}\frac{n(n-1)(z-2i)^n}{2^n}$$

$$= \sum_{n=2}^{\infty} n(n-1)\left(\frac{z-2i}{2}\right)^n \qquad \text{where} \qquad |\,z-2i\,| < 2.$$

But (D) is precisely the given series.

Complete the problem by verifying the result by the Cauchy–Hadamard formula (6), p. 683, in Sec. 15.2.

9. Radius of convergence by integration: Theorem 4, p. 688. We start with the geometric series (see Example 1, p. 680) which has radius of convergence 1:

$$\sum_{n=0}^{\infty} w^n = 1 + w + w^2 + w^3 \cdots \qquad |\,w\,| < 1.$$

Hence,

$$\sum_{n=0}^{\infty} (-2w)^n = 1 - 2w + 4w^2 - 8w^3 \cdots \qquad |\,w\,| < \frac{1}{2},$$

and then,

$$\sum_{n=1}^{\infty} (-2w)^n = -2w + 4w^2 - 8w^3 \cdots \qquad |\,w\,| < \frac{1}{2}.$$

We substitute $w = z^2$ into the last series and get

$$\text{(E)} \qquad \sum_{n=1}^{\infty} (-2)^n\, z^{2n} = -2z^2 + 4z^4 - 8z^6 + - \cdots$$

which converges for

$$|\,z^2\,| < \frac{1}{2} \qquad \text{and hence} \qquad |\,z\,| < \frac{1}{\sqrt{2}}.$$

Our aim is to produce the series given in the problem. We observe that the desired series has factors $n + 2$, $n + 1$, and n in the denominator of its coefficients. This suggests that we should use three integrations to determine the radius of convergence. We use Theorem 4, p. 688, to justify termwise integration. We divide (E) by z

$$-2z + 4z^3 - 8z^5 + - \cdots = \sum_{n=1}^{\infty} (-2)^n\, z^{2n-1}.$$

We integrate termwise (omitting the constants of integration)

$$-2\int z\,dz = -2\frac{z^2}{2}, \qquad 4\int z^3\,dz = 4\frac{z^4}{4}, \qquad -8\int z^5\,dz = -8\frac{z^6}{6}, \qquad \cdots,$$

which is

$$\sum_{n=1}^{\infty} (-2)^n\, \frac{z^{2n}}{2n} \qquad \text{where} \qquad |\,z\,| < \frac{1}{\sqrt{2}}.$$

However, we want to get the factor $1/n$ so we multiply the result by 2, that is,

$$\text{(F)} \qquad 2\sum_{n=1}^{\infty} (-2)^n\, \frac{z^{2n}}{2n} = \sum_{n=1}^{\infty} (-2)^n\, \frac{z^{2n}}{n}.$$

Next we aim for the factor $1/(n + 1)$. We multiply the series obtained in (F) by z

$$\sum_{n=1}^{\infty} (-2)^n\, \frac{z^{2n+1}}{n},$$

and integrate termwise

$$\int (-2)^n \frac{z^{2n+1}}{n}\,dz = \frac{(-2)^n}{n}\int z^{2n+1}\,dz$$
$$= \frac{(-2)^n}{n}\,\frac{z^{2n+1+1}}{2n+1+1},$$

and get the series

$$\sum_{n=1}^{\infty}(-2)^n\,\frac{z^{2n+2}}{2n(n+1)}.$$

We multiply the result by 2 (to obtain precisely the factor $1/n$) and get (G)

$$\text{(G)}\qquad 2\sum_{n=1}^{\infty}(-2)^n\,\frac{z^{2n+2}}{2n(n+1)} = \sum_{n=1}^{\infty}(-2)^n\,\frac{z^{2n+2}}{n(n+1)}.$$

We multiply the right-hand side of (G) by z:

$$\sum_{n=1}^{\infty}(-2)^n\,\frac{z^{2n+3}}{n(n+1)}$$

and integrate

$$\int (-2)^n\,\frac{z^{2n+3}}{n(n+1)}\,dz = \frac{(-2)^n}{n(n+1)}\int z^{2n+3}\,dz = \frac{(-2)^n}{n(n+1)}\,\frac{z^{2n+3+1}}{2n+3+1}.$$

We get

$$\sum_{n=1}^{\infty}(-2)^n\,\frac{z^{2n+4}}{n(n+1)2(n+2)}.$$

We have an unwanted factor 2 in the denominator but only wanted $(n+2)$, so we multiply by 2 and get

$$\sum_{n=1}^{\infty}(-2)^n\,\frac{z^{2n+4}}{n(n+1)(n+2)}.$$

However, our desired series is in powers of z^{2n} instead of z^{2n+4}. Thus we must divide by z^4 and get

$$\text{(H)}\qquad \frac{1}{z^4}\sum_{n=1}^{\infty}(-2)^n\,\frac{z^{2n+4}}{n(n+1)(n+2)} = \sum_{n=1}^{\infty}(-2)^n\,\frac{z^{2n}}{n(n+1)(n+2)}.$$

But this is precisely the desired series. Since our derivation from (E) to (H) did not change the radius of convergence (Theorem 4), we conclude that the series given in this problem has radius of convergence $|z| < 1/\sqrt{2}$, that is, center 0 and radius $1/\sqrt{2}$.

Do part (a) of the problem, that is, obtain the answer by (6), p. 683.

10, 13, 14, 18 *Hint.* For problems 10, 13, 14, and 18, the notation for the coefficients is explained on pp. 1026–1028 of Sec. 24.4, of the textbook.

17. Odd functions. The even-numbered coefficients in (2), p. 685, are zero because $f(-z) = -f(z)$ implies

$$a_{2m}(-z)^{2m} = a_{2m}(-1)^{2m}z^{2m} = a_{2m}\left[(-1)^2\right]^m z^{2m} = a_{2m}1^m z^{2m} = a_{2m}z^{2m} = -a_{2m}z^{2m}$$

But

$$a_{2m}z^{2m} = -a_{2m}z^{2m}$$

means

$$a_{2m} = -a_{2m}$$

so that

$$a_{2m} + a_{2m} = 0 \qquad \text{hence} \qquad a_{2m} = 0,$$

Complete the problem by thinking of examples.

Sec. 15.4 Taylor and Maclaurin Series

Every analytic function $f(z)$ can be represented by a **Taylor series** (Theorem 1, p. 691) and a general way of doing so is given by (1) and (2), p. 690. It would be useful if you knew some Taylor series, such as for e^z [see (12), p. 694], $\sin z$, and $\cos z$ [(14), p. 695]. Also important is the **geometric series** (11) in Example 1 and Prob. 19. The section ends with *practical methods* to develop power series by substitution, integration, geometric series, and binomial series with partial fractions (pp. 695–696, Examples 5–8, Prob. 3).

Example 2, p. 694, shows the Maclaurin series of the exponential function. Using it for defining e^z would have forced us to introduce series rather early. We tried this out several times with student groups of different interests, but found the approach chosen in our book didactically superior.

Problem Set 15.4. Page 697

3. Maclaurin series. Sine function. To obtain the Maclaurin series for $\sin 2z^2$ we start with (14), p. 695, writing t instead of z

$$\sin t = \sum_{n=0}^{\infty}(-1)^n \frac{t^{2n+1}}{(2n+1)!} = t - \frac{t^3}{3!} + \frac{t^5}{5!} - + \cdots .$$

Then we set $t = 2z^2$ and have

$$\begin{aligned}\sin 2z^2 &= \sum_{n=0}^{\infty} (-1)^n \frac{(2z^2)^{2n+1}}{(2n+1)!}\\ &= \sum_{n=0}^{\infty} (-1)^n \frac{2^{2n+1}z^{4n+2}}{(2n+1)!}\\ &= 2z^2 - \frac{2^3z^6}{3!} + \frac{2^5z^{10}}{5!} - + \cdots\\ &= 2z^2 - \frac{4}{3}z^6 + \frac{4}{15}z^{10} - + \cdots .\end{aligned}$$

The center of the series thus obtained is $z_0 = 0$ (i.e., $z = z - z_0 = z - 0$) by definition of Maclaurin series on p. 690. The radius of convergence is $R = \infty$, since the series converges for all z.

15. Higher transcendental functions. Fresnel integral. It is defined by

$$S(z) = \int_0^z \sin t^2\, dt.$$

To find the Maclaurin series of $S(z)$ we start with the Maclaurin series for $\sin w$, and set $w = t^2$. From Prob. 3 of this section we know that

$$\sin t^2 = \sum_{n=0}^{\infty} (-1)^n \frac{t^{4n+2}}{(2n+1)!} = t^2 - \frac{1}{3!}t^6 + \frac{1}{5!}t^{10} - + \cdots .$$

Theorem 4, p. 688, allows us to perform termwise integration of power series. Hence

$$\begin{aligned}\int_0^z \sin t^2 dt &= \int_0^z \sum_{n=0}^{\infty} (-1)^n \frac{t^{4n+2}}{(2n+1)!}\, dt \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{t^{4n+3}}{(2n+1)!\,(4n+3)}\Bigg|_{t=0}^{z} \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{z^{4n+3}}{(2n+1)!\,(4n+3)},\end{aligned}$$

which we obtained by setting $t = z$ as required by the upper limit of integration. The lower limit $t = 0$ contributed 0. Hence

$$S(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{4n+3}}{(2n+1)!\,(4n+3)} = \frac{1}{1!\,3}z^3 - \frac{1}{3!\,7}z^7 + \frac{1}{5!\,11}z^{11} - + \cdots .$$

Since the radius of convergence for the Maclaurin series of the sine function is $R = \infty$, so is R for $S(z)$.

19. Geometric series.
First solution: We want to find the Taylor series of $1/(1-z)$ with center $z_0 = i$. We know that we are dealing with the geometric series

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n \qquad \text{[by (11), p. 694]},$$

with $z_0 = 0$.

Thus consider

$$\frac{1}{1-z} = \frac{1}{1-z-i+i} = \frac{1}{(1-i)-(z-i)}.$$

Next we work on the $1-i$ in the denominator by removing it as a common factor. We get

$$\frac{1}{(1-i)-(z-i)} = \frac{1}{(1-i)\left[1-\frac{z-i}{1-i}\right]} = \frac{1}{1-i}\cdot\frac{1}{1-\left(\frac{z-i}{1-i}\right)}.$$

This looks attractive because

$$\frac{1}{1-\left(\frac{z-i}{1-i}\right)} \qquad \text{is of the form} \quad \frac{1}{1-w} \text{ with } \quad w = \frac{z-i}{1-i}$$

and

$$\frac{1}{1-w} = \sum_{n=0}^{\infty} w^n.$$

Thus we have the desired Taylor series, which is

$$\text{(S)} \qquad \frac{1}{1-i}\sum_{n=0}^{\infty}\left(\frac{z-i}{1-i}\right)^n = \frac{1}{1-i}\sum_{n=0}^{\infty}\frac{1}{(1-i)^n}(z-i)^n.$$

This can be further simplified by noting that

$$\frac{1}{1-i} = \frac{1+i}{2} \qquad \text{[by (7), p. 610]}$$

so that (S) becomes

$$\frac{1+i}{2}\sum_{n=0}^{\infty}\left(\frac{1+i}{2}\right)^n (z-i)^n.$$

This is precisely the answer on p. A39 of the textbook with the terms written out.

The radius of convergence of the series is

$$|w| < 1, \qquad \text{that is,} \qquad \left|\frac{z-i}{1-i}\right| < 1.$$

Now

$$\left|\frac{z-i}{1-i}\right| = \frac{|z-i|}{|1-i|} = \frac{|z-i|}{\sqrt{1+1}}.$$

Hence

$$\frac{|z-i|}{\sqrt{2}} < 1 \qquad \text{and} \qquad |z-i| < \sqrt{2}$$

so that the radius of convergence is $R = \sqrt{2}$.

Second solution: Use Example 7, p. 696 with $c = 1$ and $z_0 = i$.

Remark. The method of applying (1), p. 690, directly is a less attractive way as it involves differentiating functions of the form $1/(1-i)^n$.

21. Taylor series. Sine function. For this problem, we develop the Taylor series directly with (1), p. 690. This is like the method used in regular calculus. We have for $f(z) = \sin z$ and $z_0 = \pi/2$:

$$\begin{aligned}
f(z) &= \sin z & f(z_0) &= \sin\frac{\pi}{2} = 1;\\
f'(z) &= \cos z & f'(z_0) &= \cos\frac{\pi}{2} = 0;\\
f''(z) &= -\sin z & f''(z_0) &= -\sin\frac{\pi}{2} = -1;\\
f'''(z) &= -\cos z & f'''(z_0) &= -\cos\frac{\pi}{2} = 0;\\
f^{(4)}(z) &= \sin z & f^{(4)}(z_0) &= \sin\frac{\pi}{2} = 1;\\
f^{(5)}(z) &= \cos z & f^{(5)}(z_0) &= \cos\frac{\pi}{2} = 0;\\
f^{(6)}(z) &= -\sin z & f^{(6)}(z_0) &= -\sin\frac{\pi}{2} = -1.
\end{aligned}$$

Hence the Taylor series for $\sin z$ with $z_0 = \pi/2$:

$$f(z) = 1 - \frac{1}{2!}\left(z - \frac{\pi}{2}\right)^2 + \frac{1}{4!}\left(z - \frac{\pi}{2}\right)^4 - \frac{1}{6!}\left(z - \frac{\pi}{2}\right)^6 + - \cdots$$

$$= \sum_{n=0}^{\infty} (-1)^n \frac{1}{(2n)!}\left(z - \frac{\pi}{2}\right)^{2n},$$

The radius of convergence is $R = \infty$.

Sec. 15.5 Uniform Convergence. *Optional*

The material in this section is for general information about **uniform convergence** (defined on p. 698) of arbitrary series with variable terms (functions of z). What you should know is the content of Theorem 1, p. 699. Example 4 and Prob. 13 illustrate the **Weierstrass M-test,** p. 703.

Problem Set 15.5. Page 704

3. Power series. By Theorem 1, p. 699, a power series in powers of $z - z_0$ converges uniformly in the closed disk $|z - z_0| \leqq r$, where $r < R$ and R is the radius of convergence of the series. Hence, solving **Probs. 2–9** amounts to determining the radius of convergence.

In Prob. 3 we have a power series in powers of

$$Z = (z + i)^2 \tag{A}$$

of the form

$$\sum_{n=0}^{\infty} a_n Z^n \tag{B}$$

with coefficients $a_n = \frac{1}{3^n}$. Hence the Cauchy–Hadamard formula (6), p. 683 in Sec. 15.2, gives the radius of convergence R^* of this series in Z in the form

$$\frac{a_n}{a_{n+1}} = \frac{3^{-n}}{3^{-(n+1)}} = 3,$$

so the series (B) converges uniformly in every closed disk $|Z| \leqq r^* < R^* = 3$. Substituting (A) and taking square roots, we see that this means uniform convergence of the given power series in powers of $z + i$ in every closed disk:

$$|z + i| \leqq r < R = \sqrt{3}. \tag{C}$$

We can also write this differently by setting

$$\delta = R - r. \tag{D}$$

We know that

$$R > r.$$

Subtracting r on both sides of the inequality gives

$$R - r > r - r$$

and by (D) and simplifying

$$\delta = R - r > r - r = 0 \qquad \text{thus} \qquad \delta > 0.$$

Furthermore, (D) also gives us

$$r = R - \delta.$$

Together,

$$|z + i| \leqq R - \delta = \sqrt{3} - \delta \qquad (\delta > 0).$$

This is the form in which the answer is given on p. A39 in App. 2 of the textbook.

7. Power series. No uniform convergence. We have to calculate the radius of convergence for

$$\sum_{n=1}^{\infty} \frac{n!}{n^2}\left(z + \frac{1}{2}i\right)^n.$$

We want to use the Cauchy–Hadamard formula (6), p. 683 of Sec. 15.2. We start with

$$\frac{a_n}{a_{n+1}} = \frac{n!}{n^2} \cdot \frac{(n+1)^2}{(n+1)!},$$

which is written out

$$= \frac{n \cdots 1}{n^2} \cdot \frac{(n+1)(n+1)}{(n+1)(n \cdots 1)}$$

and, with cancellations, becomes

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

Thus

$$\lim_{n->\infty} \left|\frac{a_n}{a_{n+1}}\right| = \lim_{n->\infty} \frac{n+1}{n^2} = \lim_{n->\infty} \left(\frac{1}{n} + \frac{1}{n^2}\right) = \underbrace{\lim_{n->\infty} \frac{1}{n}}_{0} + \underbrace{\lim_{n->\infty} \frac{1}{n^2}}_{0} = 0.$$

Hence $R = 0$, which means that the given series converges only at the center:

$$z_0 = -\tfrac{1}{2}i.$$

Hence it does not converge uniformly anywhere. Indeed, the result is not surprising since

$$n! >> n^2,$$

and thus the coefficients of the series

$$1, \frac{2}{4}, \frac{6}{9}, \frac{24}{16}, \frac{120}{25}, \frac{720}{36}, \cdots \frac{n!}{n} \to \infty \qquad \text{as} \qquad n \to \infty.$$

13. Uniform convergence. Weierstrass M-test. We want to show that

$$\sum_{n=1}^{\infty} \frac{\sin^n |z|}{n^2}$$

converges uniformly for all z.

Since $|z| = r = \sqrt{x^2 + y^2}$ is a real number, $\sin |z|$ is a real number such that

$$-1 \le \sin |z| \le 1 \qquad \text{and} \qquad -1 \le \sin^m |z| \le 1,$$

where m is a natural number. Hence

$$|\sin |z|\,| \le 1 \qquad \text{and} \qquad |\sin^m |z|\,| \le 1.$$

Now

$$\left|\frac{\sin^m |z|}{m^2}\right| = \frac{|\sin^m |z||}{|m^2|} = \frac{|\sin^m |z|\,|}{m^2} \le \frac{1}{m^2} \qquad \text{for any } z.$$

Since

$$\sum_{m=1}^{\infty} \frac{1}{m^2} \quad \text{converges} \qquad \text{(see Sec. 15.1 in the proof of Theorem 8, p. 677),}$$

we know, by the Weierstrass M-test, p. 703, that the given series converges uniformly.

Solution for the Harmonic Series Problem (see p. 298 of the Student Solutions Manual) The harmonic series is

$$\text{(HS)} \qquad 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots = \sum_{m=1}^{\infty} \frac{1}{m}.$$

The harmonic series **diverges.** One *elementary way* to show this is to consider particular partial sums of the series.

$$\begin{aligned}
s_1 &= 1\\
s_2 &= 1 + \tfrac{1}{2}\\
s_4 &= 1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} = 1 + \tfrac{1}{2} + \underbrace{\left(\tfrac{1}{3} + \tfrac{1}{4}\right)}_{\text{2 terms}}\\
&> 1 + \tfrac{1}{2} + \left(\tfrac{1}{4} + \tfrac{1}{4}\right) = 1 + \tfrac{1}{2} + \tfrac{2}{4} = 1 + 2 \cdot \tfrac{1}{2}\\
s_8 &= 1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8} = 1 + \tfrac{1}{2} + \underbrace{\left(\tfrac{1}{3} + \tfrac{1}{4}\right)}_{\text{2 terms}} + \underbrace{\left(\tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8}\right)}_{\text{4 terms}}\\
&> 1 + \tfrac{1}{2} + \underbrace{\left(\tfrac{1}{4} + \tfrac{1}{4}\right)}_{\text{2 terms}} + \underbrace{\left(\tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8}\right)}_{\text{4 terms}} = 1 + \underbrace{\tfrac{1}{2} + \tfrac{2}{4} + \tfrac{4}{8}}_{\text{3 fractions of value } \frac{1}{2}} = 1 + 3 \cdot \tfrac{1}{2},\\
s_{16} &= 1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8} + \tfrac{1}{9} + \tfrac{1}{10} + \tfrac{1}{11} + \tfrac{1}{12} + \tfrac{1}{13} + \tfrac{1}{14} + \tfrac{1}{15} + \tfrac{1}{16}\\
&= 1 + \tfrac{1}{2} + \underbrace{\left(\tfrac{1}{3} + \tfrac{1}{4}\right)}_{\text{2 terms}} + \underbrace{\left(\tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8}\right)}_{\text{4 terms}} + \underbrace{\left(\tfrac{1}{9} + \tfrac{1}{10} + \tfrac{1}{11} + \tfrac{1}{12} + \tfrac{1}{13} + \tfrac{1}{14} + \tfrac{1}{15} + \tfrac{1}{16}\right)}_{\text{8 terms}}\\
&> 1 + \tfrac{1}{2} + \underbrace{\left(\tfrac{1}{4} + \tfrac{1}{4}\right)}_{\text{2 terms}} + \underbrace{\left(\tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8}\right)}_{\text{4 terms}} + \underbrace{\left(\tfrac{1}{16} + \tfrac{1}{16} + \tfrac{1}{16} + \tfrac{1}{16} + \tfrac{1}{16} + \tfrac{1}{16} + \tfrac{1}{16} + \tfrac{1}{16}\right)}_{\text{8 terms}}\\
&= 1 + \underbrace{\tfrac{1}{2} + \tfrac{2}{4} + \tfrac{4}{8} + \tfrac{8}{16}}_{\text{4 fractions of value } \frac{1}{2}} = 1 + 4 \cdot \tfrac{1}{2},
\end{aligned}$$

$$
\begin{aligned}
s_{32} &= 1 + \underbrace{\tfrac{1}{2} + \cdots + \tfrac{1}{32}}_{\text{31 terms}} = 1 + \tfrac{1}{2} + \left(\tfrac{1}{3} + \tfrac{1}{4}\right) + \underbrace{\left(\tfrac{1}{5} + \cdots + \tfrac{1}{8}\right)}_{\text{4 terms}} + \underbrace{\left(\tfrac{1}{9} + \cdots + \tfrac{1}{16}\right)}_{\text{8 terms}} + \underbrace{\left(\tfrac{1}{17} + \cdots + \tfrac{1}{32}\right)}_{\text{16 terms}} \\
&> 1 + \tfrac{1}{2} + \underbrace{\left(\tfrac{1}{4} + \tfrac{1}{4}\right)}_{\text{2 terms}} + \underbrace{\left(\tfrac{1}{8} + \cdots + \tfrac{1}{8}\right)}_{\text{4 terms}} + \underbrace{\left(\tfrac{1}{16} + \cdots + \tfrac{1}{16}\right)}_{\text{8 terms}} + \underbrace{\left(\tfrac{1}{32} + \cdots + \tfrac{1}{32}\right)}_{\text{16 terms}} \\
&= 1 + \tfrac{1}{2} + \tfrac{2}{4} + 4 \cdot \tfrac{1}{8} + 8 \cdot \tfrac{1}{16} + 16 \cdot \tfrac{1}{32} = 1 + \underbrace{\tfrac{1}{2} + \tfrac{2}{4} + \tfrac{4}{8} + \tfrac{8}{16} + \tfrac{16}{32}}_{\text{5 fractions of value } \frac{1}{2}} = 1 + 5 \cdot \tfrac{1}{2}.
\end{aligned}
$$

Thus in general

$$\boxed{s_{2^n} > 1 + n \cdot \tfrac{1}{2}.}$$

As $n \to \infty$, then

$$1 + n \cdot \tfrac{1}{2} \to \infty \qquad \text{and hence} \qquad s_{2^n} \to \infty.$$

This shows that the sequence of partial sums s_{2^n} is unbounded, and hence the sequence of *all* partial sums of the series is unbounded. Hence, the harmonic series diverges.

Another way to show that the hamonic series diverges is by the *integral test* from calculus (which we can use since $f(x)$ is continuous, positive, and decreasing on the real interval $[1, \infty]$)

$$\text{(A)} \qquad \int_1^\infty \frac{1}{x} dx = \lim_{t\to\infty} \int_1^t \frac{1}{x} dx = \lim_{t\to\infty} [\ln x]_{x=1}^t = \lim_{t\to\infty} \ln t - \underbrace{\ln 1}_{0} = \lim_{t\to\infty} \ln t \to \infty.$$

Since the integral in (A) does not exist (diverges), the related harmonic series (HS) [whose nth term equals $f(n)$] diverges.

Remark. The name *harmonic* comes from overtones in music (harmony!). The harmonic series is so important because, although its terms go to zero as $m \to \infty$, it still diverges. Go back to p. 298.

Chap. 16 Laurent Series. Residue Integration

In Chap. 16, we solve complex integrals over simple closed paths C where the integrand $f(z)$ is analytic *except* at a point z_0 (or at several such points) inside C. In this scenario we cannot use Cauchy's integral theorem (1), p. 653, but need to continue our study of complex series, which we began in Chap. 15. We generalize Taylor series to **Laurent series** which allow such singularities at z_0. Laurent series have both positive and *negative* integer powers and have no significant counterpart in calculus. Their study provides the background theory (Sec. 16.2) needed for these complex integrals with singularities. We shall use **residue integration**, in Sec. 16.3, to solve them. Perhaps most amazing is that we can use residue integration to even solve certain types of **real** definite integrals (Sec. 16.4) that would be difficult to solve with regular calculus. This completes our study of the *second approach to complex integration based on residues* that we began in Chap. 15.

Before you study this chapter you should know analytic functions (p. 625, in Sec. 13.4), Cauchy's integral theorem (p. 653, in Sec. 14.2), power series (Sec. 15.2, pp. 680–685), and Taylor series (1), p. 690. From calculus, you should know how to integrate functions in the complex several times as well as know how to factor quadratic polynomials and check whether their roots lie inside a circle or other simple closed paths.

Sec. 16.1 Laurent Series

Laurent series generalize Taylor series by allowing the development of a function $f(z)$ in powers of $z-z_0$ when $f(z)$ is singular at z_0 (for "singular," see p. 693 of Sec. 15.4 in the textbook). A Laurent series (1), p. 709, consists of positive as well as **negative** integer powers of $z - z_0$ and a constant. The Laurent series converges in an annulus, a circular ring with center z_0 as shown in Fig. 370, p. 709 of the textbook.

The details are given in the important **Theorem 1**, p. 709, and expressed by (1) and (2), which can be written in shortened form (1′) and (2′), p. 710.

Take a look at **Example 4**, p. 713, and **Example 5**, pp. 713–714. A function may have different Laurent series in different annuli with the same center z_0. Of these series, the most important Laurent series is the one that converges directly near the center z_0, at which the given function has a singularity. Its negative powers form the so-called **principal part** of the singularity of $f(z)$ at z_0 (**Example 4** with $z_0 = 0$ and **Probs. 1** and **8**).

Problem Set 16.1. Page 714

Hint. To obtain the Laurent series for probs. 1–8 use either a familiar Maclaurin series of Chap. 15 or a series in powers of $1/z$.

1. Laurent series near a singularity at 0. To solve this problem we start with the Maclaurin series for $\cos z$, that is,

$$\text{(A)} \qquad \cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + - \cdots \qquad \text{[by (14), p. 695].}$$

Next, since we want

$$\frac{\cos z}{z^4},$$

we divide (A) by z^4, that is,

$$\frac{\cos z}{z^4} = \frac{1}{z^4}\left(1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + - \cdots\right)$$
$$= \frac{1}{z^4} - \frac{1}{2z^2} + \frac{1}{24} - \frac{z^2}{720} + - \cdots .$$

The principal part consists of

$$\frac{1}{z^4} - \frac{1}{2z^2}.$$

Furthermore, the series converges for all $z \neq 0$.

7. Laurent series near a singularity at 0. We start with

$$\cosh w = 1 + \frac{w^2}{2!} + \frac{w^4}{4!} + \cdots \qquad \text{[by (15), p. 695].}$$

We set $w = 1/z$. Then

$$\cosh \frac{1}{z} = 1 + \frac{1}{2!}\frac{1}{z^2} + \frac{1}{4!}\frac{1}{z^4} + \cdots \ .$$

Multiplication by z^3 yields

$$\begin{aligned} z^3 \cosh \frac{1}{z} &= z^3 + \frac{1}{2!}\frac{1}{z^2}z^3 + \frac{1}{4!}\frac{1}{z^4}z^3 + \cdots \\ &= z^3 + \frac{1}{2!}z + \frac{1}{4!}\frac{1}{z} + \frac{1}{6!}\frac{1}{z^2}\cdots \\ &= z^3 + \frac{1}{2}z + \frac{1}{24}z^{-1} + \frac{1}{720}z^{-2} + \cdots . \end{aligned}$$

We see that the principal part is

$$\frac{1}{24}z^{-1} + \frac{1}{720}z^{-2} + \cdots .$$

Furthermore, the series converges for all $z \neq 0$, or equivalently the region of convergence is $0 < |z| < \infty$.

15. Laurent series. Singularity at $z_0 = \pi$. We use (6), p. A64, of Sec. A3.1 in App. 3 of the textbook and simplify by noting that $\cos \pi = -1$ and $\sin \pi = 0$:

$$\text{(B)} \qquad \begin{aligned} \cos z &= \cos((z-\pi)+\pi) \\ &= \cos(z-\pi)\cos\pi + \sin(z-\pi)\sin\pi \\ &= -\cos(z-\pi). \end{aligned}$$

Now

$$\cos w = 1 - \frac{w^2}{2!} + \frac{w^4}{4!} - \frac{w^6}{6!} + - \cdots .$$

We set

$$w = z - \pi$$

and get

$$\cos(z-\pi) = 1 - \frac{(z-\pi)^2}{2!} + \frac{(z-\pi)^4}{4!} - \frac{(z-\pi)^6}{6!} + - \cdots .$$

Then

$$-\cos(z-\pi) = -1 + \frac{(z-\pi)^2}{2!} - \frac{(z-\pi)^4}{4!} + \frac{(z-\pi)^6}{6!} - + \cdots.$$

We multiply by

$$\frac{1}{(z-\pi)^2}$$

and get

$$-\frac{\cos(z-\pi)}{(z-\pi)^2} = -\frac{1}{(z-\pi)^2} + \frac{1}{2!} - \frac{(z-\pi)^2}{4!} + \frac{(z-\pi)^4}{6!} - + \cdots.$$

Hence by (B)

$$\frac{\cos z}{(z-\pi)^2} = -(z-\pi)^{-2} + \frac{1}{2} - \frac{1}{24}(z-\pi)^2 + \frac{1}{720}(z-\pi)^4 - + \cdots.$$

The principal part is $-(z-\pi)^{-2}$ and the radius of convergence is $0 < |z-\pi| < \infty$ (converges for all $z \neq \pi$).

19. Taylor and Laurent Series. The geometric series is

$$\frac{1}{1-w} = \sum_{n=0}^{\infty} w^n \qquad |\,w\,| < 1 \qquad \text{[by (11), p. 694].}$$

We need

$$\frac{1}{1-z^2} \qquad \text{so we set} \qquad w = z^2.$$

Then we get the Taylor series

$$\begin{aligned}\frac{1}{1-z^2} &= \sum_{n=0}^{\infty}(z^2)^n \qquad \left|\,z^2\,\right| < 1 \\ &= \sum_{n=0}^{\infty} z^{2n} \qquad \text{or} \qquad \left|\,z^2\,\right| = |z|^2 < 1 \qquad \text{so that } |\,z\,| < 1 \\ &= 1 + z^2 + z^4 + z^6 + \cdots.\end{aligned}$$

Similarly, we obtain the Laurent series converging for $|\,z\,| > 1$ by the following trick, which you should remember:

$$\begin{aligned}\frac{1}{1-z^2} &= \frac{1}{-z^2\left(1-\dfrac{1}{z^2}\right)} = \frac{1}{-z^2}\cdot\frac{1}{1-\left(\dfrac{1}{z}\right)^2} \\ &= \frac{1}{-z^2}\sum_{n=0}^{\infty}\left(\frac{1}{z}\right)^{2n} \\ &= \frac{1}{-z^2}\left(1 + z^{-2} + z^{-4} + z^{-6} + \cdots\right) \\ &= -\frac{1}{z^2} - \frac{1}{z^4} - \frac{1}{z^6} - \frac{1}{z^8} - \cdots \\ &= -\sum_{n=0}^{\infty}\frac{1}{z^{2n+2}} \qquad |\,z\,| > 1.\end{aligned}$$

23. Taylor and Laurent series. We want all Taylor and Laurent series for

$$\frac{z^8}{1-z^4} \qquad \text{with} \qquad z_0 = 0.$$

We start with

$$\frac{1}{1-w} = \sum_{n=0}^{\infty} w^n \qquad |\, w \,| < 1 \qquad \text{[by (11), p. 694].}$$

We set $w = z^4$ and get

$$\frac{1}{1-z^4} = \sum_{n=0}^{\infty} z^{4n} \qquad |\, z \,| < 1.$$

We multiply this by z^8 to obtain the desired *Taylor series:*

$$\begin{aligned} \frac{z^8}{1-z^4} &= z^8 \sum_{n=0}^{\infty} z^{4n} = \sum_{n=0}^{\infty} z^{4n+8} \qquad |\, z \,| < 1 \\ &= z^8 + z^{12} + z^{16} + \cdots . \end{aligned}$$

From Prob. 19 we know that the Laurent series for

$$\frac{1}{1-w^2} = -\sum_{n=0}^{\infty} \frac{1}{w^{2n+2}} \qquad |\, w \,| > 1.$$

We set $w = z^2$

$$\frac{1}{1-z^4} = -\sum_{n=0}^{\infty} \frac{1}{(z^2)^{2n+2}} = -\sum_{n=0}^{\infty} \frac{1}{z^{4n+4}} \qquad \left|\, z^2 \,\right| > 1 \qquad \text{so that } |\, z \,| > 1.$$

Multiply the result by z^8:

$$\frac{z^8}{1-z^4} = -z^8 \sum_{n=0}^{\infty} \frac{1}{z^{4n+4}} = -\sum_{n=0}^{\infty} \frac{z^8}{z^{4n+4}}.$$

Now

$$\frac{z^8}{z^{4n+4}} = z^{8-(4n+4)} = z^{4-4n}.$$

Hence the desired *Laurent series* for

$$\frac{z^8}{1-z^4} \qquad \text{with center} \qquad z_0 = 0$$

is

$$\frac{z^8}{1-z^4} = -\sum_{n=0}^{\infty} z^{4-4n} = -z^4 - 1 - z^{-4} - z^{-8} - \cdots$$

so that the principal part is

$$-z^{-4} - z^{-8} - \cdots$$

and $|\, z \,| > 1$.

Note that we could have developed the Laurent series without using the result by Prob. 19 (but in the same vein as Prob. 19) by starting with

$$\frac{1}{1-z^4} = \frac{1}{-z^4\left(1-\dfrac{1}{z^4}\right)}, \text{etc.}$$

Sec. 16.2 Singularities and Zeros. Infinity

Major points of this section are as follows. We have to distinguish between the concepts of singularity and pole. A function $f(z)$ has a **singularity** at z_0 if $f(z)$ is not analytic at $z = z_0$, but every neighborhood of $z = z_0$ contains points at which $f(z)$ is analytic.

Furthermore, if there is at least one such neighborhood that does not contain any other singularity, then $z = z_0$ is called an **isolated singularity**. For isolated singularities we can develop a *Laurent series* that converges in the immediate neighborhood of $z = z_0$. We look at the principal part of that series. If it is of the form

$$\frac{b_1}{z-z_0} \qquad (\text{with } b_1 \neq 0),$$

then the isolated singularity at $z = z_0$ is a **simple pole (Example 1,** pp. 715–716**).** However, if the principal part is of the form

$$\frac{b_1}{z-z_0} + \frac{b_2}{(z-z_0)^2} + \cdots \frac{b_m}{(z-z_0)^m},$$

then we have **a pole of order *m***. It can also happen that the principal part has infinitely many terms; then $f(z)$ has an **isolated essential singularity** at $z = z_0$ (see **Example 1**, pp. 715–716, **Prob. 17**).

A third concept is that of a zero, which follows our intuition. A function $f(z)$ has a **zero** at $z = z_0$ if $f(z_0) = 0$.

Just as poles have orders so do zeros. If $f(z_0) = 0$ but the derivative $f'(z) \neq 0$, then the zero is a **simple zero** (i.e., a first-order zero). If $f(z_0) = 0$, $f'(z_0) = 0$, but $f''(z) \neq 0$, then we have a **second-order zero**, and so on (see **Prob. 3** for fourth-order zero). This relates to *Taylor series* because, when developing Taylor series, we calculate $f(z_0), f'(z_0), f''(z_0), \cdots .f^{(n)}(z_0)$ by (4), p. 691, in Sec. 15.4. In the case of a second-order zero, the first two coefficients of the Taylor series are zero. Thus zeros can be classified by Taylor series as shown by (3), p. 717.

Make sure that you understand the material of this section, in particular the concepts of pole and order of pole as you will need these for residue integration. **Theorem 4**, p. 717, relates poles and zeros and will be frequently used in Sec. 16.3.

Problem Set 16.2. Page 719

3. Zeros. We claim that $f(z) = (z + 81i)^4$ has a fourth-order zero at $z = -81i$. We show this directly:

$$f(z) = (z+81i)^4 = 0 \qquad \text{gives} \qquad z = z_0 = -81i.$$

To determine the order of that zero we differentiate until $f^{(n)}(z_0) \neq 0$. We have

$$\begin{aligned} f(z) &= (z+81i)^4, & f(-81i) &= f(z_0) = 0;\\ f'(z) &= 4(z+81i)^3, & f'(-81i) &= 0;\\ f''(z) &= 12(z+81i)^2, & f''(-81i) &= 0;\\ f'''(z) &= 24(z+81i), & f'''(-81i) &= 0;\\ f^{iv}(z) &= 24, & f^{iv}(-81i) &\neq 0. \end{aligned}$$

Hence, by definition of order of a zero, p. 717, we conclude that the order at z_0 is 4. Note that we demonstrated a special case of the theorem that states that if g has a zero of first order (simple zero) at z_0, then g^n (n a positive integer) has a zero of nth order at z_0.

5. Zeros. Cancellation. The point of this, and similar problems, is that we have to be cautious. In the present case, $z = 0$ is not a zero of the given function because

$$z^{-2}\sin^2 \pi z = z^{-2}\left((\pi z)^2 + \cdots\right) = \pi^2 + \cdots .$$

11. Zeros. Show that the assumption, in terms of a formula, is

$$\text{(A)} \qquad f(z) = (z - z_0)^n g(z) \qquad \text{with} \qquad g(z_0) \neq 0,$$

so that

$$f(z_0) = 0, \qquad f'(z_0) = 0, \qquad \cdots, \qquad f^{(n-1)}(z_0) = 0,$$

as it should be for an nth-order zero.

Show that (A) implies

$$h(z) = f^2(z) = (z - z_0)^{2n} g^2(z),$$

so that, by successive product differentiation, the derivatives of $h(z)$ will be zero at z_0 as long as a factor of $z - z_0$ is present in each term. If $n = 1$, this happens for h and h', giving a second-order zero z_0 of h. If $n = 2$, we have $(z - z_0)^4$ and obtain f, f', f'', f''' equal to zero at z_0, giving a fourth-order zero z_0 of h. And so on.

17. Singularities. We start with $\cot z$. By definition,

$$\cot z = \frac{1}{\tan z} = \frac{\cos z}{\sin z}.$$

By definition on p. 715, $\cot z$ is singular where $\cot z$ is not analytic. This occurs where $\sin z = 0$, hence for

$$\text{(B)} \qquad z = 0, \pm\pi, \pm 2\pi, \ldots = \pm n\pi, \qquad \text{where} \qquad n = 0, 1, 2, \ldots .$$

Since $\cos z$ and $\sin z$ share no common zeros, we conclude that $\cot z$ is singular where $\sin z$ is 0, as given in (B). The zeros are simple poles.

Next we consider

$$\cot^4 z = \frac{\cos^4 z}{\sin^4 z}.$$

Now $\sin^4 z = 0$ for z as given in (B). But, since $\sin^4 z$ is the sine function to the fourth power and $\sin z$ has simple zeros, the zeros of $\sin^4 z$ are of order 4. Hence, by Theorem 4, p. 717, $\cot^4 z$ has poles of order 4 at (B).

But we are not finished yet. Inspired by Example 5, p. 718, we see that $\cos z$ also has an essential singularity at ∞. We claim that $\cos^4 z$ also has an essential singularity at ∞. To show this we would have to develop the Maclaurin series of $\cos^4 z$. One way to do this is to develop the first few terms of that series by (1), p. 690, of Sec. 15.4. We get (using calculus: product rule, chain rule)

$$\text{(C)} \qquad \cos^4 w = 1 - \frac{1}{2!}4w^2 + \frac{1}{4!}40w^4 - + \cdots .$$

The odd powers are zero because in the derivation of (C) these terms contain sine terms (chain rule!) that are zero at $w_0 = 0$.

We set $w = 1/z$ and multiply out the coefficients in (C):

$$\text{(D)} \qquad \cos^4 \frac{1}{z} = 1 - 2z^{-2} + \frac{5}{3}z^{-4} - + \cdots .$$

We see that the principal part of the Laurent series (D) is (D) without the constant term 1. It is infinite and thus $\cot^4 z$ has an essential singularity at ∞ by p. 718. Since multiplication of the series by $1/\sin^4 z$ does not change the type of singularity, we conclude that $\cot^4 z$ also has an essential singularity at ∞.

Sec. 16.3 Residue Integration Method

This section deals with evaluating complex integrals (1), p. 720, taken over a simple closed path C. The important concept is that of a **residue**, which is the coefficient b_1 of a Laurent series that converges for all points near a singularity $z = z_0$ inside C, as explained on p. 720. **Examples 1** and **2** show how to evaluate integrals that have only one singularity within C.

A systematic study of residue integration requires us to consider simple poles (i.e., of order 1) and poles of higher order. For simple poles, we use (3) or (4), on p. 721, to compute residues. This is shown in **Example 3**, p. 722, and **Prob. 5**. The discussion extends to higher order poles (of order m) and leads to (5), p. 722, and **Example 4**, p. 722. *It is critical that you determine the* ***order*** *of the poles inside C correctly.* In many cases we can use Theorem 4, on p. 717 of Sec. 16.2, to determine m. However, when $h(z)$ in Theorem 4 is also zero at z_0, the theorem cannot be applied. This is illustrated in **Prob. 3**.

Having determined the residues correctly, it is fairly straightforward to use the **residue theorem (Theorem 1**, p. 723) to evaluate integrals (1), p. 720, as shown in **Examples 5** and **6**, p. 724, and **Prob. 17**.

Problem Set 16.3. Page 725

3. Use of the Laurent series. The function

$$f(z) = \frac{\sin 2z}{z^6} \qquad \text{has a singularity at } z = z_0 = 0.$$

However, since both $\sin 2z$ and z^6 are 0 for $z_0 = 0$, we cannot use Theorem 4 of Sec. 16.2, p. 717, to determine the order of that zero. Hence we cannot apply (5), p. 722, directly as we do not know the value of m.

We develop the first few terms of the Laurent series for $f(z)$. From (14) in Sec. 15.4, p. 695, we know that

$$\sin w = w - \frac{w^3}{3!} + \frac{w^5}{5!} - \frac{w^7}{7!} + - \cdots .$$

We set $w = 2z$ and get

$$\text{(A)} \qquad \sin 2z = 2z - \frac{(2z)^3}{3!} + \frac{(2z)^5}{5!} - \frac{(2z)^7}{7!} + - \cdots .$$

Since we need the Laurent series of $\sin 2z/z^6$ we multiply (A) by z^{-6} and get

$$\text{(B)} \qquad \begin{aligned} z^{-6}\sin 2z &= z^{-6}\left(2z - \frac{(2z)^3}{3!} + \frac{(2z)^5}{5!} - \frac{(2z)^7}{7!} + - \cdots\right) \\ &= \frac{2}{z^5} - \frac{8}{3!}\frac{1}{z^3} + \frac{32}{5!}\frac{1}{z} - \frac{128}{7!}z + - \cdots . \end{aligned}$$

The principal part of (B) is (see definition on p. 709)

$$\frac{2}{z^5} - \frac{8}{3!}\frac{1}{z^3} + \frac{32}{5!}\frac{1}{z}.$$

We see that

$$f(z) = \frac{\sin 2z}{z^6} \quad \text{has a pole of fifth order at} \qquad z = z_0 = 0 \qquad \text{[by (2), p. 715].}$$

Note that the pole of f is only of fifth order and not of sixth order because $\sin 2z$ has a simple zero at $z = 0$.

Using the first line in the proof of (5), p. 722, we see that the coefficient of z^{-1} in the Laurent series (C) is

$$b_1 = \frac{32}{5!} = \frac{32}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{4}{15}.$$

Hence the desired residue at 0 is $\frac{4}{15}$.

Checking our result by (5)*, p. 722.* Having determined that the order of the singularity at $z = z_0 = 0$ is 5 and that we have a pole of order 5 at $z_0 = 0$, we can use (5), p. 722, with $m = 5$. We have

$$\begin{aligned}\operatorname*{Res}_{z=z_0=0} \frac{\sin 2z}{z^6} &= \frac{1}{(6-1)!} \lim_{z \to 0} \left\{ \frac{d^{6-1}}{dz^{6-1}} \left[(z-0)^6 f(z) \right] \right\} \\ &= \frac{1}{5!} \lim_{z \to 0} \left\{ \frac{d^5}{dz^5} \left[z^6 \frac{\sin 2z}{z^6} \right] \right\} \\ &= \frac{1}{5!} \lim_{z \to 0} \left\{ \frac{d^5}{dz^5} \sin 2z \right\}.\end{aligned}$$

We need

$$\begin{aligned} g(z) &= \sin 2z; \\ g'(z) &= 2 \cos 2z; \\ g''(z) &= -4 \sin 2z; \\ g'''(z) &= -8 \cos 2z; \\ g^{(4)}(z) &= 16 \sin 2z; \\ g^{(5)}(z) &= 32 \cos 2z. \end{aligned}$$

Then

$$\lim_{z \to 0} \{32 \cos 2z\} = 32 \cdot \lim_{z \to 0} \{\cos 2z\} = 32 \cdot 1.$$

Hence

$$\operatorname*{Res}_{z\,=\,z_0\,=\,0} \frac{\sin 2z}{z^6} = \frac{1}{5!} \cdot 32 \cdot 1 = \frac{4}{15}, \qquad \text{as before.}$$

Remark. In certain problems, developing a few terms of the Laurent series may be easier than using (5), p. 722, if the differentiation is labor intensive such as requiring several applications of the quotient rule of calculus (see p. 623, of Sec. 13.3).

5. Residues. Use of formulas (3) and (4), p. 721.

Step 1. Find the singularities of $f(z)$. From

$$f(z) = \frac{8}{1+z^2} \qquad \text{we see that} \qquad 1 + z^2 = 0 \qquad \text{implies} \qquad z^2 = -1, \text{ hence } z = i \text{ and } z = -i.$$

Hence we have singularities at $z_0 = i$ and $z_0 = -i$.

Step 2. Determine the order of the singularities and determine whether they are poles. Since the numerator of f is $8 = h(z) \neq 0$ (in Theorem 4), we see that the singularities in step 1 are simple, i.e., of order 1. Furthermore, by Theorem 4, p. 717, we have two poles of order 1 at i and $-i$, respectively.

Step 3. Compute the value of the residues. We can do this in two ways.

Solution 1. By (3)*, p. 721*, we have

$$\begin{aligned}\operatorname*{Res}_{z=i} f(z) &= \lim_{z\to i}\left\{(z-i)\cdot\frac{8}{1+z^2}\right\}\\ &= \lim_{z\to i}\left\{(z-i)\cdot\frac{8}{(z-i)(z+i)}\right\}\\ &= \lim_{z\to i}\left\{\frac{8}{z-i}\right\}\\ &= \frac{8}{2i} = \frac{4}{i} = -4i.\end{aligned}$$

Also

$$\begin{aligned}\operatorname*{Res}_{z=-i} f(z) &= \lim_{z\to -i}\left\{(z-(-i))\cdot\frac{8}{(z-i)(z+i)}\right\}\\ &= \lim_{z\to -i}\left\{(z+i)\cdot\frac{8}{(z-i)(z+i)}\right\}\\ &= \frac{8}{-2i} = 4i.\end{aligned}$$

Hence the two residues are

$$\operatorname*{Res}_{z=i} f(z) = -4i \qquad \text{and} \qquad \operatorname*{Res}_{z=-i} f(z) = 4i.$$

Solution 2. By (4)*, p. 721*, we have

$$\operatorname*{Res}_{z=z_0} f(z) = \operatorname*{Res}_{z=z_0}\frac{p(z)}{q(z)} = \frac{p(z_0)}{q'(z_0)} = \left.\frac{8}{(1+z^2)'}\right|_{z=z_0} = \left.\frac{8}{2z}\right|_{z=z_0} = \frac{8}{2z_0}.$$

For $z_0 = i$ we have

$$\operatorname*{Res}_{z_0=i} f(z) = \frac{8}{2i} = -4i,$$

and for $z_0 = -i$

$$\operatorname*{Res}_{z_0=-i} f(z) = \frac{8}{-2i} = 4i,$$

as before.

15. Residue theorem. We note that

$$f(z) = \tan 2\pi z = \frac{\sin 2\pi z}{\cos 2\pi z}$$

is singular where $\cos 2\pi z = 0$. This occurs at

$$2\pi z = \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \pm\frac{5\pi}{2}\cdots,$$

and hence at

$$\text{(A)} \qquad z = \pm\tfrac{1}{4}, \pm\tfrac{3}{4}, \pm\tfrac{5}{4}\cdots.$$

Since $\sin 2\pi z \neq 0$ at these points, we can use Theorem 4, p. 717, to conclude that we have infinitely many poles at (A).

Consider the path of integration $C : |z - 0.2| = 0.2$. It is a circle in the complex plane with center 0.2 and radius 0.2. We need to be only concerned with those poles that lie inside C. There is only one pole of interest, that is,

$$z = \tfrac{1}{4} = 0.25 \qquad \text{(i.e., } |0.25 - 0.2| = 0.05 < 0.2\text{)}.$$

We use (4) p. 721, to evaluate the residue of f at $z_0 = \frac{1}{4}$. We have

$$p(z) = \sin 2\pi z, \qquad p\left(\frac{1}{4}\right) = \sin\frac{\pi}{2} = 1,$$

$$q(z) = \cos 2\pi z, \qquad q'(z) = -2\pi \sin 2\pi z \quad \text{(chain rule!)}, \qquad q'\left(\tfrac{1}{4}\right) = -2\pi \sin 2\pi \tfrac{1}{4} = -2\pi.$$

Hence

$$\operatorname*{Res}_{z_0 = \frac{1}{4}} f(z) = \frac{p(\frac{1}{4})}{q'(\frac{1}{4})} = \frac{1}{-2\pi} = -\frac{1}{2\pi}.$$

Thus, by (6) of Theorem 1, p. 723,

$$\begin{aligned}\oint_C f(z)\,dz &= \oint_{C:|z-0.2|=0.2:} \tan 2\pi z\,dz \\ &= 2\pi i \cdot \operatorname*{Res}_{z_0 = \frac{1}{4}} f(z) \\ &= 2\pi i\left(-\frac{1}{2\pi}\right) \\ &= -i.\end{aligned}$$

17. Residue integration. We use the same approach as in Prob. 15. We note that

$$\cos z = 0 \qquad \text{at} \qquad z = \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \pm\frac{5\pi}{2} \cdots .$$

Also e^z is entire, see p. 631 of Sec. 13.5.

From Theorem 4, p. 717, we conclude that we have infinitely many simple poles at

$$z = \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \pm\frac{5\pi}{2} \cdots .$$

Here the closed path is a circle:

$$C : \left|z - \frac{\pi i}{2}\right| = 4.5 \qquad \text{and only} \qquad z = \frac{\pi}{2} \qquad \text{and} \qquad z = -\frac{\pi}{2} \quad \text{lie within } C.$$

This can be seen because for

$$z = \frac{\pi}{2}: \quad \left|\frac{\pi}{2} - \frac{\pi i}{2}\right| = \left|\frac{\pi}{2} - i\frac{\pi}{2}\right| = \sqrt{\left(\frac{\pi}{2}\right)^2 + \left(-\frac{\pi}{2}\right)^2} = \sqrt{\frac{2\pi^2}{4}} = \frac{\sqrt{2}\pi}{2} = 2.2214 < 4.5.$$

Same for $z = -\pi/2$. Hence, by (4) p. 721,

$$\operatorname*{Res}_{z = \pi/2} f(z) = \frac{e^{\pi/2}}{-\sin \pi/2} = -e^{\pi/2},$$

and

$$\operatorname*{Res}_{z=-\frac{\pi}{2}} f(z) = \frac{e^{-\pi/2}}{-\sin(-\pi/2)} = \frac{e^{-\pi/2}}{\sin \pi/2} = e^{-\pi/2}.$$

Using (6), p. 723,

$$\begin{aligned}
\oint_C f(z)\,dz &= \oint_{C:|z-\pi i/2|=4.5} \frac{e^z}{\cos z}\,dz \\
&= 2\pi i \left[\operatorname*{Res}_{z=\pi/2} f(z) + \operatorname*{Res}_{z=-\pi/2} f(z) \right] \\
&= 2\pi i \,\left(-e^{\pi/2} + e^{-\pi/2}\right) \\
&= 2\pi i \left(2\sinh\left(-\frac{\pi}{2}\right)\right) \qquad \text{[by (17), p. A65, App. 3]} \\
&= 2\pi i \left(-2\sinh\frac{\pi}{2}\right) \qquad \text{[since } \sinh \text{ is an odd function]} \\
&= -4\pi i \sinh\frac{\pi}{2} \\
&= -28.919i.
\end{aligned}$$

Sec. 16.4 Residue Integration of Real Integrals

It is surprising that residue integration, a method of *complex* analysis, can also be used to evaluate certain kinds of complicated *real* integrals. The key ideas in this section are as follows. To apply residue integration, we need a closed path, that is, a contour. Take a look at the different real integrals in the textbook, pp. 725–732. For real integrals (1), p. 726, we obtain a contour by the transformation (2), p. 726. This is illustrated in Example 1 and Prob. 7.

For real integrals (4), p. 726, and (10), p. 729 (real "Fourier integrals"), we start from a finite interval from $-R$ to R on the real axis (the x-axis) and close it in complex by a semicircle S as shown in Fig. 374, p. 727. Then we "blow up" this contour and make an assumption (degree of the denominator $\geq$ degree of the numerator $+2$) under which the integral over the blown-up semicircle will be 0. Note that *we only take those poles that are in the* ***upper half-plane*** *of the complex plane* and ignore the others. Example 2, p. 728, and Prob. 11 solve integrals of the kind given by (4). Real Fourier integrals (10) are solved in Example 3, pp. 729–730, and Prob. 21.

Finally, we solve real integrals (11) whose integrand becomes infinite at some point a in the interval of integration (Fig. 377, p. 731; Example 4, p. 732; Prob. 25) and requires the concept of Cauchy principal value (13), p. 730. The pole a lies on the real axis of the complex plane.

Problem Set 16.4. Page 733

7. Integral involving sine. Here the given integral is

$$\int_0^{2\pi} \frac{a}{a - \sin\theta}\,d\theta = a \int_0^{2\pi} \frac{1}{a - \sin\theta}\,d\theta.$$

Using (2), p. 726, we get

$$a - \sin\theta = a - \frac{1}{2i}\left(z - \frac{1}{z}\right)$$

and

$$d\theta = \frac{dz}{iz} \qquad \text{[see textbook after (2)].}$$

Hence

$$\int_0^{2\pi} \frac{1}{a - \sin\theta}\, d\theta = \oint_C \frac{i\, dz}{iz\left[a - \frac{1}{2i}\left(z - \frac{1}{z}\right)\right]},$$

where C is the unit circle.

Now

$$\begin{aligned} iz\left[a - \frac{1}{2i}\left(z - \frac{1}{z}\right)\right] &= iza - \frac{iz}{2i} z + \frac{1}{2i}\frac{iz}{z} \\ &= iza - \frac{z^2}{2} + \frac{1}{2} \\ &= -\frac{1}{2}\left(z^2 - 2aiz - 1\right) \end{aligned}$$

so that the last integral is equal to

$$-2\oint_C \frac{dz}{z^2 - 2aiz - 1}.$$

We need to find the roots of $z^2 - 2aiz - 1$. Using the familiar formula for finding roots of a quadratic equation,

$$az^2 + bz + c = 0, \qquad z_1, z_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

with $a = 1, b = -2ai, c = -1$ we obtain

$$z_{1,2} = \frac{2ai \pm \sqrt{(-2ai)^2 - 4 \cdot 1 \cdot (-1)}}{2} = \frac{2ai \pm \sqrt{4\left(1 - a^2\right)}}{2} = ai \pm \sqrt{1 - a^2}.$$

By Theorem 4 of Sec. 16.2 on p. 717, we have two simple poles at

$$z_1 = ai + \sqrt{1 - a^2} \qquad \text{and at} \qquad z_2 = ai - \sqrt{1 - a^2}.$$

However, z_1 is outside the unit circle and thus of no interest (see p. 726). Hence, by (3), p. 721, in Sec. 16.3 of the textbook, we compute the residue at z_2:

$$\begin{aligned} \operatorname*{Res}_{z=z_2} f(z) &= \operatorname*{Res}_{z=z_2} \frac{1}{(z - z_1)(z - z_2)} \\ &= \lim_{z \to z_2} (z - z_2) \frac{1}{(z - z_1)(z - z_2)} \\ &= \lim_{z \to z_2} \frac{1}{z - z_1} \\ &= \left[\frac{1}{z - \left(ai - \sqrt{1 - a^2}\right)}\right]_{z = ai - \sqrt{1 - a^2}} \\ &= \frac{1}{ai - \sqrt{1 - a^2} - ai - \sqrt{1 - a^2}} \\ &= -\frac{1}{2\sqrt{1 - a^2}}. \end{aligned}$$

Thus by Theorem 1, p. 723, (Residue Theorem),

$$\begin{aligned}\int_0^{2\pi} \frac{1}{a - \sin\theta}\, d\theta &= -2a \oint_C \frac{dz}{z^2 - 2aiz - 1} \\ &= -2a \cdot 2\pi i \operatorname*{Res}_{z = z_2} f(z) \\ &= -4a\pi i \cdot \left(-\frac{1}{2\sqrt{1-a^2}} \right) \\ &= \frac{2a\pi i}{\sqrt{1-a^2}}.\end{aligned}$$

We can get rid of the i in the numerator by

$$\sqrt{1-a^2} = \sqrt{(-1)(a^2-1)} = i\sqrt{a^2-1}$$

so that the answer becomes

$$\frac{2a\pi}{\sqrt{a^2-1}} \qquad \text{(as on p. A41 of the text).}$$

11. Improper integral: Infinite interval of integration. Use of (7), p. 728. The integrand, considered as a function of complex z, is

$$f(z) = \frac{1}{(1+z^2)^2}.$$

We factor the denominator and set it to 0

$$\left(1+z^2\right)^2 = (z^2+1)(z^2+1) = (z-i)(z+i)(z-i)(z+i) = (z-i)^2(z+i)^2 = 0.$$

This shows that there are singularities (p. 715) at $z = i$ and $z = -i$, respectively.

We have to consider only $z = i$ since it lies in the upper half-plane (defined on p. 619, Sec. 13.3) and ignore $z = -i$ since it lies in the lower half-plane. This is as in Example 2, p. 728 (where only z_1 and z_2 are used and z_3 and z_4 are ignored).

Furthermore, since the numerator of $f(z)$ is not zero for $z = i$, we have a pole of order 2 at $z = i$ by Theorem 4 of Sec. 16.2 on p 717. (The ignored singularity at $z = i$ also leads to a pole of order 2.)

The degree of the numerator of $f(z)$ is 1 and the degree of the denominator is 4, so that we are allowed to apply (7), p. 728.

We have by (5*), p. 722,

$$\operatorname*{Res}_{z=i} f(z) = \frac{1}{(2-1)!} \lim_{z \to i} \left\{ \frac{d^{2-1}}{dz^{2-1}} \left[(z-i)^2 f(z) \right] \right\}.$$

Now we first do the differentiation

$$\begin{aligned}\frac{d}{dz}\left[(z-i)^2 \frac{1}{(z-i)^2(z+i)^2} \right] &= \frac{d}{dz}\left[\frac{1}{(z+i)^2} \right] \\ &= \left[(z+i)^{-2} \right]' \\ &= -2(z+i)^{-3}\end{aligned}$$

and then find the residue

$$\begin{aligned}\operatorname*{Res}_{z=i} f(z) &= \lim_{z\to i} \left[-2(z+i)^{-3}\right] \\ &= \left[\frac{-2}{(z+i)^3}\right]_{z=i} = \frac{-2}{(i+i)^3} \\ &= \frac{-2}{2^3 i^3} = \frac{-1}{4i^3} = -\frac{i}{4}.\end{aligned}$$

Hence by (7), p. 728, the ***real*** infinite integral is

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{1}{(x^2+1)^2}\, dx &= -2\pi i \cdot \operatorname*{Res}_{z=i} f(z) \\ &= 2\pi i\left(-\frac{i}{4}\right) \\ &= -\frac{\pi i^2}{2} = \frac{\pi}{2}.\end{aligned}$$

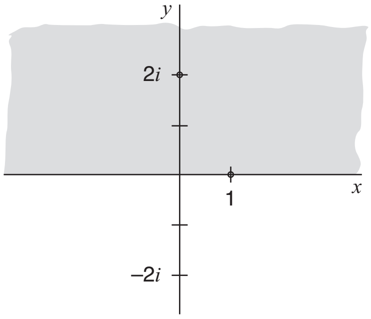
21. Improper integral: Infinite interval of integration. Simple pole in upper-half plane. Simple pole on real axis. Fourier integral. We note that the given integral is a Fourier integral of the form

$$\int_{-\infty}^{\infty} f(x) \sin sx\, dx \qquad \text{with} \qquad f(x) = \frac{1}{(x-1)(x^2+4)} \qquad \text{and} \qquad s = 1 \quad \text{[see (8), p. 729].}$$

The denominator of the integrand, expressed in z factors, is

$$(z-1)(z^2+4) = (z-1)(z-2i)(z+2i) = 0.$$

This gives singularities of $z = 1, 2i, -2i$, respectively. The pole at $z = 2i$ lies in the upper half-plane (defined in Sec. 13.3 on p. 619), while the pole at $z = 1$ lies on the contour. Because of the pole on the contour, we need to find the principal value by (14) in Theorem 1 on pp. 731–2 rather than using (10) on p. 729. (The simple pole $z = -2i$ lies in the lower half-plane and, thus, is not wanted.)



Sec. 16.4 Prob. 21. Fourier integral. Only the poles at $z = 1$ and $2i$ that lie in the upper half-plane (shaded area) are used in the residue integration

We compute the residues

$$f(z)e^{iz} \qquad (s = 1)$$

as discussed on p. 729, and in Example 3. Using (4), p. 721, we get

$$\operatorname*{Res}_{z=1} f(z)e^{iz} = \operatorname*{Res}_{z=1} \frac{1}{(z-1)(z^2+4)}e^{iz} = \left[\frac{p(z)}{q'(z)}\right]_{z=1}$$

where

$$p(z) = e^{iz}; \qquad q(z) = (z-1)(z^2+4) = z^3 - z^2 + 4z - 4; \qquad q'(z) = 3z^2 - 2z + 4.$$

Hence

$$\operatorname*{Res}_{z=1} f(z)e^{iz} = \frac{p(1)}{q'(1)} = \frac{e^i}{3-2+4} = \frac{e^i}{5}.$$

Now by (5), p. 634, in Sec. 13.6 ("Euler's formula in the complex"),

$$e^i = \cos 1 + i \sin 1$$

so that

$$\frac{e^i}{5} = \frac{1}{5}(\cos 1 + i \sin 1) = \frac{1}{5}\cos 1 + i\frac{1}{5}\sin 1.$$

Hence

$$\operatorname{Re}\left\{\operatorname*{Res}_{z=1} f(z)e^{iz}\right\} = \operatorname{Re}\left\{\frac{1}{5}\cos 1 + i\frac{1}{5}\sin 1\right\} = \frac{\cos 1}{5}.$$

Also

$$\begin{aligned}\operatorname*{Res}_{z=2i} f(z)e^{iz} &= \left[\frac{p(z)}{q'(z)}\right]_{z=2i}\\ &= \frac{e^{i\cdot 2i}}{3(2i)^2 - 2(2i) + 4}\\ &= \frac{e^{-2}}{-8-4i}\\ &= \frac{e^{-2}(-8+4i)}{8^2+4^2}\\ &= \frac{-8e^{-2}+4e^{-2}i}{80}\\ &= -\frac{e^{-2}}{10} + \frac{e^{-2}i}{20}.\end{aligned}$$

Hence

$$\operatorname{Re}\left\{\operatorname*{Res}_{z=2i} f(z)e^{iz}\right\} = -\frac{e^{-2}}{10}.$$

Using (14), p. 732, the solution to the desired ***real*** Fourier integral (with $s = 0$) is

$$\begin{aligned}pr.\ v \int_{-\infty}^{\infty} \frac{\sin x}{(x-1)(x^2+4)} &= \pi \sum \operatorname{Re}\left[\operatorname*{Res}_{z=1} f(z)e^{isz}\right] + 2\pi \sum \operatorname{Re}\left[\operatorname*{Res}_{z=2i} f(z)e^{isz}\right]\\ &= \pi\left(\frac{\cos 1}{5} - 2\frac{e^{-2}}{10}\right) = \frac{\pi}{5}\left(\cos 1 - e^{-2}\right) = 0.254448.\end{aligned}$$

Note that we wrote pr.v., that is, *Cauchy principal value* (p. 730) on account of the pole on the contour (x-axis) and the behavior of the integrand.

25. Improper integrals. Poles on the real axis. We use (14) and the approach of Example 4, p. 732. The denominator of the integrand is

$$x^3 - x = x\left(x^2 - 1\right) = x(x-1)(x+1).$$

Considering this in the complex domain, we have

$$\text{(A)} \qquad z^3 - z = z(z-1)(z+1)$$

so that there are singularities at $z = 0, 1, -1$.

Since the numerator is $x + 5$ and in the complex domain $z + 5$, we see that $z + 5$ is not zero for $z = 0, 1, -1$. Hence by Theorem 4 of Sec. 16.2 on p. 717,

$$f(z) = \frac{z+5}{z^3 - z} \qquad \text{has three simple poles at } 0, 1, -1.$$

We compute the residues as in Sec. 16.3, by using (4), p. 721,

$$p(z) = z + 5; \qquad q(z) = z^3 - z \qquad \text{so that} \qquad q'(z) = 3z^2 - 1.$$

Hence at $z = 0$

$$\operatorname*{Res}_{z=0} f(z) = \frac{p(0)}{q'(0)} = \frac{5}{3 \cdot 0^2 - 1} = -5.$$

At $z = 1$

$$\operatorname*{Res}_{z=1} f(z) = \frac{p(1)}{q'(1)} = \frac{6}{3-1} = 3.$$

Finally at $z = -1$

$$\operatorname*{Res}_{z=-1} f(z) = \frac{p(-1)}{q'(-1)} = \frac{-1+5}{3(-1)^2 - 1} = \frac{4}{2} = 2.$$

We are ready to use (14), p. 732. Note that there are no poles in the upper half-plane as (A) does not contain factors with nonzero imaginary parts. This means that the first summation in (14) is zero. Hence

$$\text{pr.v.} \int_{-\infty}^{\infty} \frac{x+5}{x^3 - x} dx = \pi i\,(-5 + 3 + 2) = \pi i \cdot 0 = 0.$$

Chap. 17 Conformal Mapping

We shift gears and introduce a third approach to problem solving in complex analysis. Recall that **so far** we covered two approaches of complex analysis. The **first method** concerned **evaluating complex integrals by Cauchy's integral formula** (Sec. 14.3, p. 660 of the textbook and p. 291 in this Manual). Specific background material needed was Cauchy's integral theorem (Sec. 14.2) and, in general, Chaps. 13 and 14. The **second method** dealt with **residue integration**, which we applied to *both* complex integrals in Sec. 16.3 (p. 719 in the textbook and p. 291 in this Manual) and real integrals in Sec. 16.4 (p. 725, p. 326 in this Manual). The background material was general power series, Taylor series (Chap. 15), and, most importantly, Laurent series which admitted negative powers (Sec. 16.1, p. 708) and thus lead to the study of poles (Sec. 16.2, p. 715).

The **new** method is a **geometric** approach to complex analysis and involves the use of conformal mappings. We need to explain two terms: (a) mapping and (b) conformal. For (a), recall from p. 621 in Sec. 13.3 that any complex function $f(z)$, where $z = x + iy$ is a complex variable, can be written in the form

$$(1) \qquad w = f(z) = u(x, y) + iv(x, y) \qquad \text{(see also p. 737 in Sec. 17.1).}$$

We want to study the geometry of complex functions $f(z)$ and consider (1).

In basic (real) calculus we graphed continuous real functions $y = f(x)$ of a real variable x as curves in the Cartesian xy-plane. This required *one* (real) plane. If you look at (1), you may notice that we need to represent geometrically *both* the variable z and the variable w as points in the complex plane. The idea is to **use two separate complex planes** for the two variables: one for the z-plane and one for the w-plane. And this is indeed what we shall do. So if we graph the points $z = x + iy$ in the z-plane (as we have done many times in Chap. 13) and, in addition, graph the corresponding $w = u + iv$ (points obtained from plugging in z into f) in the w-plane (with uv-axes), then the function $w = f(z)$ defines a correspondence (**mapping**) between the points of these two planes (for details, see p. 737). In practice, the graphs are usually not obtained pointwise, as suggested by the definition, but from mappings of sectors, rays, lines, circles, etc.

We don't just take any function $f(z)$ but we prefer *analytic* functions. In comes the concept of (b) conformality. The mapping (1) is **conformal** if it preserves angles between oriented curves both in magnitude as well as in sense. Theorem 1, on p. 738 in Sec. 17.1, links the concepts of analyticity with conformality: An analytic function $w = f(z)$ is conformal except at points z_0 (*critical points*) where its derivative $f'(z_0) = 0$.

The rest of the chapter discusses important conformal mappings and creates their graphs. Sections 17.1 (p. 737) and 17.4 (p. 750) examine conformal mappings of the major analytic functions from Chap. 13. Sections 17.3 (p. 746) and 17.4 deal with the novel linear fractional transformation, a transformation that is a fraction (see p. 746). The chapter concludes with Riemann surfaces, which allow multivalued relations of Sec. 13.7 (p. 636) to become single-valued and hence functions in the usual sense. We will see the astonishing versatility of conformal mapping in Chapter 18 where we apply it to practical problems in potential theory.

You might have to allocate more study time for this chapter than you did for Chaps. 15 and 16. You should study this chapter diligently so that you will be well prepared for the applications in Chap. 18.

As background material for Chap. 17 you should remember Chap. 13, including how to graph complex numbers (Sec. 13.1, p. 608), polar form of complex numbers (Sec. 13.2, p. 613), complex functions (pp. 620–621), e^z (Sec. 13.5, p. 630), Euler's formula (5), p. 634, $\sin z, \cos z, \sinh z, \cosh z$, and their various formulas (Sec. 13.6, p. 633), and the multivalued relations of Sec. 13.7, p. 636 (for optional Sec. 17.5). Furthermore, you should know how to find roots of polynomials and know how to algebraically manipulate fractions (in Secs. 17.2 and 17.3).

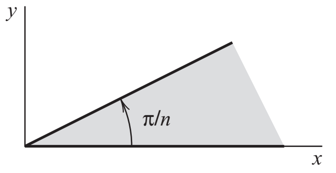
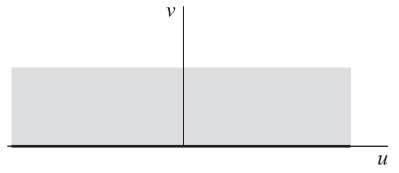
Sec. 17.1 Geometry of Analytic Functions: Conformal Mapping

We discussed mappings and conformal mappings in detail in the opening to Chap. 17 of this Manual. Related material in the textbook is: mapping (1), p. 737, and illustrated by Example 1; conformal, p. 738; conformality and analyticity in **Theorem 1**, p. 738. The section continues with four more examples of conformal mappings and their graphs. They are $w = z^n$ (**Example 2**, p. 739), $w = z + 1/z$ (Joukowski airfoil, **Example 3**, pp. 739–740), $w = e^z$ (**Example 4**, p. 740), and $w = \text{Ln } z$ (**Example 5**, p. 741). The last topic is the magnification ratio, which is illustrated in Prob. 33, p. 742.

In the examples in the text and the exercises, we consider how sectors, rays, lines, circles, etc. are mapped from the z-plane onto the w-plane by the specific given mapping. We use polar coordinates and Cartesian coordinates. Since *there is* ***no general rule*** *that fits all problems*, you have to look over, understand, and remember the specific mappings discussed in the examples in the text and supplemented by those from the problem sets. To fully understand specific mappings, make graphs or sketches.

Finally you may want to build a table of conformal mappings:

Mapping	**Region to be Mapped**	**Image of Region**	**Reference**
$w = z^n$	Sector $0 \leq \theta \leq \dfrac{\pi}{n}$	Upper half plane $v \geq 0$	Example 2, p. 739

Graph in z-plane	**Graph in w-plane**
	

Put in more mappings and graphs or sketches. The table does not have to be complete, it is just to help you remember the most important examples for exams and for solving problems.

Illustration of mapping. Turn to p. 621 of Sec. 13.3 and look at Example 1. Note that this example defines a *mapping* $w = f(z) = z^2 + 3z$. It then shows how the point $z_0 = 1 + 3i$ (from the z-plane) is being mapped onto $w_0 = f(z_0) = f(1 + 3i) = (1 + 3i)^2 + 3(1 + 3i) = 1 + 6i + 9i^2 + 3 + 9i = -5 + 15i$ (of the w-plane). A second such example is Example 2, p. 621.

More details on Example 1, p. 737. Turn to p. 737 and take a look at the example and Fig. 738. We remember that the function $f(z) = z^2$ is analytic (see pp. 622–624 of Sec. 13.3, Example 1, p. 627 of Sec. 13.4). The mapping of this function is

$$w = f(z) = z^2.$$

It has a critical point where the derivative of its underlying function f is zero, that is, where

$$f'(z) = 0 \qquad \text{here} \qquad f'(z) = (z^2)' = 2z = 0 \quad \text{hence} \quad z = 0.$$

Thus the critical point is at $z = 0$. By Theorem 1, p. 738, $f(z)$ is conformal except at $z = 0$. Indeed, at $z = 0$ conformality is violated in that the angles are doubled, as clearly shown in Fig. 378, p. 737. The same reasoning is used in Example 2, p. 739.

Problem Set 17.1. Page 741

3. Mapping. To obtain a figure similar to Fig. 378, p. 737, we follow Example 1 on that page. Using polar forms [see (6), p. 631 in Sec. 13.5]

$$z = re^{i\theta} \qquad \text{and} \qquad w = Re^{i\phi}$$

we have, under the given mapping $w = z^3$,

$$Re^{i\phi} = w = f(z) = f(re^{i\theta}) = (re^{i\theta})^3 = r^3 e^{i3\theta}.$$

We compare the moduli and arguments (for definition, see p. 613) and get

$$R = r^3 \qquad \text{and} \qquad \phi = 3\theta.$$

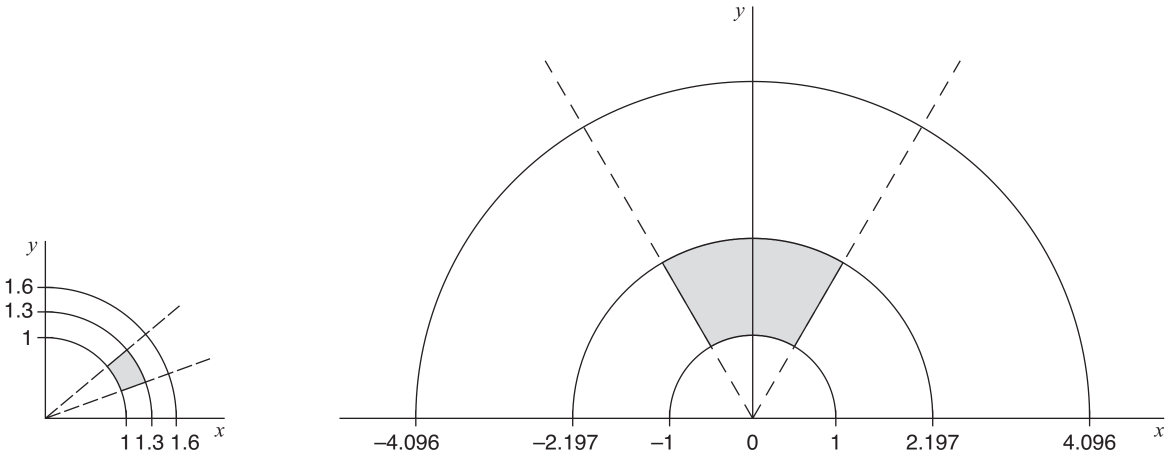
Hence circles $r = r_0$ are mapped onto circles $R = r_0^3$ and rays $\theta = \theta_0$ are mapped onto rays $\phi = 3\theta_0$. Note that the resulting circle $R = r_0^3$ is a circle bigger than $r = r_0$ when $r_0 > 1$ and a smaller one when $r_0 < 1$. Furthermore, the process of mapping a ray $\theta = \theta_0$ onto a ray $\phi = 3\theta_0$ corresponds to a rotation.

We are ready to draw the desired figure and consider the region

$$1 \leq r \leq 1.3 \qquad \text{with} \qquad \frac{\pi}{9} \leq \theta \leq \frac{2\pi}{9}.$$

It gets mapped onto the region $1^3 \leq R \leq (1.3)^3$ with $3 \cdot \pi/9 \leq \phi \leq 3 \cdot 2\pi/9$. This simplifies to

$$1 \leq R \leq 2.197 \qquad \text{with} \qquad \frac{\pi}{3} \leq \phi \leq \frac{2\pi}{3}.$$



Sec. 17.1 Prob. 3. Given region and its image under the mapping $w = z^3$

7. Mapping of curves. Rotation. First we want to show that the given mapping $w = iz$ is indeed a rotation. To do this, we express z in polar coordinates [by (6), p. 631], that is,

$$\text{(A)} \qquad z = re^{i\theta} \qquad \text{where} \qquad r > 0.$$

Then we obtain the image of (A) under the given mapping by substituting (A) directly into the mapping and simplifying:

$$\begin{aligned} w &= f(z) = f(re^{i\theta}) \\ &= iz|_{z=r\,e^{i\theta}} \\ &= e^{i\pi/2}re^{i\theta} && \text{[using } i = e^{i\pi/2} \text{ by (8), p. 631]} \\ &= re^{i(\theta+\pi/2)} \\ &= re^{i\tilde{\theta}} && \text{where} \quad \tilde{\theta} = \theta + \pi/2. \end{aligned}$$

This shows that this mapping, $w = iz$, is indeed a rotation about 0 through an angle of $\pi/2$ in the positive sense, that is, in a counterclockwise direction.

We want to determine the images of $x = 1, 2, 3, 4$, and so we consider the more general problem of determining the image of $x = c$, where c is a constant. Then for $x = c$, z becomes

$$z = x + iy = c + iy$$

so that under the mapping

$$\text{(B)} \qquad w = f(z) = iz|_{z=c+iy} = i(c + iy) = ic + i^2y = -y + ic.$$

This means that the image of points on a line $x = c$ is $w = -y + ic$. Thus $x = 1$ is mapped onto $w = -y + i$; $x = 2$ onto $w = -y + 2i$, etc. Furthermore, $z = x = c$ on the real axis and is mapped by (B) onto the imaginary axis $w = ic$. So $z = x = 1$ is mapped onto $w = i$, and $z = x = 2$ is mapped onto $w = 2i$, etc.

Similar steps for horizontal lines $y = k =$ const give us

$$z = x + ik \qquad \text{so that} \qquad w = i(x + ik) = -k + ix.$$

Hence $y = 1$ is mapped onto $w = -1 + ix$, and $y = 2$ onto $w = -2 + ix$. (Do you see a counterclockwise rotation by $\pi/2$?). Furthermore, $z = y = k$ is mapped onto $w = -k$ and $z = y = 1$ onto $w = -1$; $z = y = 2$ onto $w = -2$. Complete the problem by sketching or graphing the desired images.

11. Mapping of regions. To examine the given mapping, $w = z^2$, we express z in polar coordinates, that is, $z = re^{i\theta}$ and substitute it into the mapping

$$Re^{i\phi} = w = f(z) = f(re^{i\theta}) = z^2\big|_{z=re^{i\theta}} = (re^{i\theta})^2 = r^2(e^{i\theta})^2 = r^2e^{i2\theta}.$$

This shows that the $w = z^2$ doubles angles at $z = 0$ and that it squares the moduli. Hence for our problem, under the given mapping,

$$-\frac{\pi}{8} < \theta < \frac{\pi}{8} \qquad \text{becomes} \qquad -\frac{\pi}{4} < \phi < \frac{\pi}{4}$$

or equivalently (since $\theta = \text{Arg } z$ and $\phi = \text{Arg } w$)

$$-\frac{\pi}{8} < \text{Arg } z < \frac{\pi}{8} \qquad \text{becomes} \qquad -\frac{\pi}{4} < \text{Arg } w < \frac{\pi}{4} \qquad \text{[for definition of Arg, see (5), p. 614].}$$

Furthermore, $r = \frac{1}{2}$ maps onto $R = \frac{1}{4}$. Since

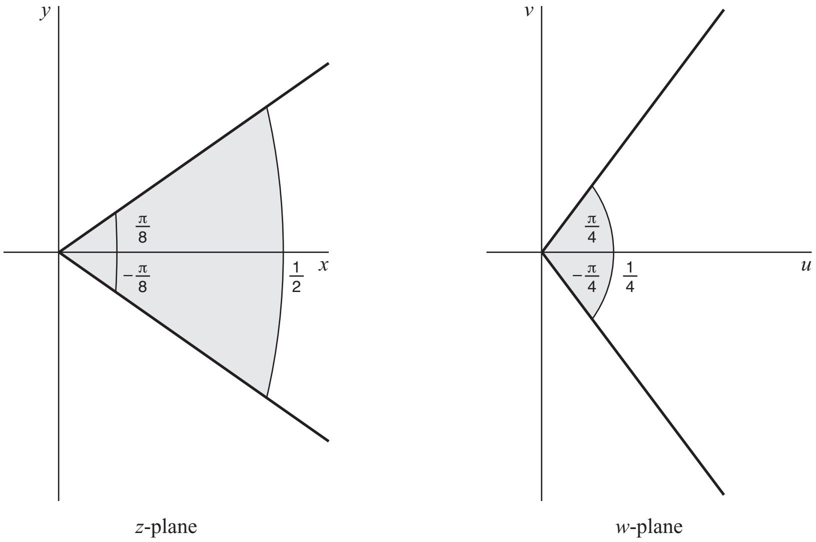
$$|z| = r \qquad \text{[by (3), p. 613 in Sec. 13.2]}$$

we get

$$|z| \leq \tfrac{1}{2}, \qquad \text{which becomes} \qquad |w| \leq \tfrac{1}{4},$$

which corresponds to the answer on p. A41 in Appendix 2 of the textbook.

Together we obtain the figures below.



Sec. 17.1 Prob. 11. Given region and its image under the mapping $w = f(z) = z^2$

15. Mapping of regions. The given region (see p. 619 in Sec. 13.3)

$$\left|z - \tfrac{1}{2}\right| \leq \tfrac{1}{2} \qquad \text{is a closed circular disk of radius } \tfrac{1}{2} \text{ with center at } x = \tfrac{1}{2}.$$

The corresponding circle can be expressed as

$$\left(x - \tfrac{1}{2}\right)^2 + y^2 = \left(\tfrac{1}{2}\right)^2.$$

Written out

$$x^2 - x + \tfrac{1}{4} + y^2 = \tfrac{1}{4}$$

and rearranged is

$$\left(x^2 + y^2\right) - x + \tfrac{1}{4} = \tfrac{1}{4}.$$

Subtract $\frac{1}{4}$ from both sides of the equation and get

$$\left(x^2 + y^2\right) - x = 0.$$

But

$$x^2 + y^2 = |z|^2 = z\overline{z} \qquad \text{[by (3), p. 613 of Sec. 13.2].}$$

Furthermore,

$$x = \frac{z + \overline{z}}{2}.$$

Substituting these last two relations into our equation yields

$$(*) \qquad z\overline{z} - \frac{z + \overline{z}}{2} = 0.$$

Now we are ready to consider the given mapping, $w = 1/z$, so that $z = 1/w$ and obtain

$$\begin{aligned}
z\overline{z} - \frac{z + \overline{z}}{2} &= \frac{1}{w}\frac{1}{\overline{w}} - \frac{\frac{1}{w} + \frac{1}{\overline{w}}}{2} \\
&= \frac{1}{w}\frac{1}{\overline{w}} - \frac{\frac{\overline{w}+w}{w\overline{w}}}{2} \\
&= \frac{1}{w}\frac{1}{\overline{w}} - \frac{\overline{w} + w}{2w\overline{w}} \\
&= \frac{1}{2}\frac{1}{w\overline{w}}\big(2 - \overline{w} - w\big) \\
&= \frac{1}{2}\frac{1}{w\overline{w}}(2 - [u - iv] - [u + iv]) && \text{[by (1), p. 737 and definition of } \overline{w}] \\
&= \frac{1}{w}\frac{1}{\overline{w}}(2 - 2u) \\
&= 0 && \text{[from the l.h.s of (*)].}
\end{aligned}$$

For the last equality to hold, we see that

$$2 - 2u = 0 \qquad \text{so that} \qquad u = 1.$$

This shows that, for the given mapping, the circle maps onto $u = 1$.

The center of the circle $\left(\frac{1}{2}, 0\right)$ maps onto

$$f(z) = \left.\frac{1}{z}\right|_{z=\frac{1}{2}} = \frac{1}{\frac{1}{2}} = 2 = u + iv \qquad \text{so that} \qquad u = 2.$$

This is > 1 so that the inside of the circle maps to $u \geq 1$.

17. Mapping of regions. Exponential function. We take the exponents of

$$-\operatorname{Ln} 2 \leq x \leq \operatorname{Ln} 4, \qquad \text{which is equivalent to} \qquad \operatorname{Ln}(2^{-1}) \leq x \leq \operatorname{Ln} 4$$

and, because the logarithm Ln x is monotone increasing with x, we obtain

$$\tfrac{1}{2} \leq e^x \leq 4.$$

Now

$$e^x = |\,e^z\,| = |\,w\,| \qquad \text{[by (10) in Sec. 13.5, p. 631].}$$

Hence the given region gets mapped onto

$$\tfrac{1}{2} \leq |\,w\,| \leq 4.$$

21. Failure of conformality. Cubic polynomial. The general cubic polynomial (CP) is

$$\text{(CP)} \qquad a_3z^3 + a_2z^2 + a_1z + a_0.$$

Conformality fails at the critical points. These are the points at which the derivative of the cubic polynomial is zero. We differentiate (CP) and set the derivative to zero:

$$(a_3z^3 + a_2z^2 + a_1z + a_0)' = 3a_3z^2 + 2a_2z + a_1 = 0.$$

We factor by the well-known quadratic formula for a general second-order polynomial $az^2 + bz + c = 0$ and obtain

$$z_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-2a_2 \pm \sqrt{4a_2^2 - 4 \cdot 3a_3 \cdot a_1}}{2 \cdot 3a_3} = \frac{-a_2 \pm \sqrt{a_2^2 - 3 \cdot a_3 \cdot a_1}}{3a_3}.$$

Thus the mapping, $w = f\,(z)$, is not conformal if $f'\,(z) = 0$. This happens when

$$z = \frac{-a_2 \pm \sqrt{a_2^2 - 3 \cdot a_3 \cdot a_1}}{3a_3}.$$

Remark. You may want to verify that our answer corresponds to the answer on p. A41 in Appendix 2 of the textbook. Set

$$a_3 = 1, \qquad a_2 = a, \qquad a_1 = b, \qquad a_0 = c.$$

Note that we can set $a_3 = 1$ in (CP) without loss of generality as we can always divide the cubic polynomial by a_3 if $0 < |a_3| \neq 1$.

33. Magnification ratio. By (4), p. 741, we need

$$\left|\,(e^z)'\,\right| = |\,e^z\,| = e^x \qquad \text{(by Sec. 13.5, p. 630).}$$

Hence $M = 1$ when $x = 0$, which is at every point on the y-axis.

Also, $M < 1$ everywhere in the left half-plane because $e^x < 1$ when $x < 0$, and $M > 1$ everywhere in the right half-plane.

By (5), p. 741, we show that the Jacobian is

$$J = \left|\,f'\,(z)\,\right|^2 = |\,(e^z)'\,|^2 = |\,e^z\,|^2 = (e^x)^2 = e^{2x}.$$

Confirm this by using partial derivatives in (5).

Sec. 17.2 Linear Fractional Transformations. (Möbius Transformations)

This new function on p. 473

$$\text{(1)} \quad \textbf{LFT} \qquad w = \frac{az+b}{cz+d} \qquad (\text{where } ad - bc \neq 0)$$

is useful in modeling and solving boundary value problems in potential theory [as in Example 2 of Sec. 18.2, where the first function on p. 765, of the textbook, is a **linear fractional transformation** (**LFT**) (1) with $a = b = d = 1$ and $c = -1$].

LFTs are versatile because—with different constants—LFTs can model translations, rotations, linear transformations, and inversions of circles as shown in (3), p. 743. They also have attractive properties (Theorem 1, p. 744). Problem 3 (in a matrix setting) and Prob. 5 (in a general setting) explore the relationship between LFT (1) and its *inverse* (4), p. 745. **Fixed points** are defined on p. 745 and illustrated in **Probs. 13** and **17.**

Problem Set 17.2. Page 745

3. Matrices. a. Using 2 × 2 matrices, prove that the coefficient matrices of (1), p. 743, and (4), p. 745, are inverses of each other, provided that

$$ad - bc = 1.$$

Solution. We start with

$$\text{(1)} \qquad w = \frac{az+b}{cz+d} \qquad (\text{where } ad - bc \neq 0),$$

and note that its coefficient matrix is

$$\text{(M1)} \qquad \mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Using formula (4*) in Sec. 7.8 on p. 304, we have that the inverse of matrix **A** is

$$\text{(M2)} \qquad \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \qquad (\text{where } \det \mathbf{A} = ad - bc).$$

We are given that the inverse mapping of (1) is

$$\text{(4)} \qquad z = \frac{dw - b}{-cw + a}$$

so that its coefficient matrix is

$$\text{(M3)} \qquad \mathbf{B} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Looking at (M2) and (M3) we see that if the only way for $\mathbf{A}^{-1} = \mathbf{B}$ is for

$$\frac{1}{\det \mathbf{A}} = \frac{1}{ad - bc} = 1, \qquad \text{that is,} \qquad ad - bc = 1.$$

Conversely, if $ad - bc = 1$, then

$$\frac{1}{\det \mathbf{A}} = \frac{1}{ad - bc} = \frac{1}{1} = 1 \qquad \text{so that} \qquad \mathbf{A}^{-1} = 1 \cdot \mathbf{B} = \mathbf{B} \qquad \text{[by (M2), (M3)].}$$

This proves **a**.

b. The composition of LFTs corresponds to the multiplication of coefficient matrices.

Hint: Start by defining two general LFTs of the form (1) that are different from each other.

5. Inverse. a. Derive (4), p. 745, from (1), p. 743.
We start with

$$(1) \qquad w = \frac{az + b}{cz + d} \qquad (\text{where } ad - bc \neq 0)$$

and multiply both sides of (1) by $cz + d$, thereby obtaining

$$(cz + d)\,w = a\,z + b.$$

Next we group the z-terms together on the left and the other terms on the right:

$$czw - az = b - dw$$

so that

$$(\text{A}) \qquad z\,(cw - a) = b - dw.$$

We divide both sides of (A) by $(cw - a)$ and get

$$(\text{A}') \qquad z = \frac{b - dw}{cw - a}.$$

This is not quite (4) yet. To obtain (4), we multiply (A′) by $\frac{-1}{-1}$ (which we can always do) and get

$$z = \frac{-(b - dw)}{-(cw - a)} = \frac{-b + dw}{-cw + a} = \frac{dw - b}{-cw + a}.$$

But this is precisely (4)! (Note that the result is determined only up to a common factor in the numerator and the denominator).

b. Derive (1) from (4).

This follows the same approach as in **a**, this time starting with (4) and deriving (1). For practice you should fill in the steps.

7. Inverse mapping. The given mapping is a linear fractional transformation. Using (1), p. 743, we have

$$w = \frac{i}{2z - 1} = \frac{0 \cdot z + i}{2z - 1} = \frac{az + b}{cz + d}$$

so that, by comparison,

$$a = 0, \qquad b = i, \qquad c = 2, \qquad d = -1.$$

We now use (4), p. 745, with the values of a, b, c, d just determined and get that the inverse mapping of (1) is

$$z = z(w) = \frac{dw - b}{-cw + a} = \frac{-1 \cdot w - i}{-2 \cdot w + 0} = \frac{-w - i}{-2w}.$$

This compares with the answer in the textbook on p. A41 since

$$\frac{-w - i}{-2w} = \frac{-(w + i)}{-(2w)} = \frac{w + i}{2w}.$$

To check that our answer is correct, we solve $z(w)$ for w and have

$$z = \frac{-w - i}{-2w} \qquad \text{and hence} \qquad z(-2w) = -w - i,$$

or

$$-2wz = -w - i.$$

Adding w gives us

$$-2wz + w = -i \qquad \text{so that} \qquad w(-2z + 1) = -i.$$

We solve the last equation for w and then factor out a minus sign both in the numerator and denominator to get

$$w = \frac{-i}{-2z + 1} = \frac{-(i)}{-(2z - 1)} = \frac{i}{2z - 1}.$$

The last fraction is precisely the given mapping with which we started, which validates our answer.

13. Fixed points. The fixed points of the mapping are those points z that are mapped onto themselves as explained on p. 475. This means for our given mapping we consider

$$\overbrace{16z^5 = w}^{\text{from given mapping}} \underbrace{= f(z) = z}_{\text{by definition of fixed point}}.$$

Hence our task is to solve

$$16z^5 = z \qquad \text{or equivalently} \qquad 16z^5 - z = 0 \qquad \text{so that} \quad z\left(16z^4 - 1\right) = 0.$$

The first root ("fixed point") is immediate, that is, $\boxed{z = 0}$. We then have to solve

(B) $$16z^4 - 1 = 0.$$

For the next fixed point, from basic elementary algebra, we use that

(C) $$x^2 - a^2 = (x - a)(x + a).$$

In (B) we set

$$16z^4 = (4z)^2 = \zeta^2 \qquad \text{to obtain} \qquad \zeta^2 - 1 = 0 \qquad \text{and} \qquad (\zeta^2 - 1)(\zeta^2 + 1) = 0 \quad \text{[by (C)].}$$

Written out we have

$$\left(4z^2 - 1\right)\left(4z^2 + 1\right) = 0. \tag{D}$$

For the first factor in (D) we use (C) again and, setting to zero, we obtain

$$\left(4z^2 - 1\right) = (2z - 1)(2z - 1) = 0$$

so that two more fixed points are

$$2z - 1 = 0 \qquad \text{giving} \qquad \boxed{z = \tfrac{1}{2}} \qquad \text{and similarly} \qquad \boxed{z = -\tfrac{1}{2}}.$$

Considering the second factor in (D)

$$\left(4z^2 + 1\right) = 0 \qquad \text{gives} \qquad z^2 = -\frac{1}{4} \qquad \text{and} \qquad z = \pm\sqrt{-\frac{1}{4}} = \pm\frac{\sqrt{-1}}{\sqrt{4}} = \pm\frac{i}{2}.$$

This means we have two more fixed points

$$\boxed{z = \tfrac{1}{2}i} \qquad \text{and} \qquad \boxed{z = -\tfrac{1}{2}i}.$$

We have found five fixed points, and, since a quintic polynomial has five roots (not necessarily distinct), we know we have found *all* fixed points of w.

Remark. We wanted to show how to solve this problem step by step. However, we could have solved the problem more elegantly by factoring the given polynomial immediately in three steps:

$$16z^5 - z = z\left(16z^4 - 1\right) = z\left(4z^2 - 1\right)\left(4z^2 + 1\right) = z\,(2z - 1)\,(2z + 1)\,(2z + i)\,(2z - i) = 0.$$

Another way is to solve the problem in polar coordinates with (15), p. 617 (whose usage is illustrated in Prob. 21 of Sec. 13.2 on p. 264, in this Manual).

17. Linear fractional transformations (LFTs) with fixed points. In general, fixed points of mappings $w = f(z)$ are defined by

$$w = f(z) = z. \tag{E}$$

LFTs are given by (1), p. 743,

$$w = \frac{az + b}{cz + d}. \tag{1}$$

Taking (E) and (1) together gives the starting point of the *general problem of finding fixed points for LFTs*, that is,

$$w = \frac{az + b}{cz + d} = z.$$

This corresponds to (5), p. 745. We obtain

$$\frac{az+b}{cz+d} - z = 0 \qquad \text{so that} \qquad az + b - z(cz+d) = 0.$$

The last equation can be written as

$$cz^2 + (d-a)z - b = 0 \qquad \text{[see formula in (5), p. 745].} \tag{F}$$

For our problem, we have to find all LFTs with fixed point $z = 0$. This means that (F) must have a root $z = 0$. We have from (F) that

$$z(cz + d - a) = b \tag{F*}$$

and, with the desired fixed point, makes the left-hand side (F*) equal to 0 so that the right-hand side of (F*) must be 0, hence

$$b = 0.$$

Substitute this into (1) gives the answer

$$w = \frac{az+b}{cz+d} = \frac{az+0}{cz+d} = \frac{az}{cz+d}. \tag{G}$$

To check our answer, let us find the fixed points of (G). We have

$$\frac{az}{cz+d} = z \qquad \text{so that} \qquad az = z(cz+d).$$

This gives

$$z(cz + d - a) = 0$$

which clearly has $z = 0$ as one of its roots ("fixed point").

Sec. 17.3 Special Linear Fractional Transformations

The important formula is (2), p. 746 (also stated in Prob. 5 below). It shows that by showing how three points z_1, z_2, z_3 are mapped onto w_1, w_2, w_3 we can derive an LFT in the form (1) of Sec. 17.2. In the textbook, we give six examples of LFTs that are useful in Chap. 18. Note that we allow points to take on the value of ∞ (**infinity**), see **Examples 2, 3,** and **4** on p. 748 and **Probs. 5** and **13**. Moreover, the approach of Sec. 17.3, as detailed in Theorem 1, p. 746, assures us that we obtain a *unique* transformation.

Remark on standard domain. By "standard domains," on p. 747, we mean domains that occur frequently in applications, either directly or after some conformal mapping to another given domain. For instance, this happens in connection with boundary value problems for PDEs in two space variables. The term is not a technical term.

Problem Set 17.3. Page 750

3. **Fixed points.** To show that a transformation and its inverse have the same fixed points we proceed as follows. If a function $w = f(z)$ maps z_1 onto w_1, we have $w_1 = f(z_1)$, and, by definition of the inverse f^{-1}, we also have $z_1 = f^{-1}(w_1)$. Now for a fixed point $z = w = z_1$, we have $z_1 = f(z_1)$, hence $z_1 = f^{-1}(z_1)$, as claimed.

5. Filling in the details of Example 2, p. 748, by formula (2), p. 746. We want to derive the mapping in Example 2, p. 748, from (2), p. 746. As required in Example 2, we set

$$z_1 = 0, z_2 = 1, z_3 = \infty; \qquad w_1 = -1, w_2 = -i, w_3 = 1$$

in

$$\text{(2)} \qquad \frac{w - w_1}{w - w_3} \cdot \frac{w_2 - w_3}{w_2 - w_1} = \frac{z - z_1}{z - z_3} \cdot \frac{z_2 - z_3}{z_2 - z_1}$$

and get

$$\text{(A)} \qquad \frac{w+1}{w-1} \cdot \frac{-i-1}{-i+1} = \frac{z-0}{z-\infty} \cdot \frac{1-\infty}{1-0}.$$

On the left-hand side, we can simplify by (7), p. 610, of Sec. 13.1, and obtain

$$\frac{-i-1}{-i+1} = \frac{-1-i}{1-i} = \frac{-1-i}{1-i} \cdot \frac{1+i}{1+i} = \frac{-1-2i+1}{1^2+1^2} = \frac{-2i}{2} = -i.$$

On the right-hand side, as indicated by Theorem 1, p. 746, we replace

$$\frac{1-\infty}{z-\infty}$$

by 1. Together we obtain, from (A),

$$\frac{w+1}{w-1} \cdot (-i) = \frac{z-0}{1-0} \cdot 1$$

so that

$$\frac{w+1}{w-1} \cdot (-i) = z.$$

This gives us the intermediate result

$$\frac{w+1}{w-1} = \frac{z}{-i} = iz.$$

Note that we used $1/i = -i$ (by Prob. 1, p. 612, and solved on p. 258 in this Manual).

We solve for w and get

$$w + 1 = iz(w-1); \qquad w + 1 = izw - iz; \qquad w - izw = -iz - 1; \qquad w(1 - iz) = -iz - 1,$$

so that

$$w = \frac{-iz-1}{-iz+1} = \frac{\frac{-iz-1}{-i}}{\frac{-iz+1}{-i}} = \frac{z - \frac{1}{-i}}{z + \frac{1}{-1}} = \frac{z + \frac{1}{i}}{z - \frac{1}{i}} = \frac{z-i}{z+i},$$

or, alternatively,

$$w = \frac{-iz-1}{-iz+1} = \frac{i(-iz-1)}{i(-iz+1)} = \frac{z-i}{z+i},$$

both of which lead to the desired result.

13. LFT for given points. Our task is to determine which LFT maps 0, 1, ∞ into ∞, 1, 0.

First solution by Theorem 1, p. 746. By (2), p. 746, with

$$z_1 = 0, z_2 = 1, z_3 = \infty; \qquad w_1 = \infty, w_2 = 1, w_3 = 0$$

we have

$$\text{(B)} \qquad \frac{w-\infty}{w-0} \cdot \frac{1-0}{1-\infty} = \frac{z-0}{z-\infty} \cdot \frac{1-\infty}{1-0}.$$

As required by Theorem 1, we have to replace, on the left-hand side,

$$\frac{w-\infty}{1-\infty} \quad \text{by} \quad 1$$

and also, on the right-hand side,

$$\frac{1-\infty}{z-\infty} \quad \text{by} \quad 1.$$

This simplifies (B)

$$1 \cdot \frac{1-0}{w-0} = \frac{z-0}{1-0} \cdot 1 \qquad \text{so that} \qquad \frac{1}{w} = \frac{z}{1} = z.$$

Hence

$$w = \frac{1}{z}$$

is the desired LFT.

Second solution by inspection. We know that

$$z \ \mapsto \ w = f(z) \qquad [\text{read } z \textit{ gets mapped onto } w = f(z)]$$

and here

$$\begin{aligned} 0 &\mapsto \infty, \\ 1 &\mapsto 1, \\ \infty &\mapsto 0. \end{aligned}$$

Looking at how these three points are mapped, we would conjecture that $w = 1/z$ and see that this mapping does fulfill the three requirements.

17. Mapping of a disk onto a disk. We have to find an LFT that maps $|z| \le 1$ onto $|w| \le 1$ so that $z = i/2$ is mapped onto $w = 0$. From p. 619, we know that $|z| \le 1$ and $|w| \le 1$ represent disks, which leads us to Example 4, pp. 748–749. We set

$$z_0 = \frac{i}{2} \qquad \text{and} \qquad c = \overline{z_0} = \overline{\left(\frac{i}{2}\right)} = \frac{-i}{2},$$

in (3), p. 749, and obtain

$$w = \frac{z - z_0}{cz - 1} = \frac{z - \frac{i}{2}}{\frac{-i}{2}z - 1} = \frac{\frac{2z-i}{2}}{\frac{-iz-2}{2}} = \frac{2z - i}{2} \cdot \frac{2}{-iz - 2} = \frac{2z - i}{-iz - 2}.$$

Complete the answer by sketching the images of the lines $x =$ const and $y =$ const.

19. Mapping of an angular region onto a unit disk. Our task is to find an analytic function, $w = f(z)$, that maps the region $0 \leq \arg z \leq \pi/4$ onto the unit disk $|\, w\, | \leq 1$. We follow Example 6, p. 749, which combines a linear fractional transformation with another transformation. We know, from Example 2, p. 739 of Sec. 17.1, that $t = z^4$ maps the given angular region $0 \leq \arg z \leq \pi/4$ onto the upper t-half-plane. (Make a sketch, similar to Fig. 382, p. 739.) (Note that the transformation $\tau = t^8$ would map the given region onto the full τ-plane, but this would of no help in obtaining the desired unit disk in the next step.)

Next we use (2) in Theorem 1, p. 746, to map that t-half-plane onto the unit disk $|\, w\, | \leq 1$ in the w-plane. We note that this is the inverse problem of the problem solved in Example 3 on p. 748 of the text.

Clearly, the real t-axis (boundary of the half-plane) must be mapped onto the unit circle $|\, w\, | = 1$. Since no specific points on the real t-axis and their images on the unit circle $|\, w\, | = 1$ are prescribed, we can obtain infinitely many solutions (mapping functions).

For instance, if we map $t_1 = -1$, $t_2 = 0$, $t_3 = 1$ onto $w_1 = -1$, $w_2 = -i$, $w_3 = 1$, respectively—a rather natural choice under which -1 and 1 are fixed points—we obtain, with these values inserted into (2) in modified form (2*), that is, inserted into

$$\text{(2*)} \qquad \frac{w - w_1}{w - w_3} \cdot \frac{w_2 - w_3}{w_2 - w_1} = \frac{t - t_1}{t - t_3} \cdot \frac{t_2 - t_3}{t_2 - t_1}$$

equation (C):

$$\text{(C)} \qquad \frac{w + 1}{w - 1} \cdot \frac{-i - 1}{-i + 1} = \frac{t + 1}{t - 1} \cdot \frac{0 - 1}{0 + 1} = \frac{t + 1}{t - 1} \cdot (-1) = \frac{-t - 1}{t - 1}.$$

We want to solve (C) for w. Cross-multiplication and equating leads to

$$(w + 1)(-i - 1)(t - 1) = (w - 1)(-i + 1)(-t - 1).$$

This gives us

$$w(-i - 1)(t - 1) + (-i - 1)(t - 1) = w(-i + 1)(-t - 1) - (-i + 1)(-t - 1)$$

and

$$w(-i - 1)(t - 1) - w(-i + 1)(-t - 1) = -(-i - 1)(t - 1) - (-i + 1)(-t - 1).$$

We get

$$w(-it + i - t + 1 - it - i + t + 1) = it - i + t - 1 - it - i + t + 1,$$

which simplifies to

$$w(-2it + 2) = -2i + 2t \qquad \text{and} \qquad 2w(-it + 1) = 2(-i + t).$$

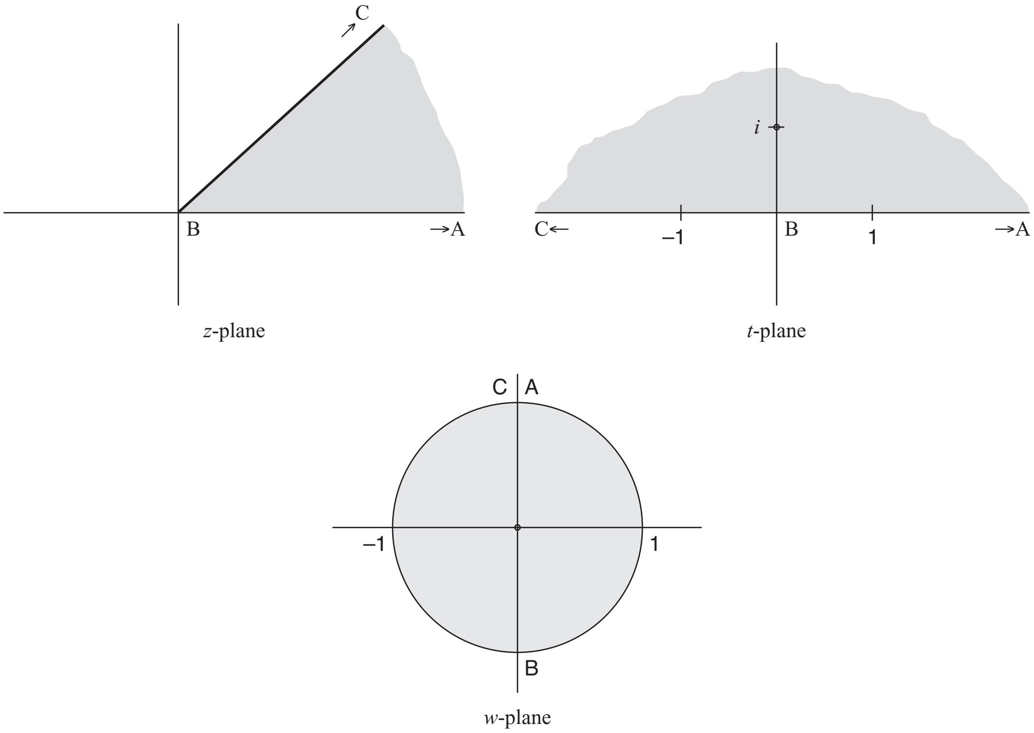
Solving for w gives us

(D)
$$w = \frac{-i + t}{-it + 1} = \frac{t - i}{-it + 1}.$$

From above we know that the mapping $t = z^4$, which substituted into (D). gives us

(E)
$$w = \frac{t - i}{-it + 1} = \frac{z^4 - i}{-iz^4 + 1},$$

which is the answer given on p. A42. Note that the mapping defined by (E) maps $t = i$ onto $w = 0$, the center of the disk.



Sec. 17.3 Prob. 19. z-, t-, and w-planes and regions for the given LFT

Sec. 17.4 Conformal Mapping by Other Functions

We continue our discussion of conformal mappings of the basic analytic functions (from Chap. 13) that we started in Sec. 17.1. It includes $w = \sin z$ (pp. 750–751, **Prob. 11**), $w = \cos z$ (p. 752, **Prob. 21**), $w = \sinh z$ and $w = \cosh z$ (p. 752), and $w = \tan z$ (pp. 752–753). *Take your time to study the examples, as they are quite tricky.*

We expand the concept of transformation by introducing the composition of transformations (see Fig. 394, p. 753, three transformations). It allows us to break a more difficult conformal mapping problem into two or three intermediate conformal mapping problems—one following the other. (*Aside:* We encountered this concept before, but in a different setting, that is, in Sec. 7.9, Composition of Linear Transformations, pp. 316–317 of the textbook, as the more theoretically inclined reader may note.)

Problem Set 17.4. Page 754

3. **Mapping $w = e^z$.** The given region to be mapped by $w = e^z$ is a (solid)

$$\text{rectangle } R \qquad \text{defined by } -\tfrac{1}{2} \le x \le \tfrac{1}{2} \qquad \text{and} \qquad -\pi \le y \le \pi.$$

Since

$$|\,w\,| = |\,e^z\,| = e^x \qquad \text{[by (10), p. 631 in Sec. 13.5],}$$

we have that the inequality

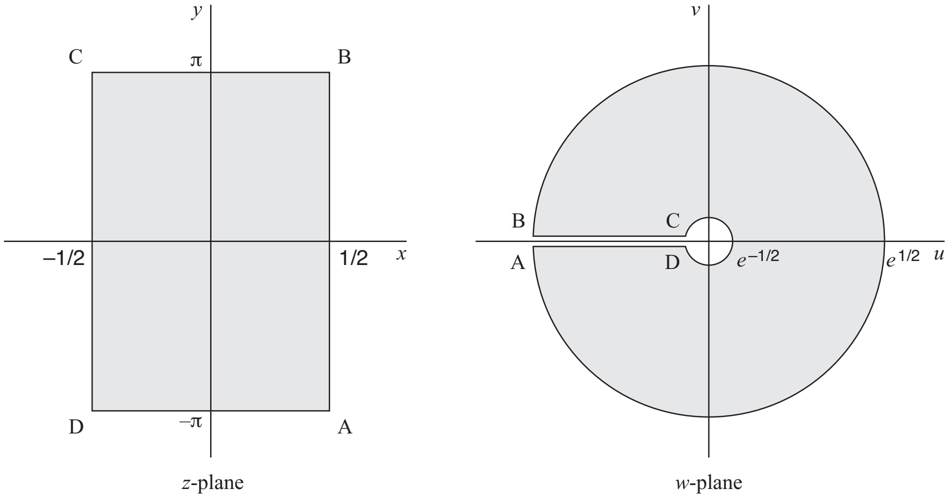
$$-\tfrac{1}{2} \le x \le \tfrac{1}{2} \qquad \text{implies that} \qquad e^{-1/2} = 0.607 \le |\,w\,| \le e^{1/2} = 1.649.$$

This is an annulus in the w-plane with center 0. The inequality

$$-\pi \le y \le \pi$$

gives no further restriction, since y ranges between $-\pi$ and π. Indeed, the side $x = -\frac{1}{2}$ of R is mapped onto the circle $e^{-1/2}$ in the w-plane and the side $x = \frac{1}{2}$ onto the circle of radius $e^{1/2}$. The images of the two horizontal sides of R lie on the real w-axis, extending from $-e^{-1/2}$ to $-e^{1/2}$ and coinciding.

Remark. Take another look at Example 4 in Sec. 17.1 on p. 740 to see how other rectangles are mapped by the complex exponential function.



Sec. 17.4 Prob. 3. Region R and its image

11. **Mapping $w = \sin z$.** The region to be mapped is

$$\text{rectangle } R \qquad \text{given by} \quad 0 < x < \frac{\pi}{2} \quad \text{and} \quad 0 < y < 2.$$

We use the approach of pp. 750–751 of the textbook, which discusses mapping by the complex sine function. We pay attention to Fig. 391 on p. 751. We use

(1) $\quad w = \sin z = \sin x \cosh y + i \cos x \sinh y \qquad$ [p. 750 or (6b), p. 634, in Sec. 13.6].

Since

$$0 < x < \frac{\pi}{2} \qquad \text{we have} \qquad \sin x > 0$$

and, because

$$\cosh y > 0, \qquad \text{we obtain} \qquad u = \sinh x \cosh y > 0.$$

This means that the entire image of R lies in the right half-plane of the w-plane. The

$$\text{origin } z = 0 \qquad \text{maps onto the origin } w = 0.$$

On the bottom edge of the rectangle ($z_A = 0$ to $z_B = \pi/2$)

$$w = \sin x \cosh 0 + \cos x \sinh 0 = \sin x$$

so it goes from $w = 0$ to 1. On the vertical right edge ($z_B = x = \pi/2$ to $z_C = \pi/2 + 2i$)

$$w = \sin \pi/2 \cosh y + i \cos \pi/2 \sinh y = \cosh y$$

so it is mapped from

$$w = 1 \qquad \text{to} \qquad w = \cosh 2 \approx 3.76.$$

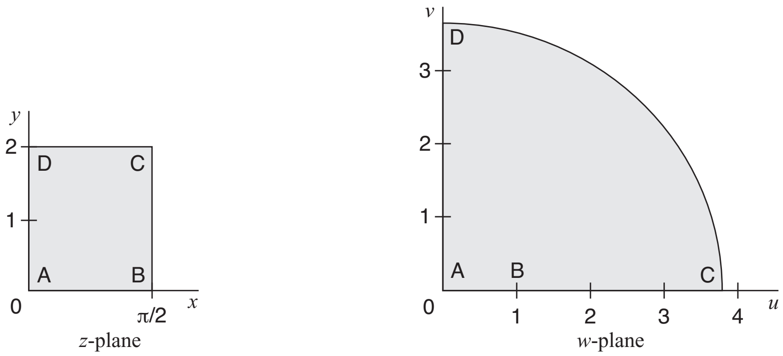
The upper horizontal side $y = 2$, $\pi/2 > x > \ 0$ is mapped onto the upper right part of the ellipse

$$\frac{u^2}{\cosh^2 2} + \frac{v^2}{\sinh^2 2} = 1 \qquad (u > 0),\ (v < 0).$$

Finally, on the left edge of R (from $z_D = 0 + 2i$ to $z_A = 0$),

$$w = \sin 0 \cosh y + i \cos 0 \sinh y = i \sinh y,$$

so it is mapped into the v-axis $u = 0$ from $i \sinh 2$ to 0. Note that, since the region to be mapped consists of the *interior* of a rectangle but not its boundary, the graphs also consist of the interior of the regions without the boundary.



Sec. 17.4 Prob. 11. Rectangle and its image under $w = \sin z$

21. Mapping $w = \cos z$. We note that the rectangle to be mapped is the same as in Prob. 11. We can solve this problem in two ways.

Method 1. Expressing cosine in terms of sine. We relate the present problem to Prob. 11 by using

$$\cos z = \sin\left(z + \tfrac{1}{2}\pi\right).$$

We set

$$t = z + \tfrac{1}{2}\pi.$$

Then the image of the given rectangle [x in $(0, \pi/2)$, y in $(0, 2)$] in the t-plane is bounded by Re t in $\left(\tfrac{1}{2}\pi, x + \tfrac{1}{2}\pi\right)$ or $\left(\tfrac{1}{2}\pi, \pi\right)$, and Im t in $(0, 2)$, i.e. shifted $\pi/2$ to the right. Now

$$w = \sin t = \sin\left(x + \tfrac{1}{2}\pi\right)\cosh y + i\cos\left(x + \tfrac{1}{2}\pi\right)\sinh y.$$

Now proceed as in Prob. 11.

Method 2. Direct solution. To solve directly, we recall that

$$w = \cos z = \cos x \cosh y - i \sin x \sinh y$$

and use z_A, z_B, z_C, and z_D as the four corners of the rectangle as in Prob. 11. Now,

$$\begin{array}{lll} z_A & \text{maps to} & \cos 0 \cosh 0 - i \sin 0 \sinh 0 = 1, \\ z_B & \text{maps to} & \cos \dfrac{\pi}{2} \cosh 0 - i \sin \dfrac{\pi}{2} \sinh 0 = 0, \\ z_C & \text{maps to} & \cos \dfrac{\pi}{2} \cosh 2 - i \sin \dfrac{\pi}{2} \sinh 2 = -i \sinh 2, \\ z_D & \text{maps to} & \cos 0 \cosh 2y - i \sin 0 \sinh 2 = \cosh 2. \end{array}$$

On the bottom edge of the rectangle ($z_A = 0$ to $z_B = \pi/2$)

$$w = \cos x \cosh 0 - i \sin x \sinh 0 = \cos x \quad \text{so it goes from} \quad w = 1 \text{ to } 0.$$

On the vertical right edge ($z_B = x = \pi/2$ to $z_C = \pi/2 + 2i$)

$$w = \cos\frac{\pi}{2}\cosh y - i \sin\frac{\pi}{2}\sinh y = -i \sinh y \quad \text{so it is mapped from} \quad w = 0 \text{ to } w = -i \sinh 2.$$

The upper horizontal side $y = 2, \pi/2 > x > 0$ is mapped onto the lower right part of the ellipse:

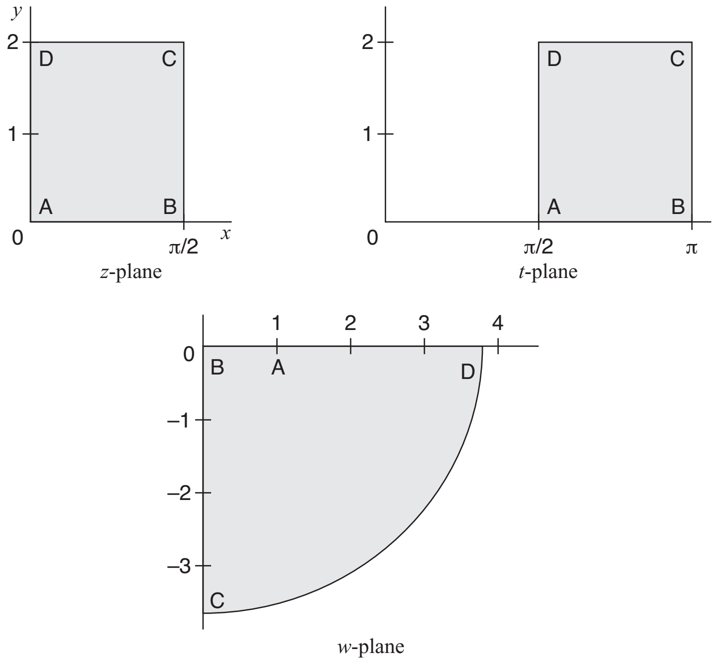
$$\frac{u^2}{\cosh^2 2} + \frac{v^2}{\sinh^2 2} = 1 \qquad (u > 0).$$

Finally, on the left edge of R (from $z_D = 0 + 2i$ to $z_A = 0$)

$$w = \cos 0 \cosh y - i \sin 0 \sinh y = \cosh y,$$

so it is mapped into the v-axis $u = 0$ from $\cosh 2$ to 0.

As in Prob. 11, our solution consists only of the *interior* of the regions depicted.



Sec. 17.4 Prob. 20. Given region in the z-plane and its images in the t- and w-planes for the mapping of $w = \cos z$

Sec. 17.5 Riemann Surfaces. *Optional*

Problem Set 17.5. Page 756

Riemann surfaces (Fig. 395, p. 755) contain an ingenious idea that allows multivalued relations, such as $w = \sqrt{z}$ and $w = \ln z$ (defined in Sec. 13.7, pp. 636–640) to become single-valued. The **Riemann surfaces** (see Fig. 395 on p. 755) consist of several sheets that are connected at certain points ("branch points"). On these sheets, the multivalued relations become single-valued. Thus, for the complex square root being double-valued, the Riemann surface needs two sheets with branch point 0.

1. Square root. We are given that z moves from $z = \frac{1}{4}$ twice around the circle $|z| = \frac{1}{4}$ and want to know what $w = \sqrt{z}$ does.

We use polar coordinates. We set

$$z = re^{i\theta} \qquad \text{[by (6), p. 631 in Sec. 13.5].}$$

On the given circle,

$$|z| = r = \tfrac{1}{4} \qquad \text{[see p. 619 and (3), p. 613],}$$

so that we actually have

$$\text{(A)} \qquad z = \tfrac{1}{4}e^{i\theta}.$$

Hence the given mapping

$$\begin{aligned} w &= \sqrt{z} \\ &= \left(\tfrac{1}{4}e^{i\theta}\right)^{1/2} \qquad \text{[by (A)].} \\ &= \tfrac{1}{2}e^{i\theta/2} \end{aligned}$$

Since z moves twice around the circle $|z| = \frac{1}{4}$,

$$\theta \qquad \text{increases by} \qquad 2 \cdot 2\pi = 4\pi.$$

Hence

$$\frac{\theta}{2} \qquad \text{increases by} \qquad \frac{4\pi}{2} = 2\pi.$$

This means that w goes *once* around the circle $|w| = \frac{1}{2}$, that is, the circle of radius $\frac{1}{2}$ centered at 0 in the w-plane.

Chap. 18 Complex Analysis and Potential Theory

We recall that **potential theory** is the area that deals with finding solutions (that have continuous second partial derivatives)—so-called *harmonic functions*—to Laplace's equation. The question that arises is how do we apply complex analysis and conformal mapping to potential theory. First, *the main idea which links potential theory to complex analysis* is to associate with the *real* potential Φ in the two–dimensional **Laplace's equation**

$$\nabla^2\Phi = \Phi_{xx} + \Phi_{yy} = 0$$

a *complex* potential F

$$F(z) = \Phi(x,\ y) + i\ \Psi(x,\ y). \tag{2}$$

This idea is so powerful because (2) *allows us to model problems in* **distinct** *areas* such as in **electrostatic fields** (Secs. 18.1, p. 759, 18.2, p. 763, 18.5, p. 777), **heat conduction** (Sec. 18.3, p. 767), and **fluid flow** (Sec. 18.4, p. 771). The main adjustment needed, in each different area, is the **interpretation of Φ and** Ψ in (2), specifically the meaning of $\Phi =$ const and its associated conjugate potential $\Psi =$ const. In electrostatic fields, $\Phi =$ const are the electrostatic equipotential lines and $\Psi =$ const are the lines of electrical force—the two types of lines intersecting at right angles. For heat flow, they are isotherms and heat flow lines, respectively. And finally, for fluid flow, they are equipotential lines and streamlines.

Second, we can apply conformal mapping to potential theory because Theorem 1, p. 763 in Sec. 18.2, asserts "closure" of harmonic functions under conformal mapping in the sense that harmonic functions remain harmonic under conformal mapping.

Potential theory is arguably the most important reason for the importance of complex analysis in applied mathematics. Here, in Chap. 18, the third approach to solving problems in complex analysis—the *geometric approach of conformal mapping* applied to solving boundary value problems in two–dimensional potential theory—comes to full fruition.

As background, it is very important that you **remember conformal mapping of basic analytic functions** (power function, exponential function in Sec. 17.1, p. 737, trigonometric and hyperbolic functions in Sec. 17.4, p. 750), and **linear fractional transformations** [(1), p. 743, and (2), p. 746]. For Sec. 18.1, you may also want to review Laplace's equation and Coulomb's law (pp. 400–401 in Sec. 9.7), for Sec. 18.5, Cauchy's integral formula (Theorem 1, p. 660 in Sec. 14.3), and the basics of how to construct Fourier series (see pp. 476–479, pp. 486–487 in Secs. 11.1 and 11.2, respectively).
The chapter ends with a brief **review of complex analysis** in part D on p. 371 of this Manual.

Sec. 18.1 Electrostatic Fields

We know from *electrostatics* that the force of attraction between two particles of opposite or the same charge is governed by Coulomb's law (12) in Sec. 9.7, p. 401. Furthermore, this force is the gradient of a function Φ known as the **electrostatic potential.** Here we are interested in the electrostatic potential Φ because, at any points in the electrostatic field that are free of charge, Φ is the solution of Laplace's equation in 3D:

$$\nabla^2\Phi = \Phi_{xx} + \Phi_{yy} + \Phi_{zz} = 0 \qquad \text{(see Sec. 12.11, pp. 593–594, pp. 596–598).}$$

Laplace's equation is so important that the study of its solutions is called **potential theory.**

Since we want to apply complex analysis to potential theory, we restrict our studies to two dimensions throughout the entire chapter. Laplace's equation in 2D becomes

$$\nabla^2\Phi = \Phi_{xx} + \Phi_{yy} = 0 \qquad \text{(see p. 759 in Sec. 18.1 of textbook).} \tag{1}$$

Then the equipotential surfaces $\Phi(x, y, z) = \text{const}$ (from the 3D case) appear as *equipotential lines* in the xy-plane (Examples 1–3, pp. 759–760).

The next part of Sec. 18.1 introduces the key idea that *it is advantageous to work with* ***complex*** *potentials instead of just real potentials*. The underlying formula for this bold step is

$$F(z) = \Phi(x, y) + i\,\Psi(x, y), \tag{2}$$

where F is the **complex potential** (corresponding to the real potential Φ) and Ψ is the *complex conjugate potential* (uniquely determined except for an additive constant, see p. 629 of Sec. 13.4).

The advantages for using complex potentials F are:

1. It is mathematically easier to solve problems with F in complex analysis because we can use conformal mappings.

2. Formula (2) has a physical meaning. The curves $\Psi = \text{const}$ ("lines of force") intersect the curves $\Phi = \text{const}$ ("equipotential lines") at right angles in the xy-plane because of conformality (p. 738). Illustrations of (2) are given in **Examples 4–6** on p. 761 and in **Probs. 3** and **15.** The section concludes with the method of superposition (**Example 7**, pp. 761–762, **Prob. 11**).

Problem Set 18.1. Page 762

3. Potential between two coaxial cylinders. The first cylinder has radius $r_1 = 10$ [cm] and potential $U_1 = 10$ [kV]. The second cylinder has radius $r_2 = 1$ [m] $= 100$ [cm] and potential $U_2 = -10$ [kV]. From Example 2, p. 759 in Sec. 18.1, we know that the potential $\Phi(r)$ between two coaxial cylinders is given by

$$\Phi(r) = a \ln r + b \qquad \text{where } a \text{ and } b \text{ are to be determined from given boundary conditions.}$$

In our problem we have from the first cylinder with $r_1 = 10$ and $U_1 = 10$

$$\Phi(r_1) = \Phi(10) = a \ln 10 + b = U_1 = 10,$$

so that

$$\Phi(10) = a \ln 10 + b = 10. \tag{C1}$$

Similarly, from the second cylinder we have

$$\Phi(100) = a \ln 100 + b = -10. \tag{C2}$$

We determine a and b. We subtract (C2) from (C1) and use that

$$a \ln 100 = a \ln(10^2) = 2a \ln 10 \tag{L}$$

to get

$$\begin{aligned} \Phi(10) - \Phi(100) &= a \ln 10 + b - (a \ln 100 + b) \\ &= a \ln 10 - a \ln 100 \\ &= a \ln 10 - 2a \ln 10 \qquad \text{[by (L)]} \\ &= -a \ln 10 \qquad \text{[from the r.h.s. of (C1) and (C2)]} \\ &= 10 - (-10) = 20 \qquad \text{[from the l.h.s. of (C1) and (C2)].} \end{aligned}$$

Solving this for a gives

$$-a \ln 10 = 20; \qquad \boxed{a = \frac{-20}{\ln 10}.}$$

We substitute this into (C1) and get

$$a \ln\, 10 + b = \left(\frac{-20}{\ln 10}\right) \ln\, 10 + b = -20 + b = 10 \qquad \text{or} \qquad \boxed{b = 30,}$$

and the real potential is

$$\Phi(r) = a \ln\, r + b = -\left(\frac{20}{\ln 10}\right) \ln\, r + 30.$$

Thus, by Example 5, p. 761, the associated complex potential is

$$F\,(z) = 30 - \left(\frac{20}{\ln 10}\right) \operatorname{Ln} z \qquad \text{where} \qquad \Phi(r) = \operatorname{Re} F\,(z)\,.$$

11. Two source lines. Verification of Example 7, pp. 761–762. The equipotential lines in Example 7, p. 761, are

$$\left|\frac{z-c}{z+c}\right| = k = \text{const} \qquad (k \text{ and } c \text{ real}).$$

Hence

$$|z - c\,| = k\,|\,z + c\,|.$$

We square both sides and get

(A) $$|\,z - c\,|^2 = K\,|\,z + c\,|^2 \qquad \text{where } K \text{ is a constant (and equal to } k^2).$$

We note that, by (3), p. 613,

$$|\,z - c\,|^2 = |\,x + iy - c\,|^2 = |\,(x - c) + iy\,|^2 = (x - c)^2 + y^2 \qquad \text{and} \qquad |\,z + c\,|^2 \;= (x + c)^2 + y^2.$$

Using this, and writing (A) in terms of the real and imaginary parts and taking all the terms to the left, we obtain

$$(x - c)^2 + y^2 - K\,[(x + c)^2 + y^2] = 0.$$

Writing out the squares gives

(B) $$x^2 - 2\,c\,x + c^2 + y^2 - K\,(x^2 + 2\,c\,x + c^2 + y^2) = 0.$$

We consider two cases. First, consider $k = 1$, hence $K = 1$, most terms in (B) cancel, and we are left with

$$-4\,c\,x = 0 \qquad \text{hence} \qquad x = 0 \;\text{ (because } c \neq 0).$$

This is the y-axis. Then

$$|\,z-c\,|^2 = |\,z+c\,|^2 = y^2 + c^2, \qquad \frac{|z-c|}{|z+c|} = 1, \qquad \text{Ln } 1 = 0.$$

This shows that the y-axis has potential 0.

We can now continue with (B), assuming that $K \neq 1$. Collecting terms in (B), we have

$$(1-K)\,(x^2+y^2+c^2) - 2\,c\,x\,(1+K) = 0.$$

Division by $1-K$ ($\neq 0$ because $K \neq 1$) gives

$$x^2 + y^2 + c^2 - 2\,L\,x = 0 \qquad \text{where} \qquad L = \frac{c\,(1+K)}{1-K}.$$

Completing the square in x, we finally obtain

$$(x-L)^2 + y^2 = L^2 - c^2.$$

This is a circle with center at L on the real axis and radius $\sqrt{L^2-c^2}$.

We simplify $\sqrt{L^2-c^2}$ as follows. First, we consider

$$\begin{aligned} L^2 - c^2 &= \left[\frac{c\,(1+K)}{1-K}\right]^2 - c^2 \qquad \text{(by inserting } L\text{)} \\ &= \frac{c^{\,2}(1+K)^2}{(1-K)^2} - c^2 \\ &= c^2\left[\frac{(1+K)^2}{(1-K)^2} - 1\right] \\ &= c^2\left[\frac{(1+K)^2}{(1-K)^2} - \frac{(1-K)^2}{(1-K)^2}\right] \\ &= c^2\,\frac{1+2K+K^2-(1-2K-K^2)}{(1-K)^2} \\ &= \frac{c^2 4K^2}{(1-K)^2}. \end{aligned}$$

Hence

$$\sqrt{L^2-c^2} = \sqrt{\frac{c^2 4K^2}{(1-K)^2}} = \frac{c2K}{1-K} = \frac{2ck^2}{1-k^2} \qquad \text{(using } K = k^2\text{)}.$$

Thus the radius equals $2ck^2/(1-k^2)$.

15. Potential in a sector. To solve the given problem, we note that

$$z^2 = (x+iy)^2 = x^2 - y^2 + 2ixy$$

gives the potential in sectors of opening $\pi/2$ bounded by the bisecting straight lines of the quadrants because

$$x^2 - y^2 = 0 \qquad \text{when} \qquad y = \pm x.$$

Similarly, higher powers of z give potentials in sectors of smaller openings on whose boundaries the potential is zero. For

$$z^3 = (x + iy)^3 = x^3 + 3ix^2y - 3xy^2 - iy^3$$

the real potential is

$$\Phi_0 = \text{Re } z^3 = x^3 - 3xy^2 = x(x^2 - 3y^2)$$

and

$$\Phi = 0 \qquad \text{when} \qquad y = \pm \frac{x}{\sqrt{3}};$$

these are the boundaries given in the problem, the opening of the sector being $\pi/3$, that is, 60°.

To satisfy the other boundary condition, multiply Φ_0 by 220 [V] and get

$$\Phi = 220\,(x^3 - 3xy^2) = \text{Re}(220\,z^3) \text{ [V]}.$$

Sec. 18.2 Use of Conformal Mapping. Modeling

Here we experience, for the first time, the full power of applying the *geometric approach of conformal mappings* to boundary value problems ("Dirichlet problems," p. 564, p. 763) in two-dimensional potential theory. Indeed, we continue to solve problems of **electrostatic potentials in a complex setting (2), p. 760** (see **Example 1**, p. 764, **Example 2**, p. 765; **Probs. 7** and **17**). However, now we apply conformal mappings (defined on p. 738 in Sec. 17.1) with the purpose of simplifying the problem by mapping a given domain onto one for which the solution is known or can be found more easily. This solution, thus obtained, is mapped back to the given domain.

Our approach of using conformal mappings is theoretically sound and, if applied properly, will give us correct answers. Indeed, Theorem 1, p. 763, assures us that if we apply any conformal mapping to a given harmonic function then the resulting function is still harmonic. [Recall that *harmonic functions* (p. 460 in Sec. 10.8) are those functions that are solutions to Laplace's equation (from Sec. 18.1) and have continuous second-order partial derivatives.]

Problem Set 18.2. Page 766

7. Mapping by *w* = sin *z*. Look at Sec. 17.4, pp. 750–751 (also Prob. 11, p. 754 of textbook and solved on p. 348 of this Manual) for the conformal mapping by

$$w = u + iv = \sin z = \sin x\, \cosh y + i \cos x\, \sinh y.$$

We conclude that the lower side (z_A to z_B) $0 < x < \pi/2$ ($y = 0$) of the given rectangle D maps onto $0 < u < 1$ ($v = 0$) because $\cosh 0 = 1$ and $\sinh 0 = 0$. The right side (z_B to z_C) $0 < y < 1$ ($x = \pi/2$) maps onto $1 < u < \cosh \pi/2$ ($v = 0$). The upper side (z_C to z_D) maps onto a quarter of the ellipse

$$\frac{u^2}{\cosh^2 1} + \frac{v^2}{\sinh^2 1} = 1$$

in the first quadrant of the w-plane. Finally, the left side (z_D to z_A) maps onto $\sinh 1 > v > 0$ $(u = 0)$.

Now the given potential is

$$\begin{aligned} \Phi^*(u, v) &= u^2 - v^2 \\ &= \sin^2 x \ \cosh^2 y - \cos^2 x \ \sinh^2 y. \end{aligned}$$

Hence $\Phi = \sin^2 x$ on the lower side ($y = 0$), and grows from 0 to 1. On the right side, $\Phi = \cosh^2 y$, which increases from 1 to $\cosh^2 1$.

On the upper side we have the potential

$$\Phi = \sin^2 x \ \cosh^2 1 - \cos^2 x \ \sinh^2 1,$$

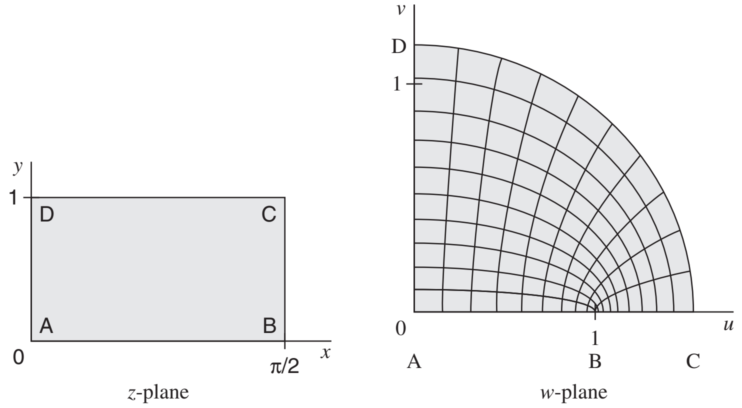
which begins with the value $\cosh^2 1$ and decreases to $-\sinh^2 1$. Finally, on the left side it begins with $-\sinh^2 1$ and returns to its value 0 at the origin.

Note that any $y = c$ maps onto an ellipse

$$\frac{u^2}{\cosh^2 c} + \frac{v^2}{\sinh^2 c} = 1$$

and any $x = k$ maps onto an hyperbola

$$\frac{u^2}{\sin^2 k} - \frac{v^2}{\cos^2 k} = 1.$$



Sec. 18.2 Prob. 7. Given region and image under conformal mapping $w = \sin z$

17. Linear fractional transformation. We want to find a linear fractional transformation (LTF) $z = g(Z)$ that maps $|Z| \leq 1$ onto $|z| \leq 1$ with $Z = i/2$ being mapped onto $z = 0$. Our task is to find an LTF with such properties. The candidate is the LTF defined by (3) on p. 749 in Sec. 17.3. Here Z plays the role of z in (3), and z plays the role of w. Thus,

(A) $$z = \frac{Z - \frac{i}{2}}{-\frac{i}{2}Z - 1}.$$

We can multiply both the numerator and denominator in (A) by 2 and get the answer on p. A43 in App. 2:

$$z = \frac{2Z - i}{-iZ - 2}. \tag{A2}$$

To complete the problem, we evaluate (A2) with $Z = 0.6 + 0.8i$ and $-0.6 + 0.8i$, respectively. We get for $Z = 0.6 + 0.8i$

$$z = \frac{2Z - i}{-iZ - 2} = \frac{2(0.6 + 0.8i) - i}{-i(0.6 + 0.8i) - 2} = \frac{1.2 + 1.6i - i}{-0.6i + 0.8 - 2} = \frac{1.2 + 0.6i}{-(0.6i + 1.2)} = -1, \tag{B}$$

which is the desired value. Similarly, you can show that for $Z = 0.6 - 0.8i$, one gets $z = 1$. Thus

$$|Z| = |0.6 \pm 0.8i| = \sqrt{0.6^2 + 0.8^2} = 1 \leq 1,$$

which means that $|Z| \leq 1$. And (B), with a similar calculation, shows that our chosen Z's get mapped by (A2) onto $z = \pm 1$, so that indeed $|z| \leq 1$. Together, this shows that (A2) is the desired LTF as described in Prob. 17 and illustrated in Fig. 407, p. 766. Convince yourself that Fig. 407 is correct.

Sec. 18.3 Heat Problems

Complex analysis can model **two-dimensional heat problems** *that are independent of time.* From the top of p. 564 in Sec.12.6, we know that the heat equation is

$$T_t = c^2 \nabla^2 T. \tag{H}$$

We assume that the heat flow is independent of time ("steady"), which means that $T_t = 0$. Hence (H) reduces to Laplace's equation

$$\nabla^2 T = T_{xx} + T_{yy} = 0. \tag{1}$$

This allows us to introduce methods of complex analysis because T [or $T(x, y)$] is the real part of the **complex heat potential**

$$F(z) = T(x, y) + i\,\Psi(x, y).$$

[Terminology: $T(x, y)$ is called the **heat potential**, $\Psi(x, y) =$ const are called **heat flow lines**, and $T(x, y) =$ const are called **isotherms**.]

It follows that we can reinterpret all the examples of Secs. 18.1 and 18.2 in electrostatics as problems of heat flow (p. 767). This is another great illustration of **Underlying Theme 3** on p. ix of the textbook of the powerful unifying principles of engineering mathematics.

Problem Set 18.3. Page 769

7. Temperature in thin metal plate. A potential in a sector (in an angular region) whose sides are kept at constant temperatures is of the form

$$\begin{aligned} T(x, y) &= a\,\theta + b \\ &= a \arctan \frac{y}{x} + b \qquad \text{(A)} \\ &= a \operatorname{Arg} z + b \qquad \text{(see similar Example 3 on pp. 768–769).} \end{aligned}$$

Here we use the fact that

$$\operatorname{Arg} z = \theta = \operatorname{Im}(\operatorname{Ln} z) \qquad \text{is a harmonic function.}$$

The two constants, a and b, can be determined from the given values on the two sides $\operatorname{Arg} z = 0$ and $\operatorname{Arg} z = \pi/2$. Namely, for $\operatorname{Arg} z = 0$ (the x-axis) we have

$$T = b = T_1.$$

Then for $\operatorname{Arg} z = \pi/2$ we have

$$T = a \cdot \frac{\pi}{2} + T_1 = T_2.$$

Solving for a gives

$$a = \frac{2(T_2 - T_1)}{\pi}.$$

Hence a potential giving the required values on the two sides is

$$T(x, y) = \frac{2(T_2 - T_1)}{\pi} \operatorname{Arg} z + T_1.$$

Complete the problem by finding the associated complex potential $F(z)$ obeying
$\operatorname{Re} F(z) = T(x, y)$ and check on p. A43 in App. 2 of the textbook.

15. Temperature in thin metal plate with portion of boundary insulated. Mixed boundary value problem. We start as in Prob. 7 by noting that a potential in an angular region whose sides are kept at constant temperatures is of the form

$$\text{(B)} \qquad T(x, y) = a \operatorname{Arg} z + b,$$

and using the fact that $\operatorname{Arg} z = \theta = \operatorname{Im}(\operatorname{Ln} z)$ is a harmonic function. We determine the values for the two constants a and b from the given values on the two sides $\operatorname{Arg} z = 0$ and $\operatorname{Arg} z = \pi/4$. For $\operatorname{Arg} z = 0$ (the x-axis) we have $T = b = -20$ and for $\operatorname{Arg} z = \pi/4$ we have

$$T = a \cdot \frac{\pi}{4} - 20 = 60 \qquad \text{so that} \qquad a = \frac{320}{\pi}.$$

Hence a potential that satisfies the conditions of the problem is

$$\text{(C)} \qquad T = \frac{320}{\pi} \operatorname{Arg} z - 20.$$

Now comes an important observation. The curved portion of the boundary (a circular arc) is insulated. Hence, on this arc, the normal derivative of the temperature T must be zero. But the normal direction is the radial direction; so the partial derivative with respect to r must vanish. Now formula (C) shows that T is independent of r, that is, the condition under discussion is automatically satisfied. (If this were not the case, the whole solution would not be valid.)

Finally we derive the complex potential F. From Sec. 13.7 we recall that

$$\text{(D)} \qquad \operatorname{Ln} z = \ln |z| + i \operatorname{Arg} z \qquad \text{[by (2), p. 637].}$$

Hence for $\operatorname{Arg} z$ to become the real part (as it must be the case because $F = T + i\ \Psi$), we must multiply both sides of (D) by $-i$. Indeed, then

$$-i\ \operatorname{Ln} z = -i\ \ln|z| + \operatorname{Arg} z.$$

Hence from this and (C) we see that the desired complex heat potential is

$$\text{(E)} \qquad \begin{aligned} F(z) &= -20 + \frac{320}{\pi}(-i\ \operatorname{Ln} z) \\ &= -20 - \frac{320}{\pi} i\ \operatorname{Ln} z, \end{aligned}$$

which, by (C) and (E), leads to the answer given on p. A43 in App. 2 of the textbook.

Sec. 18.4 Fluid Flow

The central formula of Sec. 18.4 is on p. 771:

$$\text{(3)} \qquad V = V_1 + i\ V_2 = \overline{F'(z)}.$$

It derives its importance from relating the complex **velocity** vector of the fluid flow

$$\text{(1)} \qquad V = V_1 + iV_2$$

to the **complex potential** of the fluid flow

$$\text{(2)} \qquad F(z) = \Phi(x,\ y) + i\ \Psi(x,\ y),$$

whose imaginary part Ψ gives the streamlines of the flow in the form

$$\Psi(x,\ y) = \text{const.}$$

Similarly, the real part Φ gives the equipotential lines of the flow:

$$\Phi(x,\ y) = \text{const.}$$

The use of (3), p. 771, is illustrated in different flows in **Example 1** ("flow around a corner," p. 772), **Prob. 7** ("parallel flow") and in **Example 2**, and **Prob. 15** ("flow around a cylinder").

Flows may be compressible or incompressible, rotational or irrotational, or may differ by other general properties. We reach the connection to complex analysis, that is, Laplace's equation (5) applied to Φ and Ψ of (2), written out

$$\text{(5)} \qquad \nabla^2\Phi = \Phi_{xx} + \Phi_{yy} = 0, \qquad \nabla^2\Psi = \Psi_{xx} + \Psi_{yy} = 0 \qquad \text{[on p. 772]}$$

by first assuming the flow to be incompressible and irrotational (see Theorem 1, p. 773).

Rotational flows can be modeled to some extent by complex logarithms, as shown in the textbook on pp. 776–777 in the context of a Team Project.

We encounter a third illustration in Chap. 18 of **Underlying Theme 3** of the textbook on p. ix because the model (2), p. 760, developed for electrostatic potentials is now the model for fluid flow. More details are given in the paragraph on "basic comment on modeling" on p. 766 in Sec. 18.2.

Problem Set 18.4. Page 776

7. **Parallel flow.** Our task is to interpret the flow with complex potential

$$F(z) = z.$$

We start by noting that a flow is completely determined by its complex potential

$$F(z) = \Phi(x,\ y) + i\ \Psi(x,\ y) \qquad \text{(p. 771).} \tag{2}$$

The stream function Ψ gives the streamlines $\Psi = \text{const}$ and is generally more important than the velocity potential Φ, which gives the equipotential lines $\Phi = \text{const}$. The flow can best be visualized in terms of the velocity vector V, which is obtained from the complex potential in the form (3), p. 771,

$$V = V_1 + i\ V_2 = \overline{F'(z)}. \tag{3}$$

(We need a special vector notation, in this case, because a complex function V can always be regarded as a vector function with components V_1 and V_2.)

Hence, for the given complex potential

$$F(z) = z = x + iy, \tag{A}$$

we have

$$F'(z) = 1 \qquad \text{and} \qquad \overline{F'(z)} = 1 + 0i;$$

thus,

$$V = V_1 = 1 \qquad \text{and} \qquad V_2 = 0. \tag{B}$$

The velocity vector in (B) is parallel to the x-axis and is positive, i.e., $V = V_1$ points to the right (in the positive x-direction).

Hence we are dealing with a uniform flow (a flow of constant velocity) that is parallel (the streamlines are straight lines parallel to the x-axis) and is flowing to the right (because V is positive). From (A) we see that the equipotential lines are vertical parallel straight lines; indeed,

$$\Phi(x, y) = \text{Re}\ F(z) = x = \text{const}; \qquad \text{hence} \qquad x = \text{const}.$$

Using our discussion, sketch the flow.

15. **Flow around a cylinder.** Here we are asked to change $F(z)$ in Example 2, p. 772, slightly to obtain a flow around a cylinder of radius r_0 that gives the flow in Example 2 if $r_0 \to 1$.

Solution. Since a cylinder of radius r_0 is obtained from a cylinder of radius 1 by a dilatation (a uniform stretch or contraction in all directions in the complex plane), it is natural to replace z by az with a real constant a because this corresponds to such a stretch. That is, we replace the complex potential

$$z + \frac{1}{z}$$

in Example 2, p. 772, by

$$\begin{aligned} F(z) &= \Phi(r,\theta) + i\ \Psi(r,\theta) \\ &= az + \frac{1}{az} \\ &= are^{i\,\theta} + \frac{1}{ar}e^{-i\,\theta} \qquad \text{[by (6), p. 631 applied to both terms].} \end{aligned}$$

The stream function Ψ is the imaginary part of F. Since, by Euler's formula,

$$e^{\pm i\,\theta} = \cos\,\theta \pm i\ \sin\,\theta \qquad \text{[by (5), p. 631 in Sec. 13.5]}$$

we obtain

$$\begin{aligned} \Psi(r,\theta) &= \operatorname{Im}(F) \\ &= \operatorname{Im}\left(are^{i\,\theta} + \frac{1}{ar}e^{-i\,\theta}\right) \\ &= \operatorname{Im}\left[ar\,(\cos\theta + i\ \sin\theta\,) + \frac{1}{ar}(\cos\theta - i\ \sin\theta)\right] \qquad \text{(by Euler's formula applied twice)} \\ &= \operatorname{Im}\left[\left(ar\cos\theta + \frac{1}{ar}\cos\theta\right) + \left(ar\,i\ \sin\theta - \frac{1}{ar}i\ \sin\,\theta\right)\right] \quad \text{(regrouping for imaginary part)} \\ &= ar\ \sin\theta - \frac{1}{ar}\sin\theta \\ &= \left(ar - \frac{1}{ar}\right)\sin\theta. \end{aligned}$$

The streamlines are the curves $\theta = \text{const}$. As in Example 2 of the text, the streamline $\Psi = 0$ consists of the x-axis ($\theta = 0$ and π), where $\sin\,\theta = 0$, and of the locus where the other factor of Ψ is zero, that is,

$$a\,r - \frac{1}{a\,r} = 0, \qquad \text{thus} \qquad (a\,r)^2 = 1 \qquad \text{or} \qquad a = \frac{1}{r}.$$

Since we were given that the cylinder has radius $r = r_0$, we must have

$$a = \frac{1}{r_0}.$$

With this, we obtain the answer

$$F(z) = a\,z + \frac{1}{a\,z} = \frac{z}{r_0} + \frac{r_0}{z}.$$

Sec. 18.5 Poisson's Integral Formula for Potentials

The beauty of this section is that it brings together various material from complex analysis and Fourier analysis. The section applies Cauchy's integral formula (1), p. 778 (see Theorem 1, p. 660 in Sec. 14.3), to a complex potential $F(z)$ and uses it on p. 778 to derive Poisson's integral formula (5), p. 779.

Take a look at pp. 779–780. Formula (5) yields the potential in a disk D. Ordinarily such a disk has a continuous boundary $|z| = R$, which is a circle. However, this requirement can be loosened: (5) is applicable *even if the boundary is only piecewise continuous*, such as in Figs. 405 and 406 of a typical example of a potential between two semicircular plates (Example 2 on p. 765).

From (5) we obtain the potential in a region R by mapping R conformally onto D, solving the problem in D, and then using the mapping to obtain the potential in R. The latter is given by the important formula (7), p. 780,

$$\Phi(r, \theta) = a_0 + \sum_{n=1}^{\infty} \left(\frac{r}{R}\right)^n (a_n \cos nx + b_n \sin nx) \tag{7}$$

where, typically, we consider the potential over the disk $r < R$.

In **Example 1**, p. 780, and Probs. 5–18 (with **Probs. 7** and **13** solved below) the potential $\Phi(r, \theta)$ in the unit disk is calculated. In particular note, for the unit disk $r < 1$ and given boundary function $\Phi(1, \theta)$, we have that

$$r = R \quad (= 1) \qquad \text{so that in (7)} \qquad \left(\frac{r}{R}\right)^n = \left(\frac{r}{r}\right)^n = 1$$

and (7) simplifies to a genuine Fourier series:

$$(7') \qquad \Phi(r, \theta) = a_0 + \sum_{n=1}^{\infty} (a_n \cos n\theta + b_n \sin n\theta).$$

To determine (7′) requires that we compute the Fourier coefficients of (7) by (8), p. 780, under the simplification of $r = R$. Hence the **techniques of calculating Fourier series** explained in Sec. 11.1, pp. 474–483 of the textbook and pp. 202–208 of Vol. 1 of this Manual, and furthermore, in Sec. 11.2, pp. 483–491 of the textbook and pp. 208–211 of Vol. 1 of this Manual come into play. This is illustrated in **Prob. 13** and **Example 1.**

Problem Set 18.5. Page 781

7–19. Harmonic functions in a disk. In each of **7–19 Harmonic functions in a disk.** In each of **Probs. 7–19** we are given a boundary function $\Phi(1, \theta)$. Then, using (7), p. 780, and related formula (8), we want to find the potential $\Phi(r, \theta)$ in the open unit disk $r < 1$ and compute some values of $\Phi(r, \theta)$ as well as sketch the equipotential lines. We note that, typically, these problems are solved by Fourier series as explained above.

7. Sinusoidal boundary values lead to a series (7) that, in this problem, reduces to finitely many terms (a "trigonometric polynomial"). The given boundary function

$$\Phi(1, \theta) = a \cos^2 4\theta$$

is not immediately one of the terms in (7), but we can express it in terms of a cosine function of multiple angle as follows. Indeed, in App. 3, p. A64 of the textbook, we read

$$\cos^2 x = \tfrac{1}{2} + \tfrac{1}{2} \cos 2x. \tag{10}$$

In (10), we set

$$x = 4\theta$$

and get

$$\cos^2 4\theta = \tfrac{1}{2} + \tfrac{1}{2}\cos 8\theta.$$

Hence we can write the boundary function as

$$\begin{aligned}\Phi(1,\ \theta) &= a\ \cos^2 4\theta \\ &= a\left(\frac{1}{2} + \frac{1}{2}\cos 8\theta\right) = \frac{a}{2} + \frac{a}{2}\cos 8\theta.\end{aligned}$$

From (7) we now see immedately that the potential in the unit disk satisfying the given boundary condition is

$$\Phi(r,\ \theta) = \frac{a}{2} + \frac{a}{2} r^8 \cos 8\theta.$$

Note that the answer is already in the desired form so we do not need to calculate the Fourier coefficients by (8)!

13. **Piecewise linear boundary values** given by

$$\Phi(1,\ \theta) = \begin{cases}\theta & -\frac{\pi}{2} < \theta < \frac{\pi}{2} \\ 0 & \text{otherwise}\end{cases}$$

lead to a series (7) whose coefficients are given by (8).

We follow Example 1, p. 780, and calculate the Fourier coefficients (8). Because the function $\Phi(1,\ \theta)$ is an odd function (see p. 486), we know that its Fourier series reduce to a Fourier sine series so that all the $a_n = 0$. From (8) we obtain

$$\begin{aligned}b_n &= \frac{2}{\pi}\int_0^{\pi/2} \theta \sin n\theta\, d\theta \qquad n = 1, 2, 3, \ldots \\ &= \frac{2}{\pi}\left[\frac{\sin n\theta - n\theta\cos n\theta}{n^2}\right]_0^{\pi/2} \\ &= \frac{2}{\pi}\left(\frac{\sin \frac{1}{2}n\pi - \frac{1}{2}n\pi\cos\frac{1}{2}n\pi}{n^2} - \frac{\sin n0 - n0\cos n0}{n^2}\right) \\ &= \frac{2\sin\frac{1}{2}n\pi - n\pi\cos\frac{1}{2}n\pi}{n^2\pi}.\end{aligned}$$

For $n = 1$ this simplifies to

$$b_1 = \frac{2\sin\frac{1}{2}\pi - \pi\cos\frac{1}{2}\pi}{\pi} = \frac{2\cdot 1 - \pi\cdot 0}{\pi} = \frac{2}{\pi}.$$

For $n = 2, 3, 4, \ldots$, we get the following values for the Fourier coefficients:

$$b_2 = \frac{2\sin\pi - 2\pi\cos\pi}{2^2\pi} = \frac{2\cdot 0 - 2\pi\cdot(-1)}{2^2\pi} = \frac{2\pi}{4\pi} = \frac{1}{2},$$

$$b_3 = \frac{2\sin\frac{1}{2}3\pi - 3\pi\cos\frac{1}{2}3\pi}{3^2\pi} = \frac{2\cdot(-1) - 3\pi\cdot 0}{3^2\pi} = -\frac{2}{9\pi},$$

$$b_4 = \frac{2\sin 2\pi - 4\pi\cos 2\pi}{4^2\pi} = \frac{2\cdot 0 - 4\pi\cdot 1}{4^2\pi} = \frac{-4\pi}{4^2\pi} = -\frac{1}{4},$$

$$\cdots$$

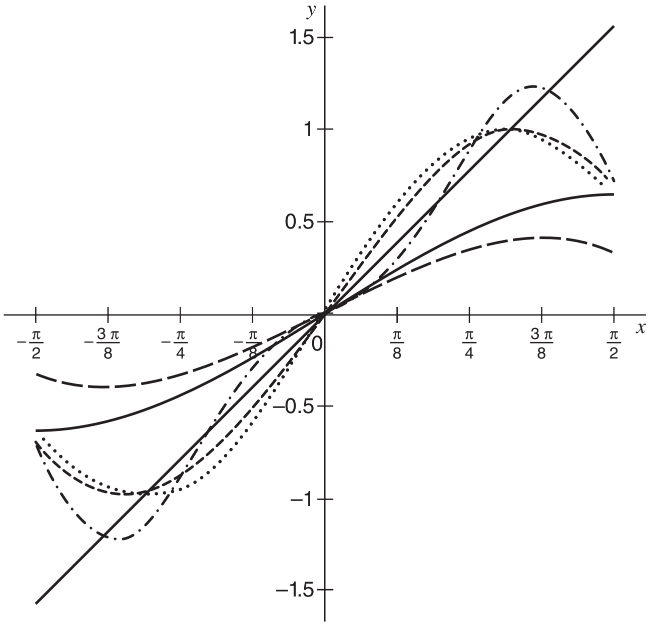
Observe that in computing b_n for n odd, the cos terms are zero, while for n even, the sin terms are zero. Hence putting it together

$$\Phi(1,\, \theta) = \frac{2}{\pi}\sin\theta + \frac{1}{2}\sin 2\theta - \frac{2}{9\pi}\sin 3\theta - \frac{1}{4}\sin 4\theta + + - - \cdots .$$

From this, we obtain the potential (7) in the disk ($R = 1$) in the form

$$\text{(A)} \qquad \Phi(r,\, \theta) = \frac{2}{\pi}r\sin\theta + \frac{1}{2}r^2\sin 2\theta - \frac{2}{9\pi}r^3\sin 3\theta - \frac{1}{4}r^4\sin 4\theta + + - - \cdots .$$

The following figure shows the given boundary potential (straight line), an approximation of it [the sum of the first, first two, first three, and first four terms (dot dash) of the series (A) with $r = 1$] along with an approximation of the potential on the circle of radius $r = \frac{1}{2}$ (the sum of those four terms for $r = \frac{1}{2}$ drawn with a long dash). Make a sketch of the disk (a circle) and indicate the boundary values around the circle.



Sec. 18.5 Prob. 13. Boundary potential and approximations for $r = 1$ and $r = \frac{1}{2}$

Sec. 18.6 General Properties of Harmonic Functions. Uniqueness Theorem for the Dirchlet Problem

Recall three concepts (needed in this section): *analytic functions* (p. 623) are functions that are defined and differentiable at every point in a domain D. Furthermore, one is able to test whether a function is analytic by the two very important Cauchy–Riemann equations on p. 625. *Harmonic functions* (p. 460) are functions that are solutions to Laplace's equation $\nabla^2\Phi = 0$ and their second-order partial derivatives are continuous. Finally, a *Dirichlet problem* (p. 564) is a boundary value problem where the *values of the function are prescribed* (*given*) ***along the boundary.***

The material is very accessible and needs some understanding of how to evaluate double integrals and also apply Cauchy's integral formula (Sec. 14.3, p. 660). We derive general properties of harmonic functions from analytic functions. Indeed, the first two mean value theorems go together, in that **Theorem 1**, (p. 781; **Prob. 3**) is for analytic functions and leads directly to **Theorem 2** (p. 782; **Prob. 7**) for harmonic functions. Similarly, **Theorems 3** and **4** are related to each other. Of the general properties of harmonic functions, the **maximum principle** of Theorem 4, p. 783, is quite important. The chapter ends on a high note with **Theorem 5**, p. 784, which states that an existing solution to a Dirichlet problem for the 2D Laplace equation must be unique.

Orientation. We have reached the end of Part D on complex analysis, a field whose diversity of topics and richness of ideas may represent a challenge to the student. Thus we include, for study purposes, **a brief review of complex analysis on p. 371** of this Manual.

Problem Set 18.6. Page 784

3. Mean value of an analytic function. Verification of Theorem 1, p. 781, for given problem. The problem is to verify that Theorem 1, p. 781, holds for

$$\text{(A)} \qquad F(z) = (3z-2)^2, \qquad z_0 = 4, \qquad |z-4| = 1.$$

Solution. We have to verify that

$$\text{(2)} \qquad F(z_0) = \frac{1}{2\pi}\int_0^{2\pi} F(z_0 + re^{i\alpha})\,d\alpha$$

holds for (A). Here we integrate $F(z) = (3z-2)^2$ around the circle, $|z-4| = 1$, of radius $r = 1$ and center $z_0 = 4$, and hence we have to verify (2) with these values. This means we have to show that

$$\text{(2*)} \qquad F(z_0) = F(4) = \frac{1}{2\pi}\int_0^{2\pi} F(z_0 + re^{i\alpha})\,d\alpha = \frac{1}{2\pi}\int_0^{2\pi} F(4 + 1\cdot e^{i\alpha})\,d\alpha.$$

Since

$$F(z_0) = F(4) = (3\cdot 4 - 2)^2 = (10)^2 = 100,$$

we have to show that the integral on the right-hand side of (2*) takes on that value of 100, that is, we must show that

$$\text{(2**)} \qquad \frac{1}{2\pi}\int_0^{2\pi} F(4 + 1\cdot e^{i\alpha})\,d\alpha = 100.$$

We go in a stepwise fashion. The path of integration is the circle, $|z-4|=1$, so that

$$z = z_0 + re^{i\alpha} = 4 + 1 \cdot e^{i\alpha} = 4 + e^{i\alpha}.$$

Hence, on this path, the integrand is

$$\begin{aligned} F(z_0 + e^{i\alpha}) &= \left(3\left[4+e^{i\alpha}\right]-2\right)^2 = \left(12+3e^{i\alpha}-2\right)^2 \\ &= \left(10+3e^{i\alpha}\right)^2 \\ &= 100 + 60e^{i\alpha} + 9e^{2i\alpha}. \end{aligned}$$

Indefinite integration over α gives

$$\begin{aligned} \int F(4+1\cdot e^{i\alpha})\,d\alpha &= \int (100 + 60e^{i\alpha} + 9e^{2i\alpha})\,d\alpha \\ &= 100\int d\alpha + 60\int e^{i\alpha}\,d\alpha + 9\int e^{2i\alpha}\,d\alpha \\ &= 100\alpha + 60\frac{1}{i}e^{i\alpha} + \frac{9}{2i}e^{2i\alpha}. \end{aligned}$$

Next we consider the definite integral

$$\int_0^{2\pi} F(4+1\cdot e^{i\alpha})\,d\alpha = \left[100\alpha + 60\frac{1}{i}e^{i\alpha} + \frac{9}{2i}e^{2i\alpha}\right]_{\alpha=0}^{\alpha=2\pi}.$$

At the upper limit of 2π this integral evaluates to

$$100\ (2\pi) + 60\frac{1}{i}e^{i2\pi} + \frac{9}{2i}e^{2i2\pi} = 200\pi + \frac{60}{i} + \frac{9}{2i} = 200\pi + \frac{129}{2i}.$$

At the lower limit of 0 it evaluates to

$$0 + 60\frac{1}{i}e^0 + \frac{9}{2i}e^0 = \frac{129}{2i}.$$

Hence the difference between the value at the upper limit and the value at the lower limit is

$$\int_0^{2\pi} F(4+1\cdot e^{i\alpha})\,d\alpha = 200\pi + \frac{129}{2i} - \frac{129}{2i} = 200\pi.$$

The integral in (2**) has a factor $1/(2\pi)$ in front, so that we put that factor in front of the last integral and obtain

$$\frac{1}{2\pi}\int_0^{2\pi} F(4+1\cdot e^{i\alpha})\,d\alpha = \frac{1}{2\pi}\cdot 200\pi = 100 \qquad \text{where} \qquad 100 = F(4).$$

Thus we have shown that (2**) holds and thereby verified Theorem 1 for (A).

7. Mean values of harmonic functions. Verification of Theorem 2, p. 782. Our problem is similar in spirit to that of Prob. 3 in that it requires us to verify another mean value theorem for a given example—here for a *harmonic* function. Turn to p. 782 and look at the two formulas [one with no number, one numbered (3)] in the proof of Theorem 2. We shall verify them for given function Φ defined on a point (x_0, y_0) and a circle. To get a better familiarity of the material, you may want to write down all the details of the solution with the integrals, as we did in Prob. 3. We verify Theorem 2 for

$$\text{(B)} \qquad \Phi(x,\, y) = (x-1)\,(y-1), \qquad (x_0,\, y_0) = (2,\, -2), \qquad z = 2 - 2i + e^{i\,\alpha}.$$

The function $\Phi(x, y)$ is indeed harmonic (for definition, see pp. 628 and 758–759). You should verify this by differentiation, that is, by showing that Φ is a solution of

$$\nabla^2\Phi = \Phi_{xx} + \Phi_{yy} = 0 \qquad [(1), \text{p. } 759].$$

We continue.

We note that

$$z_0 = x_0 + iy_0 = 2 - 2i \qquad \text{is the center of the circle in (B).}$$

In terms of the real and imaginary parts of the path, $2 - 2i + e^{i\,\alpha}$, is then [by Euler's formula (5), p. 634 in Sec. 13.6]

$$\text{(C)} \qquad x = 2 + \cos\,\alpha, \qquad y = -2 + \sin\,\alpha.$$

This is the representation we need, since Φ is a real function of the two real variables *x* and *y*. We see that

$$\begin{aligned}\Phi(z_0,\, y_0) &= \Phi(2,\, -2)\\ &= \left[(x_0 - 1)\,(y_0 - 1)\right]_{x_0 = 2,\, y_0 = -2}\\ &= (2-1)\,(-2-1) = -3.\end{aligned}$$

Hence we have to show that each of the two mean values equals -3.

Substituting (C) into (B) (which is a completely schematic process) gives

$$\text{(D)} \qquad \begin{aligned}\Phi(2 + \cos\,\alpha,\, -2 + \sin\,\alpha) &= (2 + \cos\,\alpha - 1)\,(-2 + \sin\,\alpha - 1)\\ &= (1 + \cos\,\alpha)\,(-3 + \sin\,\alpha)\\ &= -3 + 1\,\sin\,\alpha - 2\,\cos\,\alpha + \cos\,\alpha\,\sin\,\alpha.\end{aligned}$$

Consider the mean value over the circle. Now

$$\begin{aligned}&\int_0^{2\pi} (-3 + 1\,\sin\,\alpha - 2\,\cos\,\alpha + \cos\,\alpha\,\sin\,\alpha)\;d\alpha\\ &\quad = \left[-3\alpha - \cos\,\alpha - 2\sin\,\alpha + \frac{1}{2}\sin^2\alpha\right]_0^{2\pi}\\ &\quad = -6\pi - \underbrace{\cos\,2\pi}_{1} - 2\underbrace{\sin\,2\pi}_{0} + \frac{1}{2}\underbrace{\sin^2 2\pi}_{0} - \left(-0 - \underbrace{\cos 0}_{1} - \,2\underbrace{\sin 0}_{0} + \frac{1}{2}\underbrace{\sin^2 0}_{0}\right)\\ &\quad = (-6\pi - 1) - (-1)\\ &\quad = -6\pi.\end{aligned}$$

We have to multiply this result by a factor $1/(2\pi)$. (This is the factor in front of the unnumbered formula of the first integral in the proof of Theorem 2.) Doing so we get

$$\frac{1}{2\pi} \cdot (-6\pi) = -3.$$

This is the mean value of the given harmonic function over the circle considered *and completes the verification of the first part of the theorem for our given data.*

Next we work on (3)*, p.* 782. Now calculate the mean value over the disk of radius 1 and center $(2, -2)$. The integrand of the double integral in formula (3) in the proof of Theorem 2 is similar to that in (D). However, in (D) we had $r = 1$ (the circle over which we integrated), whereas now we have r being variable and we integrate over it from 0 to 1. In addition we have a factor r resulting from the element of area in polar coordinates, which is $r\,dr\,d\theta$. Hence, instead of $(1 + \cos\alpha)\,(-3 + \sin\alpha)$ in (D), we now have

$$(1 + r\ \cos\alpha)\,(-3 + r\ \sin\alpha)\,r = -3\,r + 1\,r^2\ \sin\alpha - 2\,r^2\ \cos\alpha + r^3\ \cos\alpha\ \sin\alpha.$$

The factors of r have no influence on the integration over α from 0 to $2\,\pi$ so

$$\int_0^{2\pi} \left(-3\,r + 1\,r^2\ \sin\alpha - 2\,r^2\ \cos\alpha + r^3\ \cos\alpha\ \sin\alpha\right)\,d\alpha$$

$$= \left[-3r\alpha - r^2\cos\alpha - 2r^2\sin\alpha + \frac{1}{2}r^3\sin^2\alpha\right]_{\alpha=0}^{\alpha=2\pi}$$

$$= -6r\pi - r^2\underbrace{\cos 2\pi}_{1} - 2r^2\underbrace{\sin 2\pi}_{0} + \frac{1}{2}r^3\underbrace{\sin^2 2\pi}_{0} - \left(-0 - r^2\underbrace{\cos 0}_{1} - 2r^2\underbrace{\sin 0}_{0} + \frac{1}{2}r^3\underbrace{\sin^2 0}_{0}\right)$$

$$= \left(-6r\pi - r^2\right) - \left(-r^2\right)$$

$$= -6\pi r.$$

Hence

$$\int_0^1 -6\pi r\,dr = -6\pi\int_0^1 r\,dr = -6\pi\left[\frac{r}{2}\right]_0^1 = -6\pi\cdot\frac{1}{2} = -3\pi.$$

In front of the double integral we have the factor $1/\left(\pi\ r_0^2\right) = 1/\pi$ because the circle of integration has radius 1. Hence our second result is $-3\pi/\pi = -3$. This completes the verification.

Remark. The problem requires you to only verify (3). We also verified the first formula in the proof of Theorem 2 to give you a more complete illustration of the theorem.

19. Location of maxima of a harmonic function and its conjugate. The question is whether a harmonic function Φ and a harmonic conjugate Ψ in a region R have their maximum *at the same point* of R. The answer is "not in general." We look for a counterexample that is as simple as possible. For example, a simple case would be the conjugate harmonics Ψ:

$$x = \text{Re}\ z \qquad \text{and} \qquad y = \text{Im}\ z \qquad \text{in the square} \qquad 0 \le x \le 1,\ 0 \le y \le 1.$$

Then we have

$$\max\ x = 1 \qquad \text{at all points on the right boundary}$$

and

$$\max\, y = 1 \qquad \text{at all points of the upper boundary.}$$

Hence in this case there is a point

$$(1, 1), \qquad \text{that is,} \qquad z = 1 + i,$$

where both functions Φ and Ψ have a maximum. But, if we leave out that point $(1, 1)$ in the square and consider only

$$\text{the region} \qquad R: \qquad 0 \leq x < 1,\ 0 \leq y < 1,$$

then

$$\max\, x \ \text{ and } \ \max\, y \ \text{ cannot occur at the same point.}$$

You may want to investigate the question further. What about a triangle, a square with vertices $\pm 1,\ \pm i$, and so on?

Brief Review of Part D on Complex Analysis

Since complex analysis is a rather diverse area, we include this **brief review of the essential ideas of complex analysis**. Our main point is that to get a good grasp of the field, *keep the* ***three approaches*** *(methods) of complex analysis apart and firmly planted in your mind.* This is in tune with **Underlying Theme 4** of "**Clearly identifying the conceptual structure of subject matter**" on p. x of the textbook. The three approaches were [with particularly important sections marked in boldface, page references given for the Textbook (T) and this Manual (M)]:

1. **Evaluating integrals by Cauchy's integral formula** [see **Sec. 14.3**, p. 660 (T), p. 291 (M); general background Chap. 13, p. 608 (T), p. 257 (M), and Chap. 14, p. 643 (T), p. 283 (M)]. The method required a basic understanding of analytic functions [p. 623 (T), p. 267 (M)], the Cauchy–Riemann equations [p. 625 (T), p. 269 (M)], and Cauchy's integral theorem [p. 653 (T), p. 288 (M)].

2. **Residue integration** [**applied to complex integrals** see **Sec. 16.3**, p. 719 (T), p. 322 (M); **applied to real integrals** see **Sec. 16.4**, p. 725 (T), p. 326 (M); general background Chap. 15, p. 671 (T), p. 298 (M), and Chap. 16, p. 708 (T), p. 316 (M)]. The method needed a basic understanding of radius of convergence of power series and the Cauchy–Hadamard formula [p. 683 (T), p. 303 (M)] and Taylor series p. 690 (T), p. 309 (M). This led to the very important Laurent series [which allowed negative powers, p. 709 (T), p. 316 (M)] and gave us order of singularities, poles, and zeros [p. 717 (T), p. 320 (M)].

3. **Geometric approach of conformal mapping applied to potential theory [in electrostatic fields Sec. 18.1**, p. 759 (T), p. 353 (M); **Sec. 18.2**, p. 763 (T), p. 357 (M); Sec. 18.5, p. 777 (T), p. 364 (M), in **heat conduction**, **Sec. 18.3**, p. 767 (T), p. 359 (M), in **fluid flow in Sec. 18.4**, p. 771 (T), p. 361 (M); general background in Chap. 17, p. 736 (T), p. 332 (M)]. The method required an understanding of conformal mapping [p. 738 (T), p. 333 (M)], linear fractional transformations [p. 743 (T), p. 339 (M)], and their fixed points [pp. 745, 746 (T), pp. 339, 341 (M)], and a practical understanding of how to apply conformal mappings to basic complex functions.

In general, just like in regular calculus, you have to know basic complex functions (sine, cosine, exponential, logarithm, power function, etc.) and know how they are different from their real counterparts. You have to know Euler's formula [(5), p. 634 (T), p. 277 (M)] and Laplace's equation [Theorem 3, p. 628 (T), p. 269 (M)].

PART E

Numeric Analysis

Numeric analysis in Part E (also known as *numerics* or *numerical analysis*) is an area rich in applications that include modeling chemical or biological processes, planning ecologically sound heating systems, determining trajectories of satellites and spacecrafts, and many others. Indeed, in your career as an engineer, physicist, applied mathematician, or in another field, it is likely that you will encounter projects that will require the use of some numerical methods, with the help of some software or CAS (computer algebra system), to solve a problem by generating results in terms of tables of numbers or figures.

The study of numeric analysis completes your prior studies in the sense that a lot of the material you learned before from a more *algebraic* perspective is now presented again from a *numeric* perspective. At first, we familiarize you with general concepts needed throughout numerics (floating point, roundoff, stability, algorithm, errors, etc.) and with general tasks (solution of equations, interpolation, numeric integration and differentiation) in Chap. 19. Then we continue with numerics for linear systems of equations and eigenvalue problems for matrices in Chap. 20—material previously presented in an algebraic fashion in Chaps. 7 and 8. Finally, in Chap. 21 we discuss numerical methods for differential equations (ODEs and PDEs) and thus related to Part A and Chap. 12.

Use of Technology. We have listed on pp. 788–789 **software**, computer algebra systems (CASs), programmable graphic calculators, computer guides, etc. In particular, note the **Mathematica Computer Guide** and **Maple Computer Guide (**for stepwise guidance on how to solve problems by writing programs for these two CASs) by Kreyszig and Norminton that accompany the textbook (see p. 789). However, *the problems in the problem sets in the textbook can be solved by a simple calculator, perhaps with some graphing capabilities*, except for the CAS projects, CAS experiments, or CAS problems (see Remark on Software Use on p. 788 of textbook).

Chap. 19 Numerics in General

This chapter has five sections, the first on general concepts needed throughout numerics and the other four on three basic areas, namely, solution of equations (Sec. 19.2), interpolation (Secs. 19.3 and 19.4), and numeric integration and differentiation (Sec. 19.5).

In this chapter you should also obtain a feel for the spirit, viewpoint, and nature of numerics. You will notice that numeric analysis has a flavor distinct from that of calculus.

A convenient framework on how to solve numeric problems consists of five steps:

1. Modeling the problem

2. Selecting a numeric method

3. Programming

4. Doing the computation

5. Interpreting the results

as shown on p. 791. Solving a single equation of the form $f(x) = 0$, as shown in Sec. 19.2, may serve as one of many illustrations.

From calculus, you should review *Taylor series* [in formula (1), p. 690, replace *complex z* with *real x*], limits and convergence (see pp. 671–672), and, for Sec. 19.5, review, from calculus, the basics of how one developed, geometrically, the Riemann integral.

Sec. 19.1 Introduction

This section introduces some of the general ideas and concepts that underlie all of numerics. As such it touches upon a fair amount of material in a concise fashion. Upon reading it for the first time, the material of Sec. 19.1 may seem rather abstract to you, however, with further studies of numerics, it will take on concrete meaning. For example, the concepts of algorithm and stability (p. 796 of textbook) are explained here in Sec. 19.1 but illustrated in subsequent sections. Overall, *Sec. 19.1 can be thought of as a reference section for Part E.* Hence, once in a while, it may be useful for you to refer back to this section.

Your numeric calculations require that you do computations to a certain amount of precision. It is here that **rounding** (p. 792 of the textbook) comes into play. Take a look at the **roundoff rule** at the bottom of p. 792 and at **Example 1** at the top of p. 793. The concept of rounding uses the definition of decimals on p. 791 in a fixed-point system. Note that when counting decimals, only the numbers *after* the decimal point are counted, that is,

$$78.94599, \quad -0.98700, \quad 10000.00000 \quad \text{all have 5 decimals, abbreviated 5D.}$$

Make sure that you understand Example 1. Here is a self-test. (a) Round the number 1.23454621 to seven decimals, abbreviated (7D). (b) Round the number −398.723555 to four decimals (4D). *Please close this Student Solutions Manual (!).* Check the answer on p. 27 of the Manual. If your answer is correct, great. If not, go over your answers and study Example 1 again.

The standard decimal system is not very useful for scientific computation and so we introduce the **floating-point system** on p. 791. We have

$$624.7 = \underline{0.6247 \cdot 10^3}; \qquad 0.\underbrace{0000000000000}_{\text{13 zeros}}1735 = 1735 \cdot 10^{-17} = \underline{0.1735 \cdot 10^{-13}};$$

$$-0.02000 = \underline{-0.2000 \cdot 10^{-1}},$$

where the underscored number is in floating-point form.

Each of these floating-point numbers above has four **significant digits**, also denoted by 4S. The digits are "significant" in the sense that they convey numerical information and are not just placeholders of zeros that fix the position of the decimal points, whose positions could also be achieved by multiplication of suitable powers of 10^n, respectively. This leads to our next topic of rounding with significant digits.

The **roundoff rule for significant digits** is as follows. To round a number x to k significant digits, do the following three steps:

1. Express the given number as a floating-point number:

$$x = \pm m \cdot 10^n, \quad 0.1 \le |m| < 1, \quad \text{where } n \text{ is an integer [see also (1), p. 792].}$$

Note that here m can have *any* number of digits.

2. For now, ignore the factor 10^n. Apply the roundoff rule (for decimals) on p. 792 to m only.

3. Take the result from step 2 and multiply it by 10^n. This gives us the desired number x rounded to k significant digits.

Self-test: Apply the roundoff rule for significant digits to round 102.89565 to six significant digits (6S). Check your result on p. 27.

The computations in numerics of unknown quantities are approximations, that is, they are not exact but involve errors (p. 794). Rounding, as just discussed by the roundoff rule, produces roundoff errors bounded by (3), p. 793. To gain accuracy in calculations that involve rounding, one may carry extra digits called *guarding digits* (p. 793). Severe problems in calculations may involve the *loss of significant digits* that can occur when we subtract two numbers of about the same size as shown in **Example 2** on pp. 793–794 and in **Problem 9**.

We also distinguish between **error**, defined by (6) and (6*) and **relative error** (7) and (7′), p. 794, respectively. The **error** is defined in the relationship

$$\text{True value} = \text{Approximation} + \text{Error}.$$

The **relative error** is defined by

$$\text{Relative error} = \frac{\text{Error}}{\text{True value}} \qquad (\text{where True value} \neq 0).$$

As one continues to compute over many steps, errors tend to get worse, that is, they propagate. In particular, bounds for errors add under addition and subtraction and bounds for relative errors add under multiplication and division (see Theorem 1, p. 795).

Other concepts to consider are **underflow**, **overflow** (p. 792), *basic error principle*, and *algorithm* (p. 796). Most important is the concept of **stability** because we want algorithms to be stable in that small changes in our initial data should only cause small changes in our final results.

Remark on calculations and exam. Your answers may vary slightly in some later digits from the answers given here and those in App. 2 of the textbook. You may have used different order of calculations, rounding, technology, etc. Also, for the exam, ask your professor what technology is allowed and be familiar with the use and the capabilities of that technology as it may save you valuable time on the exam and give you a better grade. It may also be a good idea, for practice, to use the same technology for your homework.

Problem Set 19.1. Page 796

9. **Quadratic equation.** We want to solve the quadratic equation $x^2 - 30\,x + 1 = 0$ in two different ways—first with 4S accuracy and then with 2S accuracy.
(a) 4S. First, we use the well-known quadratic formula

$$(4) \qquad x_1 = \frac{1}{2a}\left(-b + \sqrt{b^2 - 4ac}\right), \qquad x_2 = \frac{1}{2a}\left(-b - \sqrt{b^2 - 4ac}\right)$$

with

$$a = 1, \qquad b = -30, \qquad c = 1.$$

We get, for the square root term calculated with 4S ("significant digits," see pp. 791–792),

$$\sqrt{(-30)^2 - 4} = \sqrt{900 - 4} = \sqrt{896} = 29.933 = 29.93.$$

Hence, computing x_1 and x_2 rounded to **four** significant digits, i.e., 4S,

$$x_1 = \tfrac{1}{2} \cdot (30 + 29.93) = \tfrac{1}{2} \cdot 59.93 = 29.965 = 29.97$$

and

$$x_2 = \tfrac{1}{2} \cdot (30 - 29.93) = \tfrac{1}{2} \cdot 0.07 = 0.035.$$

It is important to notice that x_2, obtained from 4S values, is just 2S—i.e., we have lost two digits.

As an alternative method of solution for x_2, use (5), p. 794,

$$\text{(5)} \qquad x_1 = \frac{1}{2a}\left(-b + \sqrt{b^2 - 4ac}\right), \qquad x_2 = \frac{c}{ax_1}.$$

The root x_1 (where the similar size numbers are added) equals 29.97, as before. For x_2, you now obtain

$$x_2 = \frac{c}{a\,x_1} = \frac{1}{29.97} = 0.0333667 = 0.03337 \quad \text{(to \textbf{four} significant digits).}$$

(b) 2S. With 2S the calculations are as follows. We have to calculate the square root as

$$\sqrt{(-30)^2 - 4} = \sqrt{900 - 4} = \sqrt{899.6} = \sqrt{900} = 30 \qquad \text{(to \textbf{two} significant digits, i.e., 2S).}$$

Hence, by (4),

$$x_1 = \tfrac{1}{2} \cdot (30 + 30) = \tfrac{1}{2} \cdot 60 = 30$$

and

$$x_2 = \tfrac{1}{2} \cdot (30 - 30) = 0.$$

In contrast, from (5), you obtain better results for the second root. We still have $x_1 = 30$ but

$$x_2 = \frac{1}{x_1} = \frac{1}{30} = 0.033333 = 0.033 \qquad \text{(to \textbf{two} significant digits).}$$

Purpose of Prob. 9. The point of this and similar examples and problems is not to show that calculations with fewer significant digits generally give inferior results (this is fairly plain, although not always the case). The point is to show, in terms of simple numbers, what will happen in principle, regardless of the number of digits used in a calculation. Here, formula (4) illustrates the loss of significant digits, easily recognizable when we work with pencil (or calculator) and paper, but difficult to spot in a long computation in which only a few (if any) intermediate results are printed out. This explains the necessity of developing programs that are virtually free of possible cancellation effects.

19. We obtain the Maclaurin series for the exponential function by (12), p. 694, of the textbook where we replace z, a complex number, by u a real number. [For those familiar with complex numbers, note that (12) holds for any complex number $z = x + iy$ and so in particular for $z = x + iy = x + i \cdot 0 = x = \operatorname{Re} z$, thereby justifying the use of (12)! Or consult your old calculus book. Or compute it yourself.] Anyhow, we have

$$(12') \qquad f(x) = e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots + \frac{x^{10}}{10!} + \cdots .$$

[All computations to **six digits (6S)**.] We are given that the exact 6S value of $1/e$ is

$$(6S) \qquad \frac{1}{e} = \boxed{0.367879}$$

(**a**) For e^{-1}, the Maclaurin series (12′) with $x = -1$ becomes

$$(B) \qquad f(-1) = e^{-1} = 1 + \frac{(-1)}{1!} + \frac{(-1)^2}{2!} + \frac{(-1)^3}{3!} + \frac{(-1)^4}{4!} + \frac{(-1)^5}{5!} + \cdots + \frac{(-1)^{10}}{10!} + \cdots$$

$$= 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \cdots + \frac{1}{10!} + \cdots .$$

Now, using (B) with *five terms*, we get

$$1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} = \boxed{0.366667 \qquad (6S)}$$

$$(B5) \qquad \text{Error} \qquad \textit{diff}: (A) - (B5) = 0.367879 - 0.366667 = \boxed{0.001212};$$

with *eight terms*,

$$(B8) \qquad 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} + \frac{1}{8!} = 0.367882$$

$$\text{Error} \qquad \textit{diff}\,(A) - (B8) = 0.367879 - 0.367882 = -0.000003$$

while, with *ten terms*,

$$(B10) \qquad 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} + \frac{1}{8!} - \frac{1}{9!} + \frac{1}{10!} = 0.367879$$

$$\text{Error} \qquad \textit{diff}\,(A) - (B10) = 0.367879 - 0.367879 = 0.$$

(**b**) For the $1/e^1$ method, that is, computing e^x with $x = 1$ and then taking the reciprocal, we get

$$(C) \qquad f(1) = e^1 = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \cdots + \frac{1}{10!} + \cdots , .$$

so (C), with five terms is

$$(C5) \qquad e^1 = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} = 2.71667$$

giving the reciprocal

(C5*) $$\frac{1}{e^1} = \frac{1}{2.71667} = 0.368098 \qquad \text{[using the result of (C5)]}$$

Error *diff*: (A) – C5* = 0.367879 – 0.368098 = 0.000219.

This is much better than the corresponding result (B5) in (a). With seven terms we obtain

(C7) $$1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \frac{1}{6!} + \frac{1}{7!} = 2.71825$$

(C7*) $$\frac{1}{2.71825} = 0.367884$$

Error *diff*: (A) – (C7*) = 0.367879 – 0.367884 = –0.000005.

This result is almost as good as (B8) in (a), that is, the one with eight terms. With ten terms we get

(C10) $$1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \frac{1}{6!} + \frac{1}{7!} + \frac{1}{8!} + \frac{1}{9!} + \frac{1}{10!} = 2.71828$$

Sec. 19.1. Prob. 19. Table. *Computation of e^{-1} and $1/e^1$ for the MacLaurin series as a computer would do it*

No. of	Terms	Decimal	(**a**)	e^{-1}	Exact is	(**b**)	e^1	$1/e^1$	Exact is
Factorial		Terms		**Result**	0.367879			**Result**	0.367879
Terms				via	**Error:**			via	**Error:**
in (12′)				(B)	Exact – (B)			(C)	Exact – (C)
	1	1	+	1	0.632121	+	1	1	–0.632121
1	$\frac{1}{1!}$	1	–	0	–0.367879	+	2	0.5	–0.132121
2	$\frac{1}{2!}$	0.5	+	0.5	0.132121	+	2.5	0.4	–0.032121
3	$\frac{1}{3!}$	0.166667	–	0.333333	–0.034546	+	2.66667	0.375000	–0.007121
4	$\frac{1}{4!}$	0.0416667	+	0.375000	0.007121	+	2.70834	0.369230	–0.001351
5	$\frac{1}{5!}$	0.00833333	–	0.366667	–0.001212	+	2.71667	0.368098	–0.000219
6	$\frac{1}{6!}$	0.00138889	+	0.368056	0.000177	+	2.71806	0.367909	–0.000030
7	$\frac{1}{7!}$	0.000198413	–	0.367858	–0.000021	+	2.71826	0.367882	–0.000003
8	$\frac{1}{8!}$	0.0000248016	+	0.367883	0.000004	+	2.71828	0.367880	–0.000001
9	$\frac{1}{9!}$	0.00000275573	–	0.367880	0.000001	+	2.71828	0.367880	–0.000001
10	$\frac{1}{10!}$	0.000000275573	+	0.367880	0.000001	+	2.71828	0.367880	–0.000001

giving

$$\text{(C10*)} \qquad \frac{1}{2.71825} = 0.367879$$

$$\text{Error} \qquad \textit{diff}: \ (\text{A}) - (\text{C10*}) = 0.367879 - 0.367879 = 0$$

the same as (**a**). With the $1/e^1$ method, we get more accuracy for the same number of terms or we get the same accuracy with fewer terms. The $1/9!$ and $1/10!$ terms are so small that they have no effect on the result. The effect will be much greater in **Prob. 20**.

In the table, on the previous page, all computations are done to 6S accuracy. That means that each term is rounded to 6S, "added" to the previous sum, and the result is then rounded before the next term is added. For example, $1/4! = 0.0416666667$ (becomes 0.0416667), is added to 0.333333 to give 0.3749997, which becomes 0.375000. This is the way a computer does it, and it will produce a different result from that obtained by adding the first four terms and then rounding. The signs in the (**a**) and (**b**) columns indicate that the corresponding term should be added to or subtracted from the current sum.

Sec. 19.2 Solution of Equations by Iteration

The problem of finding solutions to a single equation (p. 798 of textbook)

$$f(x) = 0 \tag{1}$$

appears in many applications in engineering. This problem appeared, for example, in the context of characteristic equations (Chaps. 2, 4, 8), finding eigenvalues (Chap. 8), and finding zeros of Bessel functions (Chap. 12). We distinguish between algebraic equations, that is, when (1) is a *polynomial* such as

$$f(x) = x^3 - 5x + 3 = 0 \qquad \text{[see Probs. 21, 27]}$$

or a *transcendental* equation such as

$$f(x) = \tan x - x = 0.$$

In the former case, the solutions to (1) are called roots and the problem of finding them is called *finding roots.*

Since, in general, there are no direct formulas for solving (1), except in a few simple cases, the task of solving (1) is made for numerics.

The first method described is a **fixed-point iteration** on pp. 798–801 in the text and illustrated by **Example 1**, pp. 799–800, and **Example 2**, pp. 800–801. The main idea is to transform equation (1) from above by *some algebraic process* into the form

$$x = g(x). \tag{2}$$

This in turn leads us to choose an x_0 and compute $x_1 = g(x_0)$, $x_2 = g(x_1)$, and in general

$$x_{n+1} = g(x_n) \qquad \text{where} \qquad n = 0, 1, 2, \cdots . \tag{3}$$

We have set up an iteration because we substitute x_0 into g and get $g(x_0) = x_1$, the next value for the iteration. Then we substitute x_1 into g and get $g(x_1) = x_2$ and so forth.

A solution to (2) is called a fixed point as motivated on top of p. 799. Furthermore, Example 1 demonstrates the method and shows that the "algebraic process" that we use to transform (1) to (2) is *not unique*. Indeed, the quadratic equation in Example 1 is written in two ways (4a) and (4b) and the corresponding iterations illustrated in Fig. 426 at the bottom of p. 799. Making the "best" choice for $g(x)$ can pose a significant challenge. More on this method is given in Theorem 1 (sufficient condition for convergence), Example 2, and Prob. 1.

Most important in this section is the famous **Newton method**. The method is defined recursively by

$$(5) \qquad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad \text{where} \qquad n = 0, 1, 2, \cdots, N-1.$$

The details are given in Fig. 428 on p. 801 and in the algorithm in Table 19.1 on p. 802. Newton's method can either be derived by a geometric argument or by Taylor's formula (5*), p. 801. **Examples 3, 4, 5,** and **6** show the versatility of Newton's method in that it can be applied to transcendental and algebraic equations. **Problem 21** gives complete details on how to use the method. Newton's method converges of second order (Theorem 2, p. 804). **Example 7**, p. 805, shows when Newton's method runs into difficulties due to the problem of ill-conditioning when the denominator of (5) is very small in absolute value near a solution of (1).

Newton's method can be modified if we replace the derivative $f'(x)$ in (5) by the difference quotient

$$f'(x_n) = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$$

and simplify algebraically. The result is the **secant method** given by (10), p. 806, which is illustrated by Example 8 and **Prob. 27**. Its convergence is superlinear (nearly as fast as Newton's method). The method may be advantageous over Newton's method when the derivative is difficult to compute and computationally expensive to evaluate.

Problem Set 19.2. Page 807

1. Monotonicity and Nonmonotonicity. We consider the case of nonmonotonicity, as in Example 2 in the book, Fig. 427, p. 801. Nonmonotonicity occurs if a sequence $g(x)$ is monotone decreasing, that is,

$$\text{(A)} \qquad g(x_1) \geq g(x_2) \qquad \text{if} \qquad x_1 < x_2.$$

(Make a sketch to better understand the reasoning.) Then

$$\text{(B)} \qquad g(x) \geq g(s) \qquad \text{if and only if} \qquad x \leq s,$$

where s is such that $g(s) = s$ [the intersection of $y = x$ and $y = g_1(x)$ in Fig. 427] and

$$\text{(C)} \qquad g(x) \leq g(s) \qquad \text{if and only if} \qquad x \geq s.$$

Suppose we start with $x_1 > s$. Then $g(x_1) \leq g(s)$ by (C). If $g(x_1) = g(s)$ [which could happen if $g(x)$ is constant between s and x_1], then x_1 is a solution of $f(x) = 0$, and we are done. Otherwise $g(x_1) < g(s)$, and by the definition of x_2 [formula (3), p. 798 in the text] and since s is a fixed point [$s = g(s)$], we obtain

$$x_2 = g(x_1) < g(s) = s \qquad \text{so that} \qquad x_2 < s.$$

Hence by (B),

$$g(x_2) \geq g(s).$$

The equality sign would give a solution, as before. Strict inequality, and the use of (3) in the text, give

$$x_3 = g(x_2) > g(s) = s, \qquad \text{so that} \qquad x_3 > s,$$

and so on. This gives a sequence of values that are alternatingly larger and smaller than s, as illustrated in Fig. 427 of the text.

Complete the problem by considering monotonicty, as in Example 1, p. 799.

21. Newton's method. The equation is $f(x) = x^3 - 5x + 3 = 0$ with $x_0 = 2, 0, -2$. The derivative of $f(x)$ is

$$f'(x) = 3x^2 - 5.$$

Newton's method (5), p. 802,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^3 - 5x_n + 3}{3x_n^2 - 5}.$$

We have nothing to compute for the interation $n = 0$. For the iteration $n = 1$ we have

$$\begin{aligned} x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)} \\ &= \mathbf{2} - \frac{2^3 - 5 \cdot 2 + 3}{3 \cdot 2^2 - 5} \\ &= \mathbf{2} - \frac{8 - 10 + 3}{12 - 5} \\ &= \mathbf{2} - \frac{\mathbf{1}}{\mathbf{7}} \\ &= 2 - 0.1428571429 \\ &= 2 - 0.\overline{142857} \\ &= 1.857143 = \underline{\mathbf{1.85714}} \quad \text{(6S)}. \end{aligned}$$

From the next iteration (iteration, $n = 2$) we obtain

$$\begin{aligned} x_2 &= x_1 - \frac{f(x_1)}{f'(x_1)} \\ &= 1.85714 - \frac{(1.85714)^3 - 5 \cdot 1.85714 + 3}{3 \cdot (1.85714)^2 - 5} \\ &= 1.85714 - \frac{0.119518251}{5.34690694} = 1.85714 - \frac{\mathbf{0.119518}}{\mathbf{5.34691}} \\ &= 1.85714 - 0.02235272 = 1.85714 - 0.0223527 \\ &= 1.8347873 = \underline{\mathbf{1.83479}} \text{ (6S)}. \end{aligned}$$

The iteration $n = 3$ gives us

$$\begin{aligned}
x_3 &= x_2 - \frac{f(x_2)}{f'(x_2)} \\
&= 1.83479 - \frac{(1.83479)^3 - 5 \cdot 1.83479 + 3}{3 \cdot (1.83479)^2 - 5} \\
&= 1.83479 - \frac{0.002786766}{5.09936303} = 1.83479 - \frac{\mathbf{0.00278677}}{\mathbf{5.09936}} \\
&= 1.83479 - 0.0005464940698 = 1.83479 - 0.000546494 \\
&= 1.834243506 = \underline{\mathbf{1.83424}} \text{ (6S)}.
\end{aligned}$$

For $n = 4$ we obtain

$$\begin{aligned}
x_4 &= x_3 - \frac{f(x_3)}{f'(x_3)} \\
&= 1.83424 - \frac{(1.83424)^3 - 5 \cdot 1.83424 + 3}{3 \cdot (1.83424)^2 - 5} \\
&= 1.83424 - \frac{-0.000016219}{5.09330913} = 1.83424 - \frac{\mathbf{-0.000016219}}{\mathbf{5.09331}} \\
&= 1.83424 - (-3.184373227 \times 10^{-6}) = 1.83424 - (-3.18437 \times 10^{-6}) \\
&= 1.834243184 = \underline{\mathbf{1.83424}} \text{ (6S)}.
\end{aligned}$$

Because we have the same value for the root (6S) as we had in the previous iteration, we are finished.

Hence the iterative sequence converges to $x_4 =$**<u>1.83424</u>** (6S), which is the first root of the given cubic polynomial.

The next set of iterations starts with $x_0 = \mathbf{0}$ and converges to $x_4 =$**<u>0.656620</u>** (6S), which is the second root of the given cubic polynomial. Finally starting with $x_0 = \mathbf{-2}$ yields $x_4 =$**<u>–2.49086</u>** (6S).

The details are given in the three-part table on the next page. Note that your answer might vary slightly in the last digits, depending on what CAS or software or calculator you are using.

27. Secant method. The equation is as in Prob. 21, that is,

$$\text{(P)} \qquad x^3 - 5x + 3 = 0.$$

This time we are looking for only one root between the given values $x_0 = 1.0$ and $x_1 = 2.0$.

Solution. We use (10), p. 806, and get

$$\begin{aligned}
x_{n+1} &= x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \\
&= x_n - \left(x_n^3 - 5x_n + 3\right) \frac{x_n - x_{n-1}}{x_n^3 - 5x_n + 3 - \left(x_{n-1}^3 - 5x_{n-1} + 3\right)}.
\end{aligned}$$

The 3 in the denominator of the second equality cancels, and we get the following formula for our iteration:

$$x_{n+1} = x_n - \left(x_n^3 - 5x_n + 3\right) \frac{x_n - x_{n-1}}{x_n^3 - 5x_n - \left(x_{n-1}^3 - 5x_{n-1}\right)}.$$

Sec. 19.2. Prob. 21. Table. *Newton's method with* 6S *accuracy*

n	x_n	$f(x_n)$	$f'(x_n)$
0	2	1	7
1	1.85714	0.119518	5.34691
2	1.83479	0.00278677	5.09936
3	1.83424	−0.000016219	5.09331
4	1.83424		

n	x_n	$f(x_n)$	$f'(x_n)$
0	0	3	−5
1	0.6	0.216	−3.92
2	0.655102	0.00563268	−3.71252
3	0.656619	0.00000530425	−3.70655
4	0.656620	0.00000159770	−3.70655
5	0.656620		

n	x_n	$f(x_n)$	$f'(x_n)$
0	−2	5	7
1	−2.71429	−3.42573	17.1021
2	−2.51398	−0.318694	13.9603
3	−2.49115	−0.00389923	13.6175
4	−2.49086	0.0000492166	13.6132
5	−2.49086		

For $x_0 = 1.0$ and $x_1 = 2.0$ we have

$$
\begin{aligned}
x_2 &= x_1 - \left(x_1^3 - 5x_1 + 3\right) \frac{x_1 - x_0}{x_1^3 - 5x_1 - \left(x_0^3 - 5x_0\right)} \\
&= 2.0 - [(2.0)^3 - 5 \cdot 2.0 + 3] \cdot \frac{2.0 - 1.0}{(2.0)^3 - 5 \cdot 2.0 - [(1.0)^3 - 5 \cdot 1.0]} \\
&= 2.0 - 1.0 \cdot \frac{1.0}{-2.0 - [-4.0]} = 2.0 - 0.50 = 1.5 \text{ (exact).}
\end{aligned}
$$

Next we use $x_1 = 2.0$ and $x_2 = 1.5$ to get

$$
\begin{aligned}
x_3 &= x_2 - \left(x_2^3 - 5x_2 + 3\right) \frac{x_2 - x_1}{x_2^3 - 5x_2 - \left(x_1^3 - 5x_1\right)} \\
&= 1.5 - [(1.5)^3 - 5 \cdot 1.5 + 3] \cdot \frac{1.5 - 2.0}{(1.5)^3 - 5 \cdot 1.5 - [(2.0)^3 - 5 \cdot 2.0]} \\
&= 1.5 - (-1.125) \cdot \frac{-0.5}{-2.125} = 1.76471 \text{ (6S).}
\end{aligned}
$$

The next iteration uses $x_2 = 1.5$ and $x_3 = 1.76471$ to get $x_4 = 1.87360$ (6S). We obtain convergence at step $n = 8$ and obtain $x_8 = 1.83424$, which is one of the roots of (P). The following table shows all the steps. Note that only after we computed x_8 and found it equal (6S) to x_7 did we conclude convergence.

Sec. 19.2 Prob. 27. Table A. *Secant method with* 6S *accuracy*

n	x_n
2	1.5
3	1.76471
4	1.87360
5	1.83121
6	1.83412
7	1.83424
8	**1.83424**

For 12S values convergence occurs when $n = 10$.

Sec. 19.2 Prob. 27. Table B. *Secant method with* 12S *accuracy*

n	x_n
2	1.5
3	1.76470588235
4	1.87359954036
5	1.83120583391
6	1.83411812708
7	1.83424359586
8	1.83424318426
9	1.83424318431
10	**1.83424318431**

Note further that, for the given starting values, the convergence is monotone and is somewhat slower than that for Newton's method in Prob. 21. These properties are not typical but depend on the kind of function we are dealing with. Note that Table A, by itself, represents a complete answer.

Sec. 19.3 Interpolation

Here is an overview of the rather long Sec. 19.3. The three main topics are the problem of interpolation (pp. 808–809), Lagrange interpolation (pp. 809–812), and Newton's form of interpolation (pp. 812–819). Perhaps the main challenge of this section is to understand and get used to the (standard) notation of the formulas, particularly those of Newton's form of interpolation. Just write them out by hand and practice.

The problem of interpolation. We are given values of a function $f(x)$ as ordered pairs, say

$$(x_0, f_0),\ (x_1, f_1), (x_2, f_2), \cdots, (x_n, f_n) \quad \text{where} \quad f_j = f(x_j),\ j = 0, 1, 2, ..., n. \tag{A}$$

The function may be a "mathematical" function, such as a Bessel function, or a "measured" function, say air resistance of an airplane at different speeds. In interpolation, we want to find approximate values of $f(x)$ for new x that lie between those given in (A). The idea in **interpolation** (p. 808) is to find a polynomial $p_n(x)$ of degree n or less—the so called "interpolation polynomial"—that goes through the values in (A), that is,

$$p_n(x_0) = f_0,\ p_n(x_1) = f_1,\ p_n(x_2) = f_2, \cdots,\ p_n(x_n) = f_n. \tag{1}$$

We call $p_n(x)$ a *polynomial approximation* of $f(x)$ and use it to get those new $f(x)$'s mentioned before. When they lie within the interval $[x_0, x_n]$, then we call this interpolation and, if they lie outside the interval, extrapolation.

Lagrange interpolation. The problem of finding an interpolation polynomial p_n satisfying (1) for given data exists and is unique (see p. 809) but may be expressed in different forms. The first type of interpolation is the **Lagrange interpolation,** discussed on pp. 809–812. Take a careful look at the **linear case (2)**, p. 809, which is illustrated in Fig. 431. **Example 1** on the next page applies linear Lagrange interpolation to the natural logarithm to 3D accuracy. If you understand this example well, then the rest of the material follows the same idea, except for details and more involved (but standard) notation. **Example 2**, pp. 810–811, does the same calculations for **quadratic** Lagrange interpolation [formulas (3a), (3b), p. 810] and obtains 4D accuracy. Further illustration of the (quadratic) technique applied to the sine function and error function is shown in **Probs. 7** and **9,** respectively. This all can be **generalized** by (4a), (4b) on p. 811. Various error estimates are discussed on pp. 811–812. Example 3(B) illustrates the *basic error principle* from Sec. 19.1 on p. 796.

Newton's form of interpolation. We owe the greatest contribution to polynomial interpolation to Sir Isaac Newton (on his life cf. footnote 3, p. 15, of the textbook), whose forms of interpolation have three advantages over those of Lagrange:

1. If we want a higher degree of accuracy, then, in Newton's form, we can use all previous work and just add another term. This flexibility is not possible with Lagrange's form of interpolation.

2. Newton's form needs fewer arithmetic calculations than Lagrange's form.

3. Finally, it is easier to use the basic error principle from Sec. 19.1 for Newton's forms of interpolation.

The first interpolation of Newton is **Newton's divided difference interpolation** (10), p. 814, with the kth divided difference defined recursively by (8), p. 813. The corresponding algorithm is given in Table 19.2, p. 814, and the method illustrated by **Example 4**, p. 815, **Probs. 13** and **15**. The computation requires that we set up a divided difference table, as shown on the top of p. 815. *To understand this table, it may be useful to write out the formulas for the terms*, *using* (7), (8), *and the unnumbered equations between them on* p. 813.

If the nodes are equally spaced apart by a distance h, then we obtain **Newton's forward difference interpolation** (14), p. 816, with the kth forward difference defined by (13), p. 816. [This corresponds to (10) and (8) for the arbitrarily spaced case.] An error analysis is given by (16) and the method is illustrated by Example 5, pp. 817–818.

If we run the subscripts of the nodes backwards (see second column in table on top of p. 819), then we obtain *Newton's backward difference interpolation* (18), p. 818, and illustrated in Example 6.

Problem Set 19.3. Page 819

7. Interpolation and extrapolation. We use quadratic interpolation through three points. From (3a), (3b), p. 810, we know that

$$\begin{aligned} p_2(x) &= L_0(x)f_0 + L_1(x)f_1 + L_2(x)f_2 \\ &= \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}f_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}f_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_1)(x_2-x_1)}f_2 \end{aligned}$$

is the quadratic polynomial needed for interpolation, which goes through the three given points (x_0, f_0), (x_1, f_1), and (x_1, f_1). For our problem

k	x_k	f_k
0	$x_0 = 0$	$f_0 = \sin 0$
1	$x_1 = \dfrac{\pi}{4}$	$f_1 = \sin\dfrac{\pi}{4}$
2	$x_2 = \dfrac{\pi}{2}$	$f_2 = \sin\dfrac{\pi}{2}$

so that the desired quadratic polynomial for interpolating $\sin\, x$ with nodes at $x = 0$, $\pi/4$, and $\pi/2$ is

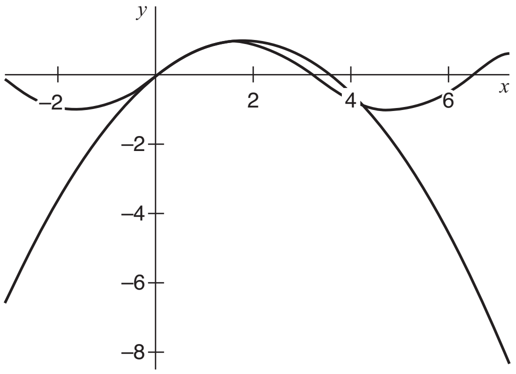
$$
\begin{aligned}
p_2(x) &= \frac{(x-\frac{\pi}{4})(x-\frac{\pi}{2})}{(0-\frac{\pi}{4})(0-\frac{\pi}{2})}\sin 0 \;+\; \frac{(x-0)(x-\frac{\pi}{2})}{(\frac{\pi}{4}-0)(\frac{\pi}{4}-\frac{\pi}{2})}\sin\frac{\pi}{4} + \frac{(x-0)(x-\frac{\pi}{2})}{(\frac{\pi}{2}-0)(\frac{\pi}{2}-\frac{\pi}{4})}\sin\frac{\pi}{2} \\
&= \frac{x^2-\frac{3}{4}x\pi+\frac{1}{8}\pi^2}{\frac{\pi^2}{8}}\sin 0 \;+\; \frac{x^2-\frac{1}{2}x\pi}{-\frac{\pi^2}{16}}\sin\frac{\pi}{4} + \frac{x^2-\frac{1}{4}x\pi}{\frac{\pi^2}{8}}\sin\frac{\pi}{2} \\
\text{(A)} \qquad &= \left(x^2-\frac{3}{4}x\pi+\frac{1}{8}\pi^2\right)\frac{8}{\pi^2}\cdot 0 + \left(x^2-\frac{1}{2}x\pi\right)\frac{-16}{\pi^2}\cdot 0.707107 + \left(x^2-\frac{1}{4}x\pi\right)\frac{8}{\pi^2}\cdot 1 \\
&= -0.3357x^2 + 1.164x.
\end{aligned}
$$

We use (A) to compute $\sin x$ for $x = -\frac{1}{8}\pi$ ("*extra*polation" since $x = -\frac{1}{8}\pi$ lies *outside* the interval $0 \le x \le \frac{\pi}{2}$), $x = \frac{1}{8}\pi$ ("*inter*polation" since $x = \frac{1}{8}\pi$ lies *inside* the interval $0 \le x \le \frac{\pi}{2}$), $x = \frac{3}{8}\pi$ (interpolation), and $\frac{5}{8}\pi$ (extrapolation) and get, by (A), the following results:

x	$p_2(x)$	$\sin x$	error $= \sin x - p_2(x)$
$-\frac{1}{8}\pi$	-0.5089	-0.3827	0.1262
$\frac{1}{8}\pi$	0.4053	0.3827	-0.0226
$\frac{3}{8}\pi$	0.9054	0.9239	0.0185
$\frac{5}{8}\pi$	0.9913	0.9239	-0.0674

We observe that the values obtained by interpolation have smaller errors than the ones obtained by extrapolation. This tends to be true and the reason can be seen in the following figure. Once we are out of the interval over which the interpolationg polynomial has been produced, the polynomial will no longer be "close" to the function—in fact, it will begin to become very large. Extrapolation should not be used without careful examination of the polynomial.

Note that a different order of computation can change the last digit, which explains a slight difference between our result and that given on p. A46 of the textbook.



Sec. 19.3 Prob. 7. Interpolation and extrapolation

9. Lagrange polynomial for the error function erf x**.** The error function [defined by (35) on p. A67 in App. 3, Sec. A3.1, and graphed in Fig. 554, p. A68, in the textbook] given by

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt$$

cannot be solved by elementary calculus and, thus, is an example where numerical methods are appropriate.

Our problem is similar in spirit to Prob. 7. From (3a), (3b), and the given data for the error function erf x, we obtain the Lagrange polynomial and simplify

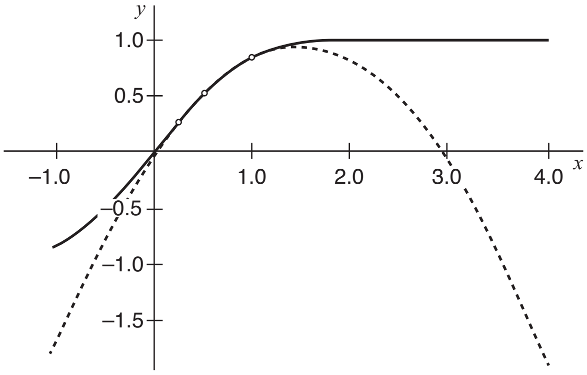
$$\begin{aligned} p_2(x) &= \frac{(x-0.5)(x-1.0)}{(-0.25)(-0.75)}0.27633 + \frac{(x-0.25)(x-1.0)}{0.25(-0.5)}0.52050 \\ \text{(B)} \qquad &\quad + \frac{(x-0.25)(x-0.5)}{0.75 \cdot 0.5}0.84270 \\ &= -0.44304x^2 + 1.30896x - 0.023220. \end{aligned}$$

We use (B) to calculate

$$p_2(0.75) = -0.44304 \cdot (0.75)^2 + 1.30896 \cdot 0.75 - 0.023220 = 0.70929.$$

This approximate value $p_2(0.75) = 0.70929$ is not very accurate. The exact 5S value is erf $0.75 = 0.71116$ so that the error is

$$\begin{aligned} \text{error} &= \text{ erf } 0.75 - p_2(0.75) \qquad \text{[by (6), p. 794]} \\ &= 0.71116 \;\; - 0.70929 = 0.00187. \end{aligned}$$



Sec. 19.3 Prob. 9. The functions erf x and Lagrange polynomial $p_2(x)$.
See also Fig. 554 on p. A68 in App. A of the textbook

13. Lower degree. Newton's divided difference interpolation. We need, from pp. 813–814,

$$a_{j+1} = f\left[x_j, x_{j+1}\right] = \frac{f_{j+1-}f_j}{x_{j+1} - x_j} = \frac{f(x_{j+1}) - f(x_j)}{x_{j+1} - x_j}$$

and

$$a_{j+2} = f\left[x_j, x_{j+1}, x_{j+2}\right] = \frac{f\left[x_{j+1}, x_{j+2}\right] - f\left[x_j, x_{j+1}\right]}{x_{j+2} - x_j}.$$

Then the desired polynomial is

(C) $$p_{j+2}(x) = f_j + (x - x_j)f\left[x_j,x_{j+1}\right] + (x - x_j)(x - x_{j+1})f\left[x_j,x_{j+1},x_{j+2}\right].$$

From the five given points (x_j, f_j) we construct a table similar to the one in Example 4, p. 815. We get

j	x_j	$f_j = f(x_j)$	$a_{j+1} = f\left[x_j,x_{j+1}\right]$	$a_{j+2} = f\left[x_j,x_{j+1},x_{j+2}\right]$
0	−4	$\underline{50}$		
			$\dfrac{18-50}{-2-(-4)} = \underline{-16.0}$	
1	−2	18		$\dfrac{-8+16}{0+4} = \underline{2.0}$
			$\dfrac{2-18}{0-(-2)} = -8.0$	
2	0	2		$\dfrac{0+8}{2+2} = 2.0$
			$\dfrac{2-2}{2-0} = 0$	
3	2	2		$\dfrac{8-0}{4-0} = 2.0$
			$\dfrac{18-2}{4-2} = 8.0$	
4	4	18		

From the table and (C), with $j = 0$, we get the following interpolation polynomial. Note that, because all the $a_{j=2}$ differences are equal, we do not need to compute the remaining differences and the polynomial is of degree 2:

$$\begin{aligned}
p_2(x) &= f_0 + (x - x_0)f\,[x_0, x_1] + (x - x_0)(x - x_1)f\,[x_0,x_1,x_2] \quad \text{(see formula on top of p. 814)}\\
&= 50 + [x - (-4)]\,(-16.0) + [x - (-4)]\,[x - (-2)](2.0)\\
&= 50 + (x + 4)(-16.0) + (x + 4)(x + 2)(2.0)\\
&= 50 - 16x - 64 + 2x^2 + 12x + 16\\
&= 2x^2 + (-16 + 12)x + (50 - 64 + 16)\\
&= 2x^2 - 4x + 2.
\end{aligned}$$

15. **Newton's divided difference formula (10), p. 814.** Using the data from Example 2, p. 810, we build the following table:

j	x_j	f_j	$f\left[x_j,x_{j+1}\right]$	$f\left[x_j,x_{j+1},x_{j+2}\right]$
0	9.0	$\underline{2.1972}$		
			$\dfrac{2.2513-2.1972}{9.5-9} = \underline{0.1082}$	
1	9.5	2.2513		$\dfrac{0.09773-0.1082}{11-9} = \underline{-0.005235}$
			$\dfrac{2.3979-2.2513}{11-9.5} = 0.09773$	
2	11.0	2.3979		

Then, using the table with the values needed for (10) underscored, the desired polynomial is

$$\begin{aligned} p_2(x) &= 2.1972 + (x-9)\cdot 0.1082 + (x-9)\,(x-9.5)\cdot 0.005235 \\ &= 1.6709925 + 0.0113525x + 0.005235x^2. \end{aligned}$$

Sec. 19.4 Spline Interpolation

We continue our study of interpolation started in Sec. 19.3. Since, for large n, the interpolation polynomial $P_n(x)$ may oscillate wildly between the nodes $x_0, x_1, x_2, \dots, x_n$, the approach of Newton's interpolation with *one* polynomial of Sec. 19.3 may not be good enough, Indeed, this is illustrated in Fig. 434, p. 821, for $n = 10$, and it was shown by reknown numerical analyst Carl Runge that, in general, this example exhibits numeric instability. Also look at Fig. 435, p. 821.

The new approach is to use n low-degree polynomials involving two or three nodes instead of one high-degree polynomial connecting all the nodes! This method of **spline interpolation,** initiated by I. J. Schoenberg is used widely in applications and forms the basis for CAD (computer-aided design), for example, in car design (Bezier curves named after French engineer P. Bezier of the Renault Automobile Company, see p. 827 in Problem Set 19.4).

Here we concentrate on cubic splines as they are the most important ones in applications because they are smooth (continuous first derivative) and also have smooth first derivatives. Theorem 1 guarantees their existence and uniqueness. The proof and its completion (Prob. 3) suggest the approach for determining splines. The best way to understand Sec. 19.4 is to study **Example 1**, p. 824. It uses (12), (13), and (14) (equidistant nodes) on pp. 823–824. A second illustration is **Prob. 13**. Figure 437 of the Shrine of the Book in Jerusalem in **Example 2** (p. 825) shows the interpolation polynomial of degree 12, which oscillates (reminiscent of Runge's example in Fig. 434), whereas the spline follows the contour of the building quite accurately.

Problem Set 19.4. Page 826

3. **Existence and uniqueness of cubic splines. Derivation of (7) and (8) from (6), p. 822, from the Proof of Theorem 1.** Formula (6), p. 822, of the unique cubic polynomial is quite involved:

$$\begin{aligned} q_j(x) &= f(x_j)c_j^2(x-x_{j+1})^2[1+2c_j(x-x_j)] \\ &\quad + f(x_{j+1})c_j^2(x-x_j)^2[1+2c_j(x-x_{j+1})] \\ &\quad + k_jc_j^2(x-x_j)(x-x_{j+1})^2 \\ &\quad + k_{j+1}c_j^2(x-x_j)^2(x-x_{j+1}). \end{aligned} \tag{6}$$

We need to differentiate (6) twice to get (7) and (8), and one might make some errors in the (paper-and-pencil) derivation. The point of the problem then is that we can minimize our chance of making errors by introducing suitable short notations.

For instance, for the expressions involving x, we may set

$$X_j = x - x_j, \qquad X_{j+1} = x - x_{j+1},$$

and, for the constant quantities occurring in (6), we may choose the short notations:

$$A = f(x_j)\,c_j^2, \quad B = 2\,c_j, \quad C = f(x_{j+1})\,c_j^2, \quad D = k_j\,c_j^2, \quad E = k_{j+1}\,c_j^2.$$

Then formula (6) becomes simply

$$q_j(x) = AX_{j+1}^2(1+BX_j) + CX_j^2\,(1-BX_{j+1}) + DX_jX_{j+1}^2 + EX_j^2X_{j+1}.$$

Differentiate this twice with respect to x, applying the product rule for the second derivative, that is,

$$(uv)'' = u''v + 2u'v' + uv'',$$

and noting that the first derivative of X_j is simply 1, and so is that of X_{j+1}. (Of course, one may do the differentiations in two steps if one wants to.) We obtain

$$\text{(I)}\qquad \begin{aligned} q_j''(x) &= A[2(1+BX_j)+4X_{j+1}B+0]+C[2(1-BX_{j+1})+4X_j(-B)+0] \\ &\quad + D(0+4X_{j+1}+2X_j)+E(2X_{j+1}+4X_j+0), \end{aligned}$$

where $4 = 2 \cdot 2$ with one 2 resulting from the product rule and the other from differentiating a square. And the zeros arise from factors whose second derivative is zero.

Now calculate q_j'' at $x = x_j$. Since

$$X_j = x - x_j, \qquad \text{we see that} \qquad X_j = 0 \qquad \text{at} \qquad x = x_j.$$

Hence, in each line, the term containing X_j disappears. This gives

$$q_j''(x_j) = A(2+4BX_{j+1}) + C(2-2BX_{j+1}) + 4DX_{j+1} + 2EX_{j+1}.$$

Also, when $x = x_j$, then

$$X_{j+1} = x_j - x_{j+1} = -\frac{1}{c_j} \qquad \text{[see (6*), p. 822, which defines } c_j\text{]}.$$

Inserting this, as well as the expressions for $A, B, \dots, E$, we obtain (7) on p. 822. Indeed,

$$q_j''(x_j) = f(x_j)c_j^2\left(2+2\cdot\frac{4c_j}{-c_j}\right) + f(x_{j+1})c_j^2\left(2-2\cdot\frac{2c_j}{-c_j}\right) + \frac{4k_jc_j^2}{-c_j} + \frac{2k_{j+1}c_j^2}{-c_j}.$$

Cancellation of some of the factors involving c_j gives

$$\text{(7)}\qquad q_j''(x_j) = -6f(x_j)c_j^2 + 6f(x_{j+1})c_j^2 - 4k_jc_j - 2k_{j+1}c_j.$$

The derivation of (8), p. 822, is similar.

For $x = x_{j+1}$, we have

$$X_{j+1} = x_{j+1} - x_{j+1} = 0,$$

so that (I) simplifies to

$$q_j''(x_{j+1}) = A(2+2BX_j) + C(2-4BX_j) + 2DX_j + 4EX_j.$$

Furthermore, for $x = x_{j+1}$, we have, by (6*), p. 822,

$$X_j = x_{j+1} - x_j = \frac{1}{c_j},$$

and, by substituting $A, \dots, E$ into the last equation, we obtain

$$q_j''(x_{j+1}) = f(x_j)c_j^2\left(2+\frac{4c_j}{c_j}\right) + f(x_{j+1})c_j^2\left(2-\frac{8c_j}{c_j}\right) + \frac{2k_jc_j^2}{c_j} + \frac{4k_{j+1}c_j^2}{c_j}.$$

Again, cancellation of some factors c_j and simplification finally gives (8), that is,

$$q_j''(x_{j+1}) = 6c_j^2 f(x_j) - 6c_j^2 f(x_{j+1}) + 2c_j k_j + 4c_j k_{j+1}. \tag{8}$$

For practice and obtaining familiarity with cubic splines, you may want to work out all the details of the derivation.

13. Determination of a spline. We proceed as in Example 1, p. 824. Arrange the given data in a table for easier work:

j	x_j	$f(x_j)$	k_j
0	0	1	0
1	1	0	
2	2	−1	
3	3	0	−6

Since there are four nodes, the spline will consist of three polynomials, $q_0(x)$, $q_1(x)$, and $q_2(x)$. The polynomial $q_0(x)$ gives the spline for x from 0 to 1, $q_1(x)$ gives the spline for x from 1 to 2, and $q_2(x)$ gives the spline for x from 2 to 3, respectively.

Step 1. Since $n = 3$ and $h = 1$, (14), p. 824, has two equations:

$$k_0 + 4k_1 + k_2 = 0 + 4k_1 + k_2 = \frac{3}{h}(f_2 - f_0) = -6, \qquad j = 1,$$
$$k_1 + 4k_2 + k_3 = k_1 + 4k_2 - 6 = \frac{3}{h}(f_3 - f_1) = 0, \qquad j = 2.$$

It is easy to show, by direct substitution, that $k_1 = -2$ and $k_2 = 2$ satisfy these equations.

Step 2 for $q_0(x)$ *Determine the coefficients of the spline from* (*13*), *p. 823.* We see that, in general, $j = 0, \ldots, n-1$, so that, in the present case, we have $j = 0$ (this will give the spline from 0 to 1), $j = 1$ (which will give the spline from 1 to 2), and $j = 2$ (which will give the spline from 2 to 3). Take $j = 0$. Then (13) gives

$$\begin{aligned}
a_{00} &= q_0(p_0) = f_0 = 1,\\
a_{01} &= q_0'(x_0) = k_0 = 0,\\
a_{02} &= \frac{1}{2}q_0''(x_0) = \frac{3}{1^2}(f_1 - f_0) - \frac{1}{1}(k_1 - 2\,k_0) = 3 \cdot (0 - 1) - (-2 - 0) = -1,\\
a_{03} &= \frac{1}{6}q_0'''(x_0) = \frac{2}{1^3}(f_0 - f_1) + \frac{1}{1^2}(k_1 + k_0) = 2 \cdot (1 - 0) + (-2 + 0) = 0.
\end{aligned}$$

With these Taylor coefficients we obtain, from (12), p. 823, the first part of the spline in the form

$$\begin{aligned}
q_0(x) &= 1 + 0(x - x_0) - 1(x - x_0)^2 + 0(x - x_0)^3\\
&= 1 + 0 - 1(x - 0)^2 + 0\,(x - 0)^3\\
&= 1 - x^2.
\end{aligned}$$

Step 2 for $q_1(x)$

$$\begin{aligned}
a_{10} &= q_1(x_1) = f_1 = 0, \\
a_{11} &= q_1'(x_1) = k_1 = -2, \\
a_{12} &= \frac{1}{2}q_1''(x_1) = \frac{3}{1^2}(f_2 - f_1) - \frac{1}{1}(k_2 + 2k_1) = 3 \cdot (-1-0) - (2-4) = -1, \\
a_{13} &= \frac{1}{6}q_1'''(x_1) = \frac{2}{1^3}(f_1 - f_2) + \frac{1}{1^2}(k_2 + k_1) = 2 \cdot (0+1) + (2-2) = 2.
\end{aligned}$$

With these coefficients and $x_1 = 0$ we obtain from (12), p. 823, with $j = 1$ the polynomial

$$\begin{aligned}
q_1(x) &= 0 - 2(x - x_1) - 1(x - x_1)^2 + 2(x - x_1)^3 \\
&= -2(x-1) - (x-1)^2 + 2(x-1)^3 \\
&= -1 + 6x - 7x^2 + 2x^3,
\end{aligned}$$

which gives the spline on the interval from 1 to 2.

Step 3 for $q_2(x)$

$$\begin{aligned}
a_{20} &= q_2(x_2) = f_2 = -1, \\
a_{21} &= q_2'(x_2) = k_2 = 2, \\
a_{22} &= \frac{1}{2}q_2''(x_2) = \frac{3}{1^2}(f_3 - f_2) - \frac{1}{1}(k_3 + 2k_2) = 3 \cdot (0+1) - (-6+4) = 5, \\
a_{23} &= \frac{1}{6}q_2'''(x_2) = \frac{2}{1^3}(f_2 - f_3) + \frac{1}{1^2}(k_3 + k_2) = 2 \cdot (-1-1) + (-6+2) = -6.
\end{aligned}$$

With these coefficients and $x_1 = 0$ we obtain, from (12), p. 823, with $j = 1$, the polynomial

$$\begin{aligned}
q_2(x) &= -1 + 2(x - x_2) + 5(x - x_2)^2 - 6(x - x_2)^3 \\
&= -1 + 2(x-2) + 5(x-2)^2 - 6(x-2)^3 \\
&= 63 - 90x + 41x^2 - 6x^3,
\end{aligned}$$

which gives the spline on the interval from 2 to 3.

To check the answer, you should verify that the spline gives the function values $f(x_j)$ and the values k_j of the derivatives in the table at the beginning. Also make sure that the first and second derivatives of the spline at $x = 1$ are continuous by verifying that

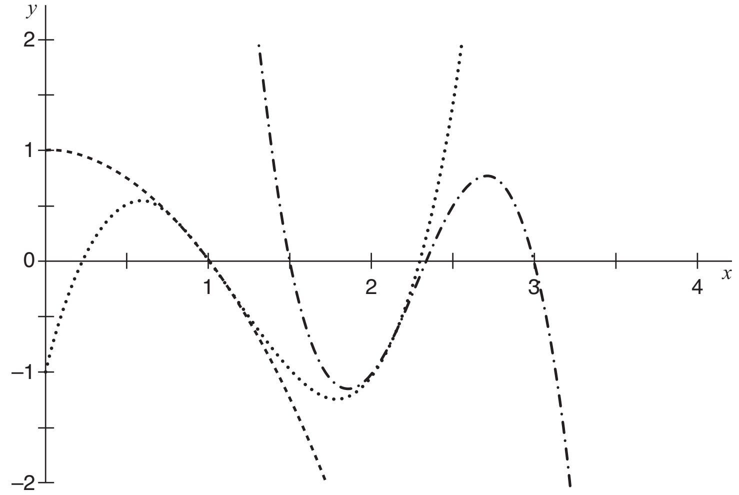
$$q_0'(1) = q_1'(1) = -2 \quad \text{and} \quad q_0''(1) = q_1''(1) = -2.$$

The third derivative is no longer continuous,

$$q_0'''(1) = 0 \quad \text{but} \quad q_1'''(1) = 12.$$

(Otherwise you would have a single cubic polynomial from 0 to 1.)

Do the same for $x = 2$.



Sec. 19.4 Prob. 13. Spline

We see that in the graph the curve q_0 is represented by the dashed line (— — —), q_1 by the dotted line (· · ·), and q_2 by the dot-dash line (— • —·).

Sec. 19.5 Numeric Integration and Differentiation

The essential idea of **numeric integration** is to approximate the integral by a sum that can be easily evaluated. There are different ways to do this approximation and the best way to understand them is to look at the diagrams.

The simplest numeric integration is the **rectangular rule** where we approximate the area under the curve by rectangles of given (often equal) width and height by a constant value (usually the value at an endpoint or the midpoint) over that width as shown in Fig. 441 on p. 828. This gives us formula (1) and is illustrated in **Prob. 1.**

We usually get more accuracy if we replace the rectangles by trapezoids in Fig. 442. p. 828, and we obtain the **trapezoidal rule** (2) as illustrated in **Example 1**, p. 829, and **Prob. 5**. We discuss various error estimates of the trapezoidal rule (see pp. 829–831) in equations (3), (4), and (5) and apply them in Example 2 and Prob. 5.

Most important in this section is **Simpson's rule** on p. 832:

$$\int_a^b f(x)\,dx \approx \frac{h}{3}(f_0 + 4f_1 + 2f_2 + 4f_3 + \cdots + 2f_{2m-2} + 4f_{2m-2} + f_{2m}), \tag{7}$$

where

$$h = \frac{b-a}{2m} \qquad \text{and} \qquad f_j \text{ stands for } f(x_j).$$

Simpson's rule is sufficiently accurate for most problems but still sufficiently simple to compute and is stable with respect to rounding. Errors are given by (9), p. 833, and (10), p. 834, **Examples 3, 4, 5, 6 ("adaptive integration")** (see pp. 833–835 of the textbook), and **Prob. 17** give various illustrations of this important practical method. The discussion on numeric integration ends with Gauss integration (11), p. 837, with Table 19.7 listing nodes and coefficients for $n = 2, 3, 4, 5$ (see Examples 7 and 8, pp. 837–838, Prob. 25).

Whereas integration is a process of "smoothing," **numeric differentiation** "makes things rough" (tends to enlarge errors) and should be avoided as much as possible by changing models—but we shall need it in Chap. 21 on the numeric solution of partial differential equations (PDEs).

Problem Set 19.5. Page 839

1. Rectangular rule (1), p. 828. This rule is generally too inaccurate in practice. Our task is to evaluate the integral of Example 1, p. 829,

$$J = \int_0^1 e^{-x^2} dx$$

by means of the rectangular rule (1) with intervals of size 0.1. The integral cannot be evaluated by elementary calculus, but leads to the error function erf x, defined by (35), p. A67, in Sec. A3.1, of App. 3 of the textbook.

Since, in (1), we take the midpoints 0.05, 0.15, . . . , we calculate

j	x_j^*	$-x_j^{*2}$	$f(x_j^*) = \exp\left(-x_j^{*2}\right)$
1	0.05	−0.0025	0.997503
2	0.15	−0.0225	0.977751
3	0.25	−0.0625	0.939413
4	0.35	−0.1225	0.884706
5	0.45	−0.2025	0.816686
6	0.55	−0.3025	0.738968
7	0.65	−0.4225	0.655406
8	0.75	−0.5625	0.569783
9	0.85	−0.7225	0.485537
10	0.95	−0.9025	0.405555
		Sum	$7.471308 = \sum_{j=1}^{10} f(x_j^*)$

Since the upper limit of integration is $b = 1$, the lower limit $a = 0$, and the number of subintervals $n = 10$, we get

$$h = \frac{b-a}{n} = \frac{1-0}{10} = \frac{1}{10} = 0.1.$$

Hence by (1), p. 828,

$$\text{Rectangular rule: } J = \int_0^1 e^{-x^2} dx \approx h \sum_{j=1}^{10} f(x_j^*) = 0.1 \cdot 7.471308 = 0.7471308 = 0.747131 \text{ (6S).}$$

We compare this with the exact 6S value of 0.746824 and obtain

$$\begin{aligned}\text{Error for rectangular rule} &= \text{True Value} - \text{Approximation}\\ &= 0.746824 - 0.747131 = -0.000307 \quad \text{[by (6), p. 794].}\end{aligned}$$

We compare our result with the one obtained in Example 1, p. 829, by the trapezoidal rule (2) on that page, that is,

$$\begin{aligned}\text{Error for trapezoidal rule} &= \text{True Value} - \text{Approximation}\\ &= 0.746824 - 0.746211 = -0.0000613 \quad \text{[by (6), p. 794].}\end{aligned}$$

This shows that the trapezoidal rule gave a more accurate answer, as was expected.

Here are some questions worth pondering about related to the rectangular rule in our calculations.

When using the rectangular rule, the approximate value was larger than the true value. Why? (Answer: The curve of the integrand is concave.)

What would you get if you took the left endpoint of each subinterval? (Answer: An upper bound for the value of the integral.)

If you took the right endpoint? (Answer: A lower bound.)

5. Trapezoidal rule: **Error estimation by halving.** The question asks us to evaluate the integral

$$J = \int_0^1 \sin\left(\frac{\pi x}{2}\right) dx$$

by the trapezoidal rule (2), p. 829, with $h = 1, 0.5,\ 0.25$ and estimate its error for $h = 0.5$ and $h = 0.25$ by halving, defined by (5), p. 830.

Step 1. *Obtain the true value of* J. The purpose of such problems (that can readily be solved by calculus) is to demonstrate a numeric method and its quality—by allowing us to calculate errors (6), p.794, and error estimates [here (5), p. 830]. We solve the indefinite integral by substitution

$$u = \frac{\pi x}{2}, \qquad \frac{du}{dx} = \frac{\pi}{2}, \qquad dx = \frac{2}{\pi} du$$

and

$$\int \sin\left(\frac{\pi x}{2}\right) dx = \int (\sin u) \frac{2}{\pi} du = \frac{2}{\pi} \int \sin u \, du = -\frac{2}{\pi} \cos u = -\frac{2}{\pi} \cos\left(\frac{\pi x}{2}\right).$$

Hence the definite integral evaluates to

$$J = \int_0^1 \sin\left(\frac{\pi x}{2}\right) dx = -\frac{2}{\pi}\left[\cos\left(\frac{\pi x}{2}\right)\right]_0^1$$

$$\text{(A)} \qquad = -\frac{2}{\pi}\left[\cos\left(\frac{\pi}{2}\right) - \cos 0\right] = -\frac{2}{\pi}(0-1) = \frac{2}{\pi} = 0.63662.$$

Step 2a. *Evaluate the integral by the trapezoidal rule* (2), *p. 821, with* $h = 1$. In the trapezoidal rule (2) we subdivide the interval of integration $a \le x \le b$ into n subintervals of equal length h, so that

$$h = \frac{b-a}{n}.$$

We also approximate f by a broken line of segments as in Fig. 442, p. 828, and obtain

$$\text{(2)} \qquad J_h = \int_a^b f(x)\, dx = h\left[\frac{1}{2} f(a) + f(x_1) + f(x_2) + \cdots + f(x_{n-1}) + \frac{1}{2} f(b)\right].$$

From (A), we know that the limits of integration are $a = 0$, $b = 1$. With $h = 1$ we get

$$n = \frac{b-a}{h} = \frac{1-0}{1} = 1 \text{ interval}; \qquad \text{that is,} \qquad \text{interval } [a, b] = [0, 1].$$

Hence (2) simplifies to

$$
\begin{aligned}
J_{1.0} &= \int_a^b f(x)\,dx \\
\text{(B)} \qquad &= h\left[\frac{1}{2}f(a) + \frac{1}{2}f(b)\right] \\
&= 1.0\left[\frac{1}{2}f(0) + \frac{1}{2}f(1)\right] \\
&= 1.0\left[\frac{1}{2}\sin 0 + \frac{1}{2}\sin\left(\frac{\pi}{2}\right)\right] = 1.0\left[\frac{1}{2}\cdot 0 + \frac{1}{2}\cdot 1\right] = \frac{1}{2} = 0.50000.
\end{aligned}
$$

From (A) and (B) we see that the error is

$$\text{Error} = \text{Truevalue} - \text{approximation} = 0.63662 - 0.50000 = 0.13662 \qquad \text{[by (6), p. 794]}$$

Step 2b. *Evaluate the integral by the trapezoidal rule* (2) *with* $h = 0.5$. We get

$$n = \frac{b-a}{h} = \frac{1-0}{0.5} = 2 \text{ intervals.}$$

The whole interval extends from 0 to 1, so that two equally spaced subintervals would be $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$. Hence

$$
\begin{aligned}
J_{0.5} &= h\left[\frac{1}{2}(f(a) + f(x_1)) + \frac{1}{2}(f(x_1) + f(b))\right] = 0.5\left[\frac{1}{2}f(0) + f\left(\frac{1}{2}\right) + \frac{1}{2}f(1)\right] \\
\text{(C)} \qquad &= 0.5\left[\frac{1}{2}\sin 0 + \sin\left(\frac{\pi\cdot\frac{1}{2}}{2}\right) + \frac{1}{2}\sin\left(\frac{\pi}{2}\right)\right] \\
&= 0.5\left[\frac{1}{2}\cdot 0 + \frac{\sqrt{2}}{2} + \frac{1}{2}\cdot 1\right] = \frac{\sqrt{2}+1}{4} = 0.60355 \qquad \left[\text{using } \sin\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}\right]
\end{aligned}
$$

with an error of $0.63662 - 0.60355 = 0.03307$.

Step 2c. *Evaluate by* (2) *with* $h = 0.25$. We get

$$n = \frac{b-a}{h} = \frac{1-0}{0.25} = 4 \text{ intervals.}$$

They each have a length of $\frac{1}{4}$ and so are $[0, \frac{1}{4}]$, $[\frac{1}{4}, \frac{1}{2}]$, $[\frac{1}{2}, \frac{3}{4}]$, and $[\frac{3}{4}, 1]$.

$$
\begin{aligned}
J_{0.25} &= h\left[\frac{1}{2}f(a) + f(x_1) + f(x_2) + f(x_3) + \frac{1}{2}f(b)\right] \\
\text{(D)} \qquad &= 0.25\left[\frac{1}{2}f(0) + f\left(\frac{1}{4}\right) + f\left(\frac{1}{2}\right) + f\left(\frac{3}{4}\right) + \frac{1}{2}f(1)\right] \\
&= 0.25\left[\frac{1}{2}\sin 0 + \sin\left(\frac{\pi}{8}\right) + \sin\left(\frac{\pi}{4}\right) + \sin\left(\frac{3\pi}{8}\right) + \frac{1}{2}\sin\left(\frac{\pi}{2}\right)\right] \\
&= 0.25\cdot(0 + 0.38268 + 0.70711 + 0.92388 + 0.50000) = 0.62842.
\end{aligned}
$$

The error is $0.63662 - 0.62842 = 0.00820$.

Step 3a. *Estimate the error by halving, that is, calculate* $\epsilon_{0.5}$ *by* (5), p. 830. Turn to pp. 829–830 of the textbook. Note that the error (3), p. 830, contains the factor h^2. Hence, in halving, we can expect the error to be multiplied by about $\left(\frac{1}{2}\right)^2 = \frac{1}{4}$. Indeed, this property is nicely reflected by the numerical values (B)–(D). Now we turn to error estimating (5), that is,

$$\epsilon_{h/2} \approx \tfrac{1}{3}(J_{h/2} - J_h). \tag{5}$$

Here we obtain

$$\epsilon_{0.5} \approx \tfrac{1}{3}(J_{0.5} - J_{1.0}) = \tfrac{1}{3}(0.60355 - 0.50000) = 0.03452.$$

The agreement of this estimate 0.03452 with the actual value of the error 0.03307 is good.

Step 3b. *Estimate the error by halving, that is, calculate* $\epsilon_{0.25}$. We get, using (5),

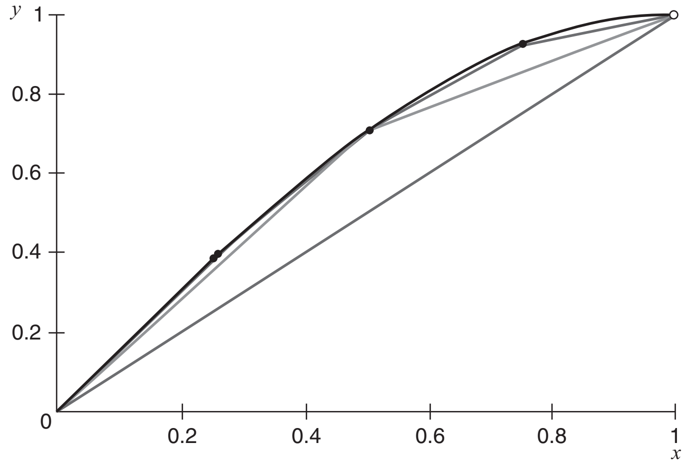
$$\epsilon_{0.25} \approx \tfrac{1}{3}(J_{0.25} - J_{0.5}) = \tfrac{1}{3}(0.62842 - 0.60355) = 0.00829,$$

which compares very well with the actual error, that is,

$$0.00820 - 0.00829 = -0.00009.$$

Although, in other cases, the difference between estimate and actual value may be larger, estimation will still serve its purpose, namely, to give an impression of the order of magnitude of the error.

Remark. Note that since we calculated the integral by (2), p. 829, for three choices of $h = 1$, 0.5, 0.25 in Steps 2a–2c, we were able to make two error estimates (5), p. 830, in steps 3a, 3b.



Sec. 19.5 Prob. 5. Given sine curve and approximating polygons in the three trapezoidal rules used. The agreement of these estimates with the actual value of the errors is very good

17. Simpson's rule for a nonelementary integral. Simpson's rule (7), p. 832, is

$$\int_a^b f(x)\,dx \approx \frac{h}{3}(f_0 + 4f_1 + 2f_2 + 4f_3 + \cdots + 2f_{2m-2} + 4f_{2m-1} + f_{2m}), \tag{7}$$

where

$$h = \frac{b-a}{2m} \qquad \text{and} \qquad f_j \text{ stands for } f(x_j).$$

The nonelementary integral is the sine integral (40) in Sec. A3.1 of App. 3 on p. A68 of the textbook:

$$\mathrm{Si}(x) = \int_0^x \frac{\sin x^*}{x^*}\, dx^*.$$

Being nonelementary means that we cannot solve the integral by calculus. For $x = 1$, its exact value (by your CAS or Table A4 on p. A98 in App. 5) is

$$\mathrm{Si}(1) = \int_0^1 \frac{\sin x}{x}\, dx = 0.9460831.$$

We construct a table with both $2m = 2$ and $2m = 4$, with values of the integrand accurate to seven digits

j	x_j	$f_j = f(x_j) = \dfrac{\sin x_j}{x_j}$	j	x_j	$f_j = f(x_j) = \dfrac{\sin x_j}{x_j}$
0	0	1.0000000	0	0	1.0000000
			1	0.25	0.9896158
1	0.5	0.9588511	2	0.5	0.9588511
			3	0.75	0.9088517
2	1.0	0.8414710	4	1.0	0.8414710

Simpson's rule, with $m = 1$, i.e., $h = 0.5$, is

$$\mathrm{Si}(1) = \frac{h}{3}(f_0 + 4f_1 + f_2) = \frac{0.5}{3}(1 + 4 \cdot 0.9588511 + 0.8414710) = 0.9461459.$$

With $m = 2$, i.e. $h = 0.25$,

$$\begin{aligned}\mathrm{Si}(1) &= \frac{h}{3}(f_0 + 4f_1 + 2f_2 + 4f_3 + f_4)\\ &= \frac{0.25}{3}(1 + 4 \cdot 0.9896158 + 2 \cdot 0.9588511 + 4 \cdot 0.9088517 + 0.8414710) = 0.9460870.\end{aligned}$$

25. Gauss integration for the error function. $n = 5$ is required. The transformation must convert the interval to $[-1, 1]$.

We can do this with $x = at + b$ so that $0 = a(-1) + b$ and $1 = a(1) + b$. We see that $a = b = \frac{1}{2}$ satisfies this so $x = \frac{1}{2}(t + 1)$.

Since $dx = \frac{1}{2}\, dt$, our integral takes the form

$$\int_0^1 e^{-x^2}\, dx = \frac{1}{2}\int_{-1}^1 e^{-1/4(t+1)^2}\, dt.$$

The exact 9S value is 0.746824133.

The nodes and coefficients are shown in Table 19.7, on p. 837, in the textbook with $n = 5$. Using them, we compute

$$J = \frac{1}{2}\sum_{j=1}^{5} A_j e^{-1/4(t_j+1)^2} = 0.746824127.$$

Note the high accuracy achieved with a rather modest amount of work.

Multiply this by $2/\sqrt{\pi}$ to obtain an approximation to the error function erf 1 (= 0.842700793 with $(9S)$) given by (35) on p. A67 in App. 3.1

$$\left(\frac{2}{\sqrt{\pi}}\right) 0.746824127 = 0.842700786.$$

Solution to Self Test on Rounding Problem to Decimals (see p. 2 of this Solutions Manual and Study Guide)

(a) $$1.23454621 + 5 \cdot 10^{-(7+1)} = 1.23454621 + 0.\underbrace{0000000}_{\text{7 zeros}}5 = 1.23454626.$$

Then we chop off the eigth digit "6" and obtain the rounded number to seven decimals (7D) 1.2345462.

(b) $$\begin{aligned} -398.723555 + 5 \cdot 10^{-(4+1)} &= -398.723555 + 5 \cdot 10^{-5} \\ &= -398.723555 + 0.\underbrace{0000}_{\text{4 zeros}}5 \\ &= -398.723605 \end{aligned}$$

Next we chop off from the fifth digit onward, that is, "05" and obtain the rounded number to four decimals (4D) −398.7236.

Solution to Self Test on Rounding Problem to Significant Digits (see p. 3 of this Solutions Manual and Study Guide)

We follow the three steps.

1. $102.89565 = 0.10289565 \cdot 10^3$;

2. We ignore the factor 10^3. Then we apply the roundoff rule for decimals to the number 0.10289565 to get

$$0.10289565 + 5 \cdot 10^{-(6+1)} = 0.10289615 \ \ \text{(6D)}.$$

3. Finally we have to reintroduce the factor 10^3 to obtain our final answer, that is,

$$0.102896 \cdot 10^3 = 102.896 \text{ (6S)}.$$

Chap. 20 Numeric Linear Algebra

Chapter 20 contains two main topics: ***solving systems of linear equations numerically*** (Secs. 20.1–20.5, pp. 844–876) and ***solving eigenvalue problems numerically*** (Secs. 20.6–20.9, pp. 876–898). Highlights are as follows.

Section 20.1 starts with the familiar **Gauss elimination method** (from Sec. 7.3), *now in the context of numerics* with partial pivoting, row scaling, and operation count and the method itself expressed in algorithmic form. This is followed by methods that are more efficient than Gauss (*Doolittle*, *Crout*, *Cholesky*) in Sec. 20.2 and iterative methods (*Gauss–Seidel*, *Jacobi*) in Sec. 20.3. We study the behavior of linear systems in detail in Sec. 20.4 and introduce the concept of a *condition number* that will help us to determine whether a system is good ("well-conditioned") or bad ("ill-conditioned"). The first part of Chap. 20 closes with the *least squares method*, an application in curve fitting, which has important uses in statistics (see Sec. 25.9).

Although we can find the roots of the characteristic equations in eigenvalue problems by methods from Sec. 19.2, such as Newton's method, there are other ways in numerics concerned with eigenvalue problems. Quite surprising is *Gerschgorin's theorem* in Sec. 20.7 because it allows us to obtain information directly, i.e., without iteration, from the elements of a square matrix, about the range in which the eigenvalues of that matrix lie. This is, of course, not as good as obtaining actual numbers for those eigenvalues but is sufficient in some problems.

Other approaches are an iterative method to determine an approximation of a dominant eigenvalue in a square matrix (the *power method*, Sec. 20.8) and a two-stage method to compute all the eigenvalues of a real symmetric matrix in Sec. 20.9.

The chapter has both easier and more involved sections. **Sections 20.2, 20.3, 20.7, 20.9 are more involved and may require more study time. You should remember formulas (4) and (8) in Sec. 20.5 and their use.**

In terms of prior knowledge, you should be familiar with matrices (Secs. 7.1, 7.2), and it would be helpful if you had some prior knowledge of *Gauss elimination with back substitution* (see Sec. 7.3, pp. 272–280). Section 20.1 moves faster than Sec. 7.3 and does not contain some details such as the three types of solutions that occur in linear systems. For the second main topic of Chap. 20 you should be familiar with the material contained in Sec. 20.6, pp. 876–879. Thus you should know what a matrix eigenvalue problem is (pp. 323–324), *remember how to find eigenvectors and eigenvalues of matrices* (pp. 324–328 in Sec. 8.1), know that similar matrices have the same eigenvalues (see Theorem 2, p. 878, also Theorem 3, p. 340 in Sec. 8.4), and refresh your knowledge of special matrices in Theorem 5, p. 879.

Sec. 20.1 Linear Systems: Gauss Elimination

Gauss elimination with back substitution is a systematic way of solving systems of linear equations (1), p. 845. We discussed this method before in Sec. 7.3 (pp. 272–282) in the context of linear algebra. This time the *context is numerics* and the current discussion is kept independent of Chap. 7, except for an occasional reference to that chapter. Pay close attention to the partial pivoting introduced here, as it is the main difference between the Gauss elimination presented in Sec. 20.1 and that of Sec. 7.3. The reason that we need pivoting in numerics is that we have only a finite number of digits available. With many systems, this can result in a severe loss of accuracy. Here (p. 846), to pivot a_{kk}, we choose as our pivoting equation the one that has the *absolutely largest* coefficients a_{jk} in column k on or below the main diagonal. The details are explained carefully in a completely worked out **Example 1**, pp. 846–847. The importance of this particular partial pivoting strategy is demonstrated in **Example 3**, pp. 848–849. In (a) the "absolutely largest" partial pivoting strategy is not followed and leads to a bad value for x_1. This corresponds to the method of Sec. 7.3. In (b) it is followed and a good value for x_1 is obtained!

Table 20.1, p. 849, presents Gauss elimination with back substitution in algorithmic form. The section ends with an operation count of $2n^3/3$ for Gauss elimination (p. 850) and $n^2 + n$ for back substitution (p. 851). *Operation count* is one way to judge the quality of a numeric method.

The solved problems show that a system of linear equations may have *no solution* (**Prob. 3**), a *unique solution* (**Prob. 9**), or *infinitely many solutions* (**Prob. 11**). This was also explained in detail on pp. 277–280 in Sec. 7.3. You may want to solve a few problems by hand until you feel reasonably comfortable with the Gaussian algorithm and the particular type of pivoting.

Problem Set 20.1. Page 851

3. System without a solution. We are given a system of two linear equations

$$\begin{aligned} &[\text{Eq. (1)}] & 7.2x_1 - 3.5x_2 &= 16.0, \\ &[\text{Eq. (2)}] & -14.4x_1 + 7.0x_2 &= 31.0. \end{aligned}$$

We multiply the first equation [Eq. (1)] by 2 to get

$$14.4x_1 - 7.0x_2 = 32.0.$$

If we add this equation to the second equation [Eq. (2)] of the given system, we get

$$0x_1 + 0x_2 = 63.0.$$

This last equation has no solution because the x_1, x_2 are each multiplied by 0, added, and equated to 63.0! Or looking at it in another way, we get the false statement that $0 = 63$. [A solution would exist if the right sides of Eq. (1) and Eq. (2) were related in the same fashion, for instance, 16.0 and −32.0 instead of 31.0.] Of course, for most systems with more than two equations, one cannot immediately see whether there will be solutions, but the Gauss elimination with partial pivoting will work in each case, giving the solution(s) or indicating that there is none. Geometrically, the result means that these equations represent two lines with the same slope of

$$\frac{7.2}{3.5} = \frac{14.4}{7.0} = 2.057143$$

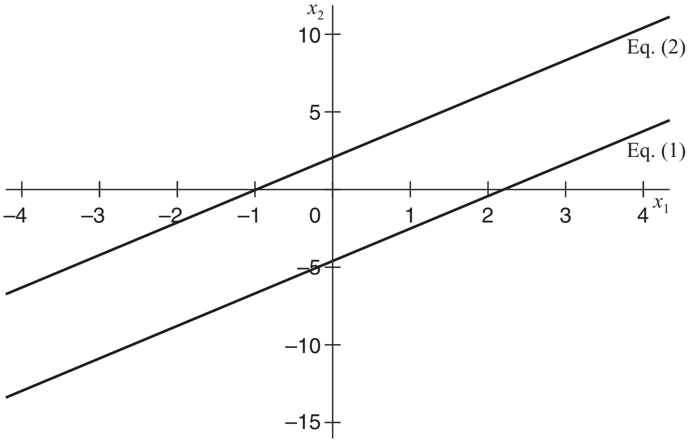
but different x_2-intercepts, that is,

$$\begin{aligned} & & -\frac{16.0}{3.5} &= -4.571429 & &\text{for Eq. (1),} \\ &\text{and} & \frac{31.0}{7.0} &= 4.428571 & &\text{for Eq. (2).} \end{aligned}$$

Hence Eq. (1) and Eq. (2) are parallel lines, as show in the figure on the next page.

9. System with a unique solution. Pivoting. ALGORITHM GAUSS, p. 849. Open your textbook to p. 849 and consider Table 20.1, which contains the algorithm for the Gauss elimination. To follow the discussion, control it for Prob. 9 in terms of matrices with paper and pencil. In each case, write down all three rows of a matrix, not just one or two rows, as is done below to save some space and to avoid copying the same numbers several times.

At the beginning, $k = 1$. Since $a_{11} = 0$, we must pivot. Between lines 1 and 2 in Table 20.1 we search for the absolutely greatest a_{j1}. This is $a_{31} (= 13)$. According to the algorithm, we have to interchange Eqs. (1) (current row) and (3) (row with the maximum), that is, Rows 1 and 3 of the *augmented* matrix. This gives



Sec. 20.1 Prob. 3. Graphic solution of a system of two parallel equations [Eq. (1) and Eq. (2)]

(A)
$$\left[\begin{array}{ccc|c} 13 & -8 & 0 & 178.54 \\ 6 & 0 & -8 & -85.88 \\ 0 & 6 & 13 & 137.86 \end{array}\right].$$

Don't forget to interchange the entries on the right side (that is, in the last column of the augmented matrix).

To get 0 as the first entry of Row 2, subtract $\frac{6}{13}$ times Row 1 from Row 2. The new Row 2 is

(A2)
$$\left[\begin{array}{ccc|c} 0 & 3.692308 & -8 & -168.28308 \end{array}\right].$$

This was $k = 1$ and $j = 2$ in lines 4 and 5 in the table.

Now comes $k = 1$ and $j = n = 3$ in line 4. The calculation is

$$m_{31} = \frac{a_{31}}{a_{11}} = \frac{0}{13} = 0.$$

Hence, the operations in line 4 simply have no effect; they merely reproduce Row 3 of the matrix in (A). This completes $k = 1$.

Next is $k = 2$. In the loop between lines 1 and 2 in Table 20.1, we have the following: Since $6 > 3.692308$, the maximum is in Row 3 so we interchange Row 2 (A2) and Row 3 in (A). This gives the matrix

(B)
$$\left[\begin{array}{ccc|c} 13 & -8 & 0 & 178.54 \\ 0 & 6 & 13 & 137.86 \\ 0 & 3.692308 & -8 & -168.28308 \end{array}\right].$$

In line 4 of the table with $k = 2$ and $j = k + 1 = 3$ we calculate

$$m_{32} = \frac{a_{32}}{a_{22}} = \frac{3.692308}{6} = 0.615385.$$

Performing the operations in line 5 of the table for $p = 3,\ 4$, we obtain the new Row 3

$$\text{(B3)} \qquad \left[\begin{array}{ccc|c} 0 & 0 & -16 & -253.12 \end{array}\right].$$

The system and its matrix have now reached triangular form.

We begin *back substitution* with line 6 of the table:

$$x_3 = \frac{a_{34}}{a_{33}} = \frac{-253.12}{-16} = 15.82.$$

(Remember that, in the table, the right sides $b_1,\ b_2,\ b_3$ are denoted by $a_{14},\ a_{24},\ a_{34}$, respectively.) Line 7 of the table with $i = 2, 1$ gives

$$x_2 = \tfrac{1}{6}(137.86 - 13 \cdot 15.82) = -11.3 \qquad (i = 2)$$

and

$$x_1 = \tfrac{1}{13}(178.54 - (-8 \cdot (-11.3)) = 6.78 \qquad (i = 1).$$

Note that, depending on the number of digits you use in your calculation, your values may be slightly affected by roundoff.

11. System with more than one solution. Homogeneous system. A homogeneous system always has the trivial solution $x_1 = 0$, $x_2 = 0, \ldots, x_n = 0$. Say the coefficient matrix of the homogeneous system has rank r. The homogeneous system has a *nontrivial* solution if and only if

$$r < n \qquad \text{where } n \text{ is the number of unknowns.}$$

The details are given in Theorem 2, p. 290 in Sec. 7.5, and related Theorem 3, p. 291.

In the present problem, we have a homogenous system with $n = 3$ equations. For such a system, we may have $r = 3$ (the trivial solution only), $r = 2$ [one (suitable) unknown remains arbitrary—infinitely many solutions], and $r = 1$ [two (suitable) variables remain arbitrary, infinitely many solutions]. Note that $r = 0$ is impossible unless the matrices are zero matrices. In most cases we have choices as to which of the variables we want to leave arbitrary; the present result will show this. To avoid misunderstandings: we need not determine those ranks, but the Gauss elimination will automatically give all solutions. *Your CAS may give only some solutions* (for example, those obtained by equating arbitrary unknowns to zero); so be careful.

The augmented matrix of the given system is

$$\left[\begin{array}{rrr|r} 3.4 & -6.12 & -2.72 & 0 \\ -1.0 & 1.80 & 0.80 & 0 \\ 2.7 & -4.86 & 2.16 & 0 \end{array}\right].$$

Because 3.4 is the largest entry in column 1, we add $1/3.4$ Row 1 (the pivot row) to Row 2 to obtain the new Row 2:

$$\left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \end{array}\right].$$

Add $-2.7/3.4$ Row 1 (the pivot row) to Row 3 of the given matrix to obtain

$$\left[\begin{array}{ccc|c} 0 & 0 & 4.32 & 0 \end{array}\right].$$

We end up with a "triangular" system of the form (after interchanging rows 2 and 3)

$$\begin{aligned} 3.4x_1 - 6.12x_2 - \ 2.72\,x_3 &= 0, \\ 4.32x_3 &= 0, \\ 0 &= 0. \end{aligned}$$

Note that the last equation contains no information. From this, we get

$$4.32x_3 = 0 \qquad \text{which implies that} \qquad \text{(S1)} \qquad x_3 = 0.$$

We substitute this into the first equation and get

$$\text{(S2)} \qquad 3.4x_1 - 6.12x_2 = 0.$$

Since the system reduced to *two* equations (S1) and (S2) in *three* unknowns, we have the choice of one parameter t.

If we set

$$\text{(S3)} \qquad x_1 = t \text{ (arbitrary)},$$

then (S2) becomes

$$\text{(S2*)} \qquad 3.4t - 6.12x_2 = 0$$

so that

$$\text{(S2**)} \qquad x_2 = \frac{3.4}{6.12}t = 0.556t.$$

Then the solution consists of equations (S3), (S2**), and (S1). This corresponds to the solution on p. A48 in App. 2 of the textbook.

If we set

$$\text{(S4)} \qquad x_2 = \tilde{t} \text{ (arbitrary, we call it } \tilde{t} \text{ instead of } t \text{ to show its independence from } t),$$

then we solve for x_1 and get

$$\text{(S5)} \qquad x_1 = \frac{6.12}{3.4}\,\tilde{t} = 1.8\,\tilde{t},$$

and the solution consists of (S5), (S4), and (S1). The two solutions are equivalent.

Sec. 20.2 Linear Systems: LU-Factorization, Matrix Inversion

The inspiration for this section is the observation that an $n \times n$ invertible matrix can be written in the form

$$\text{(2)} \qquad \mathbf{A} = \mathbf{LU},$$

where **L** is a lower triangular and **U** an upper triangular matrix, respectively.

In **Doolittle's method,** we set up a decomposition in the form (2), where m_{jk} in the matrix **L** are the multipliers of the Gauss elimination with the main diagonal 1, 1, . . . , 1 as shown in **Example 1** at the bottom of p. 853. The LU-decomposition (2), when substituted into (1), on p. 852, leads to

$$\mathbf{Ax} = \mathbf{LUx} = \mathbf{L}(\underbrace{\mathbf{Ux}}_{\mathbf{y}}) = \mathbf{Ly} = \mathbf{b},$$

which means we have written

$$(3) \qquad \text{(a)} \quad \mathbf{Ly} = \mathbf{b} \qquad \text{where} \qquad \text{(b)} \quad \mathbf{Ux} = \mathbf{y}.$$

This means we can solve first (3a) for **y** and then (3b) for **x**. Both systems (3a), (3b) are triangular, so we can solve them as in the back substitution for the Gauss elimination. Indeed, this is our approach with Doolittle's method on p. 854. The example is the same as Example 1, on p. 846 in Sec. 20.1. However, Doolittle requires only about half as many operations as Gauss elimination.

If we assign 1, 1, . . . , 1 to the main diagonal of the matrix **U** (instead of **L**) we get **Crout's method**.

A third method based on (2) is **Cholesky's method,** where the $n \times n$ matrix **A** is *symmetric, positive definite*. This means

$$\text{(symmetric)} \qquad \mathbf{A} = \mathbf{A}^{\mathsf{T}},$$

and

$$\text{(PD)} \qquad \mathbf{x}^{\mathsf{T}}\mathbf{Ax} > 0 \qquad \text{for all} \qquad \mathbf{x} \neq \mathbf{0}.$$

Under Cholesky's method, we get formulas (6), p. 855, for factorization. The method is illustrated by **Example 2**, pp. 855–856, and **Prob. 7**. Cholesky's method is attractive because it is numerically stable (Theorem 1, p. 856).

Matrix inversion by the **Gauss–Jordan elimination method** is discussed on pp. 856–857 and shown in Prob. 17.

More Details on Example 1. Doolittle's Method, pp. 853–854. In the calculation of the entries of **L** and **U** (or $\mathbf{L}^{\mathsf{T}}$ in Cholesky's method) in the factorization $\mathbf{A} = \mathbf{LU}$ with given **A**, we employ the usual matrix multiplication

Row times Column.

In all three methods in this section, the point is that the calculation can proceed in an order such that *we solve only one equation at a time*. This is possible because we are dealing with triangular matrices, so that the sums of $n = 3$ products often reduce to sums of two products or even to a single product, as we will see. This will be a discussion of the steps of the calculation, on p. 853, in terms of the matrix equation $\mathbf{A} = \mathbf{LU}$, written out

$$\mathbf{A} = \begin{bmatrix} 3 & 5 & 2 \\ 0 & 8 & 2 \\ 6 & 2 & 8 \end{bmatrix} = \mathbf{LU} = \begin{bmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & m_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}.$$

Remember that, in Doolittle's method, the main diagonal of **L** is 1, 1, 1. Also, the notation m_{jk} suggests *multiplier*, because, in Doolittle's method, the matrix **L** is the matrix of the multipliers in the Gauss elimination. Begin with Row 1 of **A**. The entry $a_{11} = 3$ is the dot product of the first row of **L** and the first column of **U**; thus,

$$3 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} u_{11} & 0 & 0 \end{bmatrix}^{\mathsf{T}} = 1 \cdot u_{11},$$

where 1 is prescribed. Thus, $u_{11} = 3$. Similarly, $a_{12} = 5 = 1 \cdot u_{12} + 0 \cdot u_{22} + 0 \cdot 0 = u_{12}$; thus $u_{12} = 5$. Finally, $a_{13} = 2 = u_{13}$. This takes care of the first row of **A**. In connection with the second row of **A** we have to consider the second row of **L**, which involves m_{21} and 1. We obtain

$$\begin{aligned}
a_{21} &= 0 = m_{21}u_{11} + 1 \cdot 0 \quad + 0 &&= m_{21} \cdot 3, &&\text{hence } m_{21} = 0,\\
a_{22} &= 8 = m_{21}u_{12} + 1 \cdot u_{22} + 0 &&= u_{22}, &&\text{hence } u_{22} = 8,\\
a_{23} &= 2 = m_{21}u_{13} + 1 \cdot u_{23} + 0 &&= u_{23}, &&\text{hence } u_{23} = 2.
\end{aligned}$$

In connection with the third row of **A** we have to consider the third row of **L**, consisting of m_{31}, m_{32}, 1. We obtain

$$\begin{aligned}
a_{31} &= 6 = m_{31}u_{11} \ + 0 \qquad + 0 &&= m_{31} \cdot 3, &&\text{hence } m_{31} = 2,\\
a_{32} &= 2 = m_{31}u_{12} + m_{32}u_{22} + 0 &&= 2 \cdot 5 + m_{32} \cdot 8, &&\text{hence } m_{32} = -1,\\
a_{33} &= 8 = m_{31}u_{13} \ + m_{32}u_{23} \ + 1 \cdot u_{33} &&= 2 \cdot 2 - 1 \cdot 2 + u_{33}, &&\text{hence } u_{33} = 6.
\end{aligned}$$

In (4), on p. 854, the first line concerns the first row of **A** and the second line concerns the first column of **A**; hence in that respect the order of calculation is slightly different from that in Example 1.

Problem Set 20.2. Page 857

7. Cholesky's method. The coefficient matrix **A** of the given system of linear equations is given by

$$\mathbf{A} = \begin{bmatrix} 9 & 6 & 12 \\ 6 & 13 & 11 \\ 12 & 11 & 26 \end{bmatrix} \qquad \text{(as explained in Sec. 7.3, pp. 272–273).}$$

We clearly see that the given matrix **A** is symmetric, since the entries off the main diagonal are mirror images of each other (see definition of symmetric on p. 335 in Sec. 8.3). The Cholesky factorization of **A** (see top of p. 856 in Example 1) is

$$\begin{bmatrix} 9 & 6 & 12 \\ 6 & 13 & 11 \\ 12 & 11 & 26 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}.$$

We do not have to check whether **A** is also positive definite because, if it is not, all that would happen is that we would obtain a complex triangular matrix **L** and would then probably choose another method. We continue.

Going through **A** row by row and applying matrix multiplication (Row times Column) as just before we calculate the following.

$$\begin{aligned}
a_{11} &= \ \ 9 = l_{11}^2 \quad + 0 + 0 &&= \ l_{11}^2, &&\text{hence} \quad l_{11} = \sqrt{a_{11}} = \sqrt{9} = 3,\\
a_{12} &= \ \ 6 = l_{11}l_{21} + 0 + 0 &&= 3\,l_{21}, &&\text{hence} \quad l_{12} = \frac{a_{21}}{l_{11}} = \frac{6}{3} = 2,\\
a_{13} &= 12 = l_{11}l_{31} + 0 + 0 &&= 3\,l_{31}, &&\text{hence} \quad l_{31} = 4.
\end{aligned}$$

In the second row of **A** we have $a_{21} = a_{12}$ (symmetry!) and need only two calculations:

$$\begin{aligned}
a_{22} &= 13 = l_{21}^2 \quad + l_{22}^2 \quad + 0 &&= (2)^2 + l_{22}^2, &&\text{hence} \quad l_{22} = 3,\\
a_{23} &= 11 = l_{21}l_{31} + l_{22}l_{32} + 0 &&= 2 \cdot 4 + 3\,l_{32}, &&\text{hence} \quad l_{32} = 1.
\end{aligned}$$

In the third row of **A** we have $a_{31} = a_{13}$ and $a_{32} = a_{23}$ and need only one calculation:

$$a_{33} = 26 = l_{31}^2 + l_{32}^2 + l_{33}^2 = (4)^2 + 1 + l_{33}^2, \qquad \text{hence} \qquad l_{33} = 3.$$

Now solve $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{b} = [17.4 \quad 23.6 \quad 30.8]^{\mathsf{T}}$. We first use **L** and solve $\mathbf{Ly} = \mathbf{b}$, where $\mathbf{y} = [y_1 \quad y_2 \quad y_3]^{\mathsf{T}}$. Since **L** is triangular, we only do back substitution as in the Gauss algorithm. Now since **L** is *lower* triangular, whereas the Gauss elimination produces an *upper* triangular matrix, begin with the first equation and obtain y_1. Then obtain y_2 and finally y_3. This simple calculation is written to the right of the corresponding equations:

$$\begin{bmatrix} 3 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 1 & 3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 17.4 \\ 23.6 \\ 30.8 \end{bmatrix} \qquad \begin{aligned} y_1 &= \tfrac{1}{3} \cdot 17.4 = 5.8, \\ y_2 &= \tfrac{1}{3}\,(23.6 - 2y_1) = 4, \\ y_3 &= \tfrac{1}{3}\,(30.8 - 4y_1 - y_2) = 1.2. \end{aligned}$$

In the second part of the procedure you solve $\mathbf{L}^{\mathsf{T}}\mathbf{x} = \mathbf{y}$ for **x**. This is another back substitution. Since $\mathbf{L}^{\mathsf{T}}$ is *upper* triangular, just as in the Gauss method after the elimination has been completed, the present back substitution is exactly as in the Gauss method, beginning with the last equation, which gives x_3, then using the second equation to get x_2, and finally the first equation to obtain x_1.

Details on the *back substitution* are as follows:

$$\begin{bmatrix} 3 & 2 & 4 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5.8 \\ 4 \\ 1.2 \end{bmatrix} \text{ written out is } \begin{aligned} &\text{(S1)} \quad 3x_1 + 2x_2 + 4x_3 = 5.8, \\ &\text{(S2)} \quad \phantom{3x_1 + {}} 3x_2 + x_3 = 4, \\ &\text{(S3)} \quad \phantom{3x_1 + 2x_2 + {}} 3x_3 = 1.2. \end{aligned}$$

Hence Eq. (S3)

$$3x_3 = 1.2 \qquad \text{gives} \qquad \text{(S4)} \qquad x_3 = \tfrac{1}{3} \cdot 1.2 = 0.4.$$

Substituting (S4) into (S2) gives

$$3x_2 + x_3 = 3x_2 + 1.2 = 4 \qquad \text{so that} \qquad \text{(S5)} \qquad x_2 = \tfrac{1}{3}(4 - 1.2) = 1.2.$$

Substituting (S4) and (S5) into (S1) yields

$$3x_1{+}2x_2{+}4x_3 = 3x_1{+}2{\cdot}1.2{+}4{\cdot}0.4 = 5.8 \ \text{ so that } \qquad \text{(S6)} \ \ x_1 = \tfrac{1}{3}(5.8{-}2.4{-}1.6) = \tfrac{1}{3}1.8 = 0.4.$$

Hence the solution is (S6), (S5), (S4):

$$x_1 = 0.4, \qquad x_2 = 1.2, \qquad x_3 = 0.4.$$

We *check the solution* by substituting it into the given linear system written as a matrix equation. Indeed,

$$\begin{aligned} \mathbf{Ax} = \begin{bmatrix} 9 & 6 & 12 \\ 6 & 13 & 11 \\ 12 & 11 & 26 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= \begin{bmatrix} 9 & 6 & 12 \\ 6 & 13 & 11 \\ 12 & 11 & 26 \end{bmatrix} \begin{bmatrix} 0.6 \\ 1.2 \\ 0.4 \end{bmatrix} \\ = \begin{bmatrix} 9 \cdot 0.6 + 6 \cdot 1.2 + 12 \cdot 0.4 \\ 6 \cdot 0.6 + 13 \cdot 1.2 + 11 \cdot 0.4 \\ 12 \cdot 0.6 + 11 \cdot 1.2 + 26 \cdot 0.4 \end{bmatrix} &= \begin{bmatrix} 5.4 + 7.2 + 4.8 \\ 3.6 + 15.6 + 4.4 \\ 7.2 + 13.2 + 10.4 \end{bmatrix} = \begin{bmatrix} 17.4 \\ 23.6 \\ 30.8 \end{bmatrix} = \mathbf{b}, \end{aligned}$$

which is correct.

Discussion. We want to show that **A** is positive definite, that is, by definition on p. 346 in Prob. 24 in Sec. 8.4, and also on p. 855:

$$\text{(PD)} \qquad \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} > 0 \qquad \text{for all} \qquad \mathbf{x} \neq \mathbf{0}.$$

We calculate

$$\begin{aligned}
\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} &= \left(\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 9 & 6 & 12 \\ 6 & 13 & 11 \\ 12 & 11 & 26 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
&= \begin{bmatrix} 9x_1 + 6x_2 + 12x_3 & 6x_1 + 13x_2 + 11x_3 & 12x_1 + 11x_2 + 26x_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
&= (9x_1 + 6x_2 + 12x_3) \cdot x_1 + (6x_1 + 13x_2 + 11x_3) \cdot x_2 + (12x_1 + 11x_2 + 26x_3) \cdot x_3 \\
&= 9x_1^2 + 12x_1x_2 + 24x_1x_3 + 22x_2x_3 + 13x_2^2 + 26x_3^2.
\end{aligned}$$

We get the quadratic form Q and want to show that (A) is true for Q:

$$\text{(A)} \qquad Q = 9x_1^2 + 12x_1x_2 + 24x_1x_3 + 22x_2x_3 + 13x_2^2 + 26x_3^2 > 0 \ \text{ for all } \ x_1, x_2, x_3 \neq 0.$$

Since Q cannot be written into a form $(\cdots)^2$, it is not trivial to show that (A) is true. Thus we look for other ways to verify (A). One such way is to use a mathematical result given in **Prob. 25, p. 346.** It states that positive definiteness (PD) holds if and only if all the principal minors of **A** are positive. This result is also known as **Sylvester's criterion.**

For the given matrix **A**, we have three principal minors. They are:

$$a_{11} = 9 > 0, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} 9 & 6 \\ 6 & 13 \end{vmatrix} = 9 \cdot 13 - 6 \cdot 6 = 101 > 0,$$

$$\begin{aligned}
\det \mathbf{A} &= 9 \begin{vmatrix} 13 & 11 \\ 11 & 26 \end{vmatrix} - 6 \begin{vmatrix} 6 & 11 \\ 12 & 26 \end{vmatrix} + 12 \begin{vmatrix} 6 & 13 \\ 12 & 11 \end{vmatrix} \\
&= 9 \cdot 217 - 6 \cdot 24 + 12 \cdot (-90) = 729 > 0.
\end{aligned}$$

Since all principal minors of **A** are positive, we conclude, by Sylvester's criterion, that **A** is indeed positive definite.

The moral of the story is that, for large **A**, showing positive definiteness is not trivial, although in some cases it may be concluded from the kind of physical (or other) application.

17. Matrix inversion. Gauss–Jordan nethod. The method suggested in this section is illustrated in detail in Sec. 7.8 by Example 1, on pp. 303–304, in the textbook, as well as in Prob. 1 on pp. 123–124 in Volume I of the Student Solutions Manual. It may be useful to look at one or both examples. *In your answer, you may want to write down the matrix operations stated here in our solution to Prob. 17 to the right of the matrix as is done in Example 1, p. 303, of the textbook.*

The matrix to be inverted is

$$\mathbf{G} = \begin{bmatrix} 1 & -4 & 2 \\ -4 & 25 & 4 \\ 2 & 4 & 24 \end{bmatrix}.$$

We start by appending the given 3×3 matrix $\mathbf{G}$ by the 3×3 unit matrix $\mathbf{I}$ to obtain the following 3×6 matrix :

$$\mathbf{G}_1 = \left[\begin{array}{ccc|ccc} 1 & -4 & 2 & 1 & 0 & 0 \\ -4 & 25 & 4 & 0 & 1 & 0 \\ 2 & 4 & 24 & 0 & 0 & 1 \end{array}\right].$$

Thus the left 3×3 submatrix is the given matrix and the right 3×3 submatrix is the 3×3 unit matrix $\mathbf{I}$. We apply the Gauss–Jordan method to $\mathbf{G}_1$ to obtain the desired inverse matrix. At the end of the process, the left 3×3 submatrix will be the 3×3 unit matrix, and the right 3×3 submatrix will be the inverse of the given matrix.

The -4 in Row 2 of $\mathbf{G}_1$ is the largest value in Column 1 so we interchange Row 2 and Row 1 and get

$$\mathbf{G}_2 = \left[\begin{array}{ccc|ccc} -4 & 25 & 4 & 0 & 1 & 0 \\ 1 & -4 & 2 & 1 & 0 & 0 \\ 2 & 4 & 24 & 0 & 0 & 1 \end{array}\right].$$

Next we replace Row 2 by Row 2 $+ \frac{1}{4}$ Row 1 and replace Row 3 by Row 3 $+ \frac{2}{4}$ Row 1. This gives us the new matrix

$$\mathbf{G}_3 = \left[\begin{array}{ccc|ccc} -4 & 25 & 4 & 0 & 1 & 0 \\ 0 & \frac{9}{4} & 3 & 1 & \frac{1}{4} & 0 \\ 0 & \frac{33}{2} & 26 & 0 & \frac{1}{2} & 1 \end{array}\right].$$

Now, because $\frac{33}{2} > \frac{9}{4}$, we swap Rows 2 and 3 of $\mathbf{G}_3$ to obtain

$$\mathbf{G}_4 = \left[\begin{array}{ccc|ccc} -4 & 25 & 4 & 0 & 1 & 0 \\ 0 & \frac{33}{2} & 26 & 0 & \frac{1}{2} & 1 \\ 0 & \frac{9}{4} & 3 & 1 & \frac{1}{4} & 0 \end{array}\right].$$

Replace Row 3 by Row 3 $- \left(\frac{9}{4}\right) / \left(\frac{33}{2}\right)$ Row 2. The new matrix is

$$\mathbf{G}_5 = \left[\begin{array}{ccc|ccc} -4 & 25 & 4 & 0 & 1 & 0 \\ 0 & \frac{33}{2} & 26 & 0 & \frac{1}{2} & 1 \\ 0 & 0 & -\frac{6}{11} & 1 & \frac{2}{11} & -\frac{3}{22} \end{array}\right].$$

This was the Gauss part. The given matrix is triangularized. Now comes the Jordan part that diagonalizes it. We know that we need 1's along the diagonal in the left-hand matrix, so we multiply Row 1 by $-\frac{1}{4}$. In addition, we also multiply Row 2 by $\frac{2}{33}$, and Row 3 by $-\frac{11}{6}$ to get

$$\mathbf{G}_6 = \left[\begin{array}{ccc|ccc} 1 & -\frac{25}{4} & -1 & 0 & -\frac{1}{4} & 0 \\ 0 & 1 & \frac{52}{33} & 0 & \frac{1}{33} & \frac{2}{33} \\ 0 & 0 & 1 & -\frac{11}{6} & -\frac{1}{3} & \frac{1}{4} \end{array}\right].$$

Eliminate the entries in Rows 1 and 2 (Col. 3) by replacing Row 2 by Row 2 $-\left(\frac{52}{33}\right)$ Row 3 and Row 1 by Row 1$+$ Row 3. This gives the matrix

$$\mathbf{G}_7 = \left[\begin{array}{ccc|ccc} 1 & -\frac{25}{4} & 0 & -\frac{11}{6} & -\frac{7}{12} & \frac{1}{4} \\ 0 & 1 & 0 & \frac{26}{9} & \frac{5}{9} & -\frac{1}{3} \\ 0 & 0 & 1 & -\frac{11}{6} & -\frac{1}{3} & \frac{1}{4} \end{array}\right].$$

Finally, we eliminate $-\frac{25}{4}$ in the second column of $\mathbf{G}_7$. We do this by replacing Row 1 of $\mathbf{G}_7$ by Row 1 $+\ \frac{25}{4}$ Row 2. The final matrix is

$$\mathbf{G}_8 = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{146}{9} & \frac{26}{9} & -\frac{11}{6} \\ 0 & 1 & 0 & \frac{26}{9} & \frac{5}{9} & -\frac{1}{3} \\ 0 & 0 & 1 & -\frac{11}{6} & -\frac{1}{3} & \frac{1}{4} \end{array}\right].$$

The last three columns constitute the inverse of the given matrix, that is,

$$\mathbf{G}^{-1} = \begin{bmatrix} \frac{146}{9} & \frac{26}{9} & -\frac{11}{6} \\ \frac{26}{9} & \frac{5}{9} & -\frac{1}{3} \\ -\frac{11}{6} & -\frac{1}{3} & \frac{1}{4} \end{bmatrix}.$$

You may want to check the result by showing that

$$\mathbf{G}\mathbf{G}^{-1} = \mathbf{I} \qquad \text{and} \qquad \mathbf{G}^{-1}\mathbf{G} = \mathbf{I}.$$

Sec. 20.3 Linear Systems: Solution by Iteration

We distinguish between direct methods and indirect methods (p. 858). **Direct methods** are those methods for which we can specify in advance how many numeric computations it will take to get a solution. The Gauss elimination and its variants (Secs. 20.1, 20.2) are examples of direct methods. **Indirect** or **iterative methods** are those methods where we start from an approximation to the true solution and, if successful, obtain better and better approximations from a computational cycle repeated as often as may be necessary for achieving a required accuracy. Such methods are useful for solving linear systems that involve large sparse systems (p. 858).

The first indirect method, the **Gauss–Seidel iteration method (Example 1, Prob. 9)** requires that we take a given linear system (1) and write it in the form (2). You see that the variables have been separated and appear on the left-hand side of the equal sign with coefficient 1. The system (2) is now prepared for iteration. Next one chooses a starting value, here $x_1^{(0)} = 100$, $x_2^{(0)} = 100$, etc. (follow the textbook on

p. 859). Equation (3) shows how Gauss–Seidel continues with these starting values. And here comes a crucial point that is particular to the method, that is, *Gauss–Seidel always uses* (*where possible*) *the most recent and therefore "most up to date" approximation for each unknown* ("successive corrections"). This is shown in the darker shaded blue area in (3) and explained in detail in the textbook as well as in Prob. 9.

The second method, **Jacobi iteration** (13), p. 862 (**Prob. 17**), is very similar to Gauss–Seidel but avoids using the most recent approximation of each unknown within an iteration cycle. Instead, as is much more common with iteration methods, all values are updated at once ("simultaneous corrections").

For these methods to converge, we require "diagonal dominance," that is, the largest (in absolute value) element in each row must be on the diagonal.

Other aspects of Gauss–Seidel include a more formal discussion [precise formulas (4), (5), (6)], ALGORITHM GAUSS–SEIDEL (see p. 860), convergence criteria (p. 861, Example 2, p. 862), and residual (12). Pay close attention to formulas (9), (10), (11) for matrix norms (**Prob. 19**) on p. 861, as they will play an important role in Sec. 20.4.

Problem Set 20.3. Page 863

9. Gauss–Seidel iteration. We write down the augmented matrix of the given system of linear equations (see p. 273 of Sec. 7.3 in the textbook):

$$\mathbf{A} = \left[\begin{array}{ccc|c} 5 & 1 & 2 & 19 \\ 1 & 4 & -2 & -2 \\ 2 & 3 & 8 & 39 \end{array}\right].$$

This is a case in which we do not need to reorder the given linear equations, since we note that the large entries 5, 4, 8 of the coefficient part of the augmented matrix stand on the main diagonal. Hence we can expect convergence.

Remark. If, say, instead the augmented matrix had been

$$\left[\begin{array}{ccc|c} 5 & 1 & 2 & 19 \\ 2 & 3 & 8 & 39 \\ 1 & 4 & -2 & -2 \end{array}\right]$$

meaning that 5, 3, −2 would be the entries of the main diagonal so that 8 and 4 would be larger entries outside the main diagonal, then we would have had to reorder the equations, that is, exchange the second and third equations. This would have led to a system corresponding to augmented matrix **A** above and expected convergence.

We continue. We divide the equations so that their main diagonal entries equal 1 and keep these terms on the left while moving the other terms to the right of the equal sign. In detail, this means that we multiply the first given equation of the problem by $\frac{1}{5}$, the second one by $\frac{1}{4}$, and the third one by $\frac{1}{8}$. We get

$$\begin{aligned} x_1 + \frac{1}{5}x_2 + \frac{2}{5}x_3 &= \frac{19}{5}, \\ \frac{1}{4}x_1 + x_2 - \frac{2}{4}x_3 &= \frac{-2}{4}, \\ \frac{2}{8}x_1 + \frac{3}{8}x_2 + x_3 &= \frac{39}{8}, \end{aligned}$$

and then moving the off-diagonal entries to the right:

$$\text{(GS)} \qquad \begin{aligned} x_1 &= \frac{19}{5} - \frac{1}{5}x_2 - \frac{2}{5}x_3, \\ x_2 &= \frac{-2}{4} - \frac{1}{4}x_1 + \frac{2}{4}x_3, \\ x_3 &= \frac{39}{8} - \frac{2}{8}x_1 - \frac{3}{8}x_2. \end{aligned}$$

We start from $x_1^{(0)} = 1$, $x_2^{(0)} = 1$, $x_3^{(0)} = 1$ (or any reasonable choice) and get

$$\begin{aligned} x_1^{(1)} &= \frac{19}{5} - \frac{1}{5}x_2^{(0)} - \frac{2}{5}x_3^{(0)} \\ &= \frac{19}{5} - \frac{1}{5}\cdot 1 - \frac{2}{5}\cdot 1 \\ &= \frac{19-1-2}{5} = \frac{16}{5} = 3.2 \ \text{(exact)}, \\ x_2^{(1)} &= \frac{-2}{4} - \frac{1}{4}x_1^{(1)} - \frac{2}{5}x_3^{(0)} \\ &= \frac{-2}{4} - \frac{1}{4}\cdot 3.2 + \frac{2}{4}\cdot 1 \\ &= -\frac{1}{2} - 0.8 + \frac{1}{2} = -0.8 \ \text{(exact)}, \\ x_3^{(1)} &= \frac{39}{8} - \frac{2}{8}x_1^{(1)} - \frac{3}{8}x_2^{(1)} \\ &= \frac{39}{8} - \frac{2}{8}\cdot 3.2 - \frac{3}{8}\cdot(-0.8) \\ &= 4.375 \ \text{(exact)}. \end{aligned}$$

Note that we always use the latest possible value in the iteration, that is, for example, in computing $x_2^{(1)}$ we use $x_1^{(1)}$ (new! and not $x_1^{(0)}$) and $x_3^{(0)}$ (no newer value available). In computing $x_3^{(1)}$ we use $x_1^{(1)}$ (new!) and $x_2^{(1)}$(new!) (see also p. 859 of the textbook).

Then we substitute $x_1^{(1)} = 3.2$, $x_2^{(1)} = -0.8$, $x_3^{(1)} = 4.375$ into system (GS) and get

$$x_1^{(2)} = 2.210000, \quad x_2^{(2)} = 1.135000, \quad x_3^{(2)} = 3.89688.$$

The results are summarized in the following table. The values were computed to 6S with two guard digits for accuracy.

Prob. 9. Gauss–Seidel Iteration Method. Table of Iterations. Five Steps.

Step	x_1	x_2	x_3
$m = 1$	3.2	−0.8	4.375
$m = 2$	2.21000	1.13500	3.89688
$m = 3$	2.01425	0.944875	4.01711
$m = 4$	2.00418	1.00751	3.99614
$m = 5$	2.00004	0.998059	4.00072

The exact solution is 2, 1, 4.

11. **Effect of starting values.** The point of this problem is to show that there is surprisingly little difference between corresponding values, as the answer on p. A49 in App. 2 shows, although the starting values differ considerably. Hence it is hardly necessary to search extensively for "good" starting values.

17. **Jacobi Iteration. Convergence related to eigenvalues.** An outline of the solution is as follows. You may want to work out some more of the details. We are asked to consider the matrix of the system of linear equations in Prob. 10 on p. 863, that is,

$$\tilde{\mathbf{A}} = \begin{bmatrix} 4 & 0 & 5 \\ 1 & 6 & 2 \\ 8 & 2 & 1 \end{bmatrix}.$$

We note that $\tilde{a}_{13} = 8$ is a large entry outside the main diagonal (see Remark in Prob. 9 above). To obtain convergence, we reorder the rows as shown, that is, we exchange Row 3 with Row 1, and get,

$$\begin{bmatrix} 8 & 2 & 1 \\ 1 & 6 & 2 \\ 4 & 0 & 5 \end{bmatrix}.$$

Then we divide the rows by the diagonal entries 8, 6, and 5, respectively, as required in (13), p. 862 (see $a_{jj} = 1$ at the end of the formula). (Equivalently, this means we take $\frac{1}{8} \cdot$ Row1, $\frac{1}{6} \cdot$ Row2, $\frac{1}{5} \cdot$ Row3):

$$\mathbf{A} = \begin{bmatrix} 1 & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{6} & 1 & \frac{1}{3} \\ \frac{4}{5} & 0 & 1 \end{bmatrix}.$$

As described in the problem, we now have to consider

$$\mathbf{B} = \mathbf{I} - \mathbf{A} = \begin{bmatrix} 0 & -\frac{1}{4} & -\frac{1}{8} \\ -\frac{1}{6} & 0 & -\frac{1}{3} \\ -\frac{4}{5} & 0 & 0 \end{bmatrix}.$$

The eigenvalues are obtained as the solutions of the characteristic equation (see pp. 326–327)

$$\det(\mathbf{B} - \lambda\mathbf{I}) = \begin{vmatrix} -\lambda & -\frac{1}{4} & -\frac{1}{8} \\ -\frac{1}{6} & -\lambda & -\frac{1}{3} \\ -\frac{4}{5} & 0 & -\lambda \end{vmatrix}$$
$$= -\lambda^3 + \frac{17}{120}\lambda - \frac{1}{15} = 0.$$

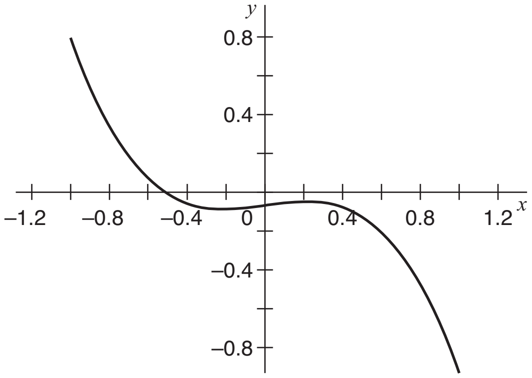
A sketch, as given below, shows that there is a real root near −0.5, but there are no further real roots because, for large $|\lambda|$, the curve comes closer and closer to the curve of $-\lambda^3$. Hence the other eigenvalues must be complex conjugates. A root-finding method (see Sec. 19.2, pp. 801–806, also Prob. 21 in the Student Solutions Manual on p. YY) gives a more accurate value of −0.5196. Division of the characteristic equation by $\lambda + 0.5196$ gives the quadratic equation

$$-\lambda^2 + 0.5196\lambda - 0.1283 = 0.$$

The roots are $0.2598 \pm 0.2466i$ [by the well-known root-finding formula (4) for quadratic equations on p. 54 of the textbook or on p. 15 in Volume I of the Student Solutions Manual]. Since all three roots are less than 1 in absolute value, that is,

$$\begin{aligned} |0.2598 \pm 0.2466i\,| &= \sqrt{(0.2598)^2 + (\pm 0.2466)^2} \qquad \text{[by (3), p. 613]} \\ &= 0.3582 < 1 \\ |-0.5196| &= 0.5196 < 1, \end{aligned}$$

the spectral radius is less than 1, by definition. This is necessary and sufficient for convergence (see at the end of the section at the top of p. 863).



Sec. 20.3 Prob. 17. Curve of the characteristic polynomial

19. Matrix norms. The given matrix is

$$\mathbf{C} = \begin{bmatrix} 10 & 1 & 1 \\ 1 & 10 & 1 \\ 1 & 1 & 10 \end{bmatrix}.$$

All the norms are given on p. 861. The Frobenius norm is

$$\begin{aligned} \text{(9)} \qquad \|\mathbf{C}\| &= \sqrt{\sum_{j=1}^{3}\sum_{k=1}^{3} c_{jk}^2} \\ &= \sqrt{10^2 + 1 + 1 + 1 + 10^2 + 1 + 1 + 1 + 10^2} = \sqrt{303} \\ &= 17.49. \end{aligned}$$

The column "sum" norm is

$$\|\mathbf{C}\| = \max_k \sum_{j=1}^{3} |c_{jk}| = 12. \tag{10}$$

Note that, to compute (10), we took the absolute value of each entry in each column and added them up. Each column gave the value of 12. So the maximum over the three columns was 12. Similarly, by (11), p. 861, the row "sum" norm is 12.

Together this problem illustrates that the three norms usually tend to give values of a similar order of magnitude. Hence, one often chooses the norm that is most convenient from a computational point of view. However, a matrix norm often results from the choice of a vector norm. When this happens, we are not completely free to choose the norm. This new aspect will be introduced in the next section of this chapter.

Sec. 20.4 Linear Systems: Ill-Conditioning, Norms

A computational problem is called **ill-conditioned** (p. 864) if *small* changes in the data cause *large* changes in the solution. The desirable counterpart, where small changes in data cause only small changes in the solution, is labeled *well-conditioned*. Take a look at Fig. 445 at the bottom of p. 864. The system in (a) is well-conditioned. The system shown in part (b) is ill-conditioned because, if we raise or lower one of the lines just a little bit, the the point of intersection (the solution) will move substantially, signifying ill-conditioning. Example 1, p. 865, expresses the same idea in an algebraic example.

Keeping these examples in mind, we move to the central concept of this section, the **condition number** $\kappa(\mathbf{A})$ of a square matrix on p. 868:

$$\kappa(\mathbf{A}) = \|\mathbf{A}\| \cdot ||\mathbf{A}||^{-1} . \tag{13}$$

Here κ is the Greek letter kappa (see back inside cover of textbook), $\|\mathbf{A}\|$ denotes the norm of matrix $\mathbf{A}$, and $\|\mathbf{A}^{-1}\|$ denotes the norm of its inverse. We need to backtrack and look at the concept of norms, which is of general interest in numerics.

Vector norms $\|\mathbf{x}\|$ for column vectors $\mathbf{x} = [x_j]$ with n components (n fixed), p. 866, are generalized concepts of length or distance and are defined by four properties (3). Most common are the l_1-norm (5), "Euclidean" or l_2-norm (6), and l_∞-norm (7)—all illustrated in Example 3, p. 866.

Matrix norms, p. 867, build on vector norms and are defined by

$$\|\mathbf{A}\| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \qquad (\mathbf{x} \neq \mathbf{0}). \tag{9}$$

We use the l_1-norm (5) for matrices—obtaining the column "sum" norm (10)—and the l_∞-norm (7) for matrices—obtaining the row "sum" norm (11)—both on p. 861 of Sec. 20.3. **Example 4**, pp. 866–867, illustrates this. We continue our discussion of the condition number.

We take the coefficient matrix $\mathbf{A}$ of a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ and calculate $\kappa(\mathbf{A})$. If $\kappa(\mathbf{A})$ is small, then the linear system is well-conditioned (**Theorem 1**, **Example 5**, p. 868).

We look at the proof of Theorem 1. We see the role of $\kappa(\mathbf{A})$ from (15), p. 868, is that a small condition number gives a small difference in the norm of $\mathbf{x} - \tilde{\mathbf{x}}$ between an approximate solution $\tilde{\mathbf{x}}$ and the unknown exact solution $\mathbf{x}$ of a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Problem 9 gives a complete example on how to compute the condition number $\kappa(\mathbf{A})$ for the well-conditioned case. Contrast this with **Prob. 19**, which solves an ill-conditioned system by Gauss elimination with partial pivoting and also computes the very large condition number $\kappa(\mathbf{A})$. See also Example 1, p. 865, and **Example 6**, p. 869.

Finally, the topic of residual [see (1), p. 865] is explored in Example 2, p. 865, and Prob. 21.

There is no sharp dividing line between well-conditioned and ill-conditioned as discussed in "Further Comments on Condition Numbers" at the bottom of p. 870.

Problem Set 20.4. Page 871

9. Matrix norms and condition numbers. From the given matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 4 \end{bmatrix}$$

we compute its inverse by (4*), p. 304, in Sec. 7.8:

$$\mathbf{A}^{-1} = \frac{1}{2 \cdot 4 - 1 \cdot 0} \begin{bmatrix} 4 & -1 \\ -0 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{8} \\ 0 & \frac{1}{4} \end{bmatrix}.$$

We want the matrix norms for **A** and $\mathbf{A}^{-1}$, that is, $\|\mathbf{A}\|$ and $\left\|\mathbf{A}^{-1}\right\|$**.** We begin with the l_1-vector norm, which is defined by (5), p. 866. We have to remember that the l_1-vector norm gives, for *matrices,* the column "sum" norm (the "sum" indicating that we take sums of absolute values) as explained in the blue box in the middle of p. 867. This gives, under the l_1-norm [summing over the absolute values of the entries of each **column** i (here $i = 1, 2$) and then selecting the maximum],

$$\|\mathbf{A}\| = \max_i \{|2| + |0|, |1| + |4|\} = \max_i \{|2|, |5|\} = 5,$$

and

$$\left\|\mathbf{A}^{-1}\right\| = \max_i \left\{\left|\tfrac{1}{2}\right| + |0|, \left|-\tfrac{1}{8}\right| + \left|\tfrac{1}{4}\right|\right\} = \max_i \left\{\left|\tfrac{1}{2}\right|, \left|\tfrac{3}{8}\right|\right\} = \tfrac{1}{2}.$$

Thus the condition number is

$$\kappa(\mathbf{A}) = \|\mathbf{A}\| \cdot \left\|\mathbf{A}^{-1}\right\| = 5 \cdot \tfrac{1}{2} = 2.5 \qquad \text{[by definition, see (13), p. 868].}$$

Now we turn to the l_∞-vector norm, defined by (7), p. 866. We have to remember that this vector norm gives for *matrices* the row "sum" norm. This gives, under the l_∞-norm [summing over the absolute values of the entries of each **row** j (in our situation $j = 1, 2$) and then selecting the maximum],

$$\|\mathbf{A}\| = \max_j \{|2| + |1|, |0| + |4|\} = \max_j \{|2|, |4|\} = 4,$$

and

$$\left\|\mathbf{A}^{-1}\right\| = \max_j \left\{\left|\tfrac{1}{2}\right| + \left|-\tfrac{1}{8}\right|, |0| + \left|\tfrac{1}{4}\right|\right\} = \max_j \left\{\left|\tfrac{5}{8}\right|, \left|\tfrac{1}{4}\right|\right\} = \tfrac{5}{8}.$$

Thus the condition number is

$$\kappa(\mathbf{A}) = \|\mathbf{A}\| \cdot \left\|\mathbf{A}^{-1}\right\| = 4 \cdot \tfrac{5}{8} = 2.5 \qquad \text{[by definition, see (13), p. 868].}$$

Since the value of the condition number is not large, we conclude that the matrix **A** is not ill-conditioned.

19. An ill-conditioned system

1. *Solving* $\mathbf{Ax} = \mathbf{b}_1$. The linear system written out is

$$(1) \qquad 4.50x_1 + 3.55x_2 = 5.2,$$

$$(2) \qquad 3.55x_1 + 2.80x_2 = 4.1.$$

The coefficient matrix $\mathbf{A}$, given in the problem, is

$$\mathbf{A} = \begin{bmatrix} 4.50 & 3.55 \\ 3.55 & 2.80 \end{bmatrix} \qquad \text{and} \qquad \mathbf{b}_1 = \begin{bmatrix} 5.2 \\ 4.1 \end{bmatrix}.$$

We use Gauss elimination with partial pivoting (p. 846) to obtain a solution to the linear system. We form the augmented matrix (pp. 845, 847):

$$[\mathbf{A}|\mathbf{b}_1] = \left[\begin{array}{cc|c} 4.50 & 3.55 & 5.2 \\ 3.55 & 2.80 & 4.1 \end{array}\right].$$

We pivot 4.5 in Row 1 and use it to eliminate 3.55 in Row 2, that is,

$$\text{Row } 2 - \frac{3.55}{4.50} \cdot \text{Row } 1, \text{ which is, Row } 2 - 0.7888888889 \cdot \text{Row } 1,$$

and get

$$\left[\begin{array}{cc|c} 4.5 & 3.55 & 5.2 \\ 0 & -0.000555555595 & -0.00222222228 \end{array}\right].$$

Back substitution (p. 847) gives us by Eq. (2)

$$x_2 = \frac{-0.00222222228}{-0.000555555595} = 3.999999992 \approx 4.$$

Substituting this into (1) yields

$$x_1 = \frac{1}{4.50}(5.2 - 3.55 \cdot x_2) = \frac{1}{4.50}(5.2 - 3.55 \cdot 4) = -2.$$

2. *Solving* $\mathbf{Ax} = \mathbf{b}_2$. The slightly modified system is

$$(1) \qquad 4.50x_1 + 3.55x_2 = 5.2,$$

$$(3) \qquad 3.55x_1 + 2.80x_2 = 4.0.$$

The coefficient matrix $\mathbf{A}$ is as before with $\mathbf{b}_2$ slightly different from $\mathbf{b}_1$, that is,

$$\mathbf{b}_2 = \begin{bmatrix} 5.2 \\ 4.0 \end{bmatrix}.$$

We form the augmented matrix

$$[\mathbf{A}|\mathbf{b}_2] = \left[\begin{array}{cc|c} 4.50 & 3.55 & 5.2 \\ 3.55 & 2.80 & 4.0 \end{array}\right]$$

and use Gauss elimination with partial pivoting with exactly the same row operation but startlingly different numbers!

$$\left[\begin{array}{cc|c} 4.5 & 3.55 & 5.2 \\ 0 & -0.00055595 & -0.1022228 \end{array}\right] \quad \text{Row } 2 - 0.788889 \cdot \text{Row } 1.$$

(There will be a small, nonzero, value in the a_{21} position due to using a finite number of digits.)

Back substitution now gives, by (3),

$$x_2 = \frac{-0.1022228}{-0.00055595} = 183.87 \approx 184$$

and hence, by (1),

$$x_1 = \frac{1}{4.50}(5.2 - 3.55 \cdot x_2) = \frac{1}{4.50}(5.2 - 3.55 \cdot 184) = -144.$$

3. *Computing the condition number of* **A**. First, we need the inverse of **A**. By (4*), p. 304, we have

$$\mathbf{A}^{-1} = \frac{1}{2.80 \cdot 4.50 - (-3.55) \cdot (-3.55)} \begin{bmatrix} 2.80 & -3.55 \\ -3.55 & 4.50 \end{bmatrix}$$
$$= -400 \begin{bmatrix} 2.80 & -3.55 \\ -3.55 & 4.50 \end{bmatrix} = \begin{bmatrix} -1120 & 1420 \\ 1420 & -1800 \end{bmatrix}.$$

The l_1-norm for matrix **A**, which we obtain by summing over the absolute values of the entries of each **column** i (here $i = 1, 2$) and then selecting the maximum

$$\|\mathbf{A}\| = \max_i \{|2.80| + |3.55|, \quad |3.55| + |4.50|\} = \max_i \{|6.35|, \quad |8.05|\} = 8.05,$$

and similarly for $\mathbf{A}^{-1}$

$$\left\|\mathbf{A}^{-1}\right\| = \max_i \{|1120| + |1420|, \quad |1420| + |1800|\} = \max_i \{|2540|, \quad |3220|\} = 3220.$$

Then by (13), p. 868, the condition number is

$$\kappa(\mathbf{A}) = \|\mathbf{A}\| \cdot \left\|\mathbf{A}^{-1}\right\| = 8.05 \cdot 3220 = 25921.$$

Furthermore, because matrix **A** is symmetric (and, consequently, so is its inverse $\mathbf{A}^{-1}$), the values of the l_∞-norm, i.e., the row "sum" norm, for both matrices **A** and $\mathbf{A}^{-1}$ are equal to their corresponding values of the l_1-norm, respectively. Hence the computation of $\kappa(\mathbf{A})$ would yield the same value.

4. *Interpretation and discussion of result.* The condition number $\kappa(\mathbf{A}) = 25921$ is very large, signifying that the given system is indeed very ill-conditioned. This was confirmed by direct

calculations in steps 1 and 2 by Gauss elimination with partial pivoting, where a small change by 0.1 in the second component from $\mathbf{b}_1$ to $\mathbf{b}_2$ causes the solution to change from $[-2, 4]^\mathsf{T}$ to $[-144, 184]^\mathsf{T}$, a change of about 1,000 times that of that component! Note that we used 10 decimals in our first set of calculations to get satisfactory results. You may want to experiment with a small number of decimals and see how you get nonsensical results. Furthermore, note that the two rows of $\mathbf{A}$ are almost proportional.

21. **Small residuals for very poor solutions.** Use (2), p. 865, defining the residual of the "approximate solution" $[-10.0 \quad -14.1]^\mathsf{T}$ of the actual solution $[-2 \quad 4]^\mathsf{T}$, to obtain

$$\begin{aligned}\mathbf{r} &= \begin{bmatrix}5.2\\4.1\end{bmatrix} - \begin{bmatrix}4.50 & 3.55\\3.55 & 2.80\end{bmatrix}\begin{bmatrix}-10.0\\14.1\end{bmatrix}\\ &= \begin{bmatrix}5.2\\4.1\end{bmatrix} - \begin{bmatrix}5.055\\3.980\end{bmatrix}\\ &= \begin{bmatrix}0.145\\0.120\end{bmatrix}.\end{aligned}$$

While the residual is not very large, the approximate solution has a first component that is 5 times that of the true solution and a second component that is 3.5 times as great. For ill-conditioned matrices, a small residue does not mean a good approximation.

Sec. 20.5 Least Squares Method

We may describe the underlying problem as follows. We obtained several points in the xy-plane, say by some experiment, through which we want to fit a straight line. We could do this visually by fitting a line in such a way that the absolute vertical distance of the points from the line would be as short as possible, as suggested by Fig. 447, p. 873. Now, to obtain an attractive algebraic model, if the *absolute value* of a point to a line is the smallest, then so is the *square* of the vertical distance of the point to the line. (The reason we do not want to use absolute value is that it is not differentiable throughout its domain.) Thus we want to fit a straight line in such a way that the sum of the squares of the distances of all those points from the line is minimal, i.e., "least"—giving us the name "**least squares method**."

The formal description of *fitting a straight line by the least squares method* is given in (2), p. 873, and solved by **two normal equations** (4). While these equations are not particularly difficult, you need some practice, such as **Prob. 1,** in order to remember how to correctly set up and solve such problems on the exam.

The least squares method also plays an important role in regression analysis in statistics. Indeed, the normal equations (4) show up again in Sec. 25.9, as (10) on p. 1105.

We extend the method to *fitting a parabola by the least squares method* and obtain three normal equations (8), p. 874. This generalization is illustrated in Example 2, p. 874, with Fig. 448 on p. 875, and in complete detail in **Prob. 9.**

Finally, the most general case is (5) and (6), p. 874.

Problem Set 20.5. Page 875

1. **Fitting by a straight line. Method of least squares**. We are given four points $(0, 2)$, $(2, 0)$, $(3, -2)$, $(5, -3)$ through which we should fit algebraically (instead of geometrically or sketching approximately) a straight line. We use the method of least squares of Example 1, on p. 873 in the textbook. This requires that we solve the normal auxiliary quantities needed in Eqs. (4), p. 873 in the textbook. When using paper and pencil or if you use your computer as a typesetting tool, you may organize the auxiliary quantities needed in (4) in a table as follows:

	x_j	y_j	x_j^2	$x_j y_j$
	0	2	0	0
	2	0	4	0
	3	−2	9	−6
	5	−3	25	−15
Sum	10	−3	38	−21

From the last line of the table we see that the sums are

$$\sum x_j = 10, \qquad \sum y_j = -3, \qquad \sum x_j^2 = 38, \qquad \sum x_j y_j = -21,$$

and $n = 4$, since we used four pairs of values. This determines the following coefficients for the variables of (4), p. 873:

$$4a + 10b = -3, \tag{1}$$

$$10a + 38b = -21, \tag{2}$$

and gives the augmented matrix

$$\left[\begin{array}{cc|c} 4 & 10 & -3 \\ 10 & 38 & -21 \end{array}\right].$$

This would be a nice candidate for Cramer's rule. Indeed, we shall solve the system by Cramer's rule (2), (3), Example 1, p. 292 in Sec. 7.6. Following that page, we have

$$D = \det \mathbf{A} = \begin{vmatrix} 4 & 10 \\ 10 & 38 \end{vmatrix} = 4 \cdot 38 - 10 \cdot 10 = 152 - 100 = 52.$$

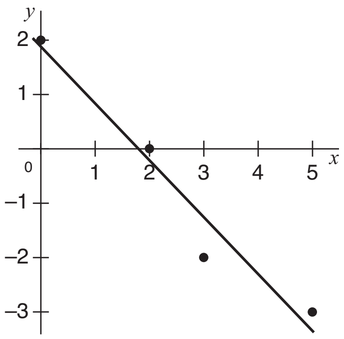
Furthermore

$$a = \frac{\begin{vmatrix} -3 & 4 \\ -21 & 10 \end{vmatrix}}{D} = \frac{-3 \cdot 10 - (-21) \cdot 4}{D} = \frac{-114 + 210}{D} = \frac{96}{52} = \frac{24}{13} = 1.846,$$

$$b = \frac{\begin{vmatrix} 4 & -3 \\ 10 & -21 \end{vmatrix}}{D} = \frac{-4 \cdot (-21) - (-3) \cdot 10}{D} = \frac{-84 - (-30)}{D} = \frac{-27}{26} = -1.038.$$

From this we immediately get our desired straight line:

$$\begin{aligned} y &= a + bx \\ &= 1.846 - 1.038x. \end{aligned}$$



Sec. 20.5 Prob. 1. Given data and straight line fitted by least squares.
(Note that the axes have equal scales)

9. Fitting by a quadratic parabola. A quadratic parabola is uniquely determined by three given points. In this problem, five points are given. We can fit a quadratic parabola by solving the normal equations (8), p. 874. We arrange the data and auxiliary quantities in (8) again in a table:

	x	y	x^2	x^3	x^4	xy	x^2y
	2	−3	4	8	16	−6	−12
	3	0	9	27	81	0	0
	5	1	25	125	625	5	25
	6	0	36	216	1296	0	0
	7	−2	49	343	2401	−14	98
Sum	23	−4	123	719	4419	−15	−85

The last line of the table gives us the following information:

$$\sum x = 23, \qquad \sum y = -4, \qquad \sum x^2 = 123, \qquad \sum x^3 = 719,$$
$$\sum x^4 = 4419, \qquad \sum xy = -15, \qquad \sum x^2 y = -85,$$

with the number of points being $n = 5$. Hence, looking at (8) on p. 874, and using the sums just obtained, we can carefully construct the augmented matrix of the system of normal equations:

$$\left[\begin{array}{ccc|c} 5 & 23 & 123 & -4 \\ 23 & 123 & 719 & -15 \\ 123 & 719 & 4419 & -85 \end{array}\right].$$

The system of normal equations is

$$\begin{aligned} 5b_0 + \ \ 23b_1 + \ \ 123b_2 &= -4, \\ 23b_0 + 123b_1 + \ \ 719b_2 &= -15, \\ 123b_0 + 719b_1 + 4419b_2 &= -85. \end{aligned}$$

We use Gauss elimination but, noting that the largest numbers are in the third row, we swap the first and third rows,

$$\left[\begin{array}{ccc|c} 123 & 719 & 4419 & -85 \\ 23 & 123 & 719 & -15 \\ 5 & 23 & 123 & -4 \end{array}\right] \begin{array}{l} \text{Row 3} \\ \\ \text{Row 1} \end{array}$$

Then we perform the following row reduction operations:

$$\left[\begin{array}{ccc|c} 123 & 719 & 4419 & -85 \\ 0 & -11.4472 & -107.317 & 0.89431 \\ 0 & -6.22764 & -56.6342 & -0.544715 \end{array}\right] \begin{array}{l} \\ \text{Row 2} - \frac{23}{123}\,\text{Row 1} \\ \text{Row 3} - \frac{5}{123}\,\text{Row 1} \end{array}$$

$$\left[\begin{array}{ccc|c} 123 & 719 & 4419 & -85 \\ 0 & -11.4472 & -107.317 & 0.89431 \\ 0 & -6.22764 & -56.6342 & -0.544715 \end{array}\right] \begin{array}{l} \\ \text{Row 2} - \frac{23}{123}\,\text{Row 1} \\ \text{Row 3} - \frac{5}{123}\,\text{Row 1} \end{array}$$

$$\left[\begin{array}{ccc|c} 123 & 719 & 4419 & -85 \\ 0 & -11.4472 & -107.317 & 0.89431 \\ 0 & 0 & 1.74965 & -1.03125 \end{array}\right] \quad \text{Row 3} - \frac{6.22764}{-11.4472} \cdot \text{Row 2}$$

Back substitution gives us, from the last row of the last matrix,

$$b_2 = \frac{-1.03125}{1.74965} = -0.589404.$$

The equation in the second row of the last matrix is

$$-11.4472b_1 - 107.317b_2 = 0.89431.$$

We use it to obtain a value for b_1:

$$\begin{aligned} -11.4472b_1 &= 0.89431 + 107.317b_2 \\ &= 0.89431 + 107.317 \cdot (-0.589404) \\ &= -62.3588 \end{aligned}$$

so that

$$b_1 = \frac{-62.3588}{-11.4472} = 5.44752.$$

Finally, from the first equation,

$$123b_0 + 719b_1 + 4419b_2 = -85,$$

we get

$$\begin{aligned} 123b_0 &= -85 - 719b_1 - 4419b_2 \\ &= -85 - 719 \cdot (5.44752) - 4419 \cdot (-0.589404) \\ &= -85 - 3916.73 + 2604.56 \\ &= -1397.19. \end{aligned}$$

Hence

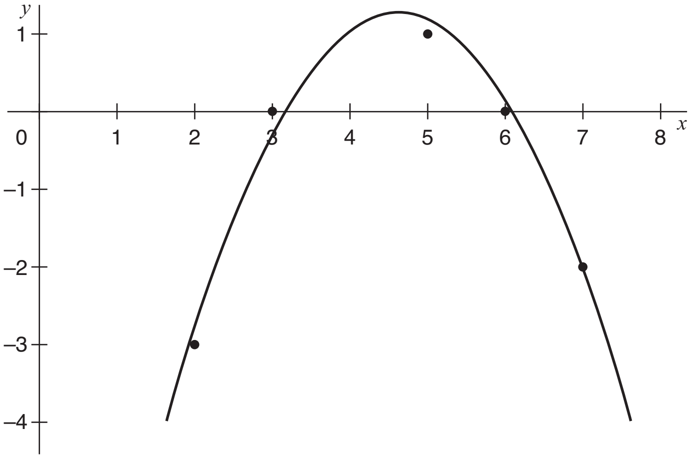
$$b_0 = \frac{-1397.17}{123} = -11.3592.$$

Rounding our answer, to 4S, we have

$$b_0 = -11.36, \qquad b_1 = 5.448, \qquad b_2 = -0.5894.$$

Hence the desired quadratic parabola that fits the data by the least squares principle is

$$y = -11.36 + 5.447x - 0.5894x^2.$$



Sec. 20.5 Prob. 9. Given points and quadratic parabola fitted by least squares

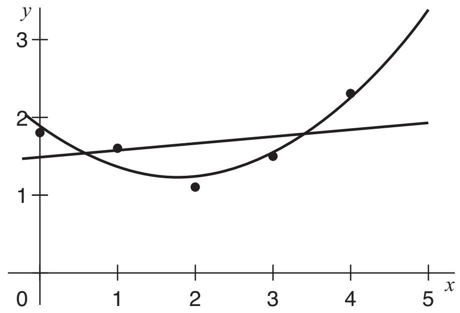
11. Comparison of linear and quadratic fit. The figure on the next page shows that a straight line obviously is not sufficient. The quadratic parabola gives a much better fit. It depends on the physical or other law underlying the data whether the fit by a quadratic polynomial is satisfactory and whether the remaining discrepancies can be attributed to chance variations, such as inaccuracy of measurement. Calculation shows that the augmented matrix of the normal equations for the straight line is

$$\begin{bmatrix} 5 & 10 & 8.3 \\ 10 & 30 & 17.5 \end{bmatrix}$$

and gives $y = 1.48 + 0.09x$. The augmented matrix for the quadratic polynomial is

$$\begin{bmatrix} 5 & 10 & 30 & 8.30 \\ 10 & 30 & 100 & 17.50 \\ 30 & 100 & 354 & 56.31 \end{bmatrix}$$

and gives $y = 1.896 - 0.741x + 0.208x^2$. For practice, you should fill in the details.



Sec. 20.5 Prob. 11. Fit by a straight line and by a quadratic parabola

Sec. 20.6 Matrix Eigenvalue Problems: Introduction

This section gives you the general facts on eigenvalues necessary for the understanding of the special numeric methods to be discussed, so that you need not consult Chap. 8. Theorem 2 on similarity of matrices is particularly important.

Sec. 20.7 Inclusion of Matrix Eigenvalues

The central issue in finding eigenvalues of an $n \times n$ matrix is to determine the roots of the corresponding characteristic polynomial of degree n. This is usually quite difficult and requires the use of an iterative numerical method, say from Sec. 19.2, or from Secs. 20.8 and 20.9 for matrices with additional properties. However, sometimes *we may only want some rough approximation* of one or more eigenvalues of the matrix, thereby avoiding costly computations. This leads to our main topic of eigenvalue inclusion.

Gerschgorin was only 30 years old when he published his beautiful and imaginative theorem, **Theorem 1, p. 879.** Take a look at **Gerschgorin's theorem** at the bottom of that page. Formula (1) says that the eigenvalues of an $n \times n$ matrix lie in the complex plane in closed circular disks. The centers of these disks are the elements of the diagonal of the matrix, and the size of these disks are determined by the sum of the elements off the diagonal in each corresponding row, respectively. Turn over to p. 880 and look at **Example 1**, which applies Gerschgorin's theorem to a 3×3 matrix and gets three disks, so called *Gerschgorin disks*, two of which overlap as shown in Fig. 449. The centers of these disks can serve as crude approximations of the eigenvalues of the matrix and the radii of the disks as the corresponding error bounds.

Problems 1 and **5** are further illustrations of Gerschgorin's theorem for real and complex matrices, respectively.

Gerschgorin's theorem (Theorem 1) and its extension (Theorem 2, p. 881) are types of theorems know as inclusion theorems. *Inclusion theorems* (p. 882) are theorems that give point sets in the complex plane that "include," i.e., contain one or several eigenvalues of a given matrix. Other such theorems are Schur's theorem (Theorem 4, p. 882), Perron's theorem (Theorem 5, p. 882) for real or complex square matrices, and Collatz inclusion theorem (Theorem 6, p. 883), which applies only to real square matrices whose elements are all positive. **Be aware that, throughout Secs. 20.7–20.9, some theorems can only be applied to certain types of matrices.**

Finally, Probs. 7, 11, and 13 are of a more theoretical nature.

Problem Set 20.7. Page 884

1. Gerschgorin disks. Real matrix.

1. *Determination of the Gerschgorin disks.* The diagonal entries of the given real matrix (which we shall denote by **A**)

$$\mathbf{A} = \begin{bmatrix} 5 & 2 & 4 \\ -2 & 0 & 2 \\ 2 & 4 & 7 \end{bmatrix}$$

are 5, 0, and 7. By Gerschgorin's theorem (Theorem 1, p. 879), these are the centers of the three desired Gerschgorin disks D_1, D_2, and D_3, respectively. For the first disk, we have the radius by (1), p. 879,

$$|a_{11} - \lambda| = |5 - \lambda| \leq |a_{12}| + |a_{13}| = |2| + |4| = 6,$$

so that

$$D_1 : |5 - \lambda| \leq 6$$

or equivalently,

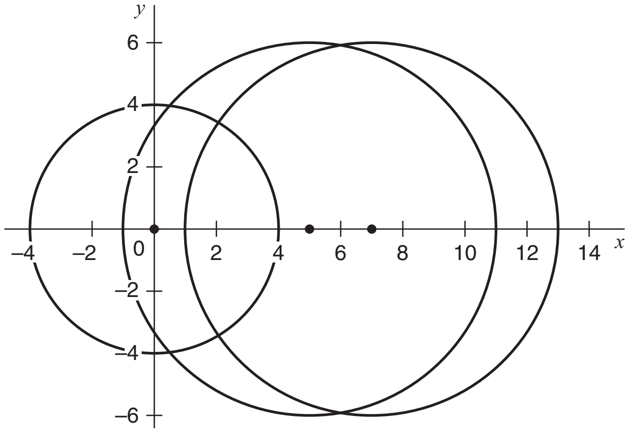
$$D_1 : \text{ center } 5, \text{ radius } 6.$$

This means, to obtain the radius of a Gerschgorin disk, we add up the absolute value of the entries in the same row as the diagonal entry (*except for the value of the diagonal entry itself*). Thus for the other two Gerschgorin disks we have

$$D_2 : \text{ center } 0, \text{ radius } 4 = (|-2| + |2|),$$

$$D_3 : \text{ center } 7, \text{ radius } 9 = (|2| + |7|).$$

Below is a sketch of the three Gerschgorin disks. Note that they intersect in the closed interval $-4 \leq \lambda \leq 13$.



Sec. 20.7 Prob. 1. Gerschgorin disks. The disks have centers $5,\ 0,\ 7$ and radii $6,\ 4,\ 6$, respectively

2. *Determination of actual eigenvalues*. We compute the characteristic polynomial $p(\lambda)$:

$$\begin{aligned} p(\lambda) = \det(\mathbf{A}-\lambda\mathbf{I}) &= \begin{vmatrix} 5-\lambda & 2 & 4 \\ -2 & -\lambda & 2 \\ 2 & 4 & 7-\lambda \end{vmatrix} \\ &= (5-\lambda)\begin{vmatrix} -\lambda & 2 \\ 4 & 7-\lambda \end{vmatrix} - 2\begin{vmatrix} -2 & 2 \\ 2 & 7-\lambda \end{vmatrix} + 4\begin{vmatrix} -2 & -\lambda \\ 2 & 4 \end{vmatrix} \\ &= (5-\lambda)[-\lambda(\,7-\lambda) - 8] - 2[-2(7-\lambda) - 4] + 4[-8 + 2\lambda] \\ &= \lambda^3 - 12\,\lambda^2 + 23\,\lambda + 36 = 0. \end{aligned}$$

We want to find the roots of the characteristic polynomial. We know the following observations:

F1. *The product of the eigenvalues of a characteristic polynomial is equal to the constant term of that polynomial.*

F2. *The sum of the eigenvalues is equal to* $(-1)^{n-1}$ *times the coefficient of the second highest term of the characteristic polynomial.* (Another example is discussed on pp. 129–130, in Volume 1, of the Student Solutions Manual).

Using these two facts, we factor the constant term 36 and get $36 = 1 \cdot 2 \cdot 2 \cdot 3 \cdot 3$. We calculate, starting with the smallest factors (both positive as given) and negative: $p(1) = 1^3 - 12 \cdot 1^2 + 23 \cdot 1 + 36 = 48 \neq 0$, $p(-1) = 0$. We found an eigenvalue! Thus a factor is $(\lambda + 1)$ and we could use long division and apply the well-known quadratic formula for finding roots. Or we can continue: $p(2) = 42$, $p(-2) = -66$, $p(4) = 0$. We found another eigenvalue. From F2, we know that the sum of the three eigenvalues must equal $(-1)^{3-1} \cdot 12 = 12$. Hence $-1 + 4 + \lambda = 12$ so the other eigenvalue must be equal to $\lambda = 9$. Hence the three eigenvalues (or the spectrum) are $-1, 4, 9$.

3. *Discussion.* The inclusion interval obtained from Gerschgorin's theorem is larger; this is typical. But the interval is the best possible in the sense that we can find, for a set of disks (with real or complex centers), a corresponding matrix such that its spectrum cannot be included in a set of smaller closed disks with the main diagonal entries of that matrix as centers.

5. Gerschgorin disks. Complex matrix. To obtain the radii of the Geschgorin disks, we compute by (1), p. 879,

$$|a_{12}| + |a_{13}| = |i| + |1+i| = \sqrt{1^2} + \sqrt{1^2 + 1^2} = 1 + \sqrt{2} \qquad \text{[by (3), p. 613]},$$

$$|a_{21}| + |a_{23}|\, |-i| + |0| = 1,$$

$$|a_{31}| + |a_{32}| = |1-i| + |0| = \sqrt{1^2 + (-1)^2} + 0 = \sqrt{2}.$$

The diagonal elements, and hence centers of the Gerschgorin disks, are

$$a_{11} = 2, \qquad a_{22} = 3, \qquad a_{33} = 8.$$

Putting it all together: The disks are D_1 : center in Prob. 1. You may want to sketch the Gerschgorin disks and determine in which closed interval they intersect.

The determination of the actual eigenvalues is as follows. Developing the determinant along the last row, with the usual checkerboard pattern in mind giving the correct plus and minus signs of the cofactors (see bottom of p. 294), we obtain

$$\begin{aligned} p(\lambda) = \det(\mathbf{A}-\lambda\mathbf{I}) &= \begin{vmatrix} 2-\lambda & i & 1+i \\ -i & 3-\lambda & 0 \\ 1-i & 0 & 8-\lambda \end{vmatrix} \\ &= (1-i)\begin{vmatrix} i & 1+i \\ 3-\lambda & 0 \end{vmatrix} - 0 + (8-\lambda)\begin{vmatrix} 2-\lambda & i \\ -i & 3-\lambda \end{vmatrix} \\ &= (1-i)[0-(1+i)(3-\lambda)] + (8-\lambda)[(2-\lambda)(3-\lambda)-1] \\ &= -\lambda^3 + 13\,\lambda^2 - 43\,\lambda + 34 = 0. \end{aligned}$$

The constant term of the characteristic polynomial is 34 and factors as follows:

$$34 = 1 \cdot 2 \cdot 17.$$

However, none of its positive and negative factors, when substituted into the characteristic polynomial, yields $p(\lambda)$ equal to zero. Hence we would have to resort to a root-finding method from Sec. 19.2, p. 802, such as Newton's method. A starting value, as suggested by Geschgorin's theorem, would be $\lambda = 1.0000$. However, the problem suggests the use of a CAS (if available). Using a CAS (here Mathematica), the spectrum $\{\lambda_1, \lambda_2, \lambda_3\}$ is

$$\begin{aligned}\lambda_1 &= 1.16308,\\ \lambda_2 &= 3.51108,\\ \lambda_3 &= 8.32584.\end{aligned}$$

Comment. We initially tried to use the approach of Prob. 1 when we determined the characteristic polyomial, factored the constant term, and then tried to determined whether any of these factors yielded zeros. This was to show that we first try a simpler approach and then go to more involved methods.

7. Similarity transformation. The matrix in Prob. 2 shows a typical situation. It may have resulted from a numeric method of diagonalization that left off-diagonal entries of various sizes but not exceeding 10^{-2} in absolute value. Gerschgorin's theorem then gives circles of radius 2×10^{-2}. These furnish bounds for the deviation of the eigenvalues from the main diagonal entries. This describes the starting situation for the present problem. Now, in various applications, one is often interested in the eigenvalue of largest or smallest absolute value. In our matrix, the smallest eigenvalue is about 5, with a maximum possible deviation of 2×10^{-2}, as given by Gerschgorin's theorem. We now wish to decrease the size of this Gerschgorin disk as much as possible. Example 2, on p. 881 in the text, shows us how we should proceed. The entry 5 stands in the first row and column. Hence we should apply, to $\mathbf{A}$, a similarity transformation involving a diagonal matrix $\mathbf{T}$ with main diagonal a, 1, 1, where a is as large as possible. The inverse of $\mathbf{T}$ is the diagonal matrix with main diagonal $1/a$, 1, 1. Leave a arbitrary and first determine the result of the similarity transformation (as in Example 2).

$$\begin{aligned}\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T} &= \begin{bmatrix} 1/a & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0.01 & 0.01 \\ 0.01 & 8 & 0.01 \\ 0.01 & 0.01 & 9 \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 0.01/a & 0.01/a \\ 0.01a & 8 & 0.01 \\ 0.01a & 0.01 & 9 \end{bmatrix}.\end{aligned}$$

We see that the Gerschgorin disks of the transformed matrix $\mathbf{B}$, by Gerschgorin's theorem, p. 879, are

Center	Radius
5	$0.02/a$
8	$0.01(a+1)$
9	$0.01(a+1)$

The last two disks must be small enough so that they do not touch or even overlap the first disk. Since $8 - 5 = 3$, the radius of the second disk, after the transformation, must be less than $3 - 0.02/a$, that is,

$$0.01\,(a+1) < 3 - 0.02/a.$$

Multiplication by $100a$ (> 0) gives

$$a^2 + a < 300\,a - 2.$$

If we replace the inequality sign by an equality sign, we obtain the quadratic equation

$$a^2 - 299a + 2 = 0.$$

Hence a must be less than the larger root 298.9933 of this equation, say, for convenience, $a = 298$. Then the radius of the second disk is $0.01(a+1) = 2.99$, so that the disk will not touch the first one, and neither will the third, which is farther away from the first. The first disk is substantially reduced in size, by a factor of almost 300, the radius of the reduced disk being

$$\frac{0.02}{298} = 0.000067114.$$

The choice of $a = 100$ would give a reduction by a factor 100, as requested in the problem. Our systematic approach shows that we can do better.

For $a = 100$ the computation is

$$\begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5.00 & 0.01 & 0.01 \\ 0.01 & 8.00 & 0.01 \\ 0.01 & 0.01 & 9.00 \end{bmatrix} \begin{bmatrix} 100 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0.0001 & 0.0001 \\ 1 & 8 & 0.01 \\ 1 & 0.01 & 9 \end{bmatrix}.$$

Remark. In general, the error bounds of the Gerschgorin disk are quite poor unless the off-diagonal entries are very small. However, for an eigenvalue in an isolated Gerschgorin disk, as in Fig. 449, p. 880, it can be meaningful to make an error bound smaller by choosing an appropriate similarity transformation

$$\mathbf{B} = \mathbf{T}^{-1}\mathbf{AT},$$

where $\mathbf{T}$ is a diagonal matrix. Do you know why this is possible? Answer: This is allowed by Theorem 2, p. 878, which ensures that *similarity transformations preserve eigenvalues*. So here we picked the smallest eigenvalue and made the error bound smaller by a factor $1/100$ as requested.

11. **Spectral radius**. By definition (see p. 324), the spectral radius of a square matrix $\mathbf{A}$ is the absolute value of an eigenvalue of $\mathbf{A}$ that is largest in absolute value. Since every eigenvalue of $\mathbf{A}$ lies in a Gerschgorin disk, for every eigenvalue of $\mathbf{A}$ we must have (make a sketch)

$$\text{(I)} \qquad |a_{jj}| + \sum |a_{jk}| \geq |\lambda_j|$$

where we sum over all off-diagonal entries in Row j (and the eigenvalues of $\mathbf{A}$ are numbered suitably).

Since (I) is true for all eigenvalues of $\mathbf{A}$, it must be true for the eigenvalue of $\mathbf{A}$ that is largest in absolute value, that is, the largest $|\lambda_j|$. But this is, by definition, the spectral radius $\rho(\mathbf{A})$. The left-hand side of (I) is precisely the row "sum" norm of $\mathbf{A}$. Hence, we have proven that

$$\text{the row "sum" norm of} \mathbf{A} \geq \rho(\mathbf{A}).$$

13. Spectral radius. The row norm was used in Prob. 11, but we could also use the Frobenius norm

$$|\lambda_j| \le \sqrt{\sum_j \sum_k c_{jk}^2} \qquad \text{[see (9), p. 861]}$$

to find the upper bound. In this case, we would get (calling the elements a_{jk}, since we called the matrix in Prob. 1 **A)**

$$\begin{aligned} |\lambda_j| &\le \sqrt{\sum_{j=1}^{3}\sum_{k=1}^{3} a_{jk}^2} \\ &= \sqrt{5^2 + 2^2 + 4^2 + (-2)^2 + 0^2 + 2^2 + 2^2 + 4^2 + 7^2} \\ &= \sqrt{122} \\ &= 11.05. \end{aligned}$$

Sec. 20.8 Power Method for Eigenvalues

The main attraction of the power method is its simplicity. For an $n \times n$ matrix **A** with a dominant eigenvalue (“dominant” means “largest in absolute value”) the method gives us an approximation (1), p. 885, usually of that eigenvalue. Furthermore, if matrix **A** is symmetric, that is, $a_{jk} = a_{kj}$ [by (1), p. 335], then we also get an error bound (2) for approximation (1). Convergence may be slow but can be improved by a *spectral shift* (Example 2, p. 887). Another use for a spectral shift is to make the method converge to the *smallest* eigenvalue as shown in Prob. 11. *Scaling* can provide a convergent sequence of eigenvectors (for more information, see Example 1, p. 886). The power method is explained in great detail in **Prob. 5**.

More details on Example 1, pp. 886–887. Application of Power Method, Error Bound (Theorem 1, p. 885). Scaling. We take a closer look at the six vectors listed at the beginning of the example:

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_1 = \begin{bmatrix} 0.890244 \\ 0.609756 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 0.890244 \\ 0.609756 \\ 1 \end{bmatrix},$$

$$\mathbf{x}_5 = \begin{bmatrix} 0.990663 \\ 0.504682 \\ 1 \end{bmatrix}, \quad \mathbf{x}_{10} = \begin{bmatrix} 0.999707 \\ 0.500146 \\ 1 \end{bmatrix}, \quad \mathbf{x}_{15} = \begin{bmatrix} 0.999991 \\ 0.500005 \\ 1 \end{bmatrix}.$$

Vector $\mathbf{x}_0$ was scaled. The others were obtained by multiplication by the given matrix **A** and subsequent scaling. We can use any of these vectors for obtaining a corresponding Rayleigh quotient q as an approximate value of an (unknown) eigenvalue of **A** and a corresponding error bound δ for q. Hence we have six possibilities using one of the given vectors, and indeed many more if we want to compute further vectors. Note that we must not use two of the given vectors because of the scaling, but just one vector. For instance, if we use $\mathbf{x}_1$, and then its product $\mathbf{Ax}_1$ we get

$$\mathbf{A} = \begin{bmatrix} 0.49 & 0.02 & 0.22 \\ 0.02 & 0.28 & 0.20 \\ 0.22 & 0.20 & 0.40 \end{bmatrix}, \quad \mathbf{x}_1 = \begin{bmatrix} 0.890244 \\ 0.609756 \\ 1 \end{bmatrix}, \quad \mathbf{Ax}_1 = \begin{bmatrix} 0.668415 \\ 0.388537 \\ 0.717805 \end{bmatrix}.$$

From these data we calculate the inner products by Theorem 1, p. 885,

$$\begin{aligned} m_0 &= \mathbf{x}_1^{\mathsf{T}}\mathbf{x}_1 &&= 2.164337, \\ m_1 &= \mathbf{x}_1^{\mathsf{T}}\mathbf{Ax}_1 &&= 1.549770, \\ m_2 &= (\mathbf{Ax}_1)^{\mathsf{T}}\mathbf{Ax}_1 &&= 1.112983. \end{aligned}$$

These now give the Rayleigh quotient q and error bound δ of q by (1), (2) p. 885:

$$\begin{aligned} q &= m_1/m_0 &&= 0.716048, \\ \delta &= \sqrt{m_2/m_0 - q^2} &&= 0.038887, \end{aligned}$$

where q approximates the eigenvalue 0.72 of **A,** so that the error of q is

$$\epsilon = 0.72 - q = 0.003952.$$

These values agree with those for $j = 2$ in the table for Example 1 on p. 887 of the textbook.

Problem Set 20.8. Page 887

5. Power method with scaling. The given matrix is

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

Use the same notation as in Example 1 in the text. From $\mathbf{x}_0 = [1 \quad 1 \quad 1]^{\mathsf{T}}$ calculate $\mathbf{Ax}_0$ and then scale it as indicated in the problem, calling the resulting vector $\mathbf{x}_1$. This is the first step. In the second step calculate $\mathbf{Ax}_1$ and then scale it, calling the resulting vector $\mathbf{x}_2$. And so on. More details are as follows:

Iteration 1: We start with

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Multiplication by the given matrix **A** gives us

$$\mathbf{Ax}_0 = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}.$$

The calculations that give approximations q (Rayleigh quotients) and error bounds are as follows.

For m_0, m_1, and m_2

$$m_0 = \mathbf{x}_0^{\mathsf{T}}\mathbf{x}_0 = \begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = 1\cdot 1 + 1\cdot 1 + 1\cdot 1 = 3,$$

$$m_1 = \mathbf{x}_0^{\mathsf{T}}\mathbf{A}\mathbf{x}_0 = \begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}2\\4\\6\end{bmatrix} = 1\cdot 2 + 1\cdot 4 + 1\cdot 6 = 12,$$

$$m_2 = (\mathbf{A}\mathbf{x}_0)^{\mathsf{T}}\mathbf{A}\mathbf{x}_0 = \begin{bmatrix}2 & 4 & 6\end{bmatrix}\begin{bmatrix}2\\4\\6\end{bmatrix} = 2\cdot 2 + 4\cdot 4 + 6\cdot 6 = 56,$$

$$q = \frac{m_1}{m_0} = \frac{12}{3} = 4.$$

We know that $\delta^2 = m_2/m_0 - q^2$; so

$$\delta^2 = \frac{m_2}{m_0} - q^2 = \frac{56}{3} - 4^2 = 18.66667 - 16 = 2.666667,$$

$$\delta = \sqrt{2.666667} = 1.632993,$$

$$q - \delta = 4 - 1.632993 = 2.367007,$$

$$q + \delta = 4 + 1.632993 = 5.632993.$$

Iteration 2: If this is not sufficient, we iterate by using a **scaling factor**. We chose the absolute largest component of $\mathbf{Ax}_0$. This is 6, so we get

$$\mathbf{x}_1 = \begin{bmatrix}\frac{2}{6}\\\frac{4}{6}\\\frac{6}{6}\end{bmatrix} = \begin{bmatrix}0.3333333\\0.6666667\\1\end{bmatrix}.$$

Again, we multiply this vector by the given matrix **A**:

$$\mathbf{Ax}_1 = \begin{bmatrix}2 & -1 & 1\\-1 & 3 & 2\\1 & 2 & 3\end{bmatrix}\begin{bmatrix}0.3333333\\0.6666667\\1\end{bmatrix} = \begin{bmatrix}1\\3.666667\\4.666667\end{bmatrix}.$$

As before, we compute the values required to obtain our next approximation to q and δ:

$$m_0 = \mathbf{x}_1^{\mathsf{T}}\mathbf{x}_1 = \begin{bmatrix}0.3333333 & 0.6666667 & 1\end{bmatrix}\begin{bmatrix}0.3333333\\0.6666667\\1\end{bmatrix} = 1.555556,$$

$$m_1 = \mathbf{x}_1^\mathsf{T}\mathbf{A}\mathbf{x}_1 = \begin{bmatrix}0.3333333 & 0.6666667 & 1\end{bmatrix}\begin{bmatrix}1\\ 3.666667\\ 4.666667\end{bmatrix} = 7.444445,$$

$$m_2 = (\mathbf{A}\mathbf{x}_1)^\mathsf{T}\,\mathbf{A}\mathbf{x}_1 = \begin{bmatrix}1 & 3.666667 & 4.666667\end{bmatrix}\begin{bmatrix}1\\ 3.666667\\ 4.666667\end{bmatrix} = 36.22223,$$

$$\begin{aligned} q &= \frac{m_1}{m_0} = \frac{7.444445}{1.555556} = 4.785713,\\ \delta^2 &= \frac{m_2}{m_0} - q^2 = \frac{36.22223}{1.555556} - (4.785713)^2,\\ &= 23.28571 - 22.90305 = 0.38266. \end{aligned}$$

It is important to notice that we have a loss of significant digits (*subtractive cancelation*) in the computation of δ. The two terms that are used in the subtraction are similar and we go from seven digits to five. This suggests that, for more than three iterations, we might require our numbers to have more digits.

$$\begin{aligned} \delta &= \sqrt{0.38266} = 0.6185952,\\ q - \delta &= 4.785713 - 0.6185952 = 4.167118,\\ q + \delta &= 4.785713 + 0.6185952 = 5.404308. \end{aligned}$$

Iteration 3: Again, if the result is not good enough, we need to move to the next iteration by using the largest value of $\mathbf{A}\mathbf{x}_1$ as our scaling factor. This is 4.666665 so we get for $\mathbf{x}_2$

$$\mathbf{x}_2 = \begin{bmatrix}1/4.666667\\ 3.666667/4.666667\\ 4.666667/4.666667\end{bmatrix} = \begin{bmatrix}0.2142857\\ 0.7857143\\ 1\end{bmatrix},$$

from which

$$\mathbf{A}\mathbf{x}_2 = \begin{bmatrix}2 & -1 & 1\\ -1 & 3 & 2\\ 1 & 2 & 3\end{bmatrix}\begin{bmatrix}0.2142857\\ 0.7857143\\ 1\end{bmatrix} = \begin{bmatrix}0.6428571\\ 4.142857\\ 4.785714\end{bmatrix}.$$

This is followed by one more scaling step for the final result of $\mathbf{x}_3$:

$$\mathbf{x}_3 = \begin{bmatrix}0.6428571/4.785714\\ 4.142857/4.785714\\ 4.785714/4.785714\end{bmatrix} = \begin{bmatrix}0.1343284\\ 0.8656717\\ 1\end{bmatrix},$$

$$m_0 = \mathbf{x}_2^{\mathsf{T}}\mathbf{x}_2 = \begin{bmatrix}0.2142857 & 0.7857143 & 1\end{bmatrix}\begin{bmatrix}0.2142857\\0.7857143\\1\end{bmatrix} = 1.663265,$$

$$m_1 = \mathbf{x}_2^{\mathsf{T}}\mathbf{A}\mathbf{x}_2 = \begin{bmatrix}0.2142857 & 0.7857143 & 1\end{bmatrix}\begin{bmatrix}0.6428571\\4.142857\\4.785714\end{bmatrix} = 8.178571,$$

$$m_2 = (\mathbf{A}\mathbf{x}_2)^{\mathsf{T}}\,\mathbf{A}\mathbf{x}_2 = \begin{bmatrix}0.6428571 & 4.142857 & 4.785714\end{bmatrix}\begin{bmatrix}0.6428571\\4.142857\\4.785714\end{bmatrix}$$

$$= 40.47959,$$

$$q = \frac{m_1}{m_0} = \frac{8.178571}{1.663265} = 4.917179,$$

$$\delta^2 = \frac{m_2}{m_0} - q^2 = \frac{40.47959}{1.663265} - (4.917179)^2$$

$$= 24.33743 - 24.17865 = 0.1587774,$$

$$\delta = \sqrt{0.1587774} = 0.3984688,$$

$$q - \delta = 4.917179 - 0.3984688 = 4.51871,$$

$$q + \delta = 4.917179 + 0.3984688 = 5.315648.$$

The results are summarized and rounded in the following table. Note how the value of the δ gets smaller so that we have a smaller error bound on q.

m_0	$\mathbf{x}_0^{\mathsf{T}}\mathbf{x}_0 = 3$	$\mathbf{x}_1^{\mathsf{T}}\mathbf{x}_1 = 1.55556$	$\mathbf{x}_2^{\mathsf{T}}\mathbf{x}_2 = 1.663$
m_1	$\mathbf{x}_0^{\mathsf{T}}\mathbf{A}\mathbf{x}_0 = 12$	$\mathbf{x}_1^{\mathsf{T}}\mathbf{A}\mathbf{x}_1 = 7.44444$	$\mathbf{x}_2^{\mathsf{T}}\mathbf{A}\mathbf{x}_2 = 8.179$
m_2	$(\mathbf{A}\mathbf{x}_0)^{\mathsf{T}}\,\mathbf{A}\mathbf{x}_0 = 56$	$(\mathbf{A}\mathbf{x}_1)^{\mathsf{T}}\,\mathbf{A}\mathbf{x}_1 = 36.22$	$(\mathbf{A}\mathbf{x}_2)^{\mathsf{T}}\,\mathbf{A}\mathbf{x}_2 = 40.48$
$q = \dfrac{m_2}{m_0}$	4	4.786	4.917
$\delta^2 = \dfrac{m_2}{m_0} - q^2$	2.667	0.3826	0.1588
δ	1.633	0.6186	0.3985
$q - \delta$	2.367	4.167	4.519
$q + \delta$	5.633	5.404	5.316

Solving the characteristic equation $-x^3 + 8x^2 - 15x$ shows that the matrix has the eigenvalues 0, 3, and 5. Corresponding eigenvectors are

$$\mathbf{z}_1 = [0 \quad 1 \quad 1]^{\mathsf{T}}, \quad \mathbf{z}_2 = [-1 \quad -1 \quad 1]^{\mathsf{T}}, \quad \mathbf{z}_3 = [-2 \quad 1 \quad -1]^{\mathsf{T}},$$

respectively. We see that the interval obtained in the first step includes the eigenvalues 3 and 5. Only in the second step and third step of the iteration did we obtain intervals that include only the largest eigenvalue, as is usually the case from the beginning on. The reason for this interesting observation is the fact that $\mathbf{x}_0$ is a linear combination of all three eigenvectors,

$$\mathbf{x}_0 = \mathbf{z}_1 - \tfrac{1}{3}(\mathbf{z}_2 + \mathbf{z}_3),$$

as can be easily verified, and it needs several iterations until the powers of the largest eigenvalue make the iterate $\mathbf{x}_j$ come close to $\mathbf{z}_1$, the eigenvector corresponding to $\lambda = 5$. This situation occurs quite frequently, and one needs more steps for obtaining satisfactory results the closer in absolute value the other eigenvalues are to the absolutely largest one.

11. Spectral shift, smallest eigenvalue. In Prob. 3,

$$\mathbf{B} = \mathbf{A} - 3\mathbf{I} = \begin{bmatrix} -1 & -1 & 1 \\ -1 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}.$$

Now the power method converges to the eigenvalue $\lambda_{\max}$ of largest absolute value. (Here we assume that the matrix does not have $-\lambda_{\max}$ as another eigenvalue.) Accordingly, to obtain convergence to the *smallest* eigenvalue, make a shift to $\mathbf{A} + k\mathbf{I}$ with a negative k. Choose k by trial and error, reasoning about as follows. The given matrix has trace $\mathbf{A} = 2 + 3 + 3 = 8$. This is the sum of the eigenvalues. From Prob. 5 we know that the absolutely largest eigenvalue is about 5. Hence the sum of the other eigenvalues equals about 3. Hence $k = -3$ suggested in the problem seems to be a reasonable choice. Our computation of the Rayleigh quotients and error bounds gives for the first step $\mathbf{x}_0 = [1 \quad 1 \quad 1]^\mathsf{T}$, $\mathbf{x}_1 = [-1 \quad 1 \quad 3]^\mathsf{T}$, $m_0 = 3$, $m_1 = 3$, $m_2 = 11$, $q = 1$, $\delta = \sqrt{\frac{11}{3} - 1} = \sqrt{\frac{8}{3}}$, and so on, namely,

q	1	0.63636	−0.28814	−1.2749	−2.0515	−2.5288	−2.7790	−2.8993	−2.9547
δ	1.6323	2.2268	2.4910	2.3770	1.9603	1.4608	1.0277	0.70234	0.47355

We see that the Rayleigh quotients seem to converge to −3, which corresponds to the eigenvalue 0 of the given matrix. It is interesting that the sequence of the δ is not monotone; δ first increases and starts decreasing when q gets closer to the limit −3. This is typical. Also, note that the error bounds are much larger than the actual errors of q. This is also typical.

Sec. 20.9 Tridiagonalization and QR-Factorization

Somewhat more recent developments in numerics provided us with a widely used method of computing *all* the eigenvalues of an $n \times n$ *real symmetric* matrix $\mathbf{A}$. Recall that, in such a special matrix, its entries off the main diagonal are mirror images, that is, $a_{jk} = a_{kj}$ [by (1), p. 335].

In the first stage, we use **Householder's tridiagonalization method** (pp. 889–892) to transform the matrix $\mathbf{A}$ into a *tri*diagonal matrix $\mathbf{B}$ ("*tri*" = "three"), that is, a matrix having all its nonzero entries *on* the main diagonal, in the position immediately *below* the main diagonal, or immediately *above* the main diagonal (Fig. 450, matrix in Third Step, p. 889). In the second stage, we apply the **QR-factorization method** (pp. 892–896) to the tridiagonal matrix $\mathbf{B}$ to obtain a matrix $\mathbf{B}_{s+1}$ whose real diagonal entries are approximations of the desired eigenvalues of $\mathbf{A}$ (whereby the nonzero entries are sufficiently small in absolute value). The purpose of the first stage is to produce many zeros in the matrix and thus speed up the convergence for the QR method in the second stage.

Perhaps the easiest way to understand Householder's tridiagonalization method is to go through **Example 1**, pp. 890–891. A further illustration of the method is given in **Prob. 3**. Similarly, another good way to understand the QR-factorization method is to work through **Example 2**, pp. 894–896 with a further demonstration of the method in **Prob. 7**. Both examples and both problems are each concerned with the same real symmetric matrices, respectively.

An **outline** of this section is as follows:

- Discussion of the problem and biographic reference to Householder's tridiagonalization method, p. 888.
- Householder's tridiagonalization method (pp. 889–892).
- Formula (1), on p. 889, is the general set of formulas for the similarity transformations $\mathbf{P}_r$ to obtain, in stages, the tridiagonal matrix $\mathbf{B}$.
- Figure 450 illustrates, visually, how a 5×5 matrix $\mathbf{A}$ gets transformed into $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ so that at the end $\mathbf{B} = \mathbf{A}_3$.
- Formulas (2) and (3), p. 889, show the general form of the similarity transformations $\mathbf{P}_r$ and associated unit vectors $\mathbf{v}_r$.
- The important formula (4), on the top of p. 890, defines the components of the unit vectors $\mathbf{v}_r$ of (2) and (3). Notice, in 4(b), sgn a_{21} is the sign function. It extracts the sign from a number, here a_{21}. This function gives "plus one" when a number is zero or positive and "minus one" when a number is negative. Thus, for example,

$$\operatorname{sgn} 8 = +1, \quad \operatorname{sgn} 0 = +1, \quad \operatorname{sgn}\ (-55) = -1.$$

- For each interation in formula (4) we increase, by 1, all subscripts of the components of the column vector(s) $\mathbf{v}_r$ $(r = 2$ for step 2). We iterate $n - 2$ times for an $n \times n$ matrix.
- Example 1, on p. 890, illustrates the method in detail.
- Proof, p. 891, of Formula (1)
- QR-factorization method (pp. 892–896).
- Biographic references to the QR-factorization method, p. 892.
- Assuming that Householder's Tridiagonalization Method has been applied first to matrix $\mathbf{A}$, we start with tridiagonal matrix $\mathbf{B} = \mathbf{B}_0$. Two different kind of matrices in *Step 1*, p. 892: orthogonal matrix $\mathbf{Q}_0$ (means $\mathbf{Q}_0^{-1} = \mathbf{Q}_0^{\mathsf{T}}$) and upper triangular matrix $\mathbf{R}_0$. Step consists first of factorization ("QR-factorization") and then computation.
- Formula (5) gives *General Step* with matrices $\mathbf{Q}_s$ and $\mathbf{R}_s$ with 5(a) factorization (QR-factorization) and 5(b) computation.
- Proof, p. 892, of Formula (5).
- Detailed outline on how to get the 5(a) factorization (QR-factorization), p. 892. The method needs orthogonal matrices $\mathbf{C}_j$ that contain 2×2 plane rotation submatrices, which for $n = 4$ can be determined by (11).
- How to get 5(b) computation from 5(a), p. 892.
- Example 2, on p. 894, illustrates the method in detail.

More Details on Example 2, p. 894. QR-Factorization Method. The tridiagonalized matrix is (p. 895)

$$\mathbf{B} = \begin{bmatrix} 6 & -\sqrt{18} & 0 \\ -\sqrt{18} & 7 & \sqrt{2} \\ 0 & \sqrt{2} & 6 \end{bmatrix}.$$

We use the abbreviations c_2, s_2, and t_2 for $\cos\theta_2$, $\sin\theta_2$, and $\tan\theta_2$, respectively. We multiply **B** from the left by

$$\mathbf{C}_2 = \begin{bmatrix} c_2 & s_2 & 0 \\ -s_2 & c_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The purpose of this multiplication is to obtain a matrix $\mathbf{C}_2\mathbf{B} = [\,b_{jk}^{(2)}]$ for which the off-diagonal entry $b_{21}^{(2)}$ is zero. Now this entry is the inner product of Row 2 of $\mathbf{C}_2$ times Column 1 of **B**, that is,

$$-s_2 \cdot 6 + c_2(-\sqrt{18}) = 0, \qquad \text{thus} \quad t_2 = -\sqrt{\tfrac{18}{6}} = -\sqrt{\tfrac{1}{2}}.$$

From this and the formulas that express cos and sin in terms of tan we obtain

$$c_2 = 1/\sqrt{1+t_2^2} = \sqrt{\tfrac{2}{3}} = 0.816496581,$$

$$s_2 = t_2/\sqrt{1+t_2^2} = -\sqrt{\tfrac{1}{3}} = -0.577350269.$$

θ_3 is determined similarly, with the purpose of obtaining $b_{32}^{(3)} = 0$ in $\mathbf{C}_3\mathbf{C}_2\mathbf{B} = [b_{jk}^{(3)}]$.

Problem Set 20.9. Page 896

3. Tridiagonalization. The given matrix

$$\mathbf{A} = \begin{bmatrix} 7 & 2 & 3 \\ 2 & 10 & 6 \\ 3 & 6 & 7 \end{bmatrix}$$

is symmetric. Hence we can apply Householder's method for obtaining a tridiagonal matrix (which will have two zeros in the location of the entries 3). Proceed as in Example 1 of the text. Since **A** is of size $n = 3$, we have to perform $n - 2 = 1$ step. (In Example 1 we had $n = 4$ and needed $n - 2 = 2$ steps.) Calculate the vector $\mathbf{v}_1$ from (4), p. 890. Denote it simply by **v** and its components by $v_1(= 0)$, v_2, v_3 because we do only one step. Similarly, denote S_1 in (4c) by S. Compute

$$S = \sqrt{a_{21}^2 + a_{31}^2} = \sqrt{2^2 + 3^2} = \sqrt{13} = 3.60555.$$

If we compute, using, say, six digits, we may expect that, instead of those two zeros in the tridiagonalized matrix, we obtain entries of the order 10^{-6} or even larger in absolute value. We always have $v_1 = 0$. From (4a) we obtain the second component

$$v_2 = \sqrt{\frac{1 + a_{21}/S}{2}} = \sqrt{\frac{1 + 2/3.60555}{2}} = 0.881675.$$

From (4b) with $j = 3$ and sgn $a_{21} = +1$ (because a_{21} is positive) we obtain the third component

$$v_3 = \frac{a_{31}}{2\,v_2\,S} = \frac{3}{2 \cdot 0.881675 \cdot 3.60555} = 0.471858.$$

With these values we now compute $\mathbf{P}_r$ from (2), where $r = 1, \ldots, n-2$, so that we have only $r = 1$ and can denote $\mathbf{P}_1$ simply by $\mathbf{P}$. Note well that $\mathbf{v}^\mathsf{T}\mathbf{v}$ would be the dot product of the vector by itself (thus the square of its length), whereas $\mathbf{v}\mathbf{v}^\mathsf{T}$ is a 3×3 matrix because of the usual matrix multiplication. We thus obtain from (2), p. 889,

$$\begin{aligned}
\mathbf{P} &= \mathbf{I} - 2\mathbf{v}\mathbf{v}^\mathsf{T} \\
&= \mathbf{I} - 2\begin{bmatrix} v_1^2 & v_1v_2 & v_1v_3 \\ v_2v_1 & v_2^2 & v_2v_3 \\ v_3v_1 & v_3v_2 & v_3^2 \end{bmatrix} \\
&= \begin{bmatrix} 1-2v_1^2 & -2v_1v_2 & -2v_1v_3 \\ -2v_2v_1 & 1-2v_2^2 & -2v_2v_3 \\ -2v_3v_1 & -2v_3v_2 & 1-2v_3^2 \end{bmatrix} \\
&= \begin{bmatrix} 1.0 & 0 & 0 \\ 0 & -0.554702 & -0.832051 \\ 0 & -0.832051 & 0.554700 \end{bmatrix}.
\end{aligned}$$

Finally use $\mathbf{P}$, and its inverse $\mathbf{P}^{-1} = \mathbf{P}$, for the similarity transformation that will produce the tridiagonal matrix

$$\begin{aligned}
\mathbf{B} = \mathbf{PAP} = \mathbf{P}&\begin{bmatrix} 7.0 & -3.605556 & -0.000001 \\ 2.0 & -10.539321 & -4.992308 \\ 3.0 & -9.152565 & -1.109404 \end{bmatrix} \\
= &\begin{bmatrix} 7.0 & -3.605556 & -0.000001 \\ -3.605556 & 13.461578 & 3.692322 \\ -0.000001 & 3.692322 & 3.538467 \end{bmatrix}.
\end{aligned}$$

The point of the use of similarity transformations is that they preserve the spectrum of $\mathbf{A}$, consisting of the eigenvalues

$$2, \quad 5, \quad 16,$$

which can be found, for instance, by graphing the characteristic polynomial of $\mathbf{A}$ and applying Newton's method for improving the values obtained from the graph.

7. QR-factorization. The purpose of this factorization is the determination of approximate values of *all* the eigenvalues of a given matrix. To save work, one usually begins by tridiagonalizing the matrix, which must be symmetric. This was done in Prob. 3. The matrix at the end of that problem

$$\mathbf{B}_0 = [b_{jk}] = \begin{bmatrix} 7.0 & -3.605551275 & 0 \\ -3.605551275 & 13.46153846 & 3.692307692 \\ 0 & 3.692307692 & 3.538461538 \end{bmatrix}$$

is tridiagonal (note that greater accuracy is being used). Hence QR can begin. We proceed as in Example 2, on p. 894, of the textbook. To save writing, we write c_2, s_2, t_2 for $\cos\theta_2$, $\sin\theta_2$, $\tan\theta_2$, respectively.

Step 1. Consider the matrix

$$\mathbf{C}_2 = \begin{bmatrix} c_2 & s_2 & 0 \\ -s_2 & c_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

with the angle of rotation θ_2 determined so that, in the product $\mathbf{W}_0 = \mathbf{C}_2\mathbf{B}_0 = [w_{jk}^{(0)}]$, the entry $w_{21}^{(0)}$ is zero. By the usual matrix multiplication (row times column) $w_{21}^{(0)}$ is the inner product of Row 2 of $\mathbf{C}_2$ times Column 1 of $\mathbf{B}_0$, that is,

$$-s_2\, b_{11}^{(0)} + c_2\, b_{21}^{(0)} = 0, \qquad \text{hence} \qquad t_2 = s_2/c_2 = b_{21}^{(0)}/b_{11}^{(0)}.$$

From this, and the formulas for cos and sin in terms of tan (usually discussed in calculus), we obtain

$$\text{(I/1)} \qquad \begin{aligned} c_2 &= 1\Big/\sqrt{1 + \left(b_{21}^{(0)}/b_{11}^{(0)}\right)^2} = 0.889000889, \\ s_2 &= \frac{b_{21}^{(0)}}{b_{11}^{(0)}}\Big/\sqrt{1 + \left(b_{21}^{(0)}/b_{11}^{(0)}\right)^2} = -0.4579054698. \end{aligned}$$

Use these values in $\mathbf{C}_2$ and calculate $\mathbf{C}_2\mathbf{B}_0 = \mathbf{W}_0 = [w_{jk}^{(0)}]$. Thus

$$\mathbf{W}_0 = [w_{jk}^{(0)}] = \mathbf{C}_2\mathbf{B}_0 = \begin{bmatrix} 7.874007873 & -9.369450382 & -1.690727888 \\ 0 & 10.31631801 & 3.282464821 \\ 0 & 3.692307692 & 3.538461538 \end{bmatrix}.$$

$\mathbf{C}_2$ has served its purpose: instead of $b_{21}^{(0)} = -3.605551276$ we now have $w_{21}^{(0)} = 0$. (Instead of $w_{21}^{(0)} = 0$, on the computer we may get -10^{-10} or another very small entry—the use of more digits in $\mathbf{B}_0$ ensured the 0.) Now use the abbreviations c_3, s_3, t_3 for $\cos\theta_3$, $\sin\theta_3$, $\tan\theta_3$. Consider the matrix

$$\mathbf{C}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_3 & s_3 \\ 0 & -s_3 & c_3 \end{bmatrix}$$

with the angle of rotation θ_3 such that, in the product matrix $\mathbf{R}_0 = [r_{jk}] = \mathbf{C}_3\mathbf{W}_0 = \mathbf{C}_3\mathbf{C}_2\mathbf{B}_0$, the entry r_{32} is zero. This entry is the inner product of Row 3 of $\mathbf{C}_3$ times Column 2 of $\mathbf{W}_0$. Hence

$$-s_3\, w_{22}^{(0)} + c_3\, w_{32}^{(0)} = 0, \qquad \text{so that} \qquad t_3 = s_3/c_3 = w_{32}^{(0)}/w_{22}^{(0)} = 0.357909449.$$

This gives, for c_3 and s_3,

$$\text{(II/1)} \qquad c_3 = 1/\sqrt{1+t_3^2} = 0.9415130836, \qquad s_3 = t_3/\sqrt{1+t_3^2} = 0.3369764287.$$

Using this, we obtain

$$\mathbf{R}_0 = \mathbf{C}_3\mathbf{W}_0 = \mathbf{C}_3\mathbf{C}_2\mathbf{B}_0 = \begin{bmatrix} 7.874007873 & -9.369450382 & -1.690727888 \\ 0 & 10.95716904 & 4.282861708 \\ 0 & 0 & 2.225394561 \end{bmatrix}.$$

(Again, instead of 0, you might obtain 10^{-10} or another very small term—similarly in the further calculations.) Finally, we multiply $\mathbf{R}_0$ from the right by $\mathbf{C}_2^\mathsf{T}\mathbf{C}_3^\mathsf{T}$. This gives

$$\begin{aligned} \mathbf{B}_1 &= \mathbf{R}_0\mathbf{C}_2^\mathsf{T}\mathbf{C}_3^\mathsf{T} = \mathbf{C}_3\mathbf{C}_2\mathbf{B}_0\mathbf{C}_2^\mathsf{T}\mathbf{C}_3^\mathsf{T} \\ &= \begin{bmatrix} 11.29032258 & -5.017347637 & 0 \\ -5.017347637 & 10.61443933 & 0.7499055128 \\ 0 & 0.7499055119 & 2.095238095 \end{bmatrix}. \end{aligned}$$

The given matrix $\mathbf{B}_0$ (and, thus, also the matrix $\mathbf{B}_1$) has the eigenvalues 16, 6, 2. We see that the main diagonal entries of $\mathbf{B}_1$ are approximations that are not very accurate, a fact that we could have concluded from the relatively large size of the off-diagonal entries of $\mathbf{B}_1$. In practice, one would perform further steps of the iteration until all off-diagonal entries have decreased in absolute value to less than a given bound. The answer, on p. A51 in App. 2, gives the results of two more steps, which are obtained by the following calculations.

Step 2. The calculations are the same as before, with $\mathbf{B}_0 = [b_{jk}^{(0)}]$ replaced by $\mathbf{B}_1 = [b_{jk}^{(1)}]$. Hence, instead of (I/1), we now have

$$\text{(I/2)} \qquad \begin{aligned} c_2 &= 1/\sqrt{1+(b_{21}^{(1)}/b_{11}^{(1)})^2} = 0.9138287756, \\ s_2 &= (b_{21}^{(1)}/b_{11}^{(1)})/\sqrt{1+(b_{21}^{(1)}/b_{11}^{(1)})^2} = -0.4060997031. \end{aligned}$$

We can now write the matrix $\mathbf{C}_2$, which has the same general form as before, and calculate the product

$$\begin{aligned} \mathbf{W}_1 &= [w_{jk}^{(1)}] = \mathbf{C}_2\mathbf{B}_1 \\ &= \begin{bmatrix} 12.35496505 & -8.895517309 & -0.3045364061 \\ 0 & 7.662236711 & 0.6852852366 \\ 0 & 0.7499055119 & 2.095238095 \end{bmatrix}. \end{aligned}$$

Now calculate the entries of $\mathbf{C}_3$ from (II/1) with $t_3 = w_{32}^{(0)}/w_{22}^{(0)}$ replaced by $t_3 = w_{32}^{(1)}/w_{22}^{(1)}$, that is,

$$\text{(II/2)} \qquad \begin{aligned} c_3 &= 1/\sqrt{1+(t_3)^2} = 0.9952448346, \\ s_3 &= t_3/\sqrt{1+(t_3)^2} = 0.09740492434. \end{aligned}$$

We can now write $\mathbf{C}_3$, which has the same general form as in step 1, and calculate

$$\mathbf{R}_1 = \mathbf{C}_3\mathbf{W}_1 = \mathbf{C}_3\mathbf{C}_2\mathbf{B}_1$$
$$= \begin{bmatrix} 12.35496505 & -8.895517309 & -0.3045364061 \\ 0 & 7.698845998 & 0.8861131001 \\ 0 & 0 & 2.018524735 \end{bmatrix}.$$

This gives the next result

$$\mathbf{B}_2 = [b_{jk}^{(2)}] = \mathbf{R}_1\mathbf{C}_2^{\mathsf{T}}\mathbf{C}_3^{\mathsf{T}} = \mathbf{C}_3\mathbf{C}_2\mathbf{B}_1\mathbf{C}_2^{\mathsf{T}}\mathbf{C}_3^{\mathsf{T}}$$
$$= \begin{bmatrix} 14.90278952 & -3.126499072 & 0 \\ -3.126499074 & 7.088284172 & 0.1966142499 \\ 0 & 0.1966142491 & 2.008926316 \end{bmatrix}.$$

The approximations of the eigenvalues have improved. The off-diagonal entries are smaller than in $\mathbf{B}_1$. Nevertheless, in practice, the accuracy would still not be sufficient, so that one would do several more steps. We do one more step, whose result is also given on p. A51 in App. 2 of the textbook.

Step 3. The calculations are the same as in step 2, with $\mathbf{B}_1 = [b_{jk}^{(1)}]$ replaced by $\mathbf{B}_2 = [b_{jk}^{(2)}]$. Hence we calculate the entries of $\mathbf{C}_2$ from

$$\text{(I/3)} \qquad \begin{aligned} c_2 &= 1\Big/\sqrt{1 + (b_{21}^{(2)}/b_{11}^{(2)})^2} = 0.9786942487, \\ s_2 &= (b_{21}^{(2)}/b_{11}^{(2)})\Big/\sqrt{1 + (b_{21}^{(2)}/b_{11}^{(2)})^2} = -0.2053230812. \end{aligned}$$

We can now write the matrix $\mathbf{C}_2$ and calculate the product

$$\mathbf{W}_2 = [w_{jk}^{(2)}] = \mathbf{C}_2\mathbf{B}_2$$
$$= \begin{bmatrix} 15.22721682 & -4.515275007 & -0.04036944359 \\ 0 & 6.295320529 & 0.1924252356 \\ 0 & 0.1966142491 & 2.008926316 \end{bmatrix}.$$

Now calculate the entries of $\mathbf{C}_3$ from (II/2) with t_2 replaced by $t_3 = w_{22}^{(2)}/w_{32}^{(2)}$, that is,

$$\text{(II/3)} \qquad \begin{aligned} c_3 &= 1\Big/\sqrt{1 + (t_3)^2} = 0.9995126436, \\ s_3 &= t_3\Big/\sqrt{1 + (t_3)^2} = 0.03121658809. \end{aligned}$$

Write $\mathbf{C}_3$ and calculate

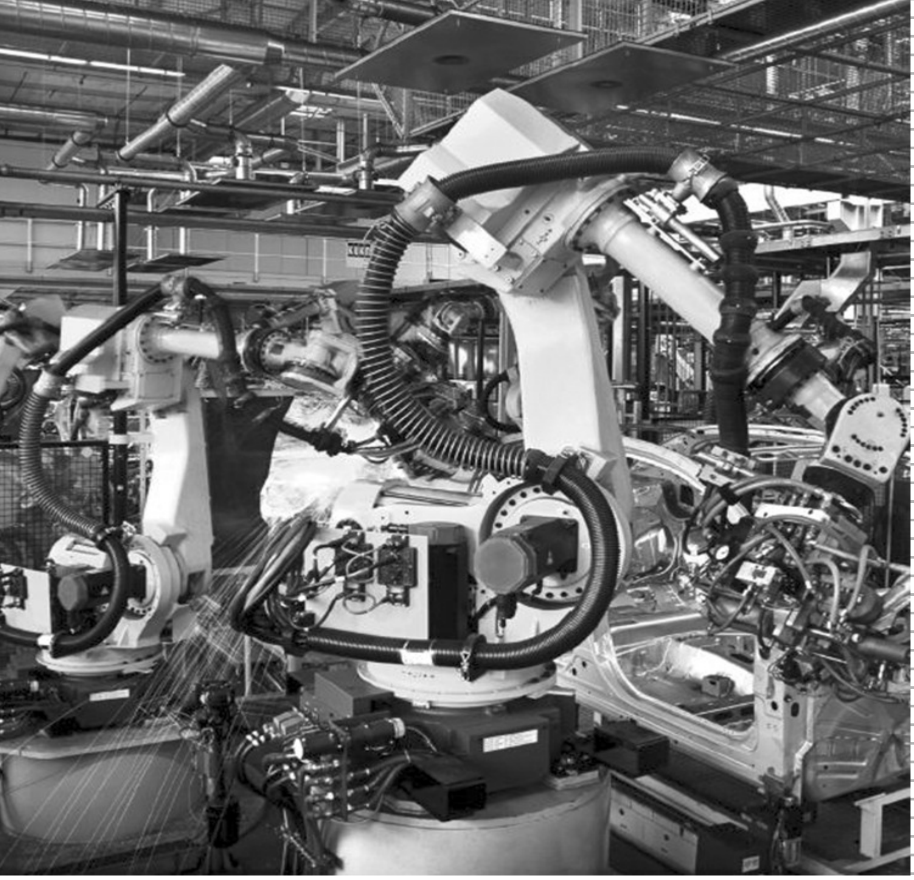
$$\mathbf{R}_2 = \mathbf{C}_3\mathbf{W}_2 = \mathbf{C}_3\mathbf{C}_2\mathbf{B}_2$$
$$= \begin{bmatrix} 15.22721682 & -4.515275007 & -0.04036944359 \\ 0 & 6.298390090 & 0.2550432812 \\ 0 & 0 & 2.001940393 \end{bmatrix}$$

and, finally,

$$\begin{aligned}\mathbf{B}_3 &= \mathbf{R}_2\mathbf{C}_2^{\mathsf{T}}\mathbf{C}_3^{\mathsf{T}} = \mathbf{C}_3\mathbf{C}_2\mathbf{B}_2\mathbf{C}_2^{\mathsf{T}}\mathbf{C}_3^{\mathsf{T}} \\ &= \begin{bmatrix} 15.82987970 & -1.293204857 & 0 \\ -1.293204856 & 6.169155576 & 0.06249374942 \\ 0 & 0.06249374864 & 2.000964734 \end{bmatrix}.\end{aligned}$$

This is a substantial improvement over the result of step 2.

Further steps would show convergence to 16, 6, 2, with roundoff errors in the last digits. Rounding effects are also shown in small deviations of $\mathbf{B}_2$ and $\mathbf{B}_3$ from symmetry. Note that, for simplicity in displaying the process, some very small numbers were set equal to zero.



PART F

Optimization, Graphs

The purpose of Part F is to introduce the main ideas and methods of unconstrained and constrained optimization (Chap. 22) and graphs and combinatorial optimization (Chap. 23). These topics of discrete mathematics are particularly well suited for modeling large-scale real-world problems and have many applications as described on p. 949 of the textbook.

Chap. 22 Unconstrained Optimization. Linear Programming

Optimization is concerned with problems and solution techniques on how to “best” (optimally) allocate limited resources in projects. Optimization methods can be applied to a wide variety of problems such as efficiently running power plants, easing traffic congestions, making optimal production plans, and others. Its methods are also applied to the latest fields of green logistics and green manufacturing.

Chapter 22 deals with two main topics: unconstrained optimization (Sec. 22.1) and a particular type of constrained optimization, that is, linear programming (Secs. 22.2–22.4). We show how to solve linear programming problems by the important **simplex method** in Secs. 22.3 (pp. 958–962) and 22.4 (pp. 962–968).

Some prior knowledge of *augmented matrix*, *pivoting,* and *row operation*—concepts that occur in the Gauss elimination method in Sec. 7.3, pp. 272–282—would be helpful, since the simplex method uses these concepts. (However, the simplex method *is different* from the Gauss elimination method.)

Sec. 22.1 Basic Concepts. Unconstrained Optimization: Method of Steepest Descent

The purpose of this section is twofold. First, we learn about what an optimization problem is (p. 951) and, second, what unconstrained optimization is (pp. 951–952), which we illustrate by the method of steepest descent.

In an **optimization problem** we want to optimize, that is, **maximize** or **minimize** some function f. This function f is called the **objective function** and consists of several variables

$$x_1, x_2, x_3, \ldots, x_n,$$

whose values we can choose, that is, *control*. Hence these variables are called *control variables*. This idea of "control" can be immediately understood if we think of an application such as the yield of a chemical process that depends on pressure x_1 and temperature x_2.

In most optimization problems, the control variables are restricted, that is, they are subject to some *constraints,* as shall be illustrated in Secs. 22.2–22.4.

However, certain types of optimization problems have no restrictions and thus fall into the category of **unconstrained optimization.** The theoretical details of such problems are explained on the bottom third of p. 951 and continued on p. 952. Within unconstrained optimization the textbook selected a particular way of solving such problems, that is, the **method of steepest descent** or **gradient method**. It is illustrated in **Example 1,** pp. 952–953 and in great details in **Prob. 3**.

Problem Set 22.1. Page 953

3. Cauchy's method of steepest descent. We are given the function

$$\text{(A)} \qquad f(\mathbf{x}) = 2x_1^2 + x_2^2 - 4x_1 + 4x_2$$

with the starting value (expressed as a column vector) $\mathbf{x}_0 = [0 \quad 0]^\mathsf{T}$. We proceed as in Example 1, p. 952, beginning with the general formulas and using the starting value later. To simplify notations, let us denote the components of the gradient of f by f_1 and f_2. Then, the gradient of f is [see also (1), p. 396]

$$\nabla f(\mathbf{x}) = [f_1 \quad f_2]^\mathsf{T} = [4x_1 - 4 \quad 2x_2 + 4]^\mathsf{T}.$$

In terms of components,

$$\text{(B)} \qquad f_1 = 4x_1 - 4, \qquad f_2 = 2x_2 + 4.$$

Furthermore,

$$\mathbf{z}(t) = [z_1 \quad z_2]^\mathsf{T} = \mathbf{x} - t\nabla f(\mathbf{x}) = [x_1 - tf_1 \quad x_2 - tf_2]^\mathsf{T},$$

which, in terms of components, is

$$\text{(C)} \qquad z_1(t) = x_1 - tf_1, \qquad z_2(t) = x_2 - tf_2.$$

Now obtain $g(t) = f(\mathbf{z}(t))$ from $f(\mathbf{x})$ in (A) by replacing x_1 with z_1 and x_2 with z_2. This gives

$$g(t) = 2z_1^2 + z_2^2 - 4z_1 + 4z_2.$$

We calculate the derivative of $g(t)$ with respect to t, obtaining

$$g'(t) = 4z_1z_1' + 2z_2z_2' - 4z_1' + 4z_2'.$$

From (C) we see that $z_1' = -f_1$ and $z_2' = -f_2$ with respect to t. We substitute this and z_1 and z_2 from (C) into $g'(t)$ and obtain

$$g'(t) = 4(x_1 - tf_1)(-f_1) + 2(x_2 - tf_2)(-f_2) + 4f_1 - 4f_2.$$

Order the terms as follows: Collect the terms containing t and denote their sum by D (suggesting "denominator" in what follows). This gives

$$\text{(D)} \qquad tD = t(4f_1^2 + 2f_2^2).$$

We denote the sum of the other terms by N (suggesting "numerator") and get

$$\text{(E)} \qquad N = -4x_1 f_1 - 2x_2 f_2 + 4f_1 - 4f_2.$$

With these notations we have $g'(t) = \ tD + N$. Solving $g'(t) = 0$ for t gives

$$t = -\frac{N}{D}.$$

Next we start the iteration process.

Step 1. For the given $\mathbf{x} = \mathbf{x}_0 = [0 \quad 0]^\mathsf{T}$ we have $x_1 = 0$, $x_2 = 0$ and from (B)

$$\begin{aligned} f_1 &= 4 \cdot 0 - 4 = -4, \qquad f_2 = 2 \cdot 0 + 4 = 4, \\ tD &= t\,(4 \cdot (-4)^2 + 2 \cdot 4^2) = 96t \\ N &= \ -4 \cdot 0 \cdot (-6) - 2 \cdot 0 \cdot 4 + 4 \cdot (-4) - 4 \cdot 4 \\ &= \ -16 - 16 = -32 \end{aligned}$$

so that

$$t = t_0 = -\frac{N}{D} = -\frac{-32}{96} = \frac{1}{3} = 0.3333333.$$

From this and (B) and (C) we obtain the next approximation $\mathbf{x}_1$ of the desired solution in the form

$$\begin{aligned} \mathbf{x}_1 = \mathbf{z}(t_0) &= [0 - t_0(-4) \quad 0 - t_0 \cdot 4]^\mathsf{T} = [4t_0 \quad -4t_0]^\mathsf{T} = \left[4 \cdot \tfrac{1}{3} \quad -4 \cdot \tfrac{1}{3}\right]^\mathsf{T} \\ &= \left[\tfrac{4}{3} \quad -\tfrac{4}{3}\right]^\mathsf{T} = [1.3333333 \quad -1.3333333]^\mathsf{T}. \end{aligned}$$

Also from (A) we find that $f(\mathbf{x}_1)$ is

$$\begin{aligned} f(\mathbf{x}_1) &= 2\left(\frac{4}{3}\right)^2 + \left(-\frac{4}{3}\right)^2 - 4\left(\frac{4}{3}\right) + 4\left(-\frac{4}{3}\right) \\ &= \frac{32+16}{9} + \frac{-16-16}{3} = -\frac{16}{3} = -5.333333. \end{aligned}$$

This completes the first step.

Step 2. Instead of $\mathbf{x}_0$ we now use $\mathbf{x}_1$, which is, in terms of components,

$$x_1 = \tfrac{4}{3}, \qquad x_2 = -\tfrac{4}{3}.$$

Then from (B) we get

$$
\begin{aligned}
f_1 &= 4 \cdot \frac{4}{3} - 4 = \frac{16}{3} - \frac{12}{3} = \frac{4}{3}, \\
f_2 &= 2 \cdot \left(-\frac{4}{3}\right) + 4 = -\frac{8}{3} + \frac{12}{3} = \frac{4}{3}, \\
tD &= t\left(4 \cdot \left(\frac{4}{3}\right)^2 + 2 \cdot \left(\frac{4}{3}\right)^2\right) = t\left(\frac{64}{9} + \frac{32}{9}\right) = t\,\frac{96}{9}, \\
N &= -4 \cdot \frac{4}{3} \cdot \frac{4}{3} - 2 \cdot \left(-\frac{4}{3}\right) \cdot \frac{4}{3} + 4 \cdot \frac{4}{3} - 4 \cdot \frac{4}{3} \\
&= \frac{-64 + 32 + 0}{9} = -\frac{32}{9},
\end{aligned}
$$

so that

$$
t = t_1 = -\frac{N}{D} = -\frac{-\frac{32}{9}}{\frac{96}{9}} = \frac{32}{9} \cdot \frac{9}{96} = \frac{1}{3} = 0.3333333.
$$

From this and (B) and (C) we obtain the next approximation $\mathbf{x}_2$ of the desired solution in the form

$$
\begin{aligned}
\mathbf{x}_2 = \mathbf{z}(t_1) &= [x_1 - t_1 f_1, \qquad x_2 - t_1 f_2]^{\mathsf{T}} \\
&= \begin{bmatrix} \frac{4}{3} - \frac{1}{3} \cdot \frac{4}{3} & \quad -\frac{4}{3} - \frac{1}{3} \cdot 4 \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} \frac{12 - 4}{9} & \quad \frac{-12 - 4}{9} \end{bmatrix}^{\mathsf{T}} \\
&= \begin{bmatrix} \frac{8}{9} & \quad -\frac{16}{9} \end{bmatrix}^{\mathsf{T}} = [0.8888889 \qquad -1.777778]^{\mathsf{T}}
\end{aligned}
$$

Also from (A) we find that $f(\mathbf{x}_2)$ is

$$
\begin{aligned}
f(\mathbf{x}_2) &= 2\left(\frac{8}{9}\right)^2 + \left(-\frac{16}{9}\right)^2 - 4\left(\frac{8}{9}\right) + 4\left(-\frac{16}{9}\right) \\
&= \frac{128 + 256}{81} + \frac{-32 - 64}{9} = \frac{384 - 864}{81} = -\frac{480}{81} = -\frac{160}{27} = -5.925926.
\end{aligned}
$$

This completes the second step.

Step 3. Instead of $\mathbf{x}_1$ we now use $\mathbf{x}_2$, which is, in terms of components,

$$
x_1 = \frac{8}{9}, \qquad x_2 = -\frac{16}{9}.
$$

Then from (B) we get

$$
\begin{aligned}
f_1 &= 4 \cdot \frac{8}{9} - 4 = \frac{32}{9} - \frac{36}{9} = -\frac{4}{9}, \\
f_2 &= 2 \cdot \left(-\frac{16}{9}\right) + 4 = -\frac{32}{9} + \frac{36}{9} = \frac{4}{9}, \\
\text{(1)} \qquad tD &= t\left(4 \cdot \left(-\frac{4}{9}\right)^2 + 2 \cdot \left(\frac{4}{9}\right)^2\right) = t\left(4 \cdot \frac{16}{81} + 2 \cdot \frac{16}{81}\right) = t\,\frac{64+32}{81} = t \cdot \frac{96}{81}, \\
N &= -4 \cdot \frac{8}{9} \cdot \left(-\frac{4}{9}\right) - 2 \cdot \left(-\frac{16}{9}\right) \cdot \frac{4}{9} + 4 \cdot \left(-\frac{4}{9}\right) - 4 \cdot \frac{4}{9} \\
&= \frac{128 + 128 - 144 - 144}{81} = -\frac{32}{81},
\end{aligned}
$$

so that

$$
t = t_2 = -\frac{N}{D} = -\frac{-\frac{32}{81}}{\frac{96}{81}} = \frac{32}{81} \cdot \frac{81}{96} = \frac{1}{3} = 0.3333333.
$$

From this and (B) and (C) we obtain the next approximation $\mathbf{x}_3$ of the desired solution in the form

$$
\begin{aligned}
\mathbf{x}_3 = \mathbf{z}(t_2) &= [x_1 - t_2 f_1, \qquad x_2 - t_2 f_2]^{\mathsf{T}} \\
&= \left[\frac{8}{9} - \frac{1}{3} \cdot \left(-\frac{4}{9}\right) \qquad -\frac{16}{9} - \frac{1}{3} \cdot \frac{4}{9}\right]^{\mathsf{T}} = \left[\frac{24+4}{27} \qquad \frac{-48-4}{27}\right]^{\mathsf{T}} \\
&= \left[\frac{28}{27} \qquad -\frac{52}{27}\right]^{\mathsf{T}} = [1.037037 \qquad -1.925926]^{\mathsf{T}}.
\end{aligned}
$$

From (A) we find that $f(\mathbf{x}_3)$ is

$$
\begin{aligned}
f(\mathbf{x}_3) &= 2\left(\frac{28}{27}\right)^2 + \left(-\frac{52}{27}\right)^2 - 4\left(\frac{28}{27}\right) + 4\left(-\frac{52}{27}\right) \\
&= \frac{2 \cdot 28^2 + 52^2 - 4 \cdot 27 \cdot 28 - 4 \cdot 27 \cdot 52}{27^2} \\
&= \frac{1568 + 2704 - 3024 - 5616}{729} \\
&= \frac{-4368}{729} = -5.991770.
\end{aligned}
$$

This completes the third step.

The results for the first seven steps, with six significant digit accuracy, are as follows.

Discussion. Table I gives a more accurate answer in more steps than is required by the problem. Table II gives the same answer—this time as fractions—thereby ensuring total accuracy. With the help of your computer algebra system (CAS) or calculator, you can readily convert the fractions of Table II to the desired number of decimals of your final answer and check your result. Thus any variation in your answer from the given answer due to rounding errors or technology used can be

Sec. 22.1. Prob. 3. Table I. *Method of steepest descent. Seven steps with 6S accuracy and one guarding digit*

n	$\mathbf{x}$		f
0	0.000000	0.000000	0.000000
1	1.333333	−1.333333	−5.333333
2	0.8888889	−1.777778	−5.925925
3	1.0370370	−1.925926	−5.991770
4	0.9876543	−1.975309	−5.999056
5	1.004115	−1.991769	−5.999894
6	0.9986283	−1.997256	−5.999998
7	1.000457	−1.999086	−5.999999

Sec. 22.1. Prob. 3. Table II. *Method of steepest descent. Seven steps expressed as fractions to ensure complete accuracy*

n	$\mathbf{x}$		f
0	0	0	0
1	$\frac{4}{3}$	$-\frac{16}{9}$	$-\frac{48}{9}$
2	$\frac{8}{9}$	$-\frac{16}{9}$	$-\frac{480}{81}$
3	$\frac{28}{27}$	$-\frac{52}{27}$	$-\frac{4,368}{729}$
4	$\frac{80}{81}$	$-\frac{160}{81}$	$-\frac{39,360}{6,561}$
5	$\frac{244}{243}$	$-\frac{484}{243}$	$-\frac{354.288}{59,049}$
6	$\frac{728}{729}$	$-\frac{1456}{729}$	$-\frac{3,188,640}{531,441}$
7	$\frac{2,188}{2,187}$	$-\frac{4,372}{2,187}$	$-\frac{28,697,808}{4,782,969}$

checked with Tables I and II. Furthermore, the last column in each table shows that the values of f converge toward a minimum value of approximately minus 6. We can readily see this and other information from the given function (A) by *completing the square,* as follows.

Recall that, for a quadratic equation,

$$ax^2 + bx + c = 0$$

completing the square amounts to writing the equation in the form

$$a(x+d)^2+e=0 \qquad \text{where} \qquad d=\frac{b}{2a} \qquad \text{and} \qquad e=c-\frac{b^2}{4a}.$$

We apply to our given function f this method twice, that is, first to the x_1-terms $2x_1^2 - 4x_1$, and then to the x_2-terms ${x_2}^2 + 4x_2$. For the x_1-terms we note that $a = 2$, $b = -4$, $c = 0$ so that

$$d=\frac{b}{2a}=\frac{-4}{2\cdot 2}=-1 \qquad \text{and} \qquad e=c-\frac{b^2}{4a}=0-\frac{16}{8}=-2.$$

This gives us

$$\text{(F)} \qquad {2x_1}^2-4x_1=2\cdot(x_1-1)^2-2.$$

Using the same approach yields

$$\text{(G)} \qquad {x_2}^2+4x_2=1\cdot(x_1+2)^2-4.$$

Adding (F) and (G) together, we see that by completing the square, f can be written as

$$\text{(H)} \qquad f(\mathbf{x})=2\cdot(x_1-1)^2+1\cdot(x_2+2)^2-6.$$

Equation (H) explains the numeric results. It shows that $f(\mathbf{x}) = -6$ occurs at $x_1 = 1$ and $x_2 = -2$, which is in reasonably good agreement with the corresponding entries for $n = 7$ in the tables. Furthermore, we see, geometrically, that the level curves $f = \text{const}$ are ellipses with principal axes in the directions of the coordinate axes (the function has no term x_1x_2) and semiaxes of length proportional to $\sqrt{2}$ and $\sqrt{1}$.

Remark. Your answer requires only three steps. We give seven steps for a better illustration of the method. Also note that in our calculation we used fractions, thereby maintaining higher accuracy, and converted these fractions into decimals only when needed.

Sec. 22.2 Linear Programming

The remaining sections of this chapter deal with *constrained* optimization which differs from *unconstrained* optimization in that, in addition to the objective function, there are also some constraints. We are only considering problems that have a *linear* objective function and whose constraints are *linear*. Methods that solve such problems are called **linear programming** (or linear optimization, p. 954). A typical example is as follows.

Consider a linear objective function, such as

$$z=f(\mathbf{x})=40x_1+88x_2,$$

subject to some constraints, consisting of linear inequalities, such as

$$\text{(1)} \qquad 2x_1+8x_2\le 60$$

$$\text{(2)} \qquad 5x_1+2x_2\le 60$$

with the usual additional constraints on the variables $x_1 \ge 0$, $x_2 \ge 0$, as given in **Example 1**, p. 954, where the goal is to find maximum $\mathbf{x} = (x_1, x_2)$ to maximize revenue z in the objective function.

The inequality (1) can be converted into an equality by introducing a variable x_3 (where $x_3 \geq 0$), thus obtaining

$$2x_1 + 8x_2 + x_3 = 60.$$

The variable x_3 has taken up the slack or difference between the two sides of the inequality. Thus x_3 is called a **slack variable** (see p. 956). We also introduce a slack variable x_4 for equation (2) as shown in Example 2, p. 956. This leads to the **normal form** of a linear optimization problem. This is an important concept because any problem has to be first converted to a normal form before a systematic method of solution (as shown in the next section) can be applied.

Problems 3, 21, and Fig. 474 of Example 1 on p. 955 explore the geometric aspects of linear programming problems.

Problem Set 22.2. Page 957

3. Region, constraints. Perhaps the easiest way to do this problem is to denote x_1 by x and x_2 by y. Then our axes are labeled in a more familiar way and we can rewrite the problem as

$$\text{(A}'\text{)} \qquad -0.5x + \ y \leq 2,$$
$$\text{(B}'\text{)} \qquad x + \ y \geq 2,$$
$$\text{(C}'\text{)} \qquad -x + 5y \geq 5.$$

Consider inequality (A′). This is also equivalent to

$$\text{(A}''\text{)} \qquad y \leq 0.5x + 2.$$

Now, if we consider the corresponding equality,

$$y = 0.5x + 2,$$

we get line ① in Fig. A. Since (A″) is an inequality of the kind $\leq$, the region determined by (A″) and hence (A′) must lie below ①. We shade this in Fig. A.

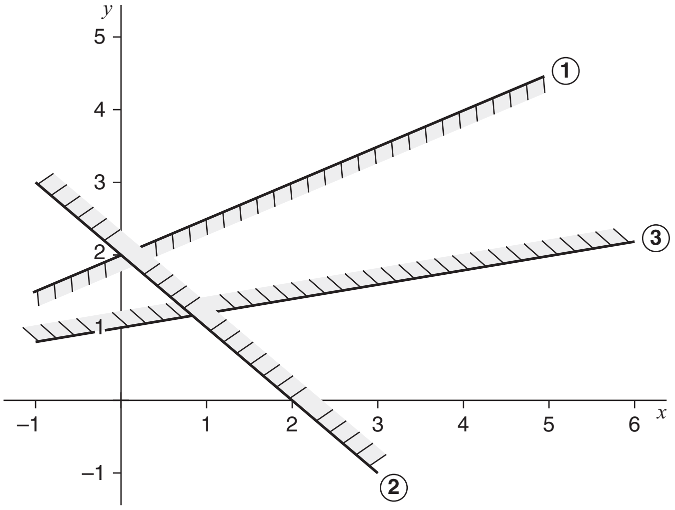
The same reasoning applies to (B′).

$$\text{(B}'\text{)} \implies \text{(B}''\text{)} \qquad y \geq -x + 2.$$

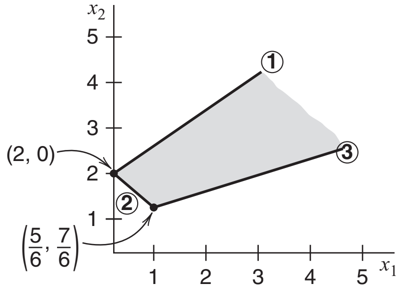
We consider $y = -x + 2$ and get line ② in Fig. A. Since B″ is an inequality $\geq$, we have that (A″) and (B′) lie above line ② as shaded.

Also (C′) $\implies$ (C″) $y \geq \frac{1}{5}x + 1$, which, as an equality, gives line ③ in Fig. A. Since we have $\geq$, the corresponding shaded region lies above line ③ as shaded in Fig. A.

Taking (A″), (B″), (C″) together gives the intersection of all three shaded regions. This is precisely the region below ①, to the right of ②, and above ③. It extends from $(0, 2)$ below ①, from $(0, 2)$ to $\left(\frac{5}{6}, \frac{7}{6}\right)$ above ②, and from $(1, 1.2)$ above ③. Together we have the infinite region with boundaries as marked in Fig. B, with the notation x_1 (for x) and x_2 (for y). Note that the region lies entirely in the first quadrant of the x_1x_2-plane, so that the conditions $x_1 \geq 0$, $x_2 \geq 0$ (often imposed by the kind of application, for instance, number of items produced, time or quantity of raw material needed, etc.) are automatically satisfied.



Sec. 22.2 Prob. 3. Fig. A Graphical development of solution



Sec. 22.2 Prob. 3. Fig. B Final solution: region determined by the three inequalities given in the problem statement

7. **Location of maximum.** Consider what happens as we move the straight line

$$z = c = \text{const},$$

beginning its position when $c = 0$ (which is shown in Fig. 474, p. 955) and increase c continuously.

21. **Maximum profit.** The profit per lamp L_1 is \$150 and that per lamp L_2 is \$100. Hence the total profit for producing x_1 lamps L_1 and x_2 lamps L_2 is

$$f(x_1, x_2) = 150x_1 + 100x_2.$$

We want to determine x_1 and x_2 such that the profit $f(x_1, x_2)$ is as large as possible.

Limitations arise due to the available workforce. For the sake of simplicity the problem is talking about two workers W_1 and W_2, but it is clear how the corresponding constraints could be made into a larger problem if teams of workers were involved or if additional constraints arose from raw material. The assumption is that, for this kind of high-quality work, W_1 is available 100 hours per month and that he or she assembles three lamps L_1 per hour or two lamps L_2 per hour. Hence W_1 needs $\frac{1}{3}$ hour for assembling lamp L_2 and $\frac{1}{2}$ hour for assembling lamp L_2. For a production of x_1 lamps L_1 and x_2 lamps L_2, this gives the restriction (constraint)

$$\text{(A)} \qquad \tfrac{1}{3}x_1 + \tfrac{1}{2}x_2 \leq 100.$$

(As in other applications, it is essential to measure time or other physical quantities by the same units throughout a calculation.) (A) with equality sign gives a straight line that intersects the x_1-axis at 300 (put $x_2 = 0$) and the x_2-axis at 200 (put $x_1 = 0$) as seen in Fig. C. If we put both $x_1 = 0$ and $x_2 = 0$, the inequality becomes $0 + 0 \leq 100$, which is true. This means that the region to be determined extends from that straight line downward.

Worker W_2 paints the lamps, namely, 3 lamps L_1 per hour or 6 lamps L_2 per hour. Hence painting a lamp L_1 takes $\frac{1}{3}$ hour, and painting lamp L_2 takes $\frac{1}{6}$ hour. W_2 is available 80 hours per month. Hence if x_1 lamps L_1 and x_2 lamps L_2 are produced per month, his or her availability gives the constraint

$$\text{(B)} \qquad \tfrac{1}{3}x_1 + \tfrac{1}{6}x_2 \leq 80.$$

(B) with the equality sign gives a straight line that intersects the x_1-axis at 240 (put $x_2 = 0$) and the x_2-axis at 480 (put $x_1 = 0$); see Fig. C. If we put $x_1 = 0$ and $x_2 = 0$, the inequality (B) becomes $0 + 0 \leq 80$, which is true. Hence the region to be determined extends from that line downward. And the region must lie in the first quadrant because we must have $x_1 \geq 0$ and $x_2 \geq 0$.

The intersection of those two lines is at (210, 60). This gives the maximum profit

$$f(210, 60) = 150 \cdot 210 + 100 \cdot 60 = \$37{,}500.$$

Next we reason graphically that (210, 60) does give the maximum profit. The straight line

$$f = 37{,}500$$

(the middle of the three lines in the figure) is given by

$$x_2 = 375 - 1.5x_1.$$

And by varying c in the line

$$f = \text{const},$$

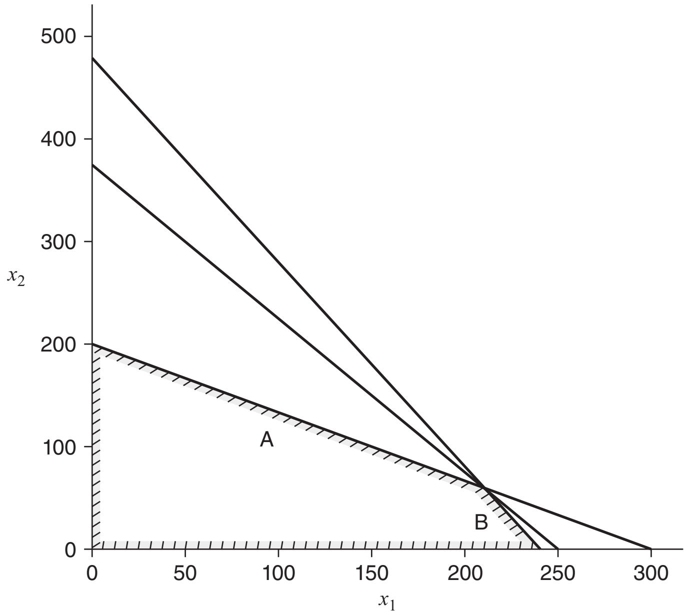
that is, in

$$x_2 = c - 1.5x_1,$$

which corresponds to moving the line up and down, it becomes clear that (210, 60) does give the maximum profit. We indicate the solution by a small circle in Fig. C.

Sec. 22.3 Simplex Method

This section forms the heart of Chap. 22 and explains the very important **simplex method**, which can briefly be described as follows. The given optimization problem has to be expressed in normal form (1), (2), p. 958, a concept explained in Sec. 22.2. Our discussion follows the example in the textbook which first appeared as Example 1, p. 954, and continued as Example 2, p. 956, both in Sec. 22.2. Now here, in Sec. 22.3, one constructs an augmented matrix as in (4), p. 959. Here z is the variable to be maximized, x_1, x_2 are the nonbasic variables, x_3, x_4 the basic variables, and b comes from the right-hand sides of the equalities of the equations of the constraints of the normal form. Basic variables are the slack variables and are characterized by the fact that their columns have only one nonzero entry (see p. 960).



Sec. 22.2 Prob. 21. Fig. C Constraints (A) (lower line) and (B)

From the initial simplex table, we select the column of the pivot by finding the first negative entry in Row 1. Then we want to find the row of the pivot, which we obtain by dividing the right-hand sides by the corresponding entries of the column just selected and take the smallest quotient. This will give us the desired pivot row. Finally use this pivot row to eliminate entries above and below the pivot, just like in the Gauss–Jordan method. This will lead to the second simplex table (5), p. 960. Repeat these steps until there are no more negative entries in the nonbasic variables, that is, the nonbasic variables become basic variables. We set the nonbasic variables to zero and read off the solution (p. 961).

Go over the details of this example with paper and pencil so that you get a firm grasp of this important method. The advantage of this method over a geometric approach is that it allows us to solve large problems in a systematic fashion.

Further detailed illustrations of the simplex method are given in **Prob. 3** (maximization) and **Prob. 7** (minimization).

Problem Set 22.3. Page 961

3. Maximization by the simplex method. The objective function to be maximized is

$$z = f(x_1, x_2) = 3x_1 + 2x_2. \tag{A}$$

The constraints are

$$\begin{aligned} 3x_1 + 4x_2 &\le \ \ 60, \\ 4x_1 + 3x_2 &\le \ \ 60, \\ 10x_1 + 2x_2 &\le 120. \end{aligned} \tag{B}$$

Begin by writing this in normal form, see (1) and (2), p. 958. The inequalities are converted to equations by introducing slack variables, one slack variable per inequality. In (A) and (B) we have the variables x_1 and x_2. Hence we denote the slack variables by x_3 [for the first inequality in (B)], x_4 [for the second inequality in (B)], and x_5 (for the third). This gives the normal form (with the objective function written as an equation)

$$\text{(C)}\qquad \begin{array}{rcl} z - 3x_1 - 2x_2 & = & 0, \\ 3x_1 + 4x_2 + x_3 \qquad\qquad & = & 60, \\ 4x_1 + 3x_2 \qquad + x_4 \qquad & = & 60, \\ 10x_1 + 2x_2 \qquad\qquad + x_5 & = & 120. \end{array}$$

This is a linear system of equations. The corresponding augmented matrix (a concept you should know!—see Sec. 7.3, p. 273) is called the *initial simplex table* and is denoted by $\mathbf{T}_0$. It is

$$\mathbf{T}_0 = \left[\begin{array}{c|cc|ccc|c} z & x_1 & x_2 & x_3 & x_4 & x_5 & b \\ \hline 1 & -3 & -2 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 3 & 4 & 1 & 0 & 0 & 60 \\ 0 & 4 & 3 & 0 & 1 & 0 & 60 \\ 0 & 10 & 2 & 0 & 0 & 1 & 120 \end{array}\right]$$

Take a look at (3) on p. 963, which has an extra line on top showing z, the variables, and b [denoting the terms on the right side in (C)]. We also added such a line in (D) and also drew the dashed lines, which separate the first row of $\mathbf{T}_0$ from the others as well as the columns corresponding to z, to the given variables, to the slack variables, and to the right sides.

Perform Operation O_1. The first column with a negative entry in Row 1 is Column 2, the entry being -3. This is the column of the first pivot. Perform Operation O_2. We divide the right sides by the corresponding entries of the column just selected. This gives

$$\tfrac{60}{3} = 20, \qquad \tfrac{60}{4} = 15, \qquad \tfrac{120}{10} = 12.$$

The smallest positive of these quotients is 12. It corresponds to Row 4. Hence select Row 4 as the row of the pivot. Perform Operation O_3, that is, create zeros in Column 2 by the row operations

$$\begin{array}{l} \text{Row 1} + \tfrac{3}{10}\,\text{Row 4}, \\ \text{Row 2} - \tfrac{3}{10}\,\text{Row 4}, \\ \text{Row 3} - \tfrac{4}{10}\,\text{Row 4}. \end{array}$$

This gives the new simplex table (with Row 4 as before), where we mark the row operations next to the augmented matrix with the understanding that these operations were applied to the prior augmented matrix $\mathbf{T}_0$;

$$\mathbf{T}_1 = \left[\begin{array}{c|cc|ccc|c} z & x_1 & x_2 & x_3 & x_4 & x_5 & b \\ \hline 1 & 0 & -\frac{7}{5} & 0 & 0 & \frac{3}{10} & 36 \\ \hdashline 0 & 0 & \frac{17}{5} & 1 & 0 & -\frac{3}{10} & 24 \\ 0 & 0 & \frac{11}{5} & 0 & 1 & -\frac{2}{5} & 12 \\ 0 & 10 & 2 & 0 & 0 & 1 & 120 \end{array}\right] \qquad \begin{array}{l} \text{Row 1} + \frac{3}{10}\,\text{Row 4} \\ \text{Row 2} - \frac{3}{10}\,\text{Row 4} \\ \text{Row 3} - \frac{4}{10}\,\text{Row 4}. \end{array}$$

This was the first step. (Note that the extra line on top of the augmented matrix showing z, the variables and b as well as the dashed lines is optional but is put in for better understanding.)

Now comes the second step, which is necessary because of the negative entry $-\frac{7}{5}$ in Row 1 of $\mathbf{T}_1$. Hence the column of the pivot is Column 3 of $\mathbf{T}_1$. We compute

$$\frac{24}{\frac{17}{5}} = \frac{120}{17} = 7.06, \qquad \frac{12}{\frac{11}{5}} = \frac{60}{11} = 5.45, \qquad \frac{120}{2} = 60$$

and compare. The second of these is the smallest. Hence the pivot row is Row 3. To create zeros in Column 3 we have to do the row operations

$$\text{Row } 1 + \frac{\frac{7}{5}}{\frac{11}{5}}\text{ Row } 3,$$

$$\text{Row } 2 - \frac{\frac{17}{5}}{\frac{11}{5}}\text{ Row } 3,$$

$$\text{Row } 4 - \frac{2}{\frac{11}{5}}\text{ Row } 3,$$

leaving Row 3 unchanged. This gives the simplex table

$$\mathbf{T}_2 = \left[\begin{array}{c|cc|ccc|c}
z & x_1 & x_2 & x_3 & x_4 & x_5 & b \\
1 & 0 & 0 & 0 & \frac{7}{11} & \frac{1}{22} & \frac{480}{11} \\ \hline
0 & 0 & 0 & 1 & -\frac{17}{11} & \frac{7}{22} & \frac{60}{11} \\
0 & 0 & \frac{11}{5} & 0 & 1 & -\frac{2}{5} & 12 \\
0 & 10 & 0 & 0 & -\frac{10}{11} & \frac{15}{11} & \frac{1200}{11}
\end{array}\right]
\quad
\begin{array}{l}
\\
\text{Row } 1 + \frac{7}{11}\text{ Row } 3 \\
\text{Row } 2 - \frac{17}{11}\text{ Row } 3 \\
\\
\text{Row } 4 - \frac{10}{11}\text{ Row } 3
\end{array}$$

Since no more negative entries appear in Row 1, we are finished. From Row 1 we see that

$$f_{\max} = \tfrac{480}{11} = 43.64.$$

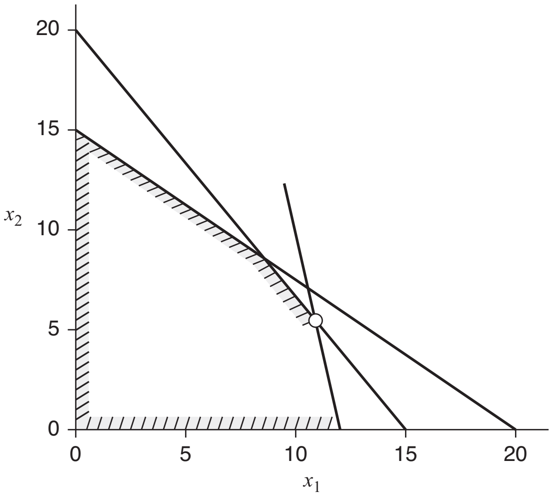
In Row 4 we divide the entry in Column 7 by the entry in Column 2 and obtain the corresponding

$$x_1 - \text{value } \frac{\frac{1200}{11}}{10} = \frac{1200}{11} \cdot \frac{1}{10} = \frac{120}{11}.$$

Similarly, in Row 3 we divide the entry in Column 7 by the entry in Column 3 and obtain the corresponding

$$x_2 - \text{value } \frac{12}{\frac{11}{5}} = \frac{12}{1} \cdot \frac{5}{11} = \frac{60}{11}.$$

You may want to convince yourself that the maximum is taken at one of the vertices of the polygon determined by the constraints. This vertex is marked by a small circle in Fig. D.



Sec. 22.3 Prob. 3. Fig. D Region determined by the constraints

7. **Minimization by the simplex method.** The given problem, in normal form [with $z = f(x_1, x_2)$ written as an equation], is

$$\begin{aligned} z - 5x_1 + 20x_2 \qquad\qquad &= 0, \\ -2x_1 + 10x_2 + x_3 \qquad &= 5, \\ 2x_1 + \ \ 5x_2 \qquad\quad + x_4 &= 10. \end{aligned}$$

From this we see that the initial simplex table is

$$\mathbf{T}_0 = \left[\begin{array}{c|cc|cc|c} z & x_1 & x_2 & x_3 & x_4 & b \\ \hline 1 & -5 & 20 & 0 & 0 & 0 \\ \hline 0 & -2 & 10 & 1 & 0 & 5 \\ 0 & 2 & 5 & 0 & 1 & 10 \end{array}\right]$$

Since we minimize (instead of maximizing), we consider the columns whose first entry is *positive* (instead of negative). There is only one such column, namely, Column 3. The quotients are

$$\tfrac{5}{10} = \tfrac{1}{2} \ \text{(from Row 2)} \quad \text{and} \quad \tfrac{10}{5} = 2 \ \text{(from Row 3)}.$$

The smaller of these is $\frac{1}{2}$. Hence we have to choose Row 2 as pivot row and 10 as the pivot. We create zeros by the row operations Row 1 − 2 Row 2 (this gives the new Row 1) and Row 3 − $\frac{1}{2}$ Row 2 (this gives the new Row 3), leaving Row 2 unchanged. The result is

$$\mathbf{T}_1 = \left[\begin{array}{c|cc|cc|c} z & x_1 & x_2 & x_3 & x_4 & b \\ \hline 1 & -1 & 0 & -2 & 0 & -10 \\ \hline 0 & -2 & 10 & 1 & 0 & 5 \\ 0 & 3 & 0 & -1/2 & 1 & 15/2 \end{array}\right] \begin{array}{l} \\ \text{Row 1} - 2 \text{ Row 2} \\ \\ \text{Row 3} - \tfrac{1}{2} \text{ Row 2} \end{array}$$

Since there are no further positive entries in the first row, we are done. From Row 1 of $\mathbf{T}_1$ we see that

$$f_{\min} = -10.$$

From Row 2, with Columns 3 and 6, we see that

$$x_2 = \tfrac{5}{10} = \tfrac{1}{2}.$$

Furthermore, from Row 3, with Columns 5 and 6, we obtain

$$x_4 = \frac{\frac{15}{2}}{1} = \frac{15}{2}.$$

Now x_4 appears in the second constraint, written as equation, that is,

$$2x_1 + 5x_2 + x_4 = 10.$$

Inserting $x_2 = \frac{1}{2}$ and $x_4 = \frac{15}{2}$ gives

$$2x_1 + 10 = 10, \qquad \text{hence} \qquad x_1 = 0.$$

Hence

$$\text{the minimum } -10 \text{ of } z = f(x_1, x_2) \text{ occurs at the point } \left(0, \tfrac{1}{2}\right).$$

Since this problem involves only two variables (not counting the slack variables), as a control and to better understand the problem, you may want to graph the constraints. You will notice that they determine a quadrangle. When you calculate the values of f at the four vertices of the quadrangle, you should obtain

$$0 \text{ at } (0, 0),\ 25 \text{ at } (5, 0),\ -7.5 \text{ at } (2.5, 1)\,, \text{ and } -10 \text{ at } \left(0, \tfrac{1}{2}\right).$$

This would confirm our result.

Sec. 22.4 Simplex Method. Difficulties

Of lesser importance are two types of difficulties that are encountered with the simplex method: *degeneracy*, illustrated in Example 1 (pp. 962–965), Problem 1 and *difficulties in starting*, illustrated in Example 2 (pp. 965–967).

Problem Set 22.4. Page 968

1. Degeneracy. Choice of pivot. Undefined quotient. The given problem is

$$z = f_1(\mathbf{x}) = 7x_1 + 14x_2$$

subject to

$$0 \le x_1 \le 6,$$
$$0 \le x_2 \le 3,$$
$$7x_1 + 14x_2 \le 84.$$

Its normal form [with $z = f(x_1, x_2)$ written as an equation] is

$$\begin{array}{rcl} z - 7x_1 - 14x_2 & = & 0, \\ x_1 \qquad + x_3 \qquad\quad & = & 6, \\ x_2 \qquad + x_4 \qquad & = & 3, \\ 7x_1 + 14x_2 \qquad\qquad + x_5 & = & 84. \end{array}$$

From this we see that the initial simplex table is

$$\mathbf{T}_0 = \left[\begin{array}{c|cc|ccc|c} z & x_1 & x_2 & x_3 & x_4 & x_5 & b \\ \hline 1 & -7 & -14 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 1 & 0 & 0 & 6 \\ 0 & 0 & 1 & 0 & 1 & 0 & 3 \\ 0 & 7 & 14 & 0 & 0 & 1 & 84 \end{array}\right]$$

The first pivot must be in Column 2 because of the entry -7 in this column. We determine the row of the first pivot by calculating

$\frac{6}{1} = 6$ (from Row 2)

ratio undefined (we cannot divide 3 by 0) (from Row 3)

$\frac{7}{1} = 7$ (from Row 4).

Since 6 is smallest, Row 2 is the pivot row. With this the next simplex table becomes

$$\mathbf{T}_1 = \left[\begin{array}{c|cc|ccc|c} z & x_1 & x_2 & x_3 & x_4 & x_5 & b \\ \hline 1 & 0 & -14 & 7 & 0 & 0 & 42 \\ \hline 0 & 1 & 0 & 1 & 0 & 0 & 6 \\ 0 & 0 & 1 & 0 & 1 & 0 & 3 \\ 0 & 0 & 14 & -7 & 0 & 1 & 42 \end{array}\right] \quad \begin{array}{l} \\ \text{Row 1} + 7\text{Row 2} \\ \\ \text{Row 3} \\ \text{Row 4} - 7\text{Row 2} \end{array}$$

We have reached a point at which $z = 42$. To find the point, we calculate

$$\begin{aligned} x_1 &= 6 \qquad \text{(from Row 2 and Column 2)}, \\ x_4 &= 3 \qquad \text{(from Row 3 and Column 4)}. \end{aligned}$$

From this and the first constraint we obtain

$$x_2 + x_4 = x_2 + 3 = 3, \qquad \text{hence} \quad x_2 = 0.$$

(More simply: x_1, x_4, x_5 are basic. x_2, x_3 are nonbasic. Equating the latter to zero gives $x_2 = 0$, $x_3 = 0$.) Thus $z = 42$ at the point $(42, 0)$ on the x_1-axis.

Column 3 of $\mathbf{T}_1$ contains the negative entry -14. Hence this column is the column of the next pivot. To obtain the row of the pivot, we calculate

ratio undefined (we cannot divide 3 by 0) (from Row 2)

$\frac{3}{1} = 3$ (from Row 3),

$\frac{32}{14} = 3$ (from Row 4).

Since both ratios gave 3 we have a choice of using Row 3 or using Row 4 as a pivot. We pick Row 3 as a pivot. We obtain

$$\mathbf{T}_1 = \left[\begin{array}{c:cc:ccc:c} z & x_1 & x_2 & x_3 & x_4 & x_5 & b \\ \hline 1 & 0 & 0 & 7 & 14 & 0 & 84 \\ \hdashline 0 & 1 & 0 & 1 & 0 & 0 & 6 \\ 0 & 0 & 1 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & -7 & -14 & 1 & 0 \end{array}\right] \qquad \begin{array}{c} \\ \text{Row 1} + 7\,\text{Row 2} \\ \\ \text{Row 3} \\ \text{Row 4} - 7\,\text{Row 2} \end{array}$$

There are no more negative entries in Row 1. Hence we have reached the maximum $z_{\max} = 84$. We see that x_1, x_2, x_5 are basic, and x_3, x_4 are nonbasic variables. $z_{\max}$ occurs at $(6, 3)$ because $x_1 = 6$ (from Row 2 and Column 2) and $x_2 = 3$ (from Row 3 and Column 3). Point $(6, 3)$ corresponds to a degenerate solution because $x_5 = 0/1 = 0$ from Row 4 and Column 6, in addition to $x_3 = 0$ and $x_4 = 0$. Geometrically, this means that the straight line

$$7x_1 + 14x_2 + x_5 = 84$$

resulting from the third constraint, also passes through $(x_1, x_2) = (6, 3)$, with $x_5 = 0$ because

$$7 \cdot 6 + 14 \cdot 3 \; + 0 = 84.$$

Observation. In Example 1, p. 962, we reached a degenerate solution before we reached the maximum (the optimal solution), and, for this reason, we had to do an additional step, that is, Step 2, on p. 964. In contrast, in the present problem we reached the maximum when we reached a degenerate solution. Hence no additional work was necessary.

Chap. 23 Graphs. Combinatorial Optimization

The field of **combinatorial optimization** deals with problems that are *discrete* [in contrast to functions in vector calculus (Chaps. 9 and 10) which are continuous and differentiable] and whose solutions are often difficult to obtain due to an extremely large number of cases that underlie the solution. Indeed, the "combinatorial nature" of the field gives us difficulties because, even for relatively small n,

$$n! = 1 \cdot 2 \cdot 3 \cdots n \qquad \text{(for } n! \text{ read "} n \textit{ factorial,} \text{" see p. 1025 in Sec. 24.4 of the textbook)}$$

is very large. For example, convince yourself, that

$$10! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 = 24 \cdot 30 \cdot 56 \cdot 90 = 3,628,800.$$

We look for optimal or suboptimal solutions to discrete problems, with a typical example being the **traveling salesman problem** on p. 976 of the textbook (turn to that page and read the description). In that problem, even for 10 cities, there are already

$$\frac{10!}{2} = \frac{3,628,800}{2} = 1,814,251 \text{ possible routes.}$$

Logistics dictates that the salesman needs some software tool for identifying an optimal or suboptimal (but acceptable) route that he or she should take!

We start gently by discussing graphs and digraphs in Sec. 23.1, p. 970, as they are useful for modeling combinatorial problems. A **chapter orientation table** summarizes the content of Chap. 23.

Table of main topics for Chap. 23 on graphs and combinatorial optimization

Section	**Main topic**	**Algorithm**
Section 23.1, pp. 970–975	Introduction to graphs and digraphs	
Section 23.2, pp. 975–980	Shortest path problem	Moore, p. 977
Section 23.3, pp. 980–984	Shortest path problem	Dijkstra, p. 982
Section 23.4, pp. 984–988	Shortest spanning trees	Kruskal, p. 985
Section 23.5, pp. 988–991	Shortest spanning trees	Prim, p. 989
Section 23.6, pp. 991–997	Flow problems in networks	
Section 23.7, pp. 998–1001	Flow problems in networks	Ford–Fulkerson, p. 998
Section 23.8, pp. 1001–1006	Assignment problems	

Applications of this chapter abound in electrical engineering, civil engineering, computer science, operations research, industrial engineering, logistics, and others. Specifics include navigation systems for cars, computer network designs and assignment problems of jobs to machines (ships to piers, etc.), among others.

The material is intuitively appealing but requires that you remember the terminology (e.g., a point in a graph is called vertex, the connecting lines are called edges, etc.).

Sec. 23.1 Graphs and Digraphs

This section discusses important concepts that are used in this chapter. A **graph** G consists of points and the lines that connect these points, as shown in Fig. 477, p. 971. We call the points *vertices* and the connecting lines *edges*. This allows us to define the graph G as two finite sets, that is, $G = (V, E)$ where

V is a set of vertices and E a set of edges. Also, we do not allow isolated vertices, loops, and multiple edges, as shown in Fig. 478, p. 971.

If, in addition, each of the edges has a direction, then graph G is called a directed graph or **digraph** (p. 972 and Fig. 479).

Another concept is *degree of a vertex* (p. 971), which measures how many edges are *incident* with that vertex. For example, in Fig. 477, vertex 1 has degree 3 because there are three edges that are "involved with" (i.e., end or start at) that vertex. These edges are denoted by $e_1 = (1, 4)$ (connecting vertex 1 with vertex 4), $e_2 = (1, 2)$ (vertex 1 with 2), and $e_5 = (1, 3)$ (vertex 1 with 3). Continuing with our example, $e_1 = (1, 4)$ indicates that vertex 1 is *adjacent to* vertex 4. Also vertex 1 is adjacent to vertex 2 and vertex 3, respectively.

Whereas in a digraph we can only traverse in the direction of each edge, in a graph (being always undirected), we can travel each edge in both directions.

While it is visually indispensable to draw graphs when discussing specific applications (routes of airlines, networks of computers, organizational charts of companies, and others; see p. 971), when using computers, it is preferable to represent graphs and digraphs by *adjacency matrices* (Examples 1, 2, p. 973, **Prob. 11**) or *incidence lists* of vertices and edges (Example 3). These matrices contain only zeroes and ones. They indicate whether pairs of vertices are connected, if "yes" by a 1 and "no" by a 0. (Since loops are not allowed in graph G, the entries in the main diagonal of these matrices are always 0.)

Problem Set 23.1. Page 974

11. **Adjacency matrix. Digraph.** The four vertices of the figure are denoted 1, 2, 3, 4, and its four edges by e_1, e_2, e_3, e_4. We observe that each edge has a direction, indicated by an arrow head, which means that the given figure is a digraph. Edge e_1 goes from vertex 1 to vertex 2, edge e_2 goes from vertex 1 to vertex 3, and so on. There are two edges connecting vertices 1 and 3. They have opposite directions (e_2 goes from vertex 1 to vertex 3, and e_3 from vertex 3 to vertex 1, respectively).

Note that, in a graph, there cannot be two edges connecting the same pair of vertices.

An adjacency matrix has entries 1 and 0 and indicates whether any two vertices in the graph are connected by an edge. If "yes," the two edges are connected, then the corresponding entry is a "1," and if no a "0." For n vertices, such an indexing scheme requires a square, $n \times n$ matrix.

Our digraph has $n = 4$ vertices so that $\mathbf{A}$ is a 4×4 matrix. Its entry $a_{12} = 1$ because the digraph has an edge (namely, e_1) that goes from vertex 1 to vertex 2. Now comes an important point worth taking some time to think about: Entry a_{12} is the entry in Row 1 and Column 2. Since e_{12} goes *from* 1 *to* 2, by definition, the row number is the number of the vertex at which an edge *begins*, and the column number is the number of the vertex at which the edge *ends*. Think this over and look at the matrix in Example 2 on p. 973. Since there are three edges that begin at 1 and end at 2, 3, 4, and since there is no edge that begins at 1 and ends at 1 (no loop), the first row of $\mathbf{A}$ is

$$0 \quad 1 \quad 1 \quad 1.$$

Since the digraph has four edges, the matrix $\mathbf{A}$ must have four 1's, the three we have just listed and a fourth resulting from the edge that goes from 3 to 1. Obviously, this gives the entry $a_{31} = 1$. Continuing in this way we obtain the matrix

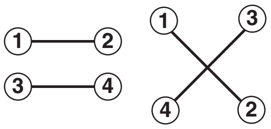
$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

which is the answer on p. A55 of the book. Note that the second and fourth row of $\mathbf{A}$ contains all zeroes since there are no directed edges that begin at vertex 2 and 4, respectively. In other words, there are no edges with initial points 2 and 4!

15. Deriving the graph for a given adjacency matrix. Since the given matrix, say **M** of the wanted *graph* G_M, is

$$\mathbf{M} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

which is a 4 × 4 matrix, the corresponding graph G_M has four vertices. Since the matrix has four 1's and each edge contributes two 1's, the graph G_M has two edges. Since $m_{12} = 1$, the graph has the edge $(1, 2)$; here we have numbered the four vertices by 1, 2, 3, 4, and 1 and 2 are the endpoints of this edge. Similarly, $m_{34} = 1$ implies that G_M has the edge $(3, 4)$ with endpoints 3 and 4. An adjacency matrix of a graph is always symmetric. Hence we must have $m_{21} = 1$ because $m_{12} = 1$, and similarly, $m_{43} = 1$ since $m_{34} = 1$. Differently formulated, the vertices 1 and 2 are adjacent, they are connected by an edge in G_M, namely, by $(1, 2)$. This results in $a_{12} = 1$ as well as $a_{21} = 1$. Similarly for $(3, 4)$. Together, this gives a graph that has two disjointed segments as shown below.



Sec. 23.1. Prob. 15. Graph G_M obtained from adjacency matrix **M**. Note that both sketches represent the same graph.

19. Incidence matrix $\widetilde{\mathbf{B}}$ of a digraph. The incidence matrix of a graph or digraph is an $n \times m$ matrix, where n is the number of vertices and m is the number of edges. Each row corresponds to one of the vertices and each column to one of the edges. Hence, in the case of a graph, each column contains two 1's. In the case of a digraph each column contains a 1 and a -1.

In this problem, we looked at the graph from **Prob. 11**. Since, for that graph, the number of vertices = number of edges = 4, the incidence matrix is square (which is not the most general case) and of dimension 4 × 4. The first column corresponds to edge e_1, which goes from vertex 1 to vertex 2. Hence by definition, $\tilde{b}_{11} = -1$ and $\tilde{b}_{21} = 1$. The second column corresponds to edge e_2, which goes from vertex 1 to vertex 3. Hence $\tilde{b}_{12} = -1$ and $\tilde{b}_{32} = 1$. Proceeding in this way we get

$$\widetilde{\mathbf{B}} = \begin{bmatrix} -1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Sec. 23.2 Shortest Path Problems. Complexity

We distinguish between walk, trail, path, and cycle as shown in Fig. 481, p. 976. A **path** requires that each vertex is visited at most once. A **cycle** is a path that ends at the same vertex from which it started. We also call such a path *closed*. Thus a cycle is a closed path.

A **weighted graph** $G = (V, E)$ is one in which each edge has a given weight or length that is positive. For example, in a graph that shows the routes of an airline, the vertices represent the cities, an edge between two cities shows that the airline flies directly between those two cities, and the weight of an edge indicates the (flight) distance in miles between such two cities.

A **shortest path** is a path such that the sum of the length of its edges is minimum; see p. 976. A shortest path problem means finding a shortest path in a weighted graph G. A *Hamiltonian cycle* (**Prob. 11**) is a cycle that contains *all* the vertices of a graph. An example of a shortest path problem is the *traveling salesman problem*; which requires the determination of a shortest Hamiltonian cycle. For more details on this important problem in combinatorial optimization, see the last paragraph on p. 976 or our opening discussion of this chapter.

Moore's BFS algorithm, p. 977 (with a backtracking rule in **Prob. 1**), is a systematic way for determining a shortest path in a connected graph, whose vertices all have length 1. The algorithm uses a **breadth first search (BFS)**, that is, at each step, the algorithm visits all neighboring (i.e., adjacent) vertices of a vertex reached. This is in contrast to a *depth first search* (*DFS*), which makes a long trail as in a maze.

Finally we discuss the **complexity of an algorithm** (see pp. 978–979) and the **order** *O,* suggesting "order." In this "big O" notation, an algorithm of complexity

$$am + b = O(m); \qquad am^2 + bm + d = O(m^2); \qquad a2^m + bm^2 + dm + k = O(2^m)$$

where a, b, d, and k are constant. This means that order O denotes the fastest growing term of the given expression. Indeed, for constant k

$$2^m >> m^2 >> m >> k \qquad \text{for large } m.$$

A more formal definition of O is given and used in **Prob. 19**. Note that, Moore's BFS algorithm is of complexity $O(m)$. (In the last equation the symbol ">>" means "*much greater than*.")

Problem Set 23.2. Page 979

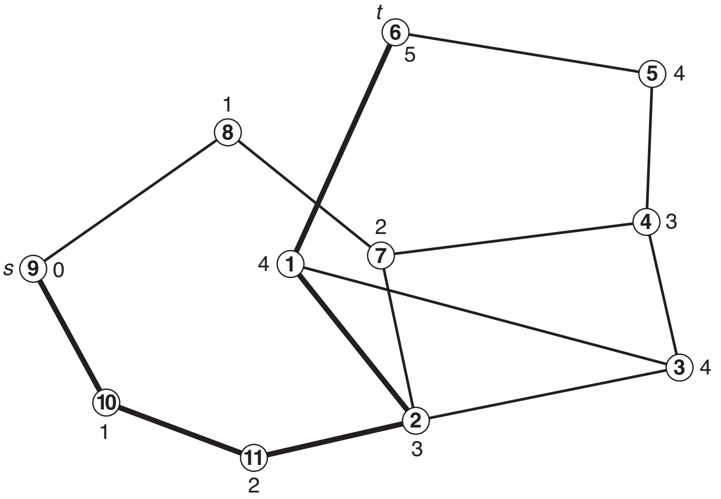
1. Shortest path. Moore's algorithm. We want to find the shortest path from s to t and its length, using Moore's algorithm (p. 977) and Example 1, p. 978. We numbered the vertices arbitrarily. This means we picked a vertex and numbered it ① and then numbered the other vertices consecutively ②, ③, ... We note that s (⑨) is a vertex that belongs to a hexagon (②, ⑦, ⑧, ⑨, ⑩, ⑪). According to step 1 in Moore's algorithm, s gets a label 0. s has two adjacent vertices (⑧ and ⑩), which get the label 1. Each of the latter has one adjacent vertex (⑦ and ⑪, respectively), which gets the label 2. These two vertices now labeled 2 are adjacent to the last still unlabeled vertex of the hexagon (②), which thus gets the label 3. This leaves five vertices still unlabeled (①, ③, ④, ⑤, ⑥). Two (①, ③) of these five vertices are adjacent to the vertex (②) labeled 3 and thus get the label 4. Vertex ①, labeled 4, is adjacent to the vertex t (⑥), which thus gets labeled 5, provided that there is no shorter way for reaching t.

There is no shorter way. We could reach t (⑥) from the right, but the other vertex adjacent to t, i.e., (⑤), gets the label 4 because the vertex (④) adjacent to it is labeled 3 since it is adjacent to a vertex of the hexagon (⑦) labeled 2. This gives the label 5 for t (⑥), as before.

Hence, by Moore's algorithm, the length of the shortest path from s to t is 5. The shortest path goes through nodes 0, 1, 2, 3, 4, 5, as shown in the diagram on the next page in heavier lines.

11. Hamiltonian cycle. For the definition of a Hamiltonian cycle, see our brief discussion before or turn to p. 976 of the textbook. Sketch the following Hamiltonian cycle (of the graph of **Prob. 1**), which we describe as follows. We start at s downward and take the next three vertices on the hexagon H, then the vertex outside H labeled 4 (③), then the vertex inside H, then t, then the vertex to the right of t (⑤), and then the vertex below it (3). Then we return to H, taking the remaining two vertices of H and return to s.

13. Postman problem. This problem models the typical behavior of a letter carrier. Naively stated, the postman starts at his post office, picks up his bags of mail, delivers the mail to all the houses, and comes back to the post office from which he/she started. (We assume that every house gets mail.)



Sec. 23.2. Prob. 1. Shortest path by Moore's algorithm

Thus the postman goes through all the streets "edges," visits each house "vertex" at least once, and returns to the vertex, which is the post office from where he/she came. Naturally, the postman wants to travel the shortest distance possible.

We solve the problem by inspection. In the present situation—with the post office s located at vertex 1—the postman can travel in four different ways:

First Route:	1—2—3—4—5—6—4—3—1
Second Route:	1—2—3—4—6—5—4—3—1
Third Route:	1—3—4—5—6—4—3—2—1
Fourth Route:	1—3—4—6—5—4—3—2—1

Each route contains 3—4 and 4—3, that is, vertices 3 and 4 are each traversed twice. The length of the first route is (with the brackets related to the different parts of the trail)

$$\begin{aligned}(l_{12} + l_{23}) + l_{34} + (l_{45} + l_{56} + l_{64}) + l_{43} + (l_{31})\\ &= (2 + 1) + 4 + (3 + 4 + 5) + 4 + (2)\\ &= 3 + 4 + 12 + 4 + 2 = 25,\end{aligned}$$

and so is that of all other three routes. Each route is optimal and represents a walk of minimum length 25.

19. Order. We can formalize the discussion of order O (pp. 978–979 in the textbook) as follows. We say that a function $g(m)$ is of the order $h(m)$, that is,

$$g(m) = O(h(m))$$

if we can find some positive constants m_0 and k such that

$$0 \leq g(m) \leq kh(m) \qquad \text{for all} \qquad m \geq m_0.$$

This means that, from a point m_0 onward, the curve of $kh(m)$ always lies above $g(m)$.

(a). To show that

$$\text{(O1)} \qquad \sqrt{1+m^2} = O(m)$$

we do the following:

$$0 \le m^2 + 1 \le m^2 + 2m + 1 \qquad \text{for all} \qquad m \ge 1.$$

So here $m_0 = 1$ throughout our derivation. Next follows

$$0 \le m^2 + 1 \ \le (m+1)^2 \qquad \text{for all} \qquad m \ge 1.$$

Taking square roots gives us

$$0 \le \sqrt{1+m^2} \ \le m + 1 \qquad \text{for all} \qquad m \ge 1.$$

Also the right-hand side of the last inequality can be bounded by

$$0 \le m + 1 \le \underbrace{m + m}_{2m} \qquad \text{for all} \qquad m \ge 1$$

so that together

$$0 \le \sqrt{1+m^2} \ \le 2m \qquad \text{for all} \qquad m \ge 1,$$

from which, by definition of order, equation (O1) follows directly where $k = 2$.

Another, more elegant, solution can be obtained by noting that

$$\sqrt{1+m^2} = m\sqrt{\frac{1}{m^2} + 1} < 2m \qquad \text{for all} \qquad m \ge 1.$$

(b). To show that

$$\text{(O2)} \qquad 0.02e^m + 100m^2 = O(e^m)$$

one wants to find a positive integer m_0 such that

$$100m^2 < e^m \qquad \text{for all} \qquad m \ge m_0.$$

Complete the derivation.

Sec. 23.3 Bellman's Principle. Dijkstra's Algorithm

In this section we consider **connected graphs** G (p. 981) with edges of positive length. Connectivity allows us to traverse from any edge of G to any other edge of G, as say in Figs. 487 and 488, on p. 983. (Figure 478, p. 971 is not connected.) Then, if we take a shortest path in a connected graph, that extends through several edges, and remove the last edge, that new (shortened) path is also a shortest path (to the prior vertex). This is the essence of *Bellman's minimality principle* (*Theorem 1,* Fig. 486, p. 981) and leads to the Bellman equations (1), p. 981. These equations in turn suggest a method to compute the length of shortest paths in G and form the heart of Dijkstra's algorithm.

Dijkstra's algorithm, p. 982, partitions the vertices of G into two sets $\mathcal{PL}$ of permanent labels and $\mathcal{TL}$ of temporary labels, respectively. At each iteration (Steps 2 and 3), it selects a temporarily labeled vertex k with the minimum distance label $\widetilde{L}_k$ from $\mathcal{TL}$, removes vertex k from $\mathcal{TL}$, and places it into $\mathcal{PL}$. Furthermore $\widetilde{L}_k$ becomes L_k. This signifies that we have found a shortest path from vertex 1 to vertex k. Then, using the idea of Bellman's equations, it updates the temporary labels in Step 3. The iterations continue until all nodes become permanently labeled, that is, until $\mathcal{TL} = \varnothing$ and $\mathcal{PL} =$ the set of all edges in G. Then the algorithm returns the lengths L_k $(k = 2, ..., n)$ of shortest paths from the given vertex (denoted by 1) to any other vertex in G. There is one more idea to consider: those vertices that were not adjacent to vertex 1, got a label of ∞ in Step 1 (an initialization step). This is illustrated in **Prob. 5**.

Note that, in Step 2, the algorithm looks for the shortest edge among all edges that originate from a node and selects it. Furthermore, the algorithm solves a more general problem than the one in Sec. 23.3, where the length of the edges were all equal to 1. *To completely understand this algorithm requires you to follow its steps when going through Example 1, p. 982, with a sketch of Fig. 487, p. 983, at hand.*

The problem of finding the shortest ("optimal") distance in a graph has many applications in various networks, such as networks of roads, railroad tracks, airline routes, as well as computer networks, the Internet, and others (see opening paragraph of Sec. 23.2, p. 975). Thus Dijkstra's algorithm is a very important algorithm as it forms a theoretical basis for solving problems in different network settings. In particular, it forms a basis for GPS navigation systems in cars, where we need directions on how to travel between two points on a map.

Problem Set 23.3. Page 983

1. Shortest path.

(a). *By inspection:*

We drop 40 because $12 + 28 = 40$ does the same.
We drop 36 because $12 + 16 = 28$ is shorter.
We drop 28 because $16 + 8 = 24$ is shorter.

(b). *By Dijkstra's algorithm.*

Dijkstra's algorithm runs as follows. (Sketch the figure yourself and keep it handy while you are working.)

Step 1

1. $L_1 = 0$, $\widetilde{L}_2 = 12$, $\widetilde{L}_3 = 40$, $\widetilde{L}_4 = 36$. Hence $\mathcal{PL} = \{1\}$, $\mathcal{TL} = \{2, 3, 4\}$. No ∞ appears because each of the vertices 2, 3, 4 is adjacent to 1, that is, is connected to vertex 1 by a single edge.

2. $L_2 = \min(\widetilde{L}_2, \widetilde{L}_3, \widetilde{L}_4) = \min(12, 40, 36) = 12$. Hence $k = 2$, $\mathcal{PL} = \{1, 2\}$, $\mathcal{TL} = \{3, 4\}$. Thus we started from vertex 1, as always, and added to the set $\mathcal{PL}$ the vertex which is closest to vertex 1, namely vertex 2. This leaves 3 and 4 with temporary labels. These must now be updated. This is Operation 3 of the algorithm (see Table 23.2 on p. 982).

3. Update the temporary label $\widetilde{L}_3$ of vertex 3,

$$\widetilde{L}_3 = \min(40, 12 + l_{23}) = \min(40, 12 + 28) = 40,$$

where 40 is the old temporary label of vertex 3, and 28 is the distance from vertex 2 to vertex 3, to which we have to add the distance 12 from vertex 1 to vertex 2, which is the permanent label of vertex 2.

Update the temporary label $\widetilde{L}_4$ of vertex 4,

$$\widetilde{L}_4 = \min(36, 12 + l_{24}) = \min(36, 12 + 16) = 28,$$

where 36 is the old temporary label of vertex 4, and 16 is the distance from vertex 2 to vertex 4. Vertex 2 belongs to the set of permanently labeled vertices, and 28 shows that vertex 4 is now closer to this set $\mathcal{PL}$ than it had been before.

This is the end of Step 1.

Step 2

1. Extend the set $\mathcal{PL}$ by including that vertex of $\mathcal{TL}$ that is closest to a vertex in $\mathcal{PL}$, that is, add to $\mathcal{PL}$ the vertex with the smallest temporary label. Now vertex 3 has the temporary label 40, and vertex 4 has the temporary label 28. Accordingly, include vertex 4 in $\mathcal{PL}$. Its permanent label is

$$L_4 = \min(\widetilde{L}_3, \widetilde{L}_4) = \min(40, 28) = 28.$$

Hence we now have $k = 4$, so that $\mathcal{PL} = \{1, 2, 4\}$ and $\mathcal{TL} = \{3\}$.

2. Update the temporary label $\widetilde{L}_3$ of vertex 3,

$$\widetilde{L}_3 = \min(40, 28 + l_{43}) = \min(40, 28 + 8) = 36,$$

where 40 is the old temporary label of vertex 3, and 8 is the distance from vertex 4 (which already belongs to PL) to vertex 3.

Step 3

Since only a single vertex, 3, is left in $\mathcal{TL}$, we finally assign the temporary label 36 as the permanent label to vertex 3.

Hence the remaining roads are

from vertex 1 to vertex 2	Length 12,
from vertex 2 to vertex 4	Length 16,
from vertex 4 to vertex 3	Length 8.

The total length of the remaining roads is 36 and these roads satisfy the condition that they connect all four communities.

Since Dijkstra's algorithm gives a shortest path from vertex 1 to each other vertex, it follows that these shortest paths also provide paths from any of these vertices to every other vertex, as required in the present problem. The solution agrees with the above solution by inspection.

5. Dijkstra's algorithm. Use of label $\widetilde{L}_j = l_{ij} = \infty$. The procedure is the same as in Example 1, p. 982, and as in **Prob. 1** just considered. You should make a sketch of the graph and use it to follow the steps.

Step 1

1. Vertex 1 gets permanent label 0. The other vertices get the temporary labels 2 (vertex 2), ∞ (vertex 3), 5 (vertex 4), and ∞ (vertex 5).

The further work is an application of Operation 2 [assigning a permanent label to the (or a) vertex closest to $\mathcal{PL}$ and Operation 3 (updating the temporary labels of the vertices that are still in the set $\mathcal{TL}$ of the temporarily labeled vertices], in alternating order.

2. $L_2 = 2$ (the minimum of 2, 5, and ∞).

3. $\widetilde{L}_3 = \min(\infty, 2 + 3) = 5.$

$\widetilde{L}_4 = \min(5, 2 + 1) = 3.$

$\widetilde{L}_5 = \min(\infty, \infty) = \infty.$

Step 2

1. $L_4 = \min(5, 3, \infty) = 3$. Thus $\mathcal{PL} = \{1, 2, 4\}, \mathcal{TL} = \{3, 5\}$. Two vertices are left in $\mathcal{TL}$; hence we have to make two updates.

2. $\widetilde{L}_3 = \min(5, 3 + 1) = 4.$

$\widetilde{L}_5 = \min(\infty, 3 + 4) = 7.$

Step 3

1. $L_3 = \min(4, 7) = 4.$

2. $\widetilde{L}_5 = \min(7, 4 + 2) = 6.$

Step 4

1. $L_5 = \widetilde{L}_5 = 6.$

Our result is as follows:

Step	Vertex added to $\mathcal{PL}$	Permanent label	Edge added to the path	Length of edge
1	1, 2	0, 2	(1, 2)	2
2	4	3	(2, 4)	1
3	3	4	(4, 3)	1
4	5	6	(3, 5)	2

The permanent label of a vertex is the length of the shortest path from vertex 1 to that vertex. Mark the shortest path from vertex 1 to vertex 5 in your sketch and convince yourself that we have omitted three edges of length 3, 4, and 5 and retained the edges that are shorter.

Sec. 23.4 Shortest Spanning Trees: Greedy Algorithm

A *tree* is a graph that is connected and has no cycles (for definition of "connected," see p. 977; for "cycle," p. 976). A **spanning tree** [see Fig. 489(b), p. 984], in a connected graph G, is a tree that contains all the vertices of G. A **shortest spanning tree** T in a connected graph G, whose vertices have positive length, is a spanning tree whose sum of the length of all edges of T is *minimum* compared to the sum of the length of all edges for any other spanning tree in G.

Sections 23.4 (p. 984) and 23.5 (p. 988) are both devoted to finding the *shortest spanning tree,* a problem also know as the *minimum spanning tree* (*MST*) *problem.*

Kruskal's greedy algorithm (p. 985; see also Example 1 and **Prob. 5**) is a systematic method for finding a shortest spanning tree. The efficiency of the algorithm is improved by using **double labeling of vertices** (look at Table 23.5 on p. 986, which is related to Example 1). Complexity considerations (p. 987) make this algorithm attractive for sparse graphs, that is, graphs with very few edges.

A **greedy algorithm** makes, at any instance, a decision that is locally optimal, that is, looks optimal at the moment, and hopes that, in the end, this strategy will lead to the desired global (or overall) optimum. Do you see that Kruskal uses such a strategy? Is Dijkstra's algorithm a greedy algorithm? (For answer see p. 20).

More details on Example 1, p. 985. Application of Kruskal's algorithm with double labeling of vertices (Table 23.3, p. 985). We reproduce the list of double labels, that is, Table 23.5, p. 986, and give some further explanations to it. Note that this table was obtained from the rather simple Table 23.4, p. 985.

	Choice 1	Choice 2	Choice 3	Choice 4	Choice 5
Vertex	(3, 6)	(1, 2)	(1, 3)	(4, 5)	(3, 4)
1		$(1, 0)$			
2		$(1, 1)$			
3	$(3, 0)$		$(1, 1)$		
4				$(4, 0)$	$(1, 3)$
5				$(4, 4)$	$(1, 4)$
6	$(3, 3)$		$(1, 3)$		

By going line by line through our table, we can see what the shortest spanning tree looks like. Follow our discussion and sketch our findings, obtaining a shortest spanning tree.

Line 1. (1, 0) shows that 1 is a root.

Line 2. (1, 1) shows that 2 is in a subtree with root 1 and is preceded by 1. [This tree consists of the single edge (1, 2).]

Line 3. (3, 0) means that 3 first is a root, and (1, 1) shows that later it is in a subtree with root 1, and then is preceded by 1, that is, joined to the root by a single edge (1, 3).

Line 4. (4, 0) shows that 4 first is a root, and (1, 3) shows that later it is in a subtree with root 1 and is preceded by 3.

Line 5. (4, 4) shows that 5 first belongs to a subtree with root 4 and is preceded by 4, and (1, 4) shows that later 5 is in a (larger) subtree with root 1 and is still preceded by 4. This subtree actually is the whole tree to be found because we are now dealing with Choice 5.

Line 6. (3, 3) shows that 6 is first in a subtree with root 3 and is preceded by 3, and then later is in a subtree with root 1 and is still preceded by 3.

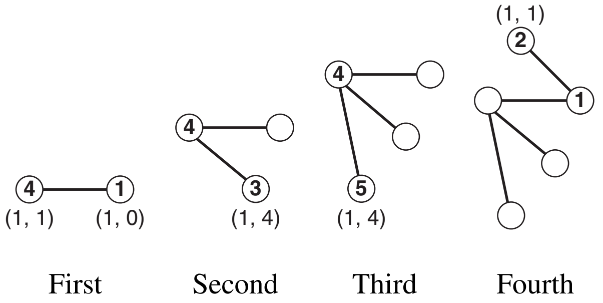
Problem Set 23.4. Page 987

5. **Kruskal's algorithm.** Trees constitute a very important type of graph. Kruskal's algorithm is straightforward. It begins by ordering the edges of a given graph G in ascending order of length. The length of an edge $(i,\ j)$ is denoted by l_{ij}. Arrange the result in a table similar to Table 23.4 on p. 985. The given graph G has $n = 5$ vertices. Hence a spanning tree in G has $n - 1 = 4$ edges, so that you can terminate your table when four edges have been chosen. Pick edges of the spanning tree to be obtained in order of length, rejecting when a cycle would be created. This gives the following table. (Look at the given graph!)

Edge	Length	Choice
(1, 4)	2	1st
(3, 4)	2	2nd
(4, 5)	3	3rd
(3, 5)	4	(Reject)
(1, 2)	5	4th

We see that the spanning tree is the one in the answer on p. A56 and has the length $L = 12$.

In the case of the present small graph we would not gain much by double labeling. Nevertheless, to understand the process as such (and also for a better understanding of the table on p. 986) do the following for the present graph and tree. Graph the growing tree as on p. 986. Double label the vertices, but attach a label only if it is new or if it changes in a step.



From these graphs we can now see what a corresponding table looks like. This table is simpler than that in the book because the root of the growing tree (subtree of the answer) does not change; it remains vertex 1.

	Choice 1	Choice 2	Choice 3	Choice 4
Vertex	(1, 4)	(3, 4)	(4, 5)	(1, 2)
1	(1, 0)			
2				(1, 1)
3		(1, 4)		
4	(1, 1)			
5			(1, 4)	

We see that vertex 1 is the root of every tree in the graph. Vertex 2 gets the label (1, 1) because vertex 1 is its root as well as its predecessor. In the label (1, 4) of vertex 3 the 1 is the root and 4 the predecessor. Label (1, 1) of vertex 4 shows that the root as well as the predecessor is 1. Finally, vertex 5 has the root 1 and the predecessor 4.

17. **Trees that are paths.** Let T be a tree with exactly two vertices of degree 1. Suppose that T is not a path. Then it must have at least one vertex v of degree $d \geq 3$. Each of the d edges, incident with v, will eventually lead to a vertex of degree 1 (at least one such vertex) because T is a tree, so it cannot have cycles (definition on p. 976 in Sec. 23.2). This contradicts the assumption that T has but *two* vertices of degree 1.

Sec. 23.5 Shortest Spanning Trees: Prim's Algorithm

From the previous section, recall that a spanning tree is a tree in a connected graph that contains all vertices of the graph. Comparison over all such trees may give a shortest one, that is, one whose sum of the length of edges is the shortest. We assume that all the lengths are positive (p. 984 of the textbook).

Another popular method to find a shortest spanning tree is by **Prim's algorithm**. This algorithm is more involved than Kruskal's algorithm and should be used when the graph has more edges and branches.

Prim's algorithm shares similarities with Dijkstra's algorithm. Both share a similar structure of three steps. They are an initialization step, a middle step where most of the action takes place, and an updating (final) step. Thus, if you studied and understood Dijkstra's algorithm, you will readily appreciate Prim's algorithm. Instead of fixing a permanent label in Dijkstra, Prim's adds an edge to a tree T in the second step. Prim's algorithm is illustrated in Example 1, p. 990. (For comparison, Dijkstra's algorithm was illustrated in Example 1, p. 982).

Here are two simple questions (open book) to test your understanding of the material. Can Prim's algorithm be applied to the graph of Example 1, p. 983? Can Dijkstra's algorithm be applied to the graph of Example 1, p. 990? Give an answer (Yes or No) and give a reason. Then turn to p. 20 to check your answer.

Problem Set 23.5. Page 990

9. Shortest spanning tree obtained by Prim's algorithm. In each step, U is the set of vertices of the tree T to be grown, and S is the set of edges of T. The beginning is at vertex 1, as always. The table is similar to that in Example 1 on p. 990. It contains the initial labels and then, in each column, the effect of relabeling. Explanations follow after the table.

			Relabeling	
Vertex	Initial	(I)	(II)	(III)
2	$l_{12} = 16$	$l_{24} = 4$	$l_{24} = 4$	–
3	$l_{13} = 8$	$l_{34} = 2$	–	–
4	$l_{14} = 4$	–	–	–
5	$l_{15} = \infty$	$l_{45} = 14$	$l_{35} = 10$	$l_{35} = 10$

1. $i(k) = 1$, $U = \{1\}$, $S = \emptyset$. Vertices 2, 3, 4 are adjacent to vertex 1. This gives their initial labels equal to the length of the edges connecting them with vertex 1 (see the table). Vertex 5 gets the initial label ∞ because the graph has no edge (1,5); that is, vertex 5 is not adjacent to vertex 1.

2. $\lambda_4 = l_{14} = 4$ is the smallest of the initial labels. Hence include vertex 4 in U and edge (1, 4) as the first edge of the growing tree T. Thus, $U = \{1, 4\}$, $S = \{(1, 4)\}$.

3. Each time we include a vertex in U (and the corresponding edge in S) we have to update labels. This gives the three numbers in column (I) because vertex 2 is adjacent to vertex 4, with $l_{24} = 4$ [the length of edge (2, 4)], and so is vertex 3, with $l_{34} = 2$ [the length of edge (3,4)]. Vertex 5 is also adjacent to vertex 4, so that ∞ is now gone and replaced by $l_{45} = 14$ [the length of edge (4, 5)].

2. $\lambda_3 = l_{34} = 2$ is the smallest of the labels in (I). Hence include vertex 3 in U and edge (3, 4) in S. We now have $U = \{1, 3, 4\}$ and $S = \{(1, 4), (3, 4)\}$.

3. Column (II) shows the next updating. $l_{24} = 4$ remains because vertex 2 is not closer to the new vertex 3 than to vertex 4. Vertex 5 is closer to vertex 3 than to vertex 4, hence the update is $l_{35} = 10$, replacing 14.

2. The end of the procedure is now quite simple. l_{24} is smaller than l_{35} in column (II), so that we set $\lambda_2 = l_{24} = 4$ and include vertex 2 in U and edge (2, 4) in S. We thus have
$U = \{1, 2, 3, 4\}$ and $S = \{(1, 4), (3, 4), (2, 4)\}$.

3. Updating gives no change because vertex 5 is closer to vertex 3, whereas it is not even adjacent to vertex 2.

2. $\lambda_5 = l_{35} = 10$. $U = \{1, 2, 3, 4, 5\}$, so that our spanning tree T consists of the edges
$S = \{(1, 4), (3, 4), (2, 4), (3, 5)\}$.
The length of the shortest spanning tree is

$$L(T) = \sum l_{ij} = l_{14} + l_{34} + l_{24} + l_{35} = 4 + 2 + 4 + 10 = 20.$$

Sec. 23.6 Flows in Networks

Overview of Sec. 23.6
We can conveniently divide this long section into the following subtopics:

0. Theme. Sections 23.6 and 23.7 cover the third major topic of **flow problems in networks**. They have many applications in electrical networks, water pipes, communication networks, traffic flow in highways, and others. A typical example is the trucking problem. A trucking company wants to transport crates, by truck, from a factory ("the source") located in one city to a warehouse ("target") located far away in another city over a network of roads. There are certain constraints. The roads, due to their construction (major highway, two-lane road), have a certain capacity, that is, they allow a certain number of trucks and cars. They are also affected by the traffic flow, that is, the number of trucks and cars on the road at different times. The company wants to determine the maximum number of crates they can ship under the given constraints.

Section 23.6 covers the terminology and theory needed to analyze such problems and illustrates them by examples. Section 23.7 gives a systematic way to determine maximum flow in a network.

1. **Network, pp. 991–992**

- We consider digraphs $G = (V, E)$ (definition, p. 972) in this section and define a network in which each edge (i, j) has assigned to it a capacity $c_{ij} > 0$. The capacity measures the maximum possible flow along (i, j). One vertex in the network is the source s and another the target t (or sink). We denote a flow along a directed edge (i, j) by f_{ij}. The flow is produced and flows naturally from the source to the target (sink), where it disappears. See p. 991.
- The edge condition means that the flow cannot exceed the capacity, that is,

$$0 \le f_{ij} \le c_{ij}.$$

- The vertex condition (Kirchhoff's law) applies to each vertex i that is not s or t. It is given by

$$\text{Inflow} = \text{Outflow}.$$

More precisely we get (2), p. 992.

2. **Paths, p. 992**

- Definition of path $P : v_1 \to v_k$ in a digraph G as a sequence of edges

$$(v_1, v_2), (v_2, v_3), \ldots, (v_{k-1}, v_k),$$

regardless of their directions in G, that forms a path as a graph.
- Related concepts of forward edge and backward edge of a path, p. 992 and Figs. 494 and 495.

3. Flow Augmenting Paths, pp. 992–993

Our goal is to maximize the flow and thus we look for a path $P : s \to t$ from the source to the sink, whose edges are not fully used so that we can push additional flow through P. This leads to

- flow augmenting path (in a network) in which

  (i) no forward edge is used to capacity

  (ii) no backward edge has flow 0,

  see definition on top of p. 993. Do you see that Conditions (i) and (ii) mean $f_{ij} < c_{ij}$ and $f_{ij} > 0$ for related edges, respectively?

4. Cut Sets, pp. 994–996

- We introduce the concept of cut set (S, T) because we want to know what is flowing from s to t. So we cut the network somewhere between s and t and see what is flowing through the edges hit by the cut. The cut set is precisely that set of edges that were hit by the cut; see upper half of p. 994.
- On the cut set we define capacity cap(S, T) to be the sum of all forward edges from source S to target T. Write it out in a formula and compare your answer with (3), p. 994.

5. Four Theorems, pp. 995–996 The section discusses the following theorems about cut sets and flows. They are:

- *Theorem 1. Net flow in cut sets.* It states that any given flow in a network G is the net flow through any cut set (S, T) of G.
- *Theorem 2. Upper bound for flows.* A flow f in a network G cannot exceed the capacity of any cut set (S, T) in G.
- *Theorem 3. Main Theorem. Augmenting path theorem for flows.* It states that a flow from s to t in a network G is maximum if and only if there does not exist a flow augmenting path $s \to t$ in G.
  The last theorem is by Ford and Fulkerson. It is
- *Theorem 4. Max-Flow Min-Cut Theorem.* It states that the maximum flow in any network G is equal to the capacity of a cut set of minimum capacity ("minimum cut set") in G.

6. Illustrations of Concepts.
An example of a network is given in Fig. 493, p. 992. Forward edge and backward edge are illustrated in Figs. 494 and 495 on the same page. **Example 1**, p. 993, and **Prob. 15** determine flow augmenting paths. Figure 498 and explanation, p. 994, as well as **Probs. 3** and **5** illustrate cut sets and capacity. Note that, in the network in Fig. 498, the first number on each edge denotes capacity and the second number flow. Intuitively, if you think of edges as roads, then capacity of the road means how many cars *can actually be* on the road and flow denotes how many cars actually *are* on the road. Finally, **Prob. 17** finds maximum flow.

Problem Set 23.6. Page 997

3. Cut sets, capacity. We are given that $S = \{1, 2, 3\}$. T consists of the other vertices that are not in S. Looking at Fig. 498, p. 994, we see that $T = \{4, 5, 6\}$. First draw Fig. 498 (without any cut) and then draw a line that separates S from T. This is the cut. Then we see that the curve cuts the edge $(1, 4)$ whose capacity is 10, the edge $(5, 2)$, which is a backward edge, the edge $(3, 5)$, whose capacity is 5, and the edge (3.6), whose capacity is 13. By definition (3), p. 994, the capacity cap (S, T) is the sum of the capacities of the forward edges from S to T. Here we have three forward edges and hence

$$\text{cap }(S, T) = 10 + 5 + 13 = 28.$$

The edge $(5, 2)$ goes from vertex 5, which belongs to T, to vertex 3, which belongs to S. This shows that edge $(5, 2)$ is indeed a backward edge, as noted above. And backward edges are not included in the capacity of a cut set, by definition.

5. Cut sets, capacity. Here $S = \{1, 2, 4, 5\}$. Looking at the graph in Fig. 499, p. 997, we see that $T = \{3, 6, 7\}$. We draw Fig. 499 and insert the cut, that is a curve that separates S from T. We see that the curve cuts edges $(2, 3)$, $(5, 3)$, and $(5, 6)$. These edges are all forward edges and thus contribute to cap (S, T). The capacities of these edges are 8, 4, and 4, respectively. Using (3), p. 994, we have

$$\text{cap }(S, T) = 8 + 4 + 4 = 16.$$

15. Flow augmenting paths. The given answer is

$$1 - 2 - 5, \qquad \Delta f = 2$$

$$1 - 4 - 2 - 5, \qquad \Delta f = 2, \text{ etc.}$$

From this, we see that the path $1 - 2 - 5$ is flow augmenting and admits an additional flow:

$$\Delta = \min(4 - 2,\ 8 - 5) = \min(2,\ 3) = 2.$$

Here $2 = 4 - 2$ comes from edge $(1,\ 2)$ and $3 = 8 - 5$ from edge $(2,\ 5)$.

Furthermore, we see that another flow augmenting path is $1 - 4 - 2 - 5$ and admits an increase of the given flow:

$$\Delta = \min(10 - 3,\ 5 - 3,\ 8 - 5) = \min(7,\ 2,\ 3) = 2.$$

And so on. Of course, if we increased the flow on $1 - 2 - 5$ by 2, then we have on edge $(2,\ 5)$ instead of $(8,\ 5)$ the new values $(8,\ 7)$ and can now increase the flow on $1 - 4 - 2 - 5$ only by $8 - 7 = 1$, the edge $(2,\ 5)$ now being the bottleneck edge.

For such a small network we can find flow augmenting paths (if they exist) by trial and error. For large networks we need an algorithm, such as that of Ford and Fulkerson in Sec. 23.7, pp. 998–1000.

17. Maximum flow. The given flow in the network depicted in this problem on p. 997 is 10. We can see this by looking at the two edges $(4, 6)$ and $(5, 6)$ that go into target t (the sink 6) and get the flow $1 + 9 = 10$. Another way is to look at the three edges $(1, 3)$, $(1, 4)$, and $(1, 2)$ that are leaving vertex 1 (the source s) and get the flow $5 + 3 + 2 = 10$.

To find the maximum flow by inspection we note the following. Each of the three edges going out from vertex 1 could carry additional flow of 3. This is computed by the difference of capacity (the first number) and flow (the second number on the edge), which, for the three edges, are

$$\Delta_{13} = 8 - 5 = 3, \qquad \Delta_{14} = 6 - 3 = 3, \qquad \Delta_{12} = 5 - 2 = 3.$$

Since the additional flow is 3, we may augment the given flow by 3 by using path $1 - 4 - 5 - 6$. Then the edges $(1, 4)$ and $(5, 6)$ are used to capacity. This increases the given flow from 10 to $10 + 3 = 13$.

Next we can use the path $1 - 2 - 4 - 6$. Its capacity is

$$\Delta = \min\ (5 - 2, 4 - 2, 4 - 1) = 2.$$

This increases the flow from 13 to $13 + 2 = 15$. For this new increased flow the capacity of the path $1 - 3 - 5 - 6$ is

$$\Delta = \min\ (3, 4, 13 - 12) = 1$$

because the first increase of 3 increased the flow in edge $(5, 6)$ from 9 to 12. Hence we can increase our flow from 15 to $15 + 1 = 16$.

Finally, consider the path $1 - 3 - 4 - 6$. The edge $(4, 3)$ is a backward edge in this path. By decreasing the existing flow in edge $(4, 3)$ from 2 to 1, we can push a flow 1 through this path. Then edge $(4, 6)$ is used to capacity, whereas edge $(1, 3)$ is still not fully used. But since both edges are going to vertex 6, that is, edges $(4, 6)$ and $(5, 6)$ are now used to capacity, we cannot augment the flow further, so that we have reached the maximum flow

$$f = 16 + 1 = 17.$$

For our solution of maximum flow $f = 17$, the flows in the edges are

$$\begin{aligned}
f_{12} &= 4 && \text{(instead of 2)}\\
f_{13} &= 7 && \text{(instead of 5)}\\
f_{14} &= 6 && \text{(instead of 3)}\\
f_{24} &= 4 && \text{(instead of 2)}\\
f_{35} &= 8 && \text{(instead of 7)}\\
f_{43} &= 1 && \text{(instead of 2)}\\
f_{45} &= 5 && \text{(instead of 2)}\\
f_{46} &= 4 && \text{(instead of 1)}\\
f_{56} &= 13 && \text{(instead of 9)}
\end{aligned}$$

You should sketch the network with the new flow and check that Kirchhoff's law

Inflow = Outflow for each vertex i that is not a source s or sink t

is satisfied at every vertex.

The answer on p. A57 presents a slightly different solution with the same final result of maximum flow $f = 17$. In that solution (although not stated) $f_{43} = 0$. For practice you may want to quickly go through that solution and show that it satisfies Kirchhoff's law at every vertex.

Sec. 23.7 Maximum Flow: Ford–Fulkerson Algorithm

We continue our discussion of flow problems in networks. Important is the Ford–Fulkerson algorithm for maximum flow given in Table 23.8, pp. 998–999 and illustrated in detail in **Example 1**, pp. 999–1000 and **Prob. 7**. For optimal learning, go through Example 1 line by line and see how the algorithm applies.

Ford–Fulkerson uses augmented paths to increase a given flow in a given network until the flow is maximum. It accomplishes this goal by constructing stepwise flow augmenting paths, one at a time, until no further paths can be constructed. This happens exactly when the flow is maximum.

Problem Set 23.7. Page 1000

7. **Maximum flow.** Example 1 in the text on pp. 999–1000 shows how we can proceed in applying the Ford–Fulkerson algorithm for obtaining flow augmenting paths until the maximum flow is reached. No algorithms would be needed for the modest problems in our problem sets. Hence the point of this, and similar problems, is to obtain familiarity with the most important algorithms for basic tasks in this chapter, as they will be needed for solving large-scale real-life problems. Keep this in mind to avoid misunderstandings. From time to time look at Example 1 in the text, which is similar and may help you to see what to do next.

1. The given initial flow is $f = 6$. This can be seen by looking at flows 2 in edge $(1, 2)$, 1 in edge $(1, 3)$, and 3 in edge $(1, 4)$, that begin at s and whose sum is 6, or, more simply, by looking at flows 5 and 1 in the two edges $(2, 5)$ and $(3, 5)$, respectively, that end at vertex 5 (the target t).

2. Label s $(= 1)$ by $\emptyset$. Mark the other edges 2, 3, 4, 5 as "unlabeled."

3. Scan 1. This means labeling vertices 2, 3, and 4 adjacent to vertex 1 as explained in Step 3 of Table 23.8 (the table of the Ford–Fulkerson algorithm), which, in the present case, amounts to the following. $j = 2$ is the first unlabeled vertex in this process, which corresponds to the first part of Step 3 in Table 23.8. We have $c_{12} > f_{12}$ and compute

$$\Delta_{12} = c_{12} - f_{12} = 4 - 2 = 2 \qquad \text{and} \qquad \Delta_2 = \Delta_{12} = 2.$$

We label 2 with the forward label $(1^+, \Delta_2) = (1^+, 2)$.

$j = 3$ is the second unlabeled vertex adjacent to 1, and we compute

$$\Delta_{13} = c_{13} - f_{13} = 3 - 1 = 2 \qquad \text{and} \qquad \Delta_3 = \Delta_{13} = 2.$$

We label 3 with the forward label $(1^+, \Delta_3) = (1^+, 2)$.

$j = 4$ is the third unlabeled vertex adjacent to 1, and we compute

$$\Delta_{14} = c_{14} - f_{14} = 10 - 3 = 7 \qquad \text{and} \qquad \Delta_4 = \Delta_{14} = 7.$$

We label 4 with the forward label $(1^+, \Delta_4) = (1^+, 7)$.

4. Scan 2. This is necessary since we have not yet reached t (vertex 5), that is, we have not yet obtained a flow augmenting path. Adjacent to vertex 2 are the vertices 1, 4, and 5. Vertices 1 and 4 are labeled. Hence the only vertex to be considered is vertex **5**. We compute

$$\Delta_{25} = c_{25} - f_{25} = 8 - 5 = 3.$$

The calculation of Δ_5 differs from the corresponding previous ones. From the table we see that

$$\Delta_5 = \min(\Delta_2, \ \Delta_{25}) = \min(2, 3) = 2.$$

The idea here is that $\Delta_{25} = 3$ is of no help because in the previous edge $(1, 2)$ you can increase the flow only by 2. Label 5 with the forward label $(2^+, \Delta_5) = (2^+, 2)$.

5. We have obtained a first flow augmenting path P: $1 - 2 - 5$.

6. We augment the flow by $\Delta_5 = 2$ and set $f = 6 + 2 = 8$.

7. Remove the labels from 2, 3, 4, 5, and go to Step 3. Sketch the given network, with the new flows $f_{12} = 4$ and $f_{25} = 7$. The other flows remain the same as before. We will now obtain a second flow augmenting path.

3. We scan 1. Adjacent are 2, 3, 4. We have $c_{12} = f_{12}$; edge $(1, 2)$ is used to capacity and is no longer to be considered. For vertex 3 we compute

$$\Delta_{13} = c_{13} - f_{13} = 3 - 1 = 2 \qquad \text{and} \qquad \Delta_3 = \Delta_{13} = 2.$$

Label 3 with the forward label $(1^+, 2)$. For vertex 4 we compute $\Delta_{14} = c_{14} - f_{14} = 10 - 3 = 7$ and $\Delta_4 = \Delta_{14} = 7$.
Label 4 with the forward label $(1^+, 7)$.

3. We need not scan 2 because we now have $f_{12} = 4$ so that $c_{12} - f_{12} = 0$; $(1, 2)$ is used to capacity; the condition $c_{12} > f_{12}$ in the algorithm is not satisfied. Scan 3. Adjacent to 3 are the vertices 4 and 5. For vertex 4 we have $c_{43} = 6$ but $f_{43} = 0$, so that the condition $f_{43} > 0$ is violated. Similarly, for vertex 5 we have $c_{35} = f_{35} = 1$, so that the condition $c_{35} > f_{35}$ is violated and we must go on to vertex 4.

3. Scan 4. The only unlabeled vertex adjacent to 4 is 2, for which we compute

$$\Delta_{42} = c_{42} - f_{42} = 5 - 3 = 2$$

and

$$\Delta_2 = \min(\Delta_4,\ \Delta_{42}) = \min(7, 2) = 2.$$

Label 2 with the forward label $(4^+, 2)$.

4. Scan 2. Unlabeled adjacent to 2 is vertex 5. Compute

$$\Delta_{25} = c_{25} - f_{25} = 8 - 7 = 1$$

and

$$\Delta_5 = \min(\Delta_2,\ \Delta_{25}) = \min(2, 1) = 1.$$

Label 5 with the forward label $(2^+, 1)$.

5. We have obtained a second flow augmenting path P: 1 – 4 – 2 – 5.

6. Augment the existing flow 8 by Δ_5 = 1 and set $f = 8 + 1 = 9$.

7. Remove the labels from 2, 3, 4, 5 and go to Step 3. Sketch the given network with the new flows, write the capacities and flows in each edge, obtaining edge (1, 2): (4, 4), edge (1, 3): (3, 1), edge (1, 4): (10, 4), edge (2, 5): (8, 8), edge (3, 5): (1, 1), edge (4, 2): (5, 4), and edge (4, 3): (6, 0). We see that the two edges going into vertex 5 are used to capacity; hence the flow $f = 9$ is maximum. Indeed, the algorithm shows that vertex 5 can no longer be reached.

Sec. 23.8 Bipartite Graphs. Assignment Problems

We consider graphs. A **bipartite graph** $G = (V, E)$ allows us to partition ("partite") a vertex set V into two ("bi") sets S and T, where S and T share no elements in common. This requirement of $S \cap T = \varnothing$ by the nature of a partition.

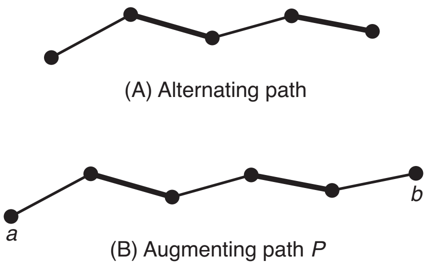
Other concepts that follow are **matching** and **maximum cardinality matching** (p. 1001 of the textbook), **exposed vertex**, **complete matching**, **alternating path**, and **augmenting path** (p. 1002).

A **matching** M in $G = (S, T; E)$ is a set M of edges of graph G such that no two of those edges have a vertex in common. In the special case, where the set M consists of the greatest possible number of edges, M is called a **maximum cardinality matching** in G. Matchings are shown in Fig. 503 at the bottom of p. 1001.

A vertex is **exposed** or not covered by M if the vertex is not an endpoint of an edge in M. If, in addition, the matching leaves no vertices exposed, then M is known as a *complete matching*. Can you see that this exists only if S and T have the same number of vertices?

An **alternating path** consists *alternately* of edges that are in M and not in M, as shown below. Closely related is an augmenting path, whereby, in the alternating path, both endpoints a and b are exposed. This leads to Theorem 1, **the augmenting path theorem for bipartite matching**. It states that the matching in a bipartite graph is of maximum cardinality $\Leftrightarrow$ there does not exist an augmenting path with respect to the matching.

The theorem forms the basis for *algorithm matching*, pp. 1003–1004, and is illustrated in Example 1. Go through the algorithm and example to convince yourself how the algorithm works. In addition to the label of the vertex, the method also requires a label that keeps track of backtracking paths.


(A) Alternating path

(B) Augmenting path *P*

Sec. 23.8. Alternating path and augmenting path P. Heavy edges are those belonging to a matching M

We *augment a given matching by an edge* by dropping from matching M the edges that are not an augmenting path P (two edges in the figure above) and adding to M the other edges of P (three in the figure, do you see it?).

Problem Set 23.8. Page 1005

1. **A graph that is not bipartite.** We proceed in the order of the numbers of the vertices. We put vertex 1 into S and its adjacent vertices 2, 3 into T. Then we consider 2, which is now in T. Hence, for the graph to be bipartite, its adjacent vertices 1 and 3 should be in S. But vertex 3 has just been put into T. This contradicts the definition of a bipartite graph on p. 1001 and shows that the graph is not bipartite.

7. **Bipartite graph.** Since graphs can be graphed in different ways, one cannot see immediately whether a graph is bipartite. Hence in the present problem we have to proceed systematically.

1. We put vertex 1 into S and all its adjacent vertices 2, 4, 6 into T. Thus

$$S = \{1\}, \qquad T = \{2, 4, 6\}.$$

2. Since vertex 2 is now in T, we put its adjacent vertices 1, 3, 5 into S. Thus

$$\text{(P)} \qquad S = \{1, 3, 5\}, \qquad T = \{2, 4, 6\}.$$

3. Next consider vertex 3, which is in S. For the graph to be bipartite, its adjacent vertices 2, 4, 6 should be in T, as is the case by (P).

4. Vertex 4 is in T. Its adjacent vertices 1, 3, 5 are in S which is true by (P).

5. Vertex 5 is in S. Hence for the graph to be bipartite, its adjacent vertices 2, 4, 6 should be in T. This is indeed true by (P).

6. Vertex 6 is in T and its adjacent vertices 1, 3, 5 are in S.

Since none of the six steps gave us any contradiction, we conclude that the given graph in this problem is bipartite. Take another look at the figure of our graph on p. 1005 to realize that, although the number of vertices and edges is small, the present problem is not completely trivial. We can sketch the graph in such a way that we can immediately see that it is bipartite.

17. K_4 **is planar** because we can graph it as a square A, B, C, D, then add one diagonal, say, A, C, inside, and then join B, D not by a diagonal inside (which would cross) but by a curve outside the square.

Answer to question on greedy algorithm (see p. 10 in Sec. 23.4 of this Student Solutions Manual and Study Guide). Yes, definitely, Dijkstra's algorithm is an example of a greedy algorithm, as in Steps 2 and 3 it looks for the shortest path between the current vertex and the next vertex.

Answer to self-test on Prim's and Dijkstra's algorithms (see p. 12 of Sec. 23.5). Yes, since both trees are spanning trees.