

Part 1

A circle no doubt has a certain appealing simplicity at the first glance, but one look at an ellipse should have convinced even the most mystical of astronomers that the perfect simplicity of the circle is akin to the vacant smile of complete idiocy. Compared to what an ellipse can tell us, a circle has little to say.

–Eric Temple Bell, *Mathematics: Queen and Servant of Science*, p. 277.

Chapter 1

The Surprising Ellipse

The Ellipse, a 52-acre park located south of the White House in President's Park, Washington D.C., is one of many geometric shapes located near the capital. According to a survey of the Ellipse (see [93]), the major axis is 1,058.26 feet and the minor axis is 902.85 feet. In case you are wondering, the area is 751,071.67 square feet and the perimeter is 3,086.87 feet. Is that enough for us to be certain that the creators intended to design an ellipse? The Colosseum in Rome, located east of the Roman Forum, also appears to be an ellipse, but is it? Some architects contend that the Colosseum is an ellipse, while others suggest that it is a curve made of circular arcs that are then connected in a smooth fashion. This is actually the subject of an important debate, because it would tell us what the Romans knew about ellipses.¹ What is an ellipse, anyway?

There are many ways to introduce the ellipse. If we take a cone and slice it in such a way that we have a bounded (nondegenerate) slice, we obtain an ellipse. We might, instead, connect two points (the foci of the ellipse) with a string, stretch the string tightly, and, using a pen to keep the length of the string constant, sketch out our ellipse. When introduced this way we say that the *ellipse* is the locus of points in the plane for which the sum of the distances to two fixed points (the *foci*) is constant. The line segments with endpoints on the ellipse that pass through the center are called *diameters*. The most important of these are the diameter through the foci or *major axis* and the shortest diameter or *minor*

¹See [73, p. 37] which says that it is likely “the parallel sequence of ovals for the Colosseum” was “laid out, one by one, as combinations of circular arcs”, while <http://www.wdl.org/en/item/4243> (accessed 12/15/2017) tells us in no uncertain terms that “It forms an ellipse, measuring approximately 190 meters long by 155 meters wide”.



Figure 1.1. The Colosseum in Rome—an ellipse?

axis. In Cartesian coordinates, ellipses are given by the equation

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

where A, B, C, D, E , and F are real numbers and $B^2 - 4AC < 0$. If we use complex numbers and think of the foci as w_1 and w_2 , assuming our string has length $2r$, where $r \geq 0$, the points on the ellipse are those complex numbers z for which

$$|z - w_1| + |z - w_2| = 2r.$$

When the two foci are the same, we have a circle of radius r . If $r = 0$, then we have a single point, which we consider a (degenerate) ellipse.

One of the reasons for the widespread use of ellipses in the sciences is the reflective property of an ellipse: A ball that travels along a ray from one focus will hit the ellipse and bounce off passing through the other focus. Both light and sound are influenced by this property—perhaps you have seen this exhibited in museums or whisper chambers; if you stand at one focus and whisper something to a friend at the other focus, your friend will hear even the quietest whisper. Another interesting consequence of this property is that ellipses are used in medicine to treat kidney stones and gallstones using a process called lithotripsy. Ellipses also

appear in the laws of planetary motion, architecture, acoustics, and optics. Given the widespread use of ellipses, we should understand them well. But most of us know much more about circles than we do about ellipses. Consider the following natural questions.

What is the area bounded by a circle? That is well known; it is πr^2 , where r is the radius of the circle. What is the area bounded by an ellipse? You may not have memorized this, but you can figure it out: The area stays the same if we rotate and shift our ellipse, so let us assume that our ellipse has its center at the origin (the point $(0, 0)$), semi-major axis of length a , and semi-minor axis of length b . Then we can write the equation of our ellipse as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

So we can find the area bounded by an ellipse if we can recall how to find the area bounded by

$$a\sqrt{1 - \frac{y^2}{b^2}}$$

and the y -axis. This gives us a chance to use calculus: Trigonometric substitution and a calculation that we are pretty sure you would like to do yourself tell us that the area bounded by our ellipse is πab . Or, if you prefer, you can use Green's theorem with the vector field $\vec{F}(x, y) = (1/2)[-y \ x]$ to get the same result. If $a = b$, then we have a circle of radius $r = a = b$ and we have the correct area, πr^2 ; that is comforting.

What about the perimeter of a circle? That is well known too. It is $2\pi r$. How about the perimeter, P , of an ellipse? We will give you a few minutes to think about it. You might try using the formula for the ellipse above, or you might try parametrizing the ellipse as $x = a \cos \theta$, $y = b \sin \theta$ for $0 \leq \theta < 2\pi$. No matter how you approach it though, you will find it to be a challenging problem. Using calculus once again you will discover that you need to compute

$$P = \int_0^{2\pi} (a^2 \sin^2 \theta + b^2 \cos^2 \theta)^{1/2} d\theta.$$

Though you have learned lots of integration techniques, you will probably soon come around to the thinking expressed in the website numerica.com.² “There is no simple exact formula: There are simple formulas but they are not exact, and there are exact formulas but they are not simple”.³



Figure 1.2. © by Sidney Harris, ScienceCartoonsPlus.com

The search for a formula (as well as estimates) for the perimeter of an ellipse has a fascinating history; a history that involves mathematicians like Gauss, Ramanujan, and—not too surprisingly—Kepler. An estimate for the perimeter of an ellipse was desirable because of its connection to the elliptical orbits of planets. In 1609 Kepler noted that $P \approx 2\pi\sqrt{ab}$ and $P \approx \pi(a + b)$. These were “the first legitimate approximations ... although [Kepler’s] arguments were not very rigorous and $2\pi\sqrt{ab}$ was intended to be only a lower bound” for P [4]. These approximations are best when a is close to b .

²<http://numerica.com/answer/ellipse.htm#elliptic> (accessed 12/15/2017)

³Semjon Adlaj [1] disputes this statement in his article, *An eloquent formula for the perimeter of an ellipse*.

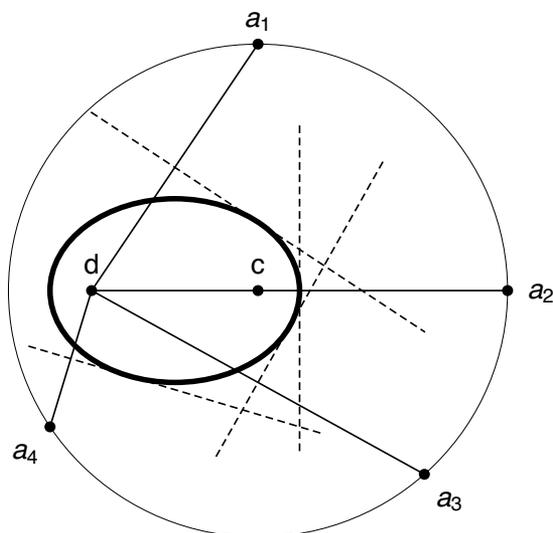


Figure 1.3. Four folds.

Returning to the Ellipse in Washington D.C. for a moment, you might check that the measurements are consistent with the claim that the Ellipse is actually an ellipse.

In fact, Colin Maclaurin [105] was the first to publish an exact expression for the perimeter, now referred to as the *Maclaurin expansion*; the perimeter P is given by

$$P = 2\pi a \left[1 - \left(\frac{1}{2}\right)^2 k^2 - \left(\frac{1 \cdot 3}{2 \cdot 4}\right)^2 \frac{k^4}{3} - \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\right)^2 \frac{k^6}{5} - \dots \right],$$

where $k^2 = 1 - (b^2/a^2)$. It is not likely that you guessed that!

Perhaps Bell was referring to this rich history when he said, “Compared to what an ellipse can tell us, a circle has little to say” or he might have meant more than that. We will tell three stories about the ellipse here together for the first time: one from linear algebra, one from complex analysis, and one from geometry. But before we tell you what these

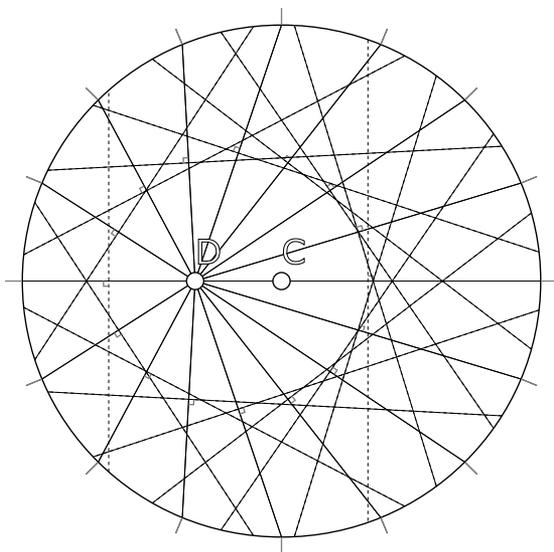


Figure 1.4. Many folds, using the applet.

stories are, we would like to present an interesting and, perhaps, unfamiliar way to construct an ellipse, a construction we first learned of in Martin Gardner’s book [53, pp. 173–183].

Here is how it works. Take out a piece of paper and choose (and label) two points, c and d . Now using c as your center, draw a circle of radius $r > |c - d|$.

Choose an arbitrary point, a_1 , on your circle and fold your paper over until the point a_1 lies on top of d . This will make a “crease” in your sheet, and you should make a fold there, creating a line segment. Choose another point, a_2 , on your circle and repeat, always folding over to the point d . Figure 1.3 shows what happens when we do this four times.

If you have a lot of time on your hands, you can do this with wax paper⁴ or you can obtain Figure 1.4 using our Ellipse Construction by Folding applet (we use this symbol ☺⁵ to tell you to go to an applet). In this figure, it looks like all the folds made using the slash marks are

⁴<https://www.youtube.com/watch?v=psuTYtDfxPE> (accessed 12/15/2017)

⁵<http://pubapps.bucknell.edu/static/aeshaffer/v1/>

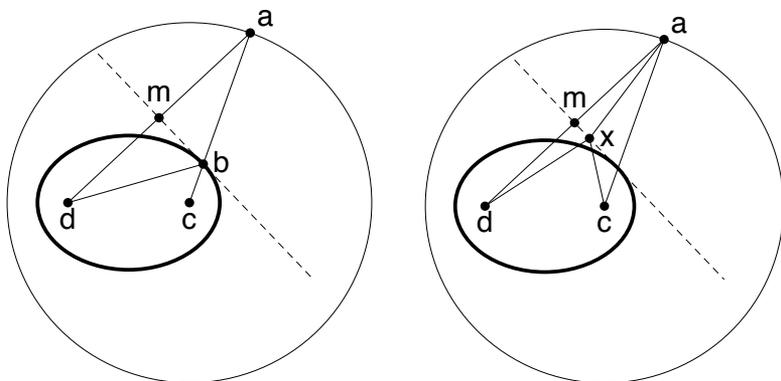


Figure 1.5. Left: Point b is on the ellipse. Right: Point x is not on the ellipse.

tangent to the ellipse with the points c and d as foci. In the applet, you can drag one of the foci or enter new ones to see that this *always* happens. Why would that be?

If we look at Figure 1.3 we see that the fold we made using points a_1 and d is the perpendicular bisector of the line segment $\overline{da_1}$. That is the crucial observation, and we are ready to see why this yields an ellipse.

Draw the fold obtained from folding point a onto d (see the picture on the left in Figure 1.5). Label the midpoint of the segment \overline{ad} by m . Now draw the radius, \overline{ac} , and call the point of intersection of the fold and the radius b . We will show that b lies on the ellipse with foci c and d and major axis of length r by showing that $|bd| + |bc| = r$, where r is the radius of the circle centered at c .

Here is how to check that our lengths have the right property. The two triangles $\triangle mdb$ and $\triangle mab$ are congruent because they share the side \overline{mb} , have right angles at m , and satisfy $|md| = |ma|$; that is the “side-angle-side” argument you learned in geometry. In particular, \overline{bd} has the same length as \overline{ab} . So if we look at the sum of the distances from b to the two fixed points c and d , we have

$$|bc| + |bd| = |bc| + |ab| = r.$$

In other words, the sum of the distances from the point b to the two fixed points c and d is always the same constant, r .

If we now take any point on the fold different from b , call it x (in the picture on the right in Figure 1.5), then the argument above shows that $|dx| + |cx| = |ax| + |cx|$. The points a , x , and c are not collinear, so by the triangle inequality, we conclude that $|dx| + |cx| > r$. So the only point that is on our ellipse and the fold is the point b , and we see that the fold is a tangent line to the ellipse.

This really does produce an ellipse! In fact, we can get every ellipse this way. (Think about how to do that.) There is another very important property hidden in this picture, and it will be useful later, so we isolate it as a proposition below. This proposition gives you what is often referred to as the optical or reflection property of an ellipse or what is sometimes more simply stated as “angle in equals angle out”.

Proposition 1.1. *Let E be an ellipse with foci c and d , and let b be a point on E that lies in the interior of the line segment $\overline{z_1 z_2}$. Then the line containing z_1 and z_2 is tangent to E if and only if $\angle cbz_1 = \angle z_2 b d$.*

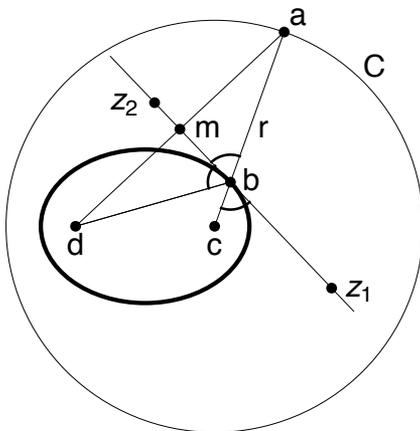


Figure 1.6. Tangents have equal angles with the lines to the foci.

Proof. We use the notation in Figure 1.6.

Given the foci and a point on the ellipse E , the length of its major axis is determined; we call it r . Draw a circle C with center c and radius r .

The point of intersection of C with the line through c and b is denoted by a . As shown above, the fold obtained by bringing a on top of d is a tangent line to this ellipse, m is the midpoint of \overline{ad} , and the length of segment \overline{db} is equal to the length of \overline{ab} . Thus, the point b is on the fold and the fold is the tangent line to E through b .

Now $\angle cbz_1$ and $\angle abz_2$ are vertical angles, so they are equal. The triangles $\triangle mbd$ and $\triangle mba$ are congruent since corresponding sides are of equal length. Thus, $\angle z_2bd = \angle mbd = \angle abm = \angle abz_2$. Putting this together, we get

$$\angle z_2bd = \angle abz_2 = \angle cbz_1,$$

which is what we wanted to show.

The converse follows from the fact that there is exactly one line through b for which $\angle z_2bd = \angle cbz_1$. \square

Figure 1.4 suggests other questions, and we list a few here. Suppose that when we fold we think of one of the points of intersection of the fold and the circle as the initial point and the other as the starting point for the next fold. Try to fold so that the second fold starts where the first ends and the third starts where the second ends. Does something special happen? Here are some related questions: Scaling things, we may suppose the boundary of the disk is the unit circle. Then we can ask: When is an ellipse inscribed in a triangle that is inscribed in the unit circle? What about a quadrilateral? Or, more generally, what about a convex polygon?

These are great questions, and they have great answers—answers that we present in the following chapters.

Chapter 2

The Ellipse Three Ways

What can an ellipse tell us? We consider three answers to this question that, on the surface, appear to be quite different. Beneath that surface, however, lies a surprising connection between matrix theory, function theory, and projective geometry.

Let us start with matrix theory and, in particular, with matrices that have complex entries. What does an ellipse know about a matrix? To tell this story, we need to review a bit of linear algebra and introduce the numerical range of a matrix.

An introductory course in linear algebra covers eigenvalues but rarely studies the numerical range in depth—if at all. Yet the numerical range could easily be included as the definition relies only on some familiarity with inner products on \mathbb{C}^n . So first recall that \mathbb{C}^n consists of elements of the form $x = [x_1 \ x_2 \ \dots \ x_n]^T$ with $x_j \in \mathbb{C}$ for all j . Now for x and y in \mathbb{C}^n , we consider the *standard inner product* $\langle x, y \rangle$, where

$$\langle x, y \rangle = \sum_{j=1}^n x_j \bar{y}_j.$$

In a general inner product space, the norm of x , or $\|x\|$, satisfies

$$\|x\|^2 = \langle x, x \rangle.$$

Thus, for $x \in \mathbb{C}^n$ with the standard inner product, we get the *Euclidean norm*

$$\|x\| = \left(\sum_{j=1}^n |x_j|^2 \right)^{1/2}.$$

Eigenvalues and the numerical range are connected and each provides us with valuable information about a matrix. Recall that for an

$n \times n$ matrix $A = (a_{ij})$ with complex entries, an *eigenvalue* of A is a complex number λ for which there exists a nonzero vector x such that $Ax = \lambda x$. We call x a corresponding *eigenvector*, but there are many of them; for example, $A(x/\|x\|) = \lambda(x/\|x\|)$ as well, so once we have an eigenvector we also have a corresponding unit eigenvector. (In fact, once we have one unit eigenvector we have many since all vectors $e^{i\theta}x/\|x\|$ with $\theta \in \mathbb{R}$ also work.) The vectors we use will usually have their entries in the complex plane, \mathbb{C} . So much for the familiar; let us now turn to the unfamiliar.

The *numerical range* of an $n \times n$ matrix A is defined by

$$W(A) = \{\langle Ax, x \rangle : x \in \mathbb{C}^n, \|x\| = 1\}.$$

The numerical range and eigenvalues are related: If λ is an eigenvalue of A with corresponding unit eigenvector y , then $\langle Ay, y \rangle = \langle \lambda y, y \rangle = \lambda$. So the eigenvalues are always in the numerical range.

The numerical range of a matrix was a natural object to study in the early days of Hilbert space theory because there was great interest in quadratic forms, and the numerical range of A is just the range of the quadratic form associated with A restricted to the unit sphere. But sometimes the numerical range provides more information about a matrix than the eigenvalues do: If you fix a point $\lambda \in \mathbb{C}$, there are several matrices for which the set of eigenvalues consists of the single point λ . But, as we will see in Theorem 6.2, there is only one matrix for which the numerical range is the set $\{\lambda\}$, and that matrix is a multiple of the identity matrix; that is, it is λI . Example 2.1 presents a matrix for which the only eigenvalue is zero, but its numerical range is more than just a singleton.

Example 2.1. *Let*

$$A_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

What is the numerical range of A_1 ?

To find the answer to this question, let $x \in \mathbb{C}^2$ of norm 1 be an arbitrary vector. Then

$$x = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$

where $\|x\|^2 = |z_1|^2 + |z_2|^2 = 1$. Thus, $|z_1| \leq 1$ and $|z_2| \leq 1$. Writing $z_1 = r e^{i\theta_1}$ with $r = |z_1|$ and $\theta_1 \in \mathbb{R}$, we see that $0 \leq r \leq 1$ and

$z_2 = \sqrt{1-r^2} e^{i\theta_2}$ for some appropriate choice of $\theta_2 \in \mathbb{R}$. Thus,

$$x = \begin{bmatrix} r e^{i\theta_1} \\ \sqrt{1-r^2} e^{i\theta_2} \end{bmatrix}.$$

So letting $\gamma = e^{i(\theta_2 - \theta_1)}$, we have $|\gamma| = 1$ and

$$\langle A_1 x, x \rangle = \gamma r \sqrt{1-r^2}.$$

Since x is an arbitrary vector in \mathbb{C}^2 , we may fix r and let γ vary. Doing so, we get a circle of radius $r\sqrt{1-r^2}$. Since $0 \leq r\sqrt{1-r^2} \leq 1/2$, we see that the numerical range of this matrix (and a very important matrix it is!) is the closed disk centered at 0 of radius 1/2 sketched in Figure 2.1.

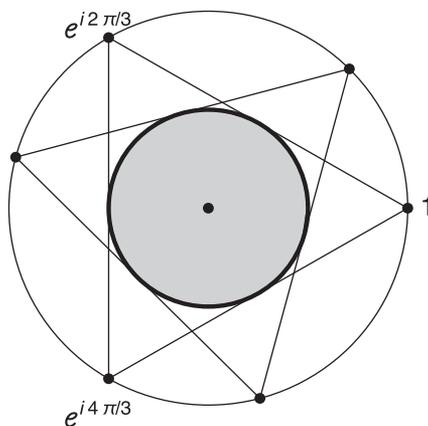


Figure 2.1. The disk in Examples 2.1, 2.4, and 2.7.

Let us look at a more interesting example.

Example 2.2. Let

$$A_2 = \begin{bmatrix} 0 & \sqrt{3}/2 \\ 0 & 1/2 \end{bmatrix}.$$

What is the numerical range of A_2 ?

It is harder to figure out what happens here (which is why it is more interesting!). In fact, $W(A_2)$ is the closed elliptical disk with foci at its eigenvalues, 0 and $1/2$, minor axis of length $\sqrt{3}/2$, and major axis of length 1 sketched in Figure 2.2. Rather than show you the computations here, we use the following theorem, which we prove in Chapter 6: For a matrix A , let A^* denote the *adjoint* of A ; that is, the matrix satisfying $\langle Ax, y \rangle = \langle x, A^*y \rangle$ for all $x, y \in \mathbb{C}^n$.¹ If $A = (a_{ij})$, the *trace* of A is denoted $\text{tr}(A)$ and defined by $\text{tr}(A) := \sum_{j=1}^n a_{jj}$. Then we have the following result [76, p. 109].

Theorem 2.3 (Elliptical range theorem). *Let A be a 2×2 matrix with eigenvalues a and b . Then the numerical range of A is an elliptical disk with foci at a and b and minor axis given by $(\text{tr}(A^*A) - |a|^2 - |b|^2)^{1/2}$.*

That is consistent with what we saw in Examples 2.1 and 2.2; we just need to think of the “two” foci as $a = 0$ and $b = 0$ and the ellipse as the circle of radius $1/2$ centered at 0.

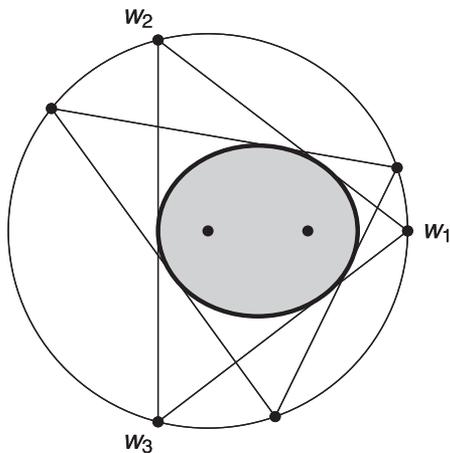


Figure 2.2. The elliptical disk in Examples 2.2, 2.5, and 2.8.

So the ellipse “knows” where the eigenvalues of the matrix are, but it knows more than that. Some of this is encoded in our second story,

¹If this is unfamiliar, you should check that if $A = (a_{ij})$, then $A^* = (\overline{a_{ji}})$.

one that at first blush seems most natural in this situation—geometry. But the theorem we will discuss here is one with which many people are not familiar. It is a beautiful and surprising theorem due to Jean-Victor Poncelet. The theorem falls into a general class of results that try to understand when, given two conics, you can find a polygon that circumscribes the smaller and is simultaneously inscribed in the larger. In addition to the examples we saw in Chapter 1, you can imagine that there are many versions of such problems—perhaps you can imagine infinitely many versions! We will discuss just a few of these in Chapter 5.

We state only the result we need here; a more general version will appear later. We are interested in ellipses that are inscribed in triangles that are inscribed in the unit circle \mathbb{T} . Not all ellipses can be thus inscribed—you may have seen this when you used the folding method to obtain an ellipse—and one of our goals will be to characterize such ellipses. For now we concentrate on providing two examples.

Example 2.4. *Consider the circle E_1 of radius $1/2$ centered at 0. Show that if α lies on the unit circle, then there is a triangle with one vertex at α and all vertices on the unit circle circumscribing E_1 ; that is, the edges of the triangle are each tangent to E_1 .*

For each point of the unit circle, the symmetry of the situation shows that if a triangle exists, it is equilateral, and once we have a triangle corresponding to a particular point, say $\alpha = 1$, then an appropriate rotation of this triangle will work for other values of α . It is not difficult to check that the triangle with vertices 1 , $e^{i2\pi/3}$, and $e^{i4\pi/3}$ is tangent to the circle at the point $z = -1/2$, and again, the symmetry of the situation makes it clear that this triangle circumscribes this circle. Looking at Figure 2.1 in this context, we are looking at triangles that circumscribe the circle.

Example 2.5. *Consider the ellipse E_2 with foci 0 and $1/2$ and major axis of length 1. Then for each point α on the unit circle, there is a triangle circumscribing the ellipse E_2 with vertices on the unit circle and one vertex at α .*

Lacking the symmetry present in Example 2.4, the justification of Example 2.5 is a fairly unpleasant computation, but it is an attractive application of a special case of Poncelet's theorem. Rather than proving

this result directly we apply Poncelet's theorem, which we state here and prove in Chapter 5.

Theorem 2.6 (Poncelet's closure theorem for triangles). *Let E denote an ellipse entirely contained in a second ellipse F . If it is possible to find one triangle that is simultaneously inscribed in F and circumscribed about E , then each point of F is a vertex of one such triangle.*

Before we discuss Example 2.5, let us think about what Poncelet's theorem is telling us. In Figure 2.3 on the left, the tangent lines we have drawn do not return to their starting point after three steps, so Poncelet's theorem tells us that no matter where we start, we will not return to our starting point after three steps. On the other hand, we see that in the picture on the right the lines return to the starting point after three steps regardless of where we begin the process.

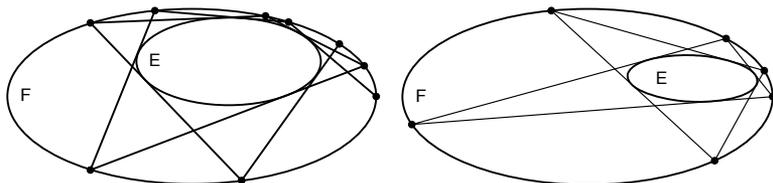


Figure 2.3. The ellipse E on the left is never inscribed in a triangle with vertices on F , while the ellipse E on the right always is.

In Example 2.5, the ellipse has Cartesian equation

$$4\left(x - \frac{1}{4}\right)^2 + \frac{16}{3}y^2 = 1.$$

Consider the triangle with vertices at $w_1 = (1, 0)$, $w_2 = (-1/4, \sqrt{15}/4)$, and $w_3 = (-1/4, -\sqrt{15}/4)$. The line segment $\overline{w_2w_3}$ touches the ellipse only at the point $(-1/4, 0)$ and is therefore tangent to it at that point. If we check the line joining w_1 to w_2 we find that this is also tangent to the ellipse at the point $(7/12, \sqrt{15}/12)$ and the final point of tangency follows again by symmetry. Thus, the ellipse is inscribed in a triangle, that triangle is inscribed in the unit circle, and we are looking at triangles that circumscribe the ellipse in Figure 2.2.

Now that we know what happens at one point, Poncelet's theorem says that at every point on \mathbb{T} there is a triangle circumscribing the ellipse that is itself inscribed in the unit circle and has one vertex at the given point. It should amaze you that once you have checked the behavior at *one* point on the outer ellipse, you know something about the behavior about *every* point on the outer ellipse.

We now turn to our final set of examples: The class of finite Blaschke products, which were first introduced by the German mathematician Wilhelm Blaschke. As the word *finite* suggests, there are also infinite Blaschke products, and though they are very important, we will not discuss them at this time. (A brief discussion of infinite Blaschke products can be found on p. 111.) Thus, we refer only to Blaschke products, which are functions of a complex variable that have the form

$$B(z) = \mu \prod_{j=1}^n \frac{z - a_j}{1 - \overline{a_j}z},$$

where n is a positive integer, $a_j \in \mathbb{D}$ for $j = 1, \dots, n$, and $\mu \in \mathbb{T}$. The number of zeros of B , counted according to multiplicity, is called the *degree* of B ; that is, the degree of B is n . Taking $a_j = 0$ for all j yields special examples of Blaschke products, namely, μz^n with $\mu \in \mathbb{T}$, and we know how important these are.

Now something truly magical will happen. There is nothing about these functions that suggests a relation—even a distant relation—to matrices or ellipses. Or is there? To see what happens we will mention the most basic properties of these functions, all of which can (and will—see Chapter 3 for the proofs) be shown easily.

Blaschke products map the open unit disk \mathbb{D} to itself and the unit circle \mathbb{T} to itself; they are analytic on an open set containing the closed unit disk and have finitely many zeros in \mathbb{D} . In fact, if an analytic function satisfies the first three of these conditions (which imply the fourth), it must be a Blaschke product—another reason why they are considered an important class of functions. We establish this characterization of finite Blaschke products in Lemma 3.2.

Blaschke products have very nice mapping properties, and one such property is that a Blaschke product is an n -to-1 map of the unit circle

onto itself. As a consequence of this the n solutions to $B(z) = \lambda$ are distinct when $\lambda \in \mathbb{T}$. What can an ellipse possibly know about a Blaschke product? Here the connection is well hidden but was recently discovered in [35].

Example 2.7. Let B_1 denote the Blaschke product $B_1(z) = z^3$. For each $\lambda \in \mathbb{T}$, the three solutions of $B_1(z) = \lambda$ form the vertices of a triangle. Denote the closed region bounded by this triangle by T_λ . Then $\bigcap_{\lambda \in \mathbb{T}} T_\lambda$ is a closed circular disk of radius $1/2$ centered at the origin.

This is easy to see if we think geometrically: For each $\lambda = e^{i\theta}$ the three solutions to $z^3 = e^{i\theta}$ are equally spaced on the unit circle. Therefore, we are finding the intersection of all triangular regions, where our triangles are equilateral triangles with vertices on the unit circle. Evidently, this is a closed circular disk centered at the origin. What is the radius? We can choose any triangle to determine the radius, and the triangle with vertices $1, e^{i2\pi/3}$, and $e^{i4\pi/3}$ works nicely. The circle is tangent to this triangle on the line joining $e^{i2\pi/3}$ and $e^{i4\pi/3}$ at the point $z = -1/2$. Thus, the radius is $1/2$ and we are back to Figure 2.1 again.

Example 2.8. Let B_2 denote the Blaschke product $B_2(z) = z^2(z - 0.5)/(1 - 0.5z)$ and $\lambda \in \mathbb{T}$. By the discussion above, B_2 is a 3-to-1 map of the unit circle onto itself. Therefore, for each $\lambda \in \mathbb{T}$, there are three distinct solutions to $B_2(z) = \lambda$, and they are the vertices of a triangle. Let T_λ denote the closed triangular region bounded by the triangle. Then $\bigcap_{\lambda \in \mathbb{T}} T_\lambda$ is the region enclosed by the ellipse with foci at the points 0 and 0.5 and major axis of length 1 .

This second example is more difficult but can be computed with the aid of the following theorem, the proof of which will have to wait until Chapter 3.

Theorem 2.9. [35] Let B be a Blaschke product of degree 3 with zeros $0, a$, and b . For $\lambda \in \mathbb{T}$, let z_1, z_2, z_3 denote the solutions of $B(z) = \lambda$. Then the lines joining z_j and z_k , for $j \neq k$, are tangent to the ellipse E given by the equation

$$|w - a| + |w - b| = |1 - \bar{a}b|. \quad (2.1)$$

Conversely, every point of E is the point of tangency of the ellipse with a line segment joining two points on \mathbb{T} that B identifies.

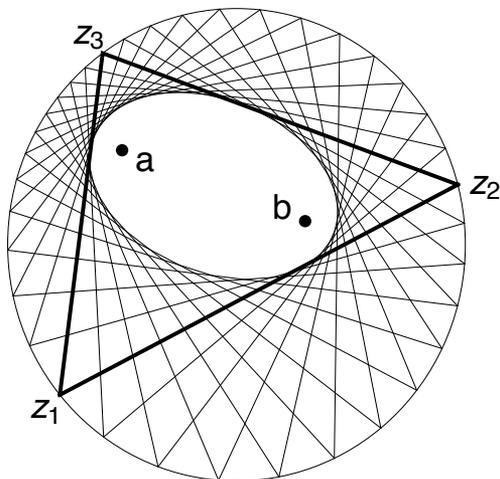


Figure 2.4. A Blaschke ellipse with foci at a and b illustrating Theorem 2.9.

The equation of the ellipse, (2.1), makes it quite clear that given a point on the ellipse, the sum of the distances from each of the foci, a and b , is a constant ($|1 - \bar{a}b|$ in this case). It follows from Theorem 2.9 that $\bigcap_{\lambda \in \mathbb{T}} T_\lambda$ is the closed elliptical disk with foci 0 and $1/2$ and major axis of length 1 . But there is only one ellipse with these foci and this major axis. Therefore, $\bigcap_{\lambda \in \mathbb{T}} T_\lambda$ is the region bounded by the ellipse in Figure 2.2.

So Examples 2.1, 2.4, and 2.7 produce the same ellipse, though in the first case the foci are the eigenvalues of a matrix, while in the last case the foci are two of the zeros of a Blaschke product. Similarly, Examples 2.2, 2.5, and 2.8 produce the same ellipse with foci at the eigenvalues of a particular matrix in the first case and foci at two of the zeros of the Blaschke product in the last case—yet there seems to be no apparent connection between numerical ranges of 2×2 matrices, Poncelet's theorem, and degree-3 Blaschke products. Our goal is to show that the appearance of the ellipse, in each case, is not a coincidence but rather the result of deep and beautiful mathematical interconnections.