Where Is the Directrix of a Circle?

David E Dobbs

The Case of the Missing Directrix

The conic sections (parabola, ellipse and hyperbola) get their name from the fact that each can be obtained as the intersection of a plane with a double-napped right circular cone. By changing the position of the plane relative to the cone, one finds that certain positions produce an intersection that is typically referred to as a degenerate or limiting case of a conic, such as a point, a line, two intersecting lines or a circle. In particular, a circle can thus be viewed as a limiting case of an ellipse (see, for instance, Dobbs and Peterson 1993, Figure 8.1, 441).

Other unifying approaches to introducing the conic sections have a similar feature. Consider, for instance, the approach that involves the shadow cast on a wall by the nonfluted lampshade of a reading lamp. If the shade is pointed almost directly at the wall, the shadow is an ellipse that is nearly circular. As the lamp is gradually tilted more and more, the elliptic shadow becomes less circular until it becomes a parabola. Increased tilting eventually produces a shadow that consists of both branches of a hyperbola. Reversing the tilting process, one finds that the hyperbola changes back into a parabola, then an ellipse and finally, when the shade is pointed directly at the wall, a circle. In this way, a circle can once again be physically obtained as a limiting case of an ellipse.

A third approach to the three basic conic sections is the one usually used in high schools today—namely, as graphs of quadratic polynomials in two variables (with real number coefficients). However, the graph of such a polynomial can also be a point, a line, two intersecting lines, two parallel lines, a circle or (most degenerate of all) an empty set. Elsewhere, I (Dobbs 1992, 803) examined the effect of subjecting a familiar equation of an ellipse, $x^2/a^2 + y^2/b^2 = 1$, with parameters $a > b > 0$, to three limiting processes of the kind studied in precalculus and calculus. The result was to produce equations whose graphs were two parallel lines, a line segment or a circle. For example, one can obtain the upper half of the above ellipse as the graph of the function $f(x) = b(1 - x^2/a^2)^{1/2}$ over the domain $-a \leq x \leq a$. Now, if a is fixed and we let b approach a from the left, we have the one-side limit $\lim_{b \to a^-} f(x) = b(1 - x^2/a^2)^{1/2} = (a^2 - x^2)^{1/2}$, a function whose graph (over the above domain) is the upper half of the circle whose equation is $x^2 + y^2 = a^2$. When the same limiting process is applied to an equation of the lower half of the ellipse, the limit is a function whose graph is the lower half of the circle. Thus, in a sense that could be very effective in a precalculus classroom, we have seen an algebraic way to view a circle as a limiting case of an ellipse.

The last approach mentioned above is part and parcel of studying conic sections through analytic geometry. In this approach, a conic section with focus F, directrix L and eccentricity e is the set of points P such that e is the ratio of the distance from P to F and the distance from P to L. The familiar conics are obtained as follows: parabolas have eccentricity $e = 1$, ellipses have eccentricity satisfying $0 < e < 1$ and hyperbolas satisfy $e > 1$. It is customary to say that circles are ellipses with eccentricity 0. This makes some sense if one views the circle $x^2 + y^2 = a^2$ as having been obtained through the limiting process considered above. Indeed, the foci of the above ellipse are the points $(c,0)$ and $(-c,0)$, where $c^2 + b^2 = a^2$ and $c = ae > 0$, and the limit process $\lim_{b \to a^-}$ sends c to 0. (More precisely, $\lim_{b \to a^-} c = \lim_{b \to a^-} (a^2 - b^2)^{1/2} = 0$.) The effect of the limiting process is to identify the foci with the centre of the limiting circle. Moreover, an ellipse whose eccentricity is a small positive number is only slightly oval and is often indistinguishable from a circle to the naked eye. The mathematical use of the term *eccentricity* comes from the fact that one can view an ellipse as having evolved from a circle whose centre has split into two foci, with the distance c from the centre of the ellipse to either focus measuring the amount that each focus has moved away from the centre. (The Latin origins of the terminology reveal this interpretation, with *ex* meaning "away from" or "out of" and *centrum* meaning "centre.")

The above point of view leads to the basic question we will study in this article. We have seen how algebra (together with functions and limits) allows us to view a circle as the limit of an ellipse, and we have also seen how that limiting process converts the foci of the ellipse to the centre of the circle. Our basic question is, What happens to the directrices of the ellipse under that limiting process? Since the circle can be viewed as a degenerate conic, it should have at least one focus and a corresponding directrix. If we view the centre of a circle as its focus, where is the corresponding directrix of the circle?

Naysayers may point out that the above ellipse has directrices $x = \pm a/e$ and that there would be no sense in considering these equations for a circle—which we have seen should have eccentricity $e = 0$—since it is said that you can't divide by 0. We will show that much mathematics has been based on the refusal to let the circle-as-conic analogy die at the hands of that tired bromide. Parts of this article could be useful in high school courses in algebra, geometry, precalculus/functions, and calculus, especially as enrichment material for the unit on one-sided or infinite limits. It is also hoped that geometry teachers will find this information effective in introducing students to the line at infinity and, more generally, to projective geometry and modern algebraic geometry.

Can Algebra Explain the Nature of $x = \infty$?

We are trying to avoid having to say that the directrices of the circle $x^2 + y^2 = a^2$ are given by $x = \pm a/0$. How can we do this? For inspiration, let's recall that the founders of calculus (especially Leibniz, with his *dy/dx* notation) managed to view the derivative of a function as a ratio whose denominator was infinitesimally small. Nowadays, we view the derivative as the limit of a certain ratio whose denominator is approaching 0. Let's take a similar approach in addressing our current difficulty. Rewrite the above equation of an ellipse as $x^2/(b + \varepsilon)^2 + y^2/b^2 = 1$, where $a = b + \varepsilon$ and $\varepsilon > 0$. The limit process $\lim_{b \to a}$ is equivalent to $\lim_{\varepsilon \to 0^+}$ (where b is fixed and a is varying). What is the effect of applying this limit process to the right-hand directrix $x = a/e$ of the above ellipse? It is

$$x = \lim_{\varepsilon \to 0^+} \frac{b+\varepsilon}{e}$$

$$= \lim_{\varepsilon \to 0^+} \frac{b+\varepsilon}{\frac{((b+\varepsilon)^2 - b^2)^{1/2}}{b+\varepsilon}}$$

$$= \lim_{\varepsilon \to 0^+} \frac{b+\varepsilon}{\left(1-\left(\frac{b}{b+\varepsilon}\right)^2\right)^{1/2}}.$$

In this limit problem, the numerator has limit b, which is a fixed positive number, and the denominator has limit 0, which is taken through positive values. Calculus teaches us how to understand such a limit. It is $x = \infty$.

We have managed to reinterpret $x = a/0$ as $x = \infty$. Are we any better off now? Yes, indeed. In this section, we will use an engineer's numerical and algebraic intuition to try to interpret the equation $x = \infty$ graphically. That effort will be instructive but ultimately unsuccessful. In the next section, we will succeed in interpreting the graph of $x = \infty$ by turning to the machinery of the real projective plane.

Let's focus here on the following question: Can $x = \infty$ be equivalent to (that is, have the same solution set as) some homogeneous linear equation $ax + by + cz = 0$ (where a, b and c are real numbers, not all of which are 0)? This leads to the question, Which ordered pairs (y,z) of real numbers can satisfy an equation $a\infty + by + cz = 0$? Let's assume that there are such points (whatever the directrix of a circle ends up being, it should not be empty!). Then $a\infty = -by - cz$ must be a real number. Our experience with products involving the ∞ symbol in calculus—along with common sense—tells us that a must equal 0. The only restriction on (y,z) is then that $by + cz = 0$. We proceed to understand the graph of this equation for different possible values of b and c.

One situation can be ruled out at once. Indeed, it cannot be the case that both b and c are 0, for the graph of $by + cz = 0$ would then consist of all ordered pairs (y,z) of real numbers. That is intuitively unacceptable because a directrix should be a line, not a plane. However, the three remaining cases each lead to a plausible interpretation. For instance, if $b = 0$ and $c \neq 0$, the graph of $by + cz = 0$ consists of the ordered pairs $(y,0)$, where y varies over the set of all real numbers. With one degree of freedom, this set could possibly be viewed as a line in some new geometry. By similar reasoning, the same type of conclusion holds for the case in which $c = 0$ and $b \neq 0$. In the final case, neither b nor c is 0. The graph of $by + cz = 0$ in this case is more complicated: it consists of all the points of the form $(y,-by/c)$, where, once again, y can be any real number. With one degree of freedom, this set could also plausibly be viewed as a kind of line.

Has the above intuitive analysis involving algebraic operations with the ∞ symbol been of any help? Not really! We have gone from having had no ready interpretation for the graph of $x = \infty$ to having three equally plausible interpretations. The feast is not preferable to the famine. We wanted one answer, not several. Since algebra (mixed with intuition from calculus) has not provided a satisfactory answer, we turn next to geometry in our quest to understand the graph of an equation such as $x = \infty$.

Projective Geometry Explains the Nature of $x = \infty$

In an intuitive sense, one might think that the graphs of equations such as $x = \infty$ and $y = \infty$ should be lines at infinity. In fact, there is an extension of the ordinary analytic geometry of the Euclidean plane where notions similar to these can be given rigorous mathematical meaning. That larger mathematical system is known as the real projective plane. Analytically, a point in the real projective plane is an ordered triple (x,y,z) of real numbers, not all of which are 0, with two such triples viewed as being the same if their corresponding components are proportional. (More precisely, this notion of *same* means that the points in the real projective plane are actually the equivalence classes arising from a certain equivalence relation on the set of certain ordered triples of real numbers.) Analytically, a line in the real projective plane is the graph of a homogeneous linear equation in the variables x, y and z. (The proportionality that defined sameness of points ensures that any two identified non-zero triples of real numbers satisfy the same homogeneous linear equations.)

The ordinary point (x,y) of the familiar analytic geometry of the real Euclidean plane can be regarded as the point $(x,y,1)$ of the real projective plane. The only other points of the real projective plane are of two types: the infinitely many points $(1,y,0)$, which are different for different values of y, and the point $(0,1,0)$. It is customary to say that these two types of points are points at infinity. Notice that the points at infinity are exactly the graph of the equation $z = 0$, which is then naturally called the line at infinity.

With the projective machinery now in hand, graphs of some familiar equations of lines in the real Euclidean plane become subsumed as subsets of projective lines in the real projective plane, as follows. The nonvertical, nonhorizontal line $y = mx + b$ (where $m \neq 0$) becomes part of the projective line given by $y = mx + bz$. Apart from the familiar points on the real Euclidean line $y = mx + b$, the only new point on this projective line is the point $(1,m,0)$. The x-axis, $y = 0$, becomes part of the projective line given by the same equation. Apart from the familiar points on the real Euclidean x-axis, the only new point on this projective line is the point $(1,0,0)$. Similarly, the y-axis, $x = 0$, is subsumed as part of the projective line $x = 0$, whose only new point is $(0,1,0)$.

Something similar happens when we try to embed the other horizontal or vertical lines of the real Euclidean plane into the projective environment. As c varies over the set of non-zero real numbers, the familiar horizontal line $y = c$ becomes part of the projective line $y = cz$, whose only new point is $(1,0,0)$. Notice that if c_1 and c_2 are unequal non-zero real numbers, then the projective lines $y = c_1z$ and $y = c_2z$ intersect at the point $(1,0,0)$, which is the same point at infinity that lies on the projective line $y = 0$. In fact, parallelism is not a useful concept in projective geometry, because you can check that *any* two distinct projective lines meet at exactly one point (which may be on the line at infinity).

The situation is similar when we extend the familiar vertical line $x = c$, with $c \neq 0$, to the projective line $x = cz$. Indeed, distinct projective lines $x = c_1z$ and $x = c_2z$ intersect at the point $(0,1,0)$, which is the same point at infinity that lies on the projective line $x = 0$.

Are we now ready to make any sense out of graphs of expressions such as $x = \infty$ and $y = \infty$? Yes! The process by which we embedded each Euclidean line as a subset of some projective line involved what algebraic geometers call homogenization: the variables x and y appearing in a Cartesian equation of a given Euclidean line L are replaced by x/z and y/z, respectively, so that cross-multiplying produces an equation of the projective line in which L is embedded. Since we arrived at the equation $x = \infty$ by using the analytic geometry of the Euclidean plane, it follows that an interpretation of $x = \infty$ in terms of the projective plane should be (after homogenization) as the graph of $x = z\infty$. What on earth is this?

Once again, our experience with calculus (or ordinary common sense) tells us that if z is a non-zero real number, then $z\infty = \pm\infty$, which is certainly not a real number. Thus, in the real projective plane, each point (x,y,z) on the graph of $x = z\infty$ must satisfy $z = 0$. In other words, the graphical interpretation of $x = \infty$—which we wanted to be a directrix and, hence, some sort of line—is that it is a subset of the line at infinity. But surely a line cannot be a proper subset of another line. The conclusion is inescapable: the graph of $x = \infty$ is the line at infinity.

By reasoning with homogenization as above, you can check that the graph of $y = \infty$ is also the line at infinity. Thus, to find a geometric answer to our basic question, we have come upon a geometry in which parallelism no longer matters and we can no longer tell horizontal from vertical. Moreover, now that we have argued that the line at infinity should be the directrix of the "circle" $x^2 + y^2 = z^2a^2$ (obtained from the Euclidean equation $x^2 + y^2 = a^2$ by homogenization), *focus* and *directrix* in projective geometry cannot continue to play their former roles. After all, *any* point P of the Euclidean plane is at a finite distance from the focus of this circle and at an infinite distance from the directrix. The ratio of these distances should surely be understood as 0 (since algebra, calculus and common sense agree that if a/∞ is to have a meaning for some real number a, that meaning must be 0). Since the circle has eccentricity 0, our earlier understanding of the terms *focus* and *directrix* would seem to imply that *each* point of the Euclidean plane lies on the circle $x^2 + y^2 = a^2$ that we started with. That conclusion is unacceptable, since the real projective plane is supposed to be a reasonable extension of the ordinary real Euclidean plane, where the only new phenomena involve the points (and line) at infinity. For this reason, we must abandon our earlier understanding of *focus* and *directrix* when working in projective geometry. In fact, the very definition of *conics* must be formulated anew in this geometry.

H S M Coxeter, probably the most distinguished geometer in Canada's history, wrote often on this subject, including an accessible introduction to projective geometry (1964). A synthetic (in other words, non-analytic) approach to the real projective plane can be found in his book *The Real Projective Plane* (1993). Coxeter writes that "in the projective plane, there is only one type of conic; the familiar distinction between the ellipse, parabola, and hyperbola can only be made by assigning a special role to the line at infinity" (p 72). Thus, one consequence of enlarging the Euclidean plane to the projective plane is that we lose part of what we had thought we knew about conics. A venerable maxim in education is that to increase our understanding of a subject, it is often necessary to take one step backward before taking two steps forward. This is exactly what has happened as we have allowed considerations of infinity to affect our view of what *point* and *line* could mean in an extension of the Euclidean plane.

A survey of textbooks reveals that the notion of *conic*, when broadened beyond the Euclidean context as indicated above, plays a central role in most treatments of projective geometry. It is heartening to note that Coxeter's *Projective Geometry* (1964, 102–03) contains a short section called "Is the Circle a Conic?" and that he provides an elegant proof that answers this question with a resounding yes. One might say that in the real projective plane, the analogies with which we began have come full circle (pun intended).

Closing Comments

A nugget of literary wisdom seems appropriate here. In "Tintern Abbey," poet William Wordsworth writes, "Other gifts have followed." He is referring to the familiar process of mellowing whereby one discovers compensations while aging, but I believe that his words also have relevance for us here. In pursuing the possible meaning of $x = \infty$, we seem to have lost a role that directrices played in some more familiar situations. However, in losing a role, we have gained the entire subject of projective conics. An accessible book by Kendig (2005), which comes packaged with a CD containing 36 applets, includes eight ways of looking at a conic. Kendig also views conics within the broader framework of algebraic curves in projective spaces with complex number coordinates. Thus, Kendig's text could be used as an introduction to modern algebraic geometry that builds on the discussion of the real projective plane in the preceding section of this article.

Modern algebraic geometry has grown to encompass much more than projective geometry. By refusing to be stopped by the bromide that you can't divide by 0, mathematicians have opened up an entire area of geometry, where new applications to science and other areas of mathematics are still being discovered.

Two other recommended books serve as introductions to the computer-facilitated methods of modern algebraic geometry: *Algebraic Geometry for Scientists and Engineers* (Abhyankar 1990) is written for scientists and engineers, and *Using Algebraic Geometry* (Cox, Little and O'Shea 2005) is written in the spirit of modern commutative algebra.

References

Abhyankar, S S. 1990. *Algebraic Geometry for Scientists and Engineers*. Providence, RI: American Mathematical Society.

Cox, D A, J Little and D O'Shea. 2005. *Using Algebraic Geometry*. 2nd ed. New York: Springer.

Coxeter, H S M. 1964. *Projective Geometry*. New York: Blaisdell.

— —. 1993. *The Real Projective Plane*. 3rd ed. With an appendix by G Beck. New York: Springer-Verlag.

Dobbs, D E. 1992. "Limits of Conics." *International Journal of Mathematics Education in Science and Technology* 23: 801–05.

Dobbs, D E, and J C Peterson. 1993. *Precalculus*. Dubuque, Iowa: Brown.

Kendig, K. 2005. *Conics*. Washington, DC: Mathematical Association of America.

David E Dobbs received his BA (majors: mathematics, statistics and philosophy) and his MA (major: geometry, minor: logic) from the University of Manitoba in his native Winnipeg. He received his doctorate in algebraic geometry from Cornell University in Ithaca, New York. He has been on staff at the University of California, Los Angeles (UCLA); Rutgers University in New Jersey; and the University of Tennessee in Knoxville, where he has held the position of professor since 1980. He has taught mathematics education courses at each of the five universities named above, and he has written almost 90 teaching notes at all levels of mathematics education. He enjoys teaching at all levels, and has directed seven PhD dissertations and several dozen senior and master's theses. His main research interest is commutative algebra, a subject on which he has written eight books and more than 200 articles.