Classical transcendental curves

Reinhard Schultz

May, 2008

In his writings on coordinate geometry, Descartes emphasized that he was only willing to work with curves that could be defined by algebraic equations. He was not able to find such equations for some important curves from classical Greek geometry, including the quadratrix/trisectrix of Hippias (*c.* 460 B.C.E.– *c.* 400 B.C.E.), and he excluded them from his setting by describing them as "mechanical." Several decades later, Leibniz took a much different view of the situation, recognizing that curves that are not definable by algebraic equations can — and in fact **should** — also be studied effectively using the methods of coordinate geometry and calculus.

Even though the Leibniz viewpoint is now universally accepted in analytic geometry and calculus, one can still ask whether certain classical Greek curves in the plane with no reasonably simple description by an algebraic equation are indeed not definable by an algebraic equation $F(x, y) = 0$, where F is a nontrivial polynomial in x and y with real coefficients. The purpose of this note is to prove that four important examples have no description of this type. One is the quadratrix/trisectrix of Hippias, another is the Archimedean spiral, which is given in polar coordinates by $r = \theta$ and in rectangular coordinates by

$$\mathbf{s}(t) \quad = \quad (t\cos t,\, t\sin t)$$

yet another is the catenary which is the graph of the function $y = \cosh x$, and the fourth is the cycloid curve, which has the following standard parametrization:

$$x(t) \quad = \quad t \, - \, \sin t \, , \qquad\qquad y(t) \quad = \quad 1 \, - \, \cos t$$

Finally, we shall consider a classical family of curves known as *epicycles* and show that some are algebraic but others are not.

In order to analyze the examples described above, we shall need some background results which state that the so-called *elementary transcendental functions* do not satisfy identities of the form $F(x,\, f(x)\,) \equiv 0$, where F is a strongly nontrivial polynomial in two variables (in other words, it is not a polynomial which is only a function of the first or second variable). This property is reflected by the use of the word "transcendental" to describe these functions, but proofs of such results do not appear in standard analytic geometry and calculus texts for several reasons (in particular, the mathematical level of such proofs is well above the levels suitable for basic courses in single variable calculus, and the results themselves are not needed for the usual applications of calculus to problems in other subjects). The following online reference contains explicit statements and proofs that exponential functions, logarithmic functions, and the six standard trigonometric functions do not satisfy the types of algebraic equations described above:

`http://math.ucr.edu/~res/math144/transcendentals.pdf`

The discussion here will be at about the same mathematical level, using some input from advanced undergraduate and beginning graduate algebra courses together with standard results from calculus and differential equations courses. The following companion document contains figures related to the discussion here:

`http://math.ucr.edu/~res/math153/transcurves2.pdf`

Additional information on Descartes' notion of mechanical curves appears in Section 13.3 of the book by Stillwell cited in Section 6.

1. Algebraic curves and analytic parametrizations

Our goal is to prove a useful result which shows that if certain types of parametrized curves satisfy an algebraic equation near a point then they do so everywhere.

PROPOSITION. *Suppose that we are given a parametrized curve*

$$\mathbf{s}(t) \quad = (\, x(t),\, y(t)\,)$$

defined on an open interval J containing t_0, where the parametric equations are real analytic functions on J. If there is a strongly nontrivial polynomial $F(x, y)$ such that $F \circ \mathbf{s}(t) = 0$ for all t sufficiently close to t_0, then this equation holds for all $t \in J$.

Proof. Standard considerations show that the composite function $F \circ \mathbf{s}(t)$ is real analytic on J (use Fact 2 in Section I.4 of `transcendentals.pdf`), and it is zero on some subinterval $(t_0 - \delta, t_0 + \delta)$. Therefore, by Fact 3 from the section cited in the previous sentence we know that $F \circ \mathbf{s}(t)$ is zero everywhere on J.■

2. The Catenary

This is the curve whose equation is

$$y \quad = \quad \cosh x \quad = \quad \frac{e^x + e^{-x}}{2}$$

and we claim that it is not algebraic. If it were, then we would have a polynomial identity $G(x, \cosh x) \equiv 0$, and by the results of `transcendentals.pdf` the latter would imply that for some positive integer N the set of all functions of the form $x^j\, e^{kax}$ — where $|j|, |k| \leq N$ — is linearly dependent.

However, as in Section I.2 of `transcendentals.pdf` we know that the given functions form a basic set of solutions for the linear differential equation

$$D^{N+1}(D^2 - a^2)^{N+1}(D^2 - 4a^2)^{N+1} \; ... \; (D^2 - N^2 a^2)^{N+1} y \quad = \quad 0$$

and therefore no such polynomial identity can exist.■

3. The Archimedean Spiral

As noted above, the standard equation defining this curve **AS** in polar coordinates is $r = \theta$, where $\theta \geq 0$, and this yields the parametrization $\mathbf{s}(\theta) = (\theta\cos\theta,\, \theta\sin\theta)$, where $\theta \geq 0$. We shall show that there is no nonzero point $\mathbf{p}$ on this curve for which one can find an open neighborhood U containing $\mathbf{p}$ and a strongly nontrivial polynomial G such that $G \equiv 0$ for all points on $\mathbf{AS} \cap U$.

By the proposition in the preceding section and the existence of real analytic parametric equations for **AS**, if one can find a point, open neighborhood and polynomial as above, then it follows that $G \equiv 0$ on all points of **AS**. Denote the polynomial in question by $G(x,y) = \sum_{i,j}\; a_{i,j}x^i y^j$. Set $H_i(y) = \sum_j\; a_{i,j}y^j$, so that $G(x,y) \;=\; \sum_i\; H_i(y)\cdot x^i$.

CLAIM 1: *Each horizontal line $y = b$ and each vertical line $x = c$ meets* **AS** *in infinitely many points.*

This seems clear if one sketches the curve (see Figure 1 in the file `transcurves2.pdf`), and on the coordinate axes we know that the points $\big((-1)^k k\pi, 0\big)$ lie on both the spiral and the x-axis while the points $\big(0, (-1)^k(k+\frac{1}{2})\pi\big)$ lie on both the spiral and the y-axis. Since the problem is symmetric in x and y, we shall only prove the statement regarding horizontal lines other than the x-axis, so that $c \neq 0$; the argument in the vertical case is similar, and in fact we shall not need this case subsequently.

The first step is to notice that if (x,b) lies on **AS** and $c \neq 0$ and $(x,b) = \mathbf{s}(\theta)$, then we have

$$\theta^2 \;=\; x^2 \;+\; b^2 \qquad \text{and} \qquad \cot\theta \;=\; \frac{x}{b}\;.$$

This is illustrated in Figure 2 from the file `transcurves2.pdf`. We need to prove that, for each nonzero real number b, there are infinitely many values of X which solve the resulting equation in x:

$$x \;=\; b\cdot\cot\sqrt{x^2+b^2}$$

If we make the change of variables $u = \sqrt{x^2+b^2}$, this equation can be rewritten in the form $b\cot u = \sqrt{u^2-b^2}$, and the goal translates to showing that there are infinitely many solutions to this curve for which $u > |b|$. In the discussion below, k will denote an arbitrary positive integer such that $k > |b|/\pi$.

CLAIM 2: *If $h(u) = b\cot u - \sqrt{u^2-b^2}$, then for each k as above there is a real number u_k such that $k\pi < u_k < (k+1)\pi$ and $h(u_k) = 0$.*

The proof of this is similar to the proof that there are infinitely many solutions to the equation $\tan x = x$. Let ε be the sign of b. Then one has the following one-sided limit formulas:

$$\lim_{u\to k\pi^+}\; \varepsilon\cdot h(u) \;=\; +\infty \qquad\qquad \lim_{u\to (k+1)\pi^-}\; \varepsilon\cdot h(u) \;=\; -\infty$$

It follows tht there is some real number u_k between $k\pi$ and $(k+1)\pi$ such that $\varepsilon\cdot h(u_k) = 0$, and the claim follows because $\varepsilon\, h(u) = 0$ if and only if $h(u) = 0$.

Completion of the proof that **AS** *is not algebraic.* Suppose that G is a polynomial in two variables such that $G(x,y) = 0$ for all (x,y) on **AS**. For each real number c we know there are infinitely many x such that $G(x,c) = 0$. If $G(x,y) = \sum_{i,j} \; a_{i,j}x^i y^j = \sum_i \; H_i(y)x^i$, then for each c the expression $G(x,c)$ is a polynomial in x which has infinitely many roots, and therefore the coefficients $H_i(y)$ must be zero for all i. Since $H_i(c) = 0$ for all c, it follows that for each i we have $a_{i,j} = 0$ for all j; but this means that all the coefficients $a_{i,j}$ must vanish. This completes the proof that **AS** is not algebraic (in fact, it is not algebraic even if one restricts to some open interval of the positive real line).■

4. The Quadratrix/Trisectrix of Hippias

The standard equation for this curve is $y = x\,\cot x$ for $x \neq 0$; this limit of the right hand side as $x \to 0$ is equal to 1 (*e.g.*, this follows from L'Hospital's Rule or more elementary considerations — see the discussion below), so it is customary to add the point $(0,1)$ so that the curve becomes continuous for $|x| < \pi$. Since the function $\cot x$ has convergent power series expansions on the intervals $(-\pi, 0)$ and $(0, \pi)$, clearly the same is true for $x\,\cot x$.

In fact, we should note that *this function also has a convergent power series expansion on an open interval containing the origin* (the remainder of this paragraph may be skipped without loss of continuity). — To prove the existence of such an expansion, first note that $x/\cot x$ is the reciprocal of the function $\tan x/x$ and that $\tan x$ has a power series expansion of the form $x \cdot g(x)$, where g is a convergent power series with a nonzero constant term because

$$\lim_{t \to 0} \frac{\tan x}{x} \quad = \quad 1$$

which can be shown either by appealing to L'Hospital's Rule or by a more direct elementary argument using the fact that $\sin x/x$ has the same limit as $x \to 0$. Therefore the function $f(x) = \tan x/x$ (extended by setting $f(0) = 1$) is given by the convergent power series $g(x)$. But now standard results on power series state that if g is representable by a convergent power series near 0 and $g(0) \neq 0$ as in this case, then the reciprocal function $1/g$ also has a convergent power series expansion near 0. Since this reciprocal function is equal to $x\,\cot x$, extended so that its value at 0 is 1, the assertion about representing this function by a convergent power series near 0 follows immediately.

We now return to our discussion of the quadratrix/trisectrix. This curve has infinitely many disconnected pieces, but in classical Greek mathematics the portion of the curve receiving attention was the connected piece in the first quadrant satisfying the additional condition $0 < x < \frac{1}{2}\pi$ (see Figure 3 in `transcurves.pdf`).

Our objective is to prove that (the coordinates for) the points on this piece of the curve do not satisfy a strongly nontrivial polynomial equation in two variables. But this follows immediately by combining the results in Sections II.2 and I.4 in `transcendentals.pdf` with the following simple observation:

> *If the function f defined on an open interval J is transcendental on J, then so are the functions $x^m \cdot f$ for all $m > 0$.* (If the latter is algebraic then there is a nontrivial polynomial identity of the form $\sum_{i,j} \; a_{i,j}x^{i+mj} \cdot f(x)^j \;=\; 0$, which we may rewrite in the form $\sum_{p,q} \; b_{p,q}x^p \cdot f(x)^q$. If $r > 0$ is the highest power of $f(x)$ which appears nontrivially in the first polynomial and $s \geq 0$ is maximal such that $a_{s,r} \neq 0$, then it follows that $b_{s+mr,r} = a_{s,r} \neq 0$.)

These considerations imply that $x \cot x$ is transcendental on the interval $0 < x < \frac{1}{2}\pi$, and from this it follows that the coordinates on this piece of the curve do not satisfy a strongly polynomial equation.■

5. The cycloid

We can analyze this using the same methods employed for the Archimedean spiral. The first step is to note two elementary properties of our parametrization of the cycloid:

$$x(t+2\pi) \;=\; x(t) \;+\; 2\pi\ , \quad y(t+2\pi) \;=\; y(t), \quad \text{for all} \quad t \;\geq\; 0$$

Also, since the explicit functions $x(t)$ and $y(t)$ are real analytic and the related power series converge for all choices of t, we may use the results from Section 1 to conclude that if there is a point $\mathbf{p}$ on the cycloid Θ, a small neighborhood U of $\mathbf{p}$, and a strongly nontrivial polynomial $G(x, y)$ such that $G \equiv 0$ on $\Theta \cap U$, then $G \equiv 0$ on all of Θ.

The next step is similar to a crucial observation from Section 2:

CLAIM: *If c lies in the interval $[0, 2]$, then there are infinitely many points of Θ which lie on the line $y = c$.*

To prove this, observe that the formula $y(t) = 1 - \cos t$ implies that if $c \in [0, 2]$ then there are infinitely many choices of t such that $y(t) = c$.

We may now proceed exactly as in Section 2. Suppose that we have a polynomial $G(x, y)$ which vanishes at all points of Θ, and write G in the standard form $G(x, y) \;=\; \sum_{i,j}\; a_{i,j}x^iy^j$; as before, let $H_i(y) = \sum_j\; a_{i,j}\, y^j$, so that $G(x, y) \;=\; \sum_i\; H_i(y) \cdot x^i$. By the preceding claim we know that for each $c \in [0, 2]$ the polynomial $G(x, c)$ has infinitely many roots, and therefore the coefficients for all powers of x must be trivial. In other words, for each i and each $c \in [0, 2]$ we have $H_i(c) = 0$. But this also means that each polynomial H_i has infinitely many roots, so that $H_i = 0$ for all i, and the latter immediately implies that $G = 0$.■

6. Epicycles

We have noted that the following book contains more information on Descartes' notion of mechanical curves:

> **J. Stillwell.** *Mathematics and Its History* (2^{nd} Ed.) Springer-Verlag, New York, 2002.
> **ISBN:** 0–387–95336–1

Section 13.3 of this book (which treats such curves) also mentions a class of curves called *epicycles* and notes that some are algebraic while others are transcendental. We shall conclude this article with an explanation of the assertions in Stillwell.

Definition. An *epicycle* is a curve in $\mathbf{R}^2$ with a parametrization of the following form:

$$\mathbf{s}(t) \;=\; b \cdot (\cos t, \sin t) \;+\; a \cdot (\cos rt, \sin rt) \qquad \text{(where)} \quad b, r \;>\; 0$$

Such curves were originally studied in classical Greek mathematics, and for centuries such curves and their higher order analogs

$$\sum_k c_k \cdot (\cos r_k t, \sin r_k t)$$

were thought to give the orbits of planets about the earth in the standard geocentric models of the universe. Usually we shall normalize our epicycles by setting $b = 1$; it is also customary to order the summands of higher order cycloids so that $c_1 > \cdots > c_m$. For the higher order epicycles that arose in astronomy the ratios of successive terms c_i/c_{i+1} are generally much smaller than 1.

Some illustrations of epicycles are contained in the following online sites:

`http://www.mathpages.com/home/kmath639/kmath639.htm`

`http://www.math.harvard.edu/ knill/seminars/fashion/fashion.pdf`

The first observations are fairly elementary, but afterwards we shall use freely use results from elementary differential geometry and point set theory, and ultimately we shall also use some fairly advanced (graduate level) material which goes beyond the usual introductory courses.

Ellipses and epicycles. The quickest proof that some epicycles are algebraic is given by the following:

PROPOSITION. *If a and b are positive, then the ellipse defined by the equation*

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

is an epicycle.

This derivation of this is fairly straightforward and appears on pages 2–3 of the following online reference:

`http://wwwdata.unibg.it/dati/bacheca/63/21692.pdf`

The main result. The rest of this section will be devoted to proving the following criterion for determining whether an epicycle is algebraic or transcendental.

THEOREM. *Let a and r be positive real number such that $a < \frac{2}{3}$ and $ra < \frac{2}{3}$. Then the epicycle*

$$\mathbf{s}(\theta) = (\cos\theta, \sin\theta) + a \cdot (\cos r\theta, \sin r\theta)$$

is algebraic if r is rational and transcendental if r is irrational.

The inequalities for r and a reflect a point made earlier; namely, the radius of the second circle is considerably smaller than the radius of the first (which is 1).

Sketch of proof. We shall begin with some general observations. The epicycles in the theorem are given by parametric equations whose coordinates are real analytic functions defined over the entire real line. Furthermore, the inequality constraints imply that the tangent vectors are given by

$$\mathbf{s}'(\theta) \quad = \quad (-\sin\theta, \cos\theta) \ + \ ra \cdot (-\sin r\theta, \cos r\theta)$$

and since the vectors $(-\sin u, \cos u)$ have length 1 it follows that

$$|\mathbf{s}'(\theta)| \quad \geq \quad 1 \ - \ ra \quad > \quad \tfrac{1}{3}$$

so that $\mathbf{s}'(\theta) \neq \mathbf{0}$ for all θ. By a standard consequence of the Inverse Function Theorem, we know that $\mathbf{s}$ is locally 1–1 for all choices of θ. It will also be useful to have a factorization of the curve $\mathbf{s}$ as $T \circ \gamma$, where $T : \mathbf{C}^2 \to \mathbf{C} \ \cong \ \mathbf{R}^2$ is the linear transformation $T(z_1, z_2) = z_1 + a\, z_2$ and $\mathbf{R} \to \mathbf{C}^2$ is given by $\gamma(\theta) = (e^{i\theta}, e^{ir\theta})$.

Not surprisingly, the proof of the theorem splits into two cases depending on whether r is rational or transcendental.

Suppose that r is rational. In this case write $r = p/q$, where the numerator and denominator are both positive. It follows immediately that $\mathbf{s}(\theta + 2q\pi) \ = \ \mathbf{s}(\theta)$ for all θ, so that the image of the entire curve is given by the image of the closed interval $[0, 2q\pi]$. We know that the curve $\mathbf{s}$ is locally 1–1, and in fact the Inverse Function Theorem and compactness imply that the image of $\mathbf{s}$ is a finite union of subsets that are homeomorphic to closed intervals and are nowhere dense in $\mathbf{R}^2$. Since a finite union of closed nowhere dense subsets is nowhere dense, this means that the image of $\mathbf{s}$ is also nowhere dense.

It A denotes the image of the curve γ described above, then it follows immediately that $A \subset \mathbf{C}^2 \ \cong \ \mathbf{R}^4$ is the set defined by the equations $|z_1|^2 = 1$, $|z_2|^2 = 1$ and $z_1^p = z_2^q$; in fact, the first equation is a consequence of the others, but this is not important for our purposes. Therefore A is the zero set of a finite number of real polynomial functions in four real variables, so that the points on the epicycle are the image of A under the linear transformation T. Note that T has maximum rank when viewed as either a real or complex linear transformation because it is onto.

More generally, one can ask the following question: *Suppose that $A \subset \mathbf{R}^n$ is the zero set of a finite number of real polynomials and $T : \mathbf{R}^n \to \mathbf{R}^m$ is an onto linear transformation. Is the image $T[A]$ also describable in terms of polynomials?*

It is not difficult to construct examples where the image is not the zero set of finitely many polynomials; in particular, this is the case if we take A to be the plane hyperbola $x_1 x_2 = 1$ and $T : \mathbf{R}^2 \to \mathbf{R}$ to be the projection $T(x_1, x_2) = x_1$, so that the image is the set of all nonzero points on the real line. However, we do have the following fundamental result:

TARSKI–SEIDENBERG THEOREM. *Let $A \subset \mathbf{R}^n$ be the zero set of finitely many polynomial functions, and let $T : \mathbf{R}^m \to \mathbf{R}^n$ be an onto linear transformation. Then $T[A]$ is a finite union of* semi – algebraic sets *B_i such that each B_i is an intersection $C_i \cap D_i$, where C_i is the zero set of some finite collection of polynomials, and D_i is the set of points defined by some finite collection of polynomial inequalities (it is possible that the collection of polynomials defining one of C_i or D_i is empty).*

Another instructive example is given by the parabola $x_1 = x_2^2$. In this case the image of A is the union of the sets where $x_1 = 0$ and $x_1 > 0$.

Here are some online references for the Tarski-Seidenberg Theorem:

`http://planetmath.org/encyclopedia/TarskiSeidenbergLojasiewiczTheorem.html`

`http://math.usask.ca/~marshall/ts.ps`

Returning to our discussion of epicycles for which the coefficient r is rational, we need to examine the consequences of the Tarski-Seidenberg Theorem for this example. Take some real number θ_0 such that both coordinates of $\mathbf{s}'(\theta_0)$ are nonzero; for example, since $ra < \frac{2}{3}$ we can take $\theta_0 = \frac{1}{4}\pi$. For each i such that $\mathbf{s}(\theta_0) \in D_i$, pick one of the polynomials g_i defining the set C_i. Then we know that the intersection of Image($\mathbf{s}$) and the given D_i's satisfies the polynomial equation $G = \prod g_i = 0$, where as before i runs over all indices such that $\mathbf{s}(\theta_0) \in D_i$.

By the preceding paragraph we know that $G \circ \mathbf{s}(\theta) = 0$ for all θ sufficiently close to θ_0. We may now apply the results of Section 1 to conclude that $G \circ \mathbf{s}(\theta) = 0$ for all real θ. Therefore the epicycle under consideration is algebraic.■

Suppose now that r is transcendental. In this case we need the following basic result:

KRONECKER–WEYL THEOREM. *Let $r > 0$ be irrational, and let $A \subset \mathbf{C}^2$ be the set of all points having the form $(e^{i\theta}, e^{ir\theta})$ for some real number θ. Then the closure $\overline{A}$ of A is the set of all (z_1, z_2) such that $|z_1|^2 = |z_2|^2 = 1$.*

The closure is the set of all points in A together with the set of all points that are limits of sequences of A. If f is a continuous real valued function and $f = 0$ on A, then $f = 0$ on the closure of A.

Here is an online reference for the theorem:

`http://mathworld.wolfram.com/Kronecker-WeylTheorem.html`

A proof is sketched on page 22 of the following book:

S. Tabachnikov. *Geometry and Billiards.* Student Mathematical Library, Vol. 30. American Math. Soc., Providence, 1995. **ISBN:** 2–856–29030–2.

Using the Kronecker-Weyl Theorem, we may now dispose of the irrational case as follows: Suppose there is a polynomial $G(x, y)$ such that $G \equiv 0$ on the image of $\mathbf{s}$, which is equal to $T[A]$. As noted above, by continuity we can then conclude that $G \equiv 0$ on $T[\overline{A}]$. However, direct examination shows that the latter is the set of all $z \in \mathbf{C}$ such that $1 - a < |z| < 1 + a$. One can now argue as in previous examples to show that the only polynomial which vanishes on such a set is the zero polynomial.■