Kepler's Second Law and the Topology of Abelian Integrals (According to Newton)

V. I. Arnol'd

This doctrine in species [power series] has the same relationship to Algebra as the doctrine in decimal numbers has to common Arithmetic.

—Isaac Newton, *Method of Fluxions*

1. Mathematics and Physics.

As I was leafing through Newton's *Principia Mathematica* (*Mathematical Principles of Natural Philosophy*)—the great book that laid the foundation of theoretical and mathematical physics and whose tercentenary is celebrated this year—I came across a couple of purely mathematical pages containing an astonishingly modern topological proof of a remarkable theorem on the transcendence of Abelian integrals.

Lost among a tangle of investigations in celestial mechanics, this theorem was little noted by mathematicians of Newton's era. The most likely explanation for this is that Newton's topological argument was a good two hundred years ahead of the mathematical level of that time.

Essentially, Newton's proof is based on the analysis of a certain very modern construction (equivalent to a Riemann surface for algebraic curves), and thus it was difficult to understand, not only for his contemporaries, but also for the twentieth-century mathematicians brought up in the spirit of set theory and real analysis, who are afraid of multivalued functions.

2. The Statement of Newton's Theorem.

A curve in the plane is said to be *algebraic* if it satisfies an equation $P(x, y) = 0$ where P is a nonzero polynomial. For example, the circle given by $x^2 + y^2 = 1$ is an algebraic curve. Further examples of algebraic curves are ellipses, hyperbolas, the lemniscate $y^2 = x^2 - x^4$ (see Figure 1; don't confuse it with the lemniscate of Bernoulli). The sine curve is not algebraic (why?).

A function is *algebraic* if its graph is an algebraic curve. This definition refers both to ordinary (single-valued) functions and to *multivalued* functions, i.e., functions like $y = \pm\sqrt{1-x^2}$ or $y = \operatorname{Arcsin} x$. For example, $y = \pm\sqrt{1-x^2}$ is a two-valued algebraic function.

The Russian original is published in *Kvant* 1987, no. 12, pp. 17–21.

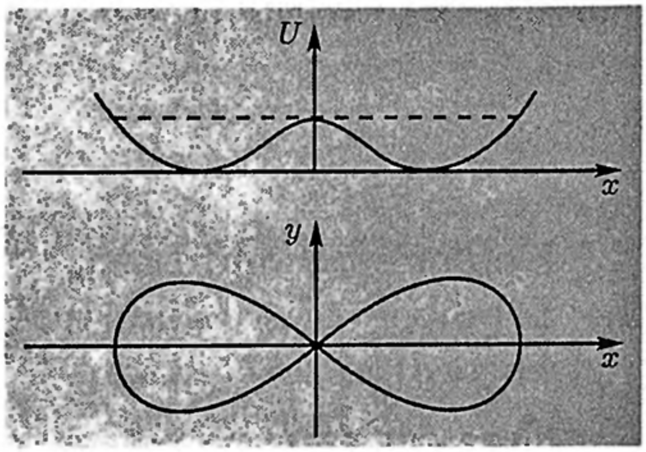

FIGURE 1. The lemniscate is the algebraic curve $y^2 = x^2 - x^4$. It can be interpreted as the energy-level curve in the phase plane of a particle moving along a line in the force field given by a cubic polynomial with two symmetric potential wells.

Consider an *algebraic oval*, i.e., a closed, convex algebraic curve. An oval is said to be *algebraically squarable* if the area of an arbitrary segment[1] can be expressed algebraically. More precisely, the area S of the segment cut by the straight line $ax + by = c$ (see Figure 2) must be an algebraic function of the line, i.e., it must satisfy an algebraic equation $P(S; a, b, c) = 0$, with P a nonzero polynomial in four variables.

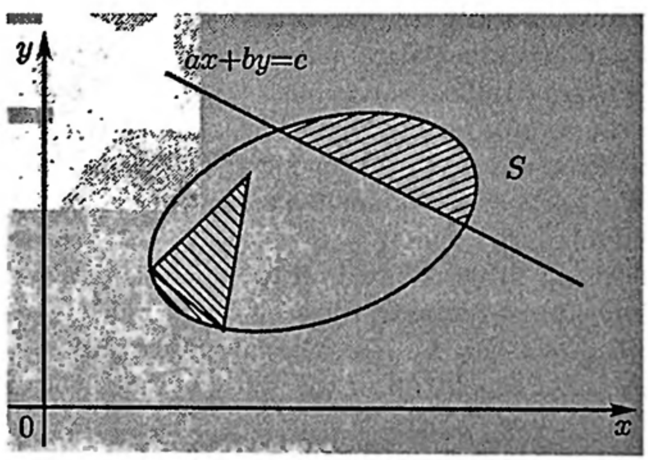

FIGURE 2. The area S of the segment cut from the oval by the line $ax + by = c$ is a function of (a, b, c). The area of a sector is a function of two lines. Functions associated in this way with algebraic ovals are called Abelian integrals.

REMARK. If an oval is algebraically squarable, then the area of a sector cut from the oval by an angle whose vertex is inside the oval is also an algebraic function of the two lines that constitute the angle. This is because the area of the triangle that is the difference between the segment and the sector is always algebraic.

[1]In this article, a segment of an oval is part of its interior cut off by a straight line.

Newton set himself the task of finding all algebraically squarable ovals. His result was the following theorem.

THEOREM. *Any algebraically squarable oval has singular points.*[2] *All smooth ovals are algebraically nonsquarable.*

EXAMPLE. An ellipse is algebraically nonsquarable. This implies that Kepler's equation that defines the position of a planet on the Keplerian ellipse as a function of time, according to Kepler's second law (the area swept out by the radius vector of the moving point is proportional to time), is transcendental, i.e., has no solution in algebraic functions.

Actually, it was this example that led Newton to his general theorem. The theorem is surprising, because at first glance there appears to be no connection whatsoever between algebraic squarability and singular points.

REMARK. In modern notation, Kepler's equation can be written as $x - e\sin x = t$. This equation has played an important role in the history of mathematics. Since the time of Newton, mathematicians have looked for a solution that expresses x as a series expansion in powers of the eccentricity e. The corresponding series converges for $|e| \leq 0.6627434\ldots$.

In attempting to explain the origin of this constant, Augustin Cauchy created complex analysis. Numerous fundamental mathematical notions and results, such as Bessel functions, Fourier series, the topological index of vector fields, and the argument principle in the theory of complex functions, also made their first appearance in investigations connected with Kepler's equation.

3. Newton's Proof.

Choose a point O inside the oval and rotate the ray $y = tx$ around this point. If the oval is algebraically squarable, then the area of the sector swept out by the radius vector of a point moving along the oval (see Figure 3) must be an algebraic function of t, the tangent of the angle between the ray and the x-axis.

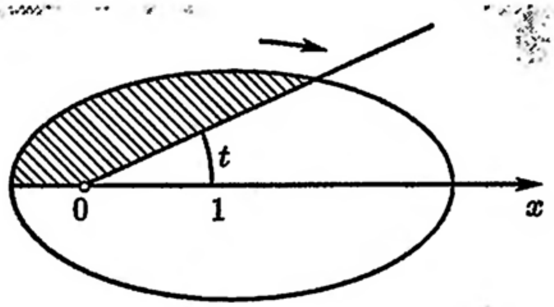

FIGURE 3. The area swept out by the radius vector cannot be an algebraic function of the tangent t of the angle between the ray and the horizontal axis, because the area assumes infinitely many values for a single value of t.

[2]A point of a curve is *singular* if in a neighborhood of this point the curve does not represent the graph of an infinitely differentiable function. Examples: a self-intersection point (Figure 4), a cuspidal point (Figure 8).

Let us make the ray go around the oval again and again. After every complete turn, the total accumulated area swept out by the ray will increase by a fixed quantity—the entire area bounded by the oval. Hence, the area swept out by the ray when viewed as a multivalued function of t has infinitely many values for a fixed position of the ray.

But an algebraic function cannot be infinitely multivalued, since the number of roots of a nonzero polynomial does not exceed its degree.

Therefore, the area swept out is not an algebraic function, and thus the oval under study is not algebraically squarable.

Newton remarks that a similar argument proves that the arc length of a smooth algebraic oval is also nonalgebraic.

4. Examples of Algebraically Squarable Ovals.

Are we forced, then, to conclude that algebraically squarable ovals do not exist at all? Of course not! Newton himself knew of examples of ovals whose area of segments is algebraic, and he mentioned some of these in relation to his theorem in *Principia Mathematica*.

The simplest example is provided by the oval in Figure 4, $y^2 = x^2 - x^3$. We can find parametric equations of this curve as follows. Take a line $y = tx$ passing through the origin. It meets the oval in one additional point, which satisfies $t^2 = 1 - x$, whence

$$x = 1 - t^2, \quad y = t - t^3.$$

From this parametric representation one immediately infers that the area integral $\int y\,dx$ is a polynomial in t. Hence the area of an arbitrary segment cut from the oval by a straight line is algebraic; i.e., the oval in question is algebraically squarable.

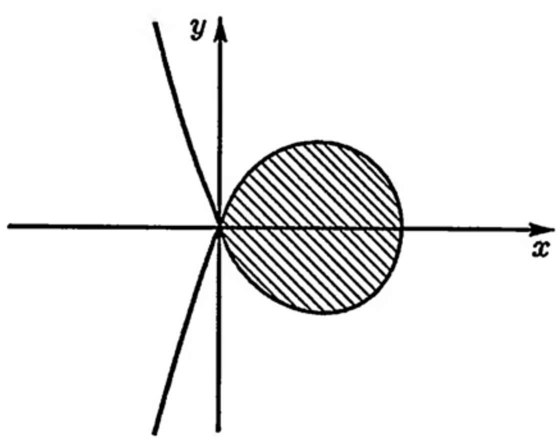

FIGURE 4. Locally algebraically squarable oval with one singular point (node).

Although this oval is not smooth, Newton's argument nonetheless applies to it. It shows that the area swept out by the radius vector cannot be expressed by a unique algebraic function. Does this contradict what we said before? No. The solution is that in this case, each time the radius vector passes through the singular point (node of the oval), the value of the area expressed by a certain algebraic function changes by a jump to another algebraic function.

5. Global and Local Algebraic Squarability.

As the previous example shows, a function can be locally algebraic but not globally algebraic (see Figure 5). In this sense one can say that the oval studied in Section 4 is *locally algebraically squarable.* To put it more precisely, an oval is said to be locally algebraically squarable if the area of the segment cut from it by a straight line is locally an algebraic function, i.e., if for every line $ax + by = c$ close enough to a given fixed line, the areas S of the portions cut by these lines from the oval satisfy certain algebraic equations $P(S; a, b, c) = 0$, where P is a polynomial.

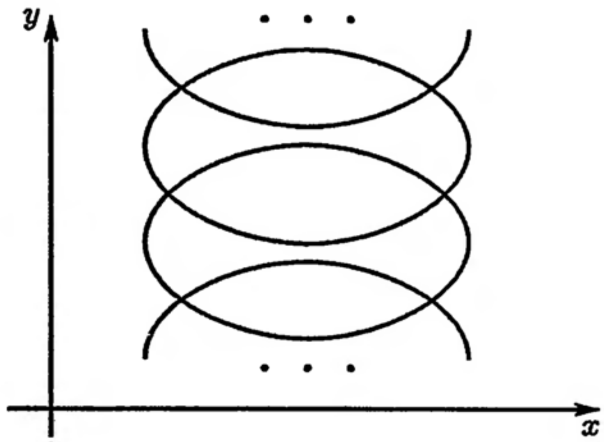

FIGURE 5. Graph of a function that is locally algebraic but not algebraic. Such functions can assume an infinite number of values at one point.

In practice, local algebraic squarability is almost as beneficial as full-fledged global squarability. So Newton posed a natural question: *Is it possible for a smooth algebraic oval to be locally algebraically squarable*? That is, can the area S of the segment cut by the line $ax + by = c$ be an algebraic function of (a, b, c) in a neighborhood of some point?

Following the method indicated in Section 4, we can find a pair of polynomials that give parametric equations of a locally algebraically squarable oval having a tangent at every point. For example, the polynomials $x = (t^2 - 1)^2$, $y = t^3 - t$ define the oval depicted in Figure 6, and the corresponding function $y = y(x)$ has two continuous derivatives at zero. The oval thus obtained looks perfectly smooth, and yet it is locally algebraically squarable (but not globally—this possibility is still ruled out by Newton's argument).

PROBLEM. Construct a locally algebraically squarable oval with one singular point such that in a neighborhood of this point the oval coincides with the graph of a function having 1987 continuous derivatives, while near all other points it is represented by functions with an infinite number of continuous derivatives.

6. Newton's Theorem on Local Nonalgebraicity.

We already know that a locally algebraically squarable oval can have an arbitrarily large, yet finite, *smoothness*, i.e., it can be everywhere represented by functions having any prescribed number of derivatives. However, every such oval

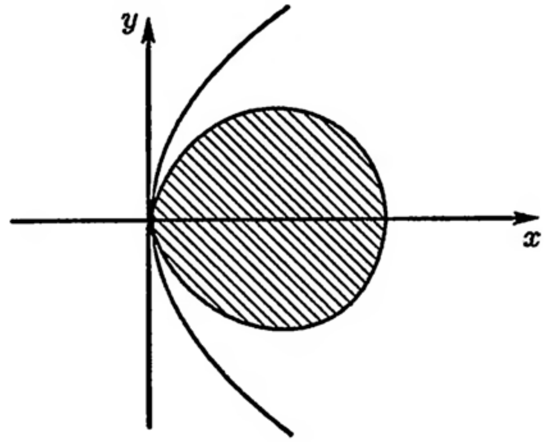

FIGURE 6. A locally algebraically squarable oval. Every polynomial that takes the value zero at all points of the oval also vanishes on the infinite branches.

considered so far has had at least one point at which the derivative of some order was discontinuous.

Newton assumed that the only curves that are truly smooth are those that are defined in a neighborhood of every point by a function that can be represented by a convergent power series

$$y = a_1x + a_2x^2 + a_3x^3 + \cdots$$

(the system of coordinates used has the origin at the chosen point of the curve). Nowadays, such curves are called *analytic*.

From the theorem of Section 2, Newton deduces a far stronger assertion:

THEOREM. *An analytic oval cannot be even locally algebraically squarable.*

7. A Proof of Local Algebraic Nonsquarability of Analytic Ovals.

If an oval were locally, but not globally, algebraically squarable, then the area swept out by the rotating ray would be given by one algebraic function on one side of a certain point of the curve and by another function on the other side of that point. But for an analytic oval the area depends analytically on the direction of the ray. This implies that both algebraic functions mentioned have one and the same convergent power-series expansion in a neighborhood of the point in question. Therefore, they coincide identically in this neighborhood and hence everywhere (remember that the number of roots that a nonzero polynomial can have cannot exceed its degree).

Thus, if a locally algebraically squarable analytic oval existed, it would also be globally squarable. Since this is impossible (see Sections 2 and 3), an analytic oval cannot be algebraically squarable even locally.

8. Can You Recognize an Analytic Oval?

A curve is said to be *infinitely smooth* if it locally coincides with the graph of a function that is infinitely differentiable.

THEOREM. *An infinitely smooth algebraic curve is analytic.*

Newton was aware of this fact, since he knew how to express the equation of any branch of an algebraic curve in a neighborhood of any point in the form of a rapidly convergent series

$$y = a_1 x^{\frac{1}{n}} + a_2 x^{\frac{2}{n}} + a_3 x^{\frac{3}{n}} + \cdots$$

(where the origin is placed at the point under consideration).

Newton stated a theorem about the convergence of this series in the following way: "... when x is small enough, the more the quotient is extended, the more it approaches the truth, to the end that its difference from the exact value of y finally shall come to be less than any given quantity you please and that the quotient extended to infinity shall be equal to y."

The series is constructed with the help of "Newton's polygon."[3] This approach, one of the most effective means of local analysis, has unfortunately been dropped from the modern system of mathematical education.

Any term of the series with a noninteger exponent has only a bounded number of derivatives. Thus a series having at least one term with a noninteger exponent defines a curve that cannot be infinitely smooth in a neighborhood of the point in question.

Hence, in the series expansion of an infinitely smooth algebraic curve only integer exponents are involved, and this means that the curve is analytic.

COROLLARY. *An infinitely smooth algebraic oval is even* locally *algebraically nonsquarable.*

Thus, an infinitely smooth closed convex curve cannot be locally algebraically squarable if it is algebraic.

9. Can a Smooth Nonalgebraic Oval Be Algebraically Squarable?

A negative answer to this question is implied by the facts already proved, since the following theorem holds.

THEOREM. *Any locally algebraically squarable oval is algebraic.*

Newton made use of this assertion, treating it as an evident fact. Here is a probable reconstruction of his argument.

LEMMA. *The envelope curve of any algebraic family of straight lines is algebraic.*

In other words, if all the tangents to a curve satisfy an algebraic equation, then the curve itself is algebraic as well.

PROOF OF THE LEMMA. Consider two close tangent lines whose angles of inclination to the x-axis have tangents t and $t+h$ (see Figure 7). For any fixed h and variable t, the intersection point of these two lines ranges over an algebraic curve. The degree of this curve, i.e., the degree of its defining polynomial, is bounded by a constant independent of h. This follows from the fact that the condition of consistency of two algebraic equations can be expressed by equating to zero a certain polynomial that depends on the coefficients of the given equations. In his discussion of this fact (still on the same two pages of *Principia Mathematica*!) Newton also

[3] See *Kvant* 1977, no. 6, p. 19.

mentions that two algebraic curves of degree m and n can have no more than mn points in common.

As h tends to zero, the intersection point of two tangent lines approaches the initial curve. Hence, the latter, as the limit of a family of algebraic curves of bounded degree, is algebraic.

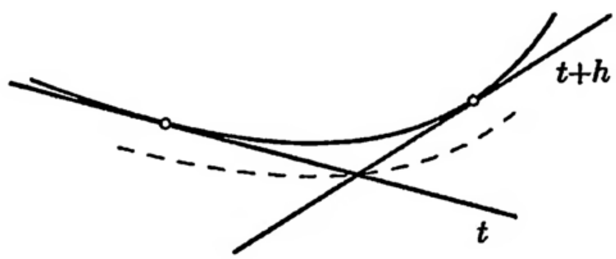

FIGURE 7. As the step h goes to zero, the dotted line formed by the intersection points of the tangents tends to the initial curve. It follows that the envelope of an algebraic family of lines is an algebraic curve.

PROOF OF THE THEOREM. The tangents of an oval are simply its secants that cut segments of zero area. Therefore, the tangents $ax + by = c$ of a locally algebraically squarable oval satisfy an algebraic relation $P(0; a, b, c) = 0$ (see Section 2). By the lemma, the oval is then algebraic.

10. Algebraically Nonsquarable Curves with Singularities.

We have seen that all infinitely smooth ovals are (even locally) algebraically nonsquarable. Moreover, Newton's argument implies local algebraic nonsquarability of infinitely smooth nonconvex curves without self-intersections and also many curves with singularities.

If all singular points of a curve are cuspidal points, then the curve is locally algebraically nonsquarable. The equations $y^2 = x^3 - x^4$ and $y^2 = (1 - x^2)^3$ furnish examples of such curves (see Figure 8). Further examples of the same kind are given by the singular curves $y = x^{p/q}$, where q is odd.

Newton remarks that a curve is locally algebraically nonsquarable if it does not have "adjoint branches going to infinity" that approach a point of the oval. By this he probably meant situations like those depicted in Figures 4 and 6.

In fact, the words "going to infinity" are unnecessary: One must forbid all kinds of self-intersections.

A correct condition that inhibits self-intersections of a closed curve defined by the equation $P(x, y) = 0$ reads as follows: A polynomial P vanishes at exactly two points of any circle of sufficiently small radius centered at a point of the curve.

Newton's method allows us to prove the following theorem.

THEOREM. *Any algebraic curve with no self-intersections in the above sense is (even locally) algebraically nonsquarable.*

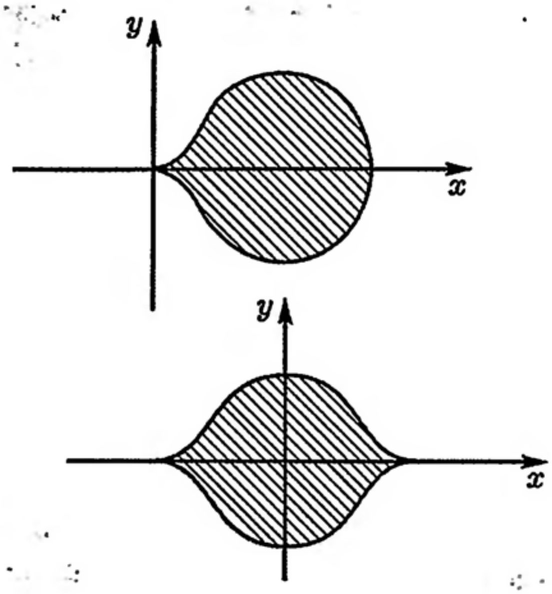

FIGURE 8. Curves without self-intersections, but with cuspidal points. Newton's argument applies to these curves as well and shows their local algebraic nonsquarability.

On the other hand, a self-intersecting closed curve actually can turn out to be locally algebraically squarable. For some reason, Newton omitted this possibility when he wrote about the branches "going to infinity."

Thus, the area of the segment of the lemniscate $y^2 = x^2 - x^4$ (Figure 1),

$$\int y\,dx = \int x\sqrt{1-x^2}\,dx = -\tfrac{1}{3}(1-x^2)^{\frac{3}{2}},$$

is an algebraic function.

Still, even for self-intersecting curves, algebraic squarability occurs very rarely.

Newton's argument shows that the total area bounded by a self-intersecting locally algebraically squarable closed curve equals zero if it is computed with sign. For example, the lemniscate is algebraically squarable only because the contributions of both of its loops to the total area are exactly opposite. If one deforms the lemniscate in such a way that the areas of the two loops become unequal, the curve will lose the property of local algebraic squarability.

At present, no complete description of all self-intersecting algebraically squarable curves is known.

11. Newton's Proof and Mathematics Today.

Today, the ideas that constitute the background of Newton's proof are called *analytic continuation* and *monodromy*. They are fundamental in the theory of Riemann surfaces as well as in a number of branches of modern topology, algebraic geometry, and the theory of differential equations, associated primarily with the name of Henri Poincaré—those branches of mathematics where analysis merges with geometry rather than with algebra.

Modern mathematics provides the tools necessary for proving a multidimensional analogue of Newton's theorem on plane curves. A sphere in three-dimensional space is algebraically squarable. This follows from Archimedes' theorem stating that the area of a spherical segment is proportional to its height. Therefore, all ellipsoids in three-dimensional (and, more generally, any odd-dimensional) space are algebraically squarable.

On the other hand, in four-dimensional (and any even-dimensional) space there are no algebraically squarable smooth hypersurfaces. This generalization of Newton's theorem is due to V. A. Vasil'ev (1987).

The neglected proof by Newton of the algebraic nonsquarability of ovals was the first "impossibility proof" to appear in mathematics after those of the ancient Greeks. It established a pattern for future proofs of nonsolvability of algebraic equations in radicals (Abel) and nonsolvability of differential equations in elementary functions or quadratures (Liouville). It was not without justification that Newton compared his proof to the proof of the irrationality of square roots in Euclid's *Elements*.

Comparing Newton's texts with the commentaries of his successors, a modern mathematician is amazed to find that Newton's original reasoning was so much more intelligible, cogent, and up-to-date than the commentators' translations of his geometric ideas into the formal language of Leibniz's calculus. The two centuries that elapsed from Newton to Riemann and Poincaré seem to me a mathematical desert filled with barren calculations ....

Translated by S. V. DUZHIN