Newton's Principia
Read 300 Years Later

V. I. Arnol'd and V. A. Vasil'ev

Abstract. Analyzing Kepler's law in two dimensions, Newton discovered an astonishingly modern topological proof of the transcendence of Abelian integrals. Newton's theorem was not really understood by mathematicians at that time, since it was based on the topology of Riemann surfaces. Thus, it was incomprehensible both for Newton's contemporaries and for 20th century mathematician who were bred on set theory and the theory of functions of a real variable, and who were afraid of multivalued functions.

In our paper we describe Newton's theorem and also some other new mathematical theorems, partially (more or less explicitly) contained in the Principia and partially suggested by Newton's text.

1. Newton's Theorem on the Nonintegrability of Ovals

An algebraic oval in the plane R^2 is part of a real algebraic curve (i.e., the zero set of a polynomial), homeomorphic to a circle. An oval is called nonsingular if it is C^∞-diffeomorphic to the circle.

An oval is called algebraically integrable if the area of its segment is an algebraic function of the secant line (i.e., there exists a nontrivial polynomial $F(V, a, b, c)$ which vanishes if V is the area cut by the line $ax + by = c$). An oval is locally algebraically integrable if the area of its segments coincides with an algebraic function in a neighborhood of any line (but these functions may be different for different lines).
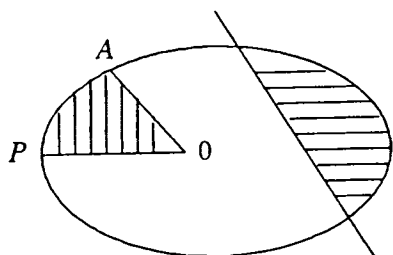

Fig. 1

Theorem 1 (Newton 1687, see [N], Lemma XXVIII). *There exists no algebraically integrable convex nonsingular algebraic curve.*

(There exist compact integrable algebraic curves, C^∞-smooth at all their points but one, at which they have any prescribed finite number of derivatives. Example: the curve

$$y^m = x^{(n-1)^m}(a^n - x^m)$$

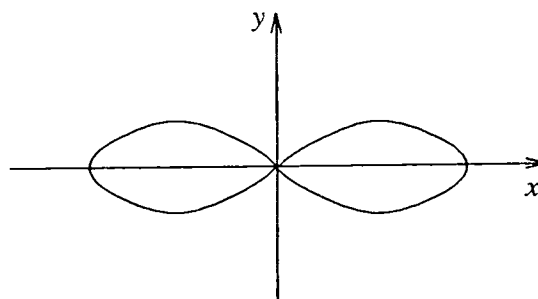where m, n are even, see [BR]. If $m = n = 2$, it is the Huygens lemniscate; see Fig. 2.)

Fig. 2

Newton's Proof. Let us fix a point O inside the oval, and a ray with the origin O. Consider the function on the oval (or, equivalently, on the space of all the rays issuing from the origin), whose value at the point A equals the area of the sector bounded by the fixed ray, the radius OA and the oval, see Fig. 1. If the oval is integrable, this function is algebraic: a sector consists of the segment and a triangle, and the areas of both of these depend algebraically on the point A: this follows from the algebraicity and the integrability of the oval. Let us move the point A along the oval. After any complete cycle, the area of the sector increases by the area bounded by the oval. In particular, for the same point A this function has an infinite number of values which contradicts its algebraicity.

Moreover, the same reasoning proves the following stronger statement.

Theorem 2 (see [A₁], [A₅]). There exists no locally algebraically integrable nonsingular convex algebraic oval.

Does there exist any obstacle to applying the preceding argument to a singular oval, e.g. to the curve of Fig. 2? Indeed, we have used the fact that the area of a sector is an analytic function of the point A. In the case of singular curves the area is, in general, nonanalytic: when A crosses the singular point, the area function may jump from one local branch to another.

Lemma. Let the area of the sector $(POA$ on Fig. 1) in a C^∞-smooth algebraic convex oval be a locally algebraic function of the point A. Then this function is even globally algebraic.

Proof. For a C^∞-smooth oval, the area of the sector, considered as a function of the point A, has an asymptotic* expansion

$$(1) \qquad V = a_0 + a_1 t + a_2 t^2 + \cdots$$

near any point. Indeed, Newton discovered that any branch of an algebraic function has an asymptotic Puiseux expansion

$$V = c_0 + c_1 t^{1/p} + c_2 t^{2/p} + \cdots, \quad p \in \mathbf{Z}_+,$$

near any point (it follows from the method of the "Newton parallelogram"). But, if such an expansion for the graph of the area contains a term with a noninteger degree of t, then our oval will not be smooth. It will also be nonsmooth if the expansions (1) for the two branches of this graph at two sides of our point are different.

However, a locally algebraic function, having an expansion (1) at any point, is globally algebraic. Indeed, in the opposite case there would exist a point on the two sides of which the graph would coincide with two different algebraic curves. Both curves having the same expansion (1) now implies their order of tangency at this point would be infinite. This contradicts Bézout's theorem (formulated by Newton in the same paragraph of Principia): the number of (perhaps, confluent) intersection points of different irreducible algebraic curves is majorized by the product of their degrees; this proves the lemma.

Newton was led to his theorem by a particular case: the position of a planet on the Kepler ellipse cannot depend algebraically on the time (or on the area of the sector, swept by the radii, which is proportional to the time according to the two-dimensional Kepler law). Newton also noted that the length of an arc of an oval, cut by a line, cannot be an algebraic function of the line.

Some of the above requirements on the oval may be omitted. Newton himself does not require either nonsingularity or convexity. He only indicates that the oval should not touch any conjugate branch connecting it with infinity. (This requirement was added in 1713 and appears only in the second edition of the Principia.) Huygens wrote in a letter to Leibniz (1691) that Newton's argument is wrong, because it may be applied even to a triangle (which is, of course, integrable). Leibniz had answered that a triangle may hardly be considered as an oval. He suggested a more dangerous counterexample— the Huygens lemniscate curve (described in one of his preceding letters to Leibniz). The lemniscate curve evidently satisfies Newton's definition of an oval but it is algebraically integrable.*

Leibniz also conjectured the transcendence of the area of almost any segment, which a line with rational (or algebraic) coefficients cuts from an oval, given by an algebraic equation with rational coefficients. For instance, he conjectured the transcendence of the area π bounded by a circle of radius 1. The general problem of Leibniz contains Hilbert's seventh problem, but, unlike Hilbert's problem, Leibniz' problem seems to be still unsolved.

The requirement of algebraicity of ovals in Theorem 1 is unnecessary, since a smooth integrable oval is algebraic.

Indeed, let us consider the zero set of the area function on the space of lines in $\mathbf{R}^2$. This set is a (semi)algebraic curve C on the dual projective plane of the plane where the initial oval lives. This initial oval is the envelope of the lines belonging to the algebraic set C of lines. Such an envelope (the dual curve of C) is algebraic.

The algebraicity of the dual curve K of an algebraic curve C was evident to Newton. Indeed, the envelope is the limit for $h \to 0$ of the curves K_h, formed by the intersections of the lines belonging to C, having the angular coefficients t and $t + h$. The degree of the curves K_h being independent of h, the limit curve K is algebraic.

The independence of the degree on h follows from the calculation of the degree of a resultant (that is, the very calculation, which proves the "Bézout" theorem, majorizing the number of intersection points of two curves of degrees m and n by mn; this theorem is explicitly formulated and used by Newton on the same page of the Principia that we discuss here).

Newton's theorem can be extended to nonconvex curves, and also to multidimensional hypersurfaces. The proofs are based on monodromy theory (or Picard–Lefschetz theory), i.e., on the study of the ramification

---

*In fact, this series converges in the neighborhood of the point, i.e., the area function is analytic; but in the proof that follows we only use the existence of an asymptotic series.

*Newton's argument proves the transcendence of the area function for any oval, which is an immersed circle closed on its Riemann surface and which bounds a nonzero area on the plane.

of integrals along circles depending continuously on a parameter (see [Ph], [M], [AVG], [AVGL], [D], [F]).

2. Generalizations of Newton's Theorem to Hypersurfaces in Even-dimensional Spaces

An ovaloid (a convex compact hypersurface) in $\mathbf{R}^n$ is called locally algebraically integrable, if the volume of the segment cut off the ovaloid by a hyperplane, coincides with an algebraic function near any fixed hyperplane.

Theorem 3 (V. A. Vasil'ev, see [AVGL]). *For n even, there does not exist any smooth convex locally algebraically integrable ovaloid in* $\mathbf{R}^n$.

We can choose a linear function x in $\mathbf{R}^n$, whose restriction to the ovaloid has the nondegenerate (Morse) minimum m and maximum M. Let us consider a path, going around the interval mM on the complexification on the axis x (Fig. 3).
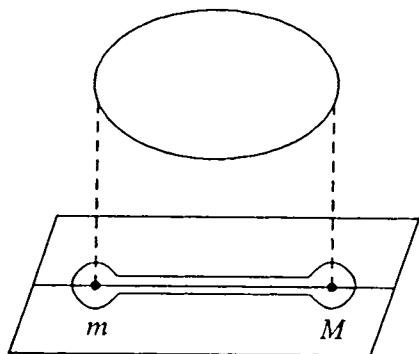

Fig. 3

Lemma 1. *The increment along the path of Fig. 3 of the volume $V(t)$ of the segment $x \leq t$ (or of the analytic continuation of this volume along the paths in the neighborhood of the interval mM in the complex plane t) is twice as large as the volume bounded by the ovaloid.*

Theorem 3 follows immediately from this lemma, which is based on the following fact.

Lemma 2. *For n even, the analytic continuation of the function $V(t)$ along the small circles centered at the points m, M, equals correspondingly the functions $-V(t)$, $V(M) - V(t)$.*

In other words, the power series of the function $V(t)$ in the neighborhoods of the points m, M contain only half-integer (but not integer!) degrees of $t - m$, $M - t$.

Conjecture 1. *For n even, there exist neither convex nor nonconvex smooth locally algebraically integrable ovaloids in* $\mathbf{R}^n$.

This is true for the usual ovals. The proof is similar to that of Theorem 3. We choose a more complicated path in the complex line x; see Fig. 4. This path turns

around the critical values of the restriction of x to the oval, consecutively in the order of the corresponding critical points on the oval. When t returns to its initial value along this path, the analytic continuation of the area function increases by twice the area bounded by the oval.
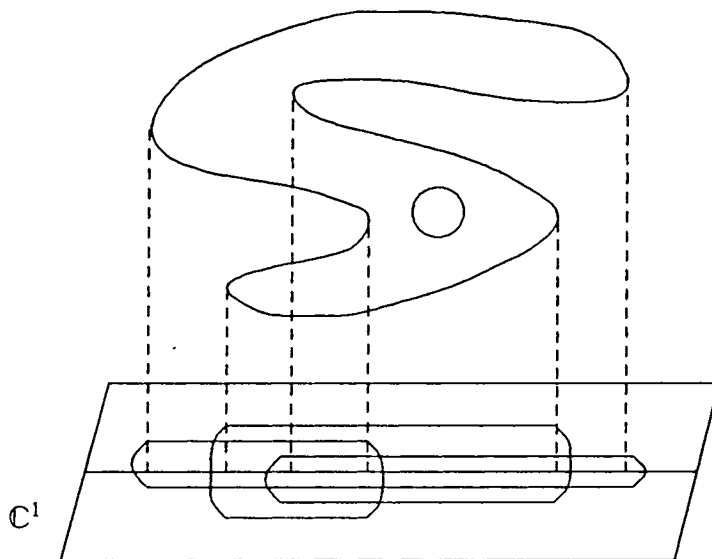

Fig. 4

Conjecture 1'. *A path with this property exists for any even n.*

3. Integrable Ovaloids in Odd-Dimensional Spaces

Unlike the even-dimensional spaces, the odd-dimensional spaces contain algebraically integrable ovaloids. A sphere in $\mathbf{R}^3$ is integrable (by a theorem of Archimedes). The same holds for any ellipsoid in any odd-dimensional space; see [A₁].

Conjecture 2. *Any irreducible smooth locally algebraically integrable ovaloid in $\mathbf{R}^{2k+1}$ is an ellipsoid.*

Theorem 4. *Almost all algebraic ovaloids of degree $d \geq 3$ in $\mathbf{R}^{2k+1}$, $k \geq 1$, are not algebraically integrable.*

The smooth algebraically integrable irreducible ovaloids of degrees ≥ 3 (if they do exist) are very special algebraic surfaces. For instance, their tangent planes at their complex parabolic points should be tangent to the (complexified) ovaloid along the curves of parabolic points, and so on (see [AVGL]).

The proof of Theorem 4 depends on the ramification properties of the volume of the segment continued analytically to the complex projective space of the cutting hyperplanes. The analytic continuation of the volume function along a path X_t in the space of hyperplanes is equal to the integral of the holomorphic differential n-form dx along an n-chain X_t. This chain is bounded

by the union of the complexified ovaloid hypersurface and the hyperplane X_t. The ramification hypersurface consists of those planes which are not general with respect to the complexified ovaloid hypersurface. The ramification at the tangents at the generic parabolic points of the complexified ovaloid is logarithmical. Hence the continuation of the volume function is infinitely multivalued. This implies Theorem 4.

The difference between the cases of even and odd n is due to the n-dependent sign in the Picard-Lefschetz formula, describing the ramification of the relative homology class of the chain X_t. The same sign is responsible for the existence of a sharp back front of a wave in three space dimensions (and for its absence in a two-dimensional space). The existence of a sharp back front makes it possible to communicate acoustically in the spaces of odd dimensions (and makes it impossible to communicate in even-dimensional spaces).

The relation of Newton's theorem to the theory of hyperbolic PDE's is deeper than it seems. The same mathematical structure is even more transparent in another of Newton's creations—in his attraction theory.

4. Newton's Theorem on the Attraction by Spheres and Hyperbolic Surfaces

First we recall the following results of Newton.

Theorem 5 ([N], Theorem XXX). *If toward the individual points of a spherical surface are directed forces decreasing inversely proportional to the distances from these points, then a particle inside this surface is not attracted to any side.*

Indeed, for any infinitely narrow cone with the wedge at this particle the intersections of the sphere with the opposite parts of the cone attract the particle with equal forces, since the areas of these intersections are proportional to the squares of distances.
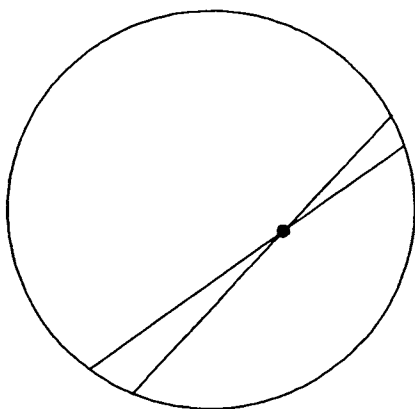

Fig. 5

Theorem 6 ([N], Theorem XXXI). *"With the same assumptions, I affirm that a particle outside a spherical surface is attracted to the center with the force inversely proportional to its squared distance from the center."*

Proof. A spherically symmetric noncompressible (of divergence zero) vector field decreases inversely proportional to the squared distance to the center (since its flows through all the spheres are the same). The attraction field of any particle is noncompressible. Hence, the attraction field of any body is noncompressible outside this body. Thus, the attraction field of any sphere is noncompressible outside the sphere. Being obviously spherically symmetric, it coincides with the attraction field of a particle in the center.

These theorems together with the corresponding proofs hold in n-dimensional space, if the attraction force is inversely proportional to the $n - 1$-th power of the distance.

Moreover, these theorems may be extended to the case of any ellipsoid in $\mathbf{R}^n$, if the density of the distribution of the matter on its surface is inversely proportional to the length of the gradient of the quadratic form defining this ellipsoid in the corresponding point. In this case, inside the ellipsoid the attraction is absent, and outside it is constant on the ellipsoids confocal with the initial one. (Ivory's theorem, [I].)

Newton's theorems on the attraction of ellipsoids may be extended to hyperbolic surfaces of arbitrary degree in $\mathbf{R}^n$.

Definition. An algebraic hypersurface of degree d in $\mathbf{R}^n$ is called hyperbolic with respect to a point x, if any real line containing this point intersects the surface exactly d times (possibly, at infinity). Such points x form the hyperbolicity set of the surface. This set is a union of some connected components of the complement to the hypersurface (see [ABG]); such components are called hyperbolicity domains.

For example, an ellipsoid has one hyperbolicity domain, and a two-sheeted hyperboloid—two such domains.

A smooth hyperbolic surface given by a polynomial equation $F = 0$ partitions $\mathbf{R}^n$ into the components which we shall call zones. Let us order them according to the minimal number of intersections with the hypersurface of a path connecting a point of a component with a point of a fixed hyperbolicity domain. This hyperbolicity domain will be called the 0-th zone.

The standard charge on the hyperbolic surface $F = 0$ is defined by the form dx/dF (i.e., as the limit of a homogeneous charge between the surfaces $F = 0$ and $F = \varepsilon$ with density $1/\varepsilon$ and the signs equal to ± 1 depending on the parity of the number of the corresponding component of the surface).

Theorem 7 (see [A₂]). *The standard charge on a hyperbolic surface does not attract the points in the hyperbolicity domain. Moreover, the same holds for the product of the standard charge and a polynomial of degree $d - 2$ (where d is the degree of the surface)*

For an ellipsoid $d = 2$, hence only the standard density is admissible, but for $d = 4$ we have many admissible densities.

If the degree of the polynomial-multiplier is m units more than the critical value $d - 2$, then the potential in the hyperbolicity domain is a polynomial of degree at most m; see [G].

The Newton-Coulomb potential (the attractive power in $\mathbf{R}^n$ is proportional to r^{1-n}) may be replaced here by any Green's function G in $\mathbf{R}^n - \{0\}$, which is homogeneous of degree $s - n$ (or, for $s = n$, proportional to the logarithm on any ray) and satisfies the equation $G(-x) = (-1)^s G(x)$ (s being a natural number). In this case the critical degree $d - 2$ is replaced by $d - s$.

Theorem 8 ([G]). *A G-potential of the charge, which is the product of the standard one and of a polynomial of degree $d - s + m$, coincides with a polynonmial of degree $\leq m$ in the hyperbolicity domain.*

Is there any trace of the algebraicity of the potential in the other domains? (The simplest example is the logarithmic potential of a uniform circle in the outer domain.)

Theorem 9 (V. A. Vasil'ev). *In the k-th zone of a hyperbolic curve of degree d in $\mathbf{R}^2$, any partial derivative of the Newton potential of the standard charge coincides with a sum of two algebraic functions, having at most C_d^k values. Moreover, the same holds for the partial derivatives of order $q + 2 - d$ of the potential of the charge which is the product of the standard one and of a polynomial of degree q.*

For instance, for the circle all the derivatives of the potential are single-valued functions so that our majorization of the number of the values $(C_2^1)^2 = 4$ is not attained. It is related with the fact that the functions F and G are not in general position: the singular lines of the function $G = \ln|x|$ in C^2 are asymptotic with respect to the equation of the circle. In the case of a typical ellipse our majorization is the best possible: the analytic continuations of the derivatives of the potential are 4-valued functions.

The idea of the proof is very close to that for Theorems 1–4 (although it provides the opposite answer). Namely, for any point $x \in \mathbf{R}^2 - \{F = 0\}$, consider two complex lines $\{\xi \mid |x - \xi| = 0\}$ (i.e., $(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 = 0$) and the intersection set of these lines with the surface $\{F = 0\} \subseteq C^2$. This set consists of $2d$ points (possibly, infinitely distant): d points on any line. If x belongs to the k-th zone and is "generic" (i.e., all these $2d$

intersection points are distinct and finite), then the μ-th partial derivative $(|\mu| > q - d + r)$ of the potential is given by the integral of a suitable differential form along a chain which consists of $2k$ small standard circles near some $2k$ of these intersection points—k points on any line. The motion of the point x in the complex domain may only permute these points and the corresponding circles.
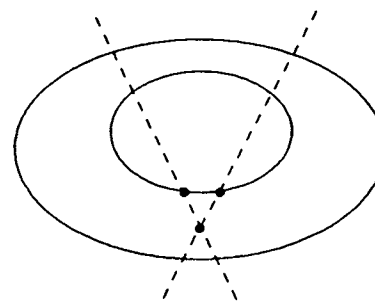

Fig. 6

5. Magnetic Field Analogues of Newton's and Ivory's Theroems

In one other generalization of the theorems of Newton and Ivory, the attracting ellipsoid is replaced by a hyperboloid of an arbitrary signature. In this case, the potential function is replaced by a differential form of a suitable degree (depending on the signature). Consider for instance the one-sheeted hyperboloid in three-dimensional Euclidean space. It is fibered in a natural way into two families of curves—the meridians and the parallels, diffeomorphic correspondingly to the lines and to the circles (these curves are the traces on the hyperboloid of the family of the ellipsoids and of the two-sheeted hyperboloids, confocal to it; they are also called the elliptic coordinate curves). The family of meridians may be extended to the family of elliptic coordinate lines fibering into the lines of the interior of the hyperboloid. The family of parallels may be extended to an analogous fibration of the exterior part of the hyperboloid into the closed elliptic coordinate lines. (See Fig. 7)

Theorem 10 ([A₃]). *There exists an electric current along the meridians (along the parallels) of the hyperboloid, whose magnetic field vanishes in the interior domain and is directed along the parallels in the exterior one (vanishes in the exterior domain and is directed along the meridians in the interior one).*

In [VS] this result is extended to the case of hyperboloids of arbitrary signatures in spaces of arbitrary dimensions.

It would be interesting to transfer these "magnetic" generalizations of Newton's and Ivory's theorems to the case of the "generalized hyperbolic" surfaces of higher degree.
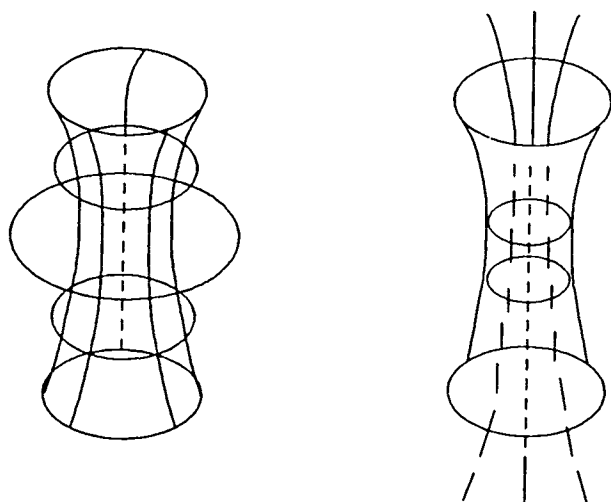

Fig. 7

6. Duality Between the Attraction Laws

Consider a point moving on a plane under the action of an attractive force directed toward the origin and proportional to the a-th degree of the distance from the origin. It turns out, that for any such attraction law there exists a dual one; the orbits of the motion under the dual force fields are related by a simple conformal mapping. For instance, the universal attraction law and Hooke's law are dual to each other. Let us introduce on the plane a complex coordinate w.

Theorem 11 ([A₄], [A₅]). *Any orbit of the motion of the point w on the plane of complex numbers in a central attraction field in which the force is proportional to w^a, is transformed into an orbit of the motion in a central field proportional to z^A by the map $z = w^\alpha$, provided that*

$$(2) \qquad (a+3)(A+3) = 4, \qquad \alpha = (a+3)/2.$$

The proof is an immediate calculation. (See Fig. 8)

Theorem 11 is not formulated in the Principia. But is was guessed thanks to Newton's formula for the angle between the consequent apocenters of an almost round orbit.

Example. If $a = 1$ (Hooke's law) then formula (2) gives $A = -2$ (the universal attraction law) and $\alpha = 2$. We get **Theorem 12** (Bohlin, see [B]). *The transformation $wpAw^2$ transforms an ellipse centered at the origin of the complex plane into an ellipse having a focus at the origin.*

Proof. The Zhukovsky function $w = \xi + \xi^{-1}$ transforms a circle $|\xi| = r > 1$ into a Hooke's ellipse centered at 0 with foci ± 2. But $w^2 = \xi^2 + \xi^{-2} + 2$—the squaring of w translates the focus of the ellipse to the origin.

Remark 1. The motion of a free point (along a straight line on the plane) may be considered as the motion in the zero field of arbitrary degree. Applying Theroem 11, we obtain special orbits of the motion in a central field of an arbitrary degree A on the plane. These orbits are obtained from the straight lines by the mapping $wpAw^\alpha$, $\alpha = 2/(A+3)$.

In the case of the universal attraction law these special trajectories are parabolas ($\alpha = 2$). In the general case, the equation of these generalized parabolic orbits in the polar coordinates is $r^\alpha = \sec\alpha\phi$.
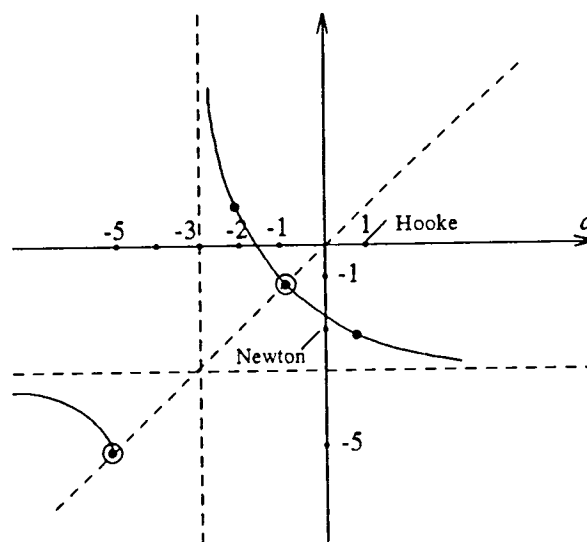

Fig. 8

Remark 2. Newton had considered in the Principia the values $a = 1, -1, -2, -3, -5$. The values $a = -1$ and -5 are special for the duality law (2) as the self-dual ones. For $a = -5$, the formula (2) gives $\alpha = -1$.

Corollary. *The orbits of motions in a central field, whose force is inversely proportional to the 5-th degree of the distance from the origin, are transformed by the inversions to the orbits of the same kind.*

Example. The straight lines of motion in the zero field which do not contain the origin are transformed by the inversion into the circles containing the origin. Hence the motion along a circle containing the attracting point is possible in a field inversely proportional to the 5-th degree of the distance—this corollary is also due to Newton [N].

References

[A₁] Arnol'd, V. I., A topological proof of transcendence of Abelian integrals in Newton's Principia. Quant 1987, No. 12, 1–15.

[A₂] Arnol'd, V. I., On the Newtonian Potential of Hyperbolic Layers. Selecta Math. Soviet., 1985, Vol. 4, No. 2, 103–106.

[A₃] Arnol'd, V. I., Magnetic field analogues of the theorems of Newton and Ivory. Uspekhi Mat. Nauk, 1983, Vol. 38, No. 5, 145, 145–146 (In Russian).

[A₄] Arnol'd, V. I., The 300th anniversary of mathematical natural philosophy and celestial mechanics. Priroda, 1987, No. 8, 5–15 (In Russian).

[A₅] Arnol'd, V. I., Huygens and Barrow, and Newton and Hooke—the first steps of the mathematical analysis and catastrophe theory, from the evolvents to the quasicrystals. Birkhauser.

[ABG] Atiyah, M. F., Bott, R. and Garding, L., Lacunas for hyperbolic differential operators with constant coefficients. I, II. Acta Math., 1970, 124, 109–189; 1973, 131, 145–206.

[AVG] Arnol'd, V. I., Varchenko, A. N. and Gusein-Zadeh, C. M., Singularities of differential mappings. II. Moscow, Nauka, 1984. (English translation:)

[AVGL] Arnol'd, V. I., Vasil'ev, V. A., Goryunov, V. V. and Lyashko, O. V., Singularities. II. Moscow, VINITI, 1989. (English translation:)

[B] Bohlin, K., Bull. Astr., 1911, Vol. 28, No. 144.

[BR] Brougham, H. and Routh, E. J., Analytical view of Sir Isaac Newton's Principia. London, 1855.

[D] Deligne, P. Équations différentielles à points singuliers réguliers. Lect. Notes Math., Springer, 1970, No. 163.

[F] Faltings, G., Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. Invent. Math., 1983, No. 73, 349–366.

[G] Givental, A. B., Polynomiality of the electrostatistical potentials. Uspecki Mat. Nauk, 1984, Vol. 39, No. 5, 253–254 (In Russian).

[I] Ivory, J. On the attractions of homogeneous ellipsoids. Philos. Trans., 1809, 99, 345–372.

[M] Malgrange, B., Integrales asymptotiques et monodromie. Ann. Sci. École Norm. Sup., 1974, Vol. 7, 405–430.

[N] Newton, I., Philosophiae Naturalis Principia Mathematica, London, 1687.

[P] Petrovsky, I. G., On the diffusion of waves and the lacunas for hyperbolic equations. Matem. Sbornik, 1945, 17(59), 289–370.

[Ph] Pham, F., Introduction à l'étude topologique des singularités de Landau. Paris, Gauthier-Villars, 1967, ii + 142 pp.

[VS] Vaïnshteïn, A. D. And Shapiro, B. Z., Multidimensional analogues of the Newton and Ivory theorems. Funtional Anal. Appl., 1985, Vol. 19, No. 1.

[V] Vasil'ev, V. A., Sharpness and the local Petrovsky condition for hyperbolic equations with constant coefficients. Isvestija Acad. Sci. USSR, Ser. Math. 1986, Vol. 50, No. 2, 242–283.

V. I. Arnol'd is Professor of Mathematics at the Moscow State University in the USSR. He was co-recipient of the Crafoord Prize of the Swedish Academy of Sciences and is a member of the USSR, USA and French academies of science. His research interests include algebraic and symplectic geometry and variational calculus. V. A. Vasil'ev, a student of Arnol'd, received his Ph.D. from Moscow State University in 1984. His thesis investigated the topic of asymptotic behavior of exponential integrals in the complex plane.