Fields Institute Communications
Volume 00, 0000

From Hilbert's Superposition Problem to Dynamical Systems

V. I. Arnold
Steklov Mathematical Institute,
Gubkina str. 8, Moscow, 117966, Russia
and
CEREMADE, Université Paris 9 – Dauphine
75775 Paris, Cedex 16-e, France

Abstract. This is the first of the series of three lectures given by Vladimir Arnold in June 1997 at the meeting in the Fields Institute dedicated to his 60th birthday.

> Some people, even though they study, but without enough zeal, and therefore live long.
>
> *Archibishop Gennady of Novgorod in a letter to Metropolitan Simon, ca 1500.*

Today, I shall try to explain the diversity of subjects I was working on. In fact, I was following one line from the very beginning and there was essentially one problem I was working on all my life. This fact seems strange even to me but I shall try to explain it.

When you are collecting mushrooms, you only see the mushroom itself. But if you are a mycologist, you know that the real mushroom is in the earth. There's an enormous thing down there, and you just see the fruit, the body that you eat.

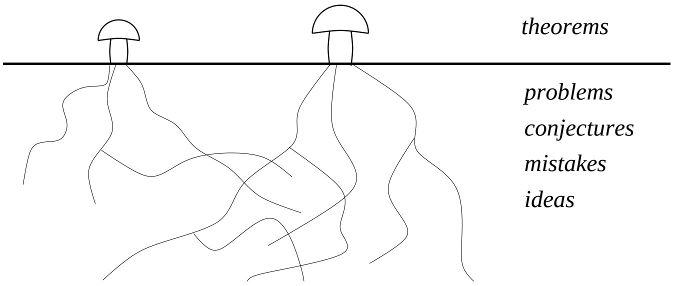

Figure 1: The Mathematical Mushroom

1991 *Mathematics Subject Classification.* Primary 01A65; Secondary 26B40, 54H20.
Partially supported by the Russian Basic Research Foundation, project 96–01–01104.
Written down from VHS tape by S. V. Duzhin. Revised by the author in December 1997.

In mathematics, the upper part of the mushroom corresponds to theorems that you see, but you don't see the things which are below, that is: *problems*, *conjectures*, *mistakes*, *ideas*, and so on.

You might have several unrelated mushrooms being unable to see what their relation is unless you know what is behind. And that's what I am now trying to describe. This is difficult, because to study the visible part of the mathematical mushroom you use the left half of the brain, the logic, while for the other part the left brain has no role at all, since this part is highly illogical. It is hence difficult to communicate it to others. But in this series of lectures I shall try to do it.

Today, I shall mostly discuss the history, in the second lecture — the hidden part of the mushroom, the irresponsible ideas, providing the main motivation for the research. And then some theorems will appear in the last lecture.

The first serious mathematical problem with which I started was formulated by Hilbert. It is a problem on superpositions emerging from one of the main mathematical problems: solution of algebraic equations.

The roots of a quadratic equation

$$z^2 + pz + q = 0$$

can be expressed by a simple formula in terms of p and q. Similar formulas are also available in degrees 3 and 4. If the degree is 5, then you know from Abel's theorem or in other terms from the monodromy of the corresponding algebraic function and the fact that the alternating group in five variables is not solvable — that there is no such formula.[1] However, there is a classical result that if you know how to solve one very special equation

$$z^5 + az + 1 = 0,$$

i. e. you know one particular algebraic function $z(a)$, then you can solve all the equations of degree 5. For quadratic equations you need square roots, for cubic equations — square roots and cubic roots (which can be considered a simple special function), for quartic equations also the root of degree 4, but in this case ($\deg = 5$) you need a more complicated special function, and this function $z(a)$ suffices. This was classically known.

And then people, for example Hermite, tried to solve the equation of degree 6 using a function of one variable. But no one succeeded.

How this was supposed to be done? You kill the terms of the equation one by one using some substitutions and to find those substitutions you solve auxiliary equations. In degrees 2, 3 and 4 all the auxiliary equations can be solved and thus all the terms can be killed.

But in degree 5 there remains one coefficient you cannot kill. And in degree 6 two coefficients remain and you get the following normal equation which is sufficient to solve all the equations of degree 6:

$$z^6 + az^2 + bz + 1 = 0.$$

[1] I have lectured on this topological version of Abel's theorem to Moscow high-school children. This course, supplied with exercises, was later published by one of the listeners, V. Alekseev, in the form of a nice book "Abel's theorem in problems". Unfortunately, it was never translated into English. My student A. G. Khovansky in his thesis has extended these topological ideas to differential algebra. He proved by topological arguments the nonsolvability of some differential equations in terms of combinations of elementary functions and of arbitrary single-valued (holomorphic) functions in any number of variables. The idea was that the monodromy of the complex solution is too complicated to be the monodromy of such a combination.

Thus, there is a special function of two variables $z(a,b)$ which solves all the equations of degree 6. By the way, no one has ever proved that you really need two variables here — the conjecture is that there is no such function of one variable which would suffice, but no one has ever proved this.

For degree 7 the same procedure leaves you 3 coefficients

$$z^7 + az^3 + bz^2 + cz + 1 = 0,$$

which defines an algebraic function of 3 variables $z(a,b,c)$.[2] Hilbert asked whether functions of three variables really do exist.

If you have a function in two variables $z(a,b)$ and you put inside, say instead of a, a function in two variables $a(u,v)$ and continue in this way — you can get a function in any number of variables. You cannot do a similar thing with one variable, but, using only functions in two variables, you can construct functions with an arbitrary number of variables. Hilbert asked whether you really need functions in three variables to solve the universal equation of degree 7 written above and, more generally, whether you can represent *any* function in three variables as a superposition of functions in two variables — i. e. whether functions in three variables really do exist.

It is easy to see that using *discontinuous* functions it is always possible to find an expression in functions of just two variables representing any given function of three variables. Hilbert asked whether by combinations of *continuous* functions you can get any continuous function in two variables.

It is strange, by the way, that Hilbert formulated this problem of algebraic geometry in terms of functions of real variables — but he has done it. In 1956 I was an undergraduate student and Kolmogorov, my supervisor, was working on this problem. He proved that "functions in 4 variables do not exist": any continuous function in 4 variables or more can be reduced to continuous functions in 3 variables. But he was not able to reduce the number of variables from 3 to 2, and he gave this problem to me.

Kolmogorov had proved that it is sufficient to represent any function on a tree in Euclidean space — actually, to find a universal tree, such that any continuous function on this tree can be represented as a sum of three continuous functions, depending on one coordinate each. If you can do this, then "there are no functions of three variables" and you can reduce any continuous function to the continuous functions of two variables — and the function $z(a,b,c)$ is reducible too. This was a problem I managed to solve. It was essentially simple — I shall show you the idea because I will need it in a minute for dynamical systems. In the simplest example of Fig. 2:

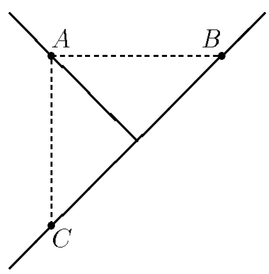

Figure 2: The Simplest Tree

[2] Starting from degree 9, one can kill one more coefficient. The known possibilities to kill more coefficients occur along a rather strange infinite sequence of degrees.

the claim is that on this tree any function can be represented as $f(x)+g(y)$. How to do this? You choose any point A of the tree, you take the value of the function at this point and you decompose it arbitrarily. Then at point B lying on the same horizontal level as A the second function is known, and the sum is known, and you get the value of the first function. And at point C lying on the same vertical line as A the first function is known, and the sum is known, and you get the value of the second function. That's all. If the tree is more complicated, you will have more branches, or even an infinite number of branches, you will have to work more, and in fact to make the infinite processes converge you need 3 variables, not 2, but here is the principal idea — that's how it worked.

Now I shall discuss this problem returning to polynomials, and I shall reformulate the Hilbert problem in the way I would like it to be formulated. The function $z(a,b,c)$ that satisfies $z^7+az^3+bz^2+cz+1=0$ is an algebraic function in three variables. You can construct algebraic functions in three variables from algebraic functions in two variables by superpositions. The problem is whether this particular algebraic function $z(a,b,c)$ can be represented as a combination of algebraic functions in two variables.[3] I would say that this was the *genuine* Hilbert problem. He did not formulate it in this way — unfortunately, and probably because of that this problem is still open: no one knows whether there is such a representation. I think this is a very nice problem, and many times I have attempted to do something in this direction.

Of course, you also have other types of functions, for instance, you have continuous functions, but you also have smooth functions. For smooth functions this problem has been attacked by Vitushkin in the beginning of the 50's. Vitushkin has proved that you have to lose some number of derivatives. For example, if you have a C^3 function in three variables, you cannot represent it by C^3 functions in two variables. The best you may hope to do is to express it by C^2 functions in two variables. The proof was based on a technology which he called the *theory of multidimensional variations* and which is in fact a version of integral geometry of the Chern classes describing the integrals over cycles in Grassmann varieties.

His technology was based on some evaluations of topological complexity in real algebraic geometry. This is also one of the main problems in mathematics. In the simplest case, for the curves, you have a polynomial equation in 2 variables, say of degree n, and you want to know the topology of the variety defined by this equation (in higher dimensions, by a system of such equations in the affine or projective space). This question was also formulated by Hilbert as a part of his 16th problem. For many years people have been working on this problem: Hilbert himself obtained some results, Harnack found the number of ovals for the curves. For high dimensional varieties the problem was studied by Petrovsky and his student Oleinik. They found the bounds for the Betti numbers of algebraic varieties defined by (systems of) polynomial equations in terms of the degrees and the dimensions. This was the crucial part of Vitushkin's proof of his statement about smooth functions. Of course, the fact that for generic functions in 3 variables you need some not very smooth functions in 2 variables does not imply anything for algebraic functions. Algebraic functions are such a small portion of all functions that you still can have such a representation for them.

[3] The given function being an entire algebraic function (without poles), it is natural to consider the representations using only entire algebraic functions. Thus one should distinguish two representation problems: that admitting only entire and that admitting arbitrary algebraic functions.

By the way, this theory by Oleinik and Petrovsky of 40's and 50's was later rediscovered in the West by Thom and by Milnor. Although they did quote Petrovsky and Oleinik, the results are mostly attributed to Milnor and Thom, who introduced the modern terminology related to Smith theory and to the interaction between the homology of real and complex manifolds. But stronger results were already present in those papers by Petrovsky and Oleinik and were heavily used by Vitushkin. I studied this as an undergraduate student because of its relation to the Hilbert problem.

I tried to do something on this problem later, and this was the motivation for me to study the algebraic function $z(a_1, ..., a_n)$ defined by the equation $z^n + a_1 z^{n-1} + ... + a_n = 0$. This function has a complicated discriminant hypersurface in the base space of coefficients $\mathbb{C}^n = \{(a_1, ..., a_n)\}$. The discriminant hypersurface is the set of all points where the function z is not nice, in particular, not smooth. In the 60's, around 1967, I started to think about how to use the topology of this object to deduce from it an obstacle to the representation of algebraic functions in terms of algebraic functions of a smaller number of variables. I thought that the topology of our algebraic function for higher n is complicated and, if there were an expression in terms of functions of fewer variables, then it should be simpler.

So I have studied the topology of this space — the complement to the discriminant — which is in fact the configuration space of sets of n points in $\mathbb{C}$ and the Eilenberg-MacLane space $K(\pi, 1)$ of the braid group. In one of the first papers on this subject — "On cohomology classes of algebraic functions, which are preserved by the Tchirnhausen transformation" (1970) — I mentioned an interesting analogy between the theory of fiber bundles and that of algebraic functions. The complement to the discriminant is the counterpart of the Grassmannian. The analogy (existing both in the complex and the real case) goes very far, for instance to the Pontryagin-Thom cobordism theory. These ideas were later used by many people — and recently have been even formalized (by A. Szücs and R. Rimányi, 1996).

This was the beginning of my work in singularity theory. And in fact all those works on ADE singularities, Coxeter groups and so on are a byproduct of the study of this special function $z(a_1, ..., a_n)$, of the question of how complicated is the topology of the discriminant.

Thinking on this, I decided to find the cohomology ring of the braid group. I have computed the first dozen of those groups (mostly torsion) and obtained a lot of information. Then D. B. Fuchs computed all those groups modulo 2. Later came the theorem of May–Segal on the relation of all these groups to the second space of loops of the 3-sphere $\Omega^2(S^3)$ which has the same homology as the braid group. In fact, this space $\Omega^2(S^3)$ is the Quillenization of the complement to the discriminant. All this was done in an attempt to find some higher dimensional properties of the braids which could prevent algebraic functions to be representable as combinations of the algebraic functions of fewer variables.

It is interesting that perhaps the most useful mushroom coming from this root, is the application of my results by Smale in his theory of complexity of the computations. In the topological complexity theory by Smale he discovered (using a theory which was essentially developed by Albert Schwarz years before) that the structure of the cohomology ring of the complementary space of the discriminant is an obstacle to compute numerically the roots of a complex polynomial with few branchings in the algorithm. For polynomials of degree n Smale proved (using essentially my computations of the cohomology) that the complexity is at least $n^{2/3}$ (you really need this number of branchings). One can obtain stronger results using the information about braids found in the marvellous paper by Fuchs "Cohomology of the braid group modulo 2" published in "Functional Analysis". By the way, the English translation of this paper was titled "Cohomology of the group kosmod 2", because 'braid group' is 'gruppa kos' in Russian. Everybody thought that the paper was related rather to space research and probably because of that it was not appreciated at that time as it should be. But later it was understood and now Vassiliev using the results of Fuchs has increased the number of topologically necessary branchings to $n - \log(n)$, which means that the topological complexity is almost n. (Smale had developed algorithms with n branchings).

But the origin of all this is in Hilbert's problem 13! Later, inspired by the analogy between between algebraic functions and fiber bundles, I constructed a theory of characteristic classes of entire algebraic functions and found a class, invariant under the substitutions, so I was able to prove some impossibility theorems using this cohomology. Later these works have been continued by V. Lin who proved the strongest result in this direction, also providing a correct basis for the Chebotarev's ideas dating back to the 40's on the topology of ramification of algebraic functions of several variables. Unfortunately what we were able to do was just to prove that there is no formula representing the function we need, $z(a_1, ...a_n)$, and only this function. Let me explain the difficulty that arises for cubic equations. The Cardano–Tartaglia formula gives you the roots you wish, but it also gives you some other, parasite roots, because you have some signs in the formula, some multivaluedness. The difficulty is how to understand such functions as, say, $\sqrt{z} - \sqrt{2z}$. What is the number of values of this function? There are several theories. For me, this function has four different values. If we understand the algebraic functions and their combinations in this way, then we can prove, using this cohomology theory, that there are functions which cannot be represented in the desired way. But then the cubic equation is not solvable too, which is not nice.

I think, however, that more work on this problem might bring some invariants of algebraic varieties and mappings of these varieties into each other which correspond to the superposition in such a way that one gets in this topological structure, in these algebraic invariants some memory of the number of variables one had in the functions participating in the superposition. There's perhaps some kind of a mixed Hodge structure whose weight filtration provides the information on the dimension of the smooth algebraic manifold from which a given cycle was born. Unfortunately, I cannot formulate this as an exact theorem, but I hope that such a theorem might exist and my conjecture is that our special function $z(a, b, c)$ cannot be represented as a combination of algebraic functions in two variables for some essentially topological reasons. I think, the representation remains impossible even if we replace all the algebraic functions in the superposition by non-holomorphic complex functions which are *topologically* equivalent to algebraic ones.

Now I'm coming to mechanics. As I have mentioned, my supervisor was Kolmogorov. He formulated the problem at the seminar and went to Paris for a semester. When he returned, I explained him what I had invented. He told me that I had solved the Hilbert problem and added, "well, now it would be very dangerous for you to ask me for the next problem, I think this will be harmful for you. I would be glad to discuss with you any kind of mathematics, but do not ask me the next problem. Choose it yourself, this will be much better for you."

Perhaps I should explain one more thing here. He took my first article, for the Doklady (the Russian *Comptes Rendus*) and he told me that the supervisor must write the first article of a student, the student being never able to write correctly, because it's a very different art from the art of solving problems and proving theorems. "I shall show you once, he said. A good student never needs the second experience of this kind". And indeed my first article was completely, word by word, rewritten by Kolmogorov. I wonder whether Kolmogorov had been involved in the writing of the first paper of Gelfand, who was also his student. This is one of the few papers signed by Gelfand alone, with no collaborators. Gelfand, whose brilliant papers and highly influential Seminar I always admired very much, mastered a special and enviable art of day-to-day collaboration with extremely gifted mathematicians (mostly his former students), resulting in important and beautiful joint papers. I dare to guess that these papers were physically written in most cases by the collaborators.

I have never collaborated with Gelfand. Recently, at the Zürich Congress he had asked me what is the reason for this. My answer was that I preferred to preserve good personal relations with him.

However, in the third volume of collected works of Gelfand there is a paper which is not by Gelfand and where he is not even a coauthor. It's my paper signed *Arnold*. I shall now explain what is it about and how is it related to my story.

When Kolmogorov suggested me to choose a problem myself, I wanted to choose something completely orthogonal to all the works of Kolmogorov. This was difficult, because he was working on so many subjects, but still I tried to invent my own problem. I had the list of Hilbert problems, written down one by one in my notebook. (Gelfand once saw it and laughed a lot). I was completely ignorant of the existence of anything else in mathematics at that time and so it was difficult for me to imagine a problem. You can see this from the problem I had chosen.

For a tree in Euclidean space I was able to represent any continuous function as the sum of continuous functions depending on one coordinate each. I decided to study other curves: what would happen if the curve was not a tree? [4] So I started to study the curves with cycles.

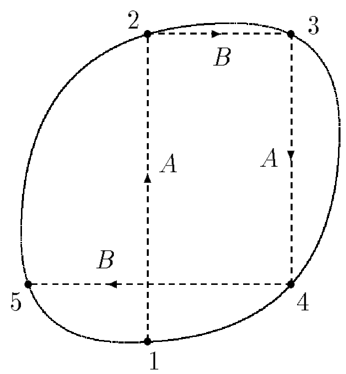

Figure 3: Dynamics on the oval

We choose a point and at this point we can decompose the function into the sum of functions of x and of y in an arbitrary way. We go upwards and in the new point of the curve we know the function of x. Since we know the sum, we can find the

[4] This problem has been recently (1996) reexamined in a nice paper by Skopenkov who listed all the obstacles to the representation of any continuous function on a plane curve in the form of the sum of two continuous functions of the coordinates. The same problem reappeared in singularity theory of the late 70's in the works of Dufour and Voronin studying the representations of functions on the germs of curves with cusps by the sums of smooth and of holomorphic functions of coordinates.

function of y at that point. We draw a horizontal line and find the decomposition at its other end, and so we continue. We thus get a dynamical system on this curve. We have two involutions of the curve: the vertical involution A and the horizontal involution B and hence we have a mapping $T = AB$, a diffeomorphism of the circle preserving the orientation.

After some experiments I was able to find that for these diffeomorphisms the rotation number exists, there might be resonances, periodic orbits and so on. Then I found that Poincaré had already studied diffeomorphisms of the circle onto itself preserving the orientation and created a theory for this. Then I read Poincaré and observed that for the ellipse, for instance, this transformation is equivalent to a rotation by an angle which depends on the ellipse and is in general incommensurable with 2π and hence represents an ergodic dynamical system.

In the resonance case, when there is a periodic orbit, the periodic points are obstacles to the solution of the initial problem, because the alternating sum of values of the function over a period must be 0, otherwise you cannot decompose it. If there are no periodic orbits, like in this irrational elliptical case, then you can formally continue, but you will have the convergence problem. For one orbit you can calculate everything, but then it is a question whether you get a continuous function. If you write the Fourier series for the mapping of a circle equivalent to a rotation, you immediately get the problem of resonances and the small denominator problem for the rotation number.

Just at this time Kolmogorov was giving at Moscow University a course on his works on small denominators and on Hamiltonian systems and on what is now called KAM theory. And so my attempt to invent something independent was completely unsuccessful!

I came to Kolmogorov with my theorems. "Well, he said. Here is my paper in Doklady '54. I think it will be good if you continue with this problem, try to think on applications to celestial mechanics and rigid body rotation. I am very glad that you have chosen a good problem." But I was completely upset by this, because it was just the opposite of my plan of a complete independence. However, it was an interesting nice problem. A few days later I learned from the Vakhania thesis which was defended at mechmath that this problem was in fact considered before me by Sobolev in a classified work of 1942 on missiles' rotation and oscillations, related to hydrodynamics. Resonances are dangerous there, since they can destroy the tank. This theory is also related to Sobolev's equation which is called so because it was first written by Poincaré in 1910.[5]

In this way I have started to learn some mathematics. I have read some other people's works and finally I discovered some papers by Siegel who was a personal friend of Kolmogorov when they stayed in Göttingen in the 1930's. Kolmogorov was not aware that Siegel had later worked on the small denominators problem. Siegel's paper was published in 1941, but was unknown to Kolmogorov. He knew about the works of Poincaré, of Denjoy and of Birkhoff, but not about Siegel. So he told me that we were in a very good company: "Siegel is really serious", he said.

I had discovered this Siegel theorem related to the normal forms for the circle rotations due to the system of education in Moscow University which was different from that in America. I think, it followed the German tradition that when you have a result and you wish to publish it, you have first to check the literature and

[5]Applications of the modern KAM theory to the corresponding hydrodynamical problems have been discovered recently by A. Babin, A. Mahalov and Nicolaenko.

to see whether someone else had ever studied it. We were told this at our first introductory course in library work where they taught us how to find, starting from zero information, everything you need. There was no Internet, of course, at that time, but still we were able to find the references, and this is how I discovered that Siegel existed.

The circle diffeomorphisms problem is related to many other problems, and I shall give you some examples. One of them is a problem which was also studied in classified works on the stability of the shells. This stability problem is very important, because a shell must be very thin, if you want to launch it far. But you cannot, by the architecture of the system, avoid non-convexity. In the convex part, you have good theorems by Cauchy that the metrics determines the shape and thus the shell is inflexible. But in the parts of hyperbolic curvature, no one knows the answer. Even for the idealized problem of the isometry, if you have, say, a torus in 3-space (this is one of the problems I like in mathematics), no one knows whether it is flexible, whether you can deform it without deforming the metrics. Only in some particular cases, for example, for the rotationally symmetric torus lying between two parallel planes, the inflexibility has recently been proved, as I was told, but the general case is still open. By the way, some polyhedra are flexible, and there is a theorem that the flexion of a polyhedron homeomorphic to the sphere does not change its volume. Problems of the interior geometry of surfaces in the Euclidean space are in fact closely related to the theory we are now discussing. Many years ago I have conjectured that any germ of a function vanishing at the origin and having there a critical point of finite multiplicity is diffeomorphic to the Gaussian curvature of the graph of a smooth function $z = f(x, y)$, but until now this conjecture is neither proved nor disproved.[6]

Returning to the shell stability, people who were constructing those shells have observed that the geometry of the characteristics, which are the asymptotic lines of the shell surface, can present obstacles to inflexibility. The asymptotic lines define a dynamical system similar to that which lead us to the diffeomorphisms of the circle. This dynamical system is in fact related to the characteristics of the string equation $\partial^2 f/\partial x \partial y = 0$. To represent a function as a sum of a function of x and a function of y means to solve the string equation. Our representability problem is thus the Dirichlet problem for the string equation. In the case of shells having the shape of a piece of one-sheet hyperboloid you have the Dirichlet problem for a hyperbolic equation. Professor Goldenweiser discovered that the resonances of the dynamical system on the boundary circle, depending on the shape of the hyperboloid, are responsible for the flexibility. This is not a theorem, as far as I know — there are some formal obstacles to inflexibility, but no mathematical proof of flexibility. People who studied these problems were doing real work, they were really constructing the shells. I have seen those shells: they are really flexible, but no one can prove that they are. It depends on the resonances. If you have resonances, then they are flexible in your hands — but I have not seen any mathematical proof of this.

I have written a paper on this subject, applying the technology of small denominators that Kolmogorov had invented in 1954 and adding some new results. Working on the circle analytic diffeomorphisms, I came to some conjectures in what is now called holomorphic dynamics, which I was unable to prove. One of them

[6] The problem is solved positively by the author in January 1998 (paper to appear in "Topological Methods in Nonlinear Analysis", 1998).

(claiming that an analytic circle diffeomorphism with a good rotation number is analytically conjugate to a rotation, the bad numbers forming a set of measure zero) was proved by M. Herman some twenty years later.

One of the others still remains a challenge and I shall formulate it here once more. It is a part of a general project of the "resonances materialization", providing the topological reasons of the series divergence in perturbation theory.

Consider an analytic diffeomorphism of a circle onto itself (defined by a holomorphic mapping of the neighbouring annulus onto another neighbouring annulus). Suppose that the mapping is analytically conjugate to an irrational rotation and that the closure of the maximal annulus where the conjugating holomorphic diffeomorphism is defined lies strictly inside the annulus where the initial holomorphic mapping is defined.

The conjecture is that *there exist periodic orbits of the initial holomorphic mapping in arbitrarily thin neighbourhoods of the boundary of the maximal annulus.* One is even tempted to conjecture that *the points of such orbits exist in any neighbourhood of any point of the boundary of the maximal annulus.*

As far as I know, these conjectures are neither proved nor disproved even for the standard circle mappings $x \mapsto x + a + b \sin x \pmod{2\pi}$, for which the conjectures have been initially formulated in 1958, or for the generic mappings. In 1958, I have also formulated similar conjectures for the boundary of the Siegel disk (centered at a fixed point).

In the case where there exists no analytical conjugation to a rotation (and where the maximal annulus is reduced to a circle and the Siegel disk to the fixed point) I have conjectured at least the generic presence of close periodic orbits. To be more precise, fix a bad rotation number. Then for generic analytic mappings with this rotation number one should expect the presence of periodic orbits in any neighbourhood of the invariant circle (of the fixed point).

Dealing with these problems, I observed, that to define what should and what should not be called generic in dynamical systems theory is highly nontrivial. Indeed, both the topological and the probabilistic approaches provide pathological answers (studied with some details by Halmos, Rokhlin and others). So the "physical genericity" notion should be different from what the mathematicians suggested.

The topological definitions (using the "Baire second category sets" etc) have the following defect. A phenomenon happenning with positive probability (in the sense of the measure of the set of the corresponding parameters values) might be neglectable from the point of view of the topological genericity (for instance, if this set of parameter values has an everywhere dense open complement).

This happens, for example, in the very natural families of circle diffeomorphisms (like $x \mapsto x + a + b \sin x$). Such a diffeomorphism is close to a rotation if b is small. From the topological point of view they are "generically" structurally stable. The structurally stable circle diffeomorphisms have attracting and repelling periodic orbits. They correspond to the resonances and have rational rotation numbers. The complementary set of the nonresonant ergodic diffeomorphisms is topologically neglectable.

But the ergodicity of the diffeomorfism happens with probability 99%, if b is small, while the "generic" behaviour is highly unprobable!

The alternative probabilistic approach has a different defect — the corresponding measure is always concentrated on the sets of functions with some specified smoothness. All the sets of functions which are smoother are then neglectable (have zero probability).

To overcome these difficulties I have then proposed to call generic those events, which happen when the parameter of the topologically generic finite-dimensional family of systems belongs to a positive measure set in the finite-dimensional parameter space.

For years I was thinking that this "physical genericity" definition was introduced by Kolmogorov in his Amsterdam talk.However recently Yu. S. Ilyashenko has explained me that Kolmogorov used a rather dual definition and that I was perhaps the first one to introduce (in 1959) the physical genericity notion described above (and now called "prevalence").

I was trying to apply this philosophy to many problems, for instance to the study of the chaotical dynamics of the area-preserving mapping in the neighbourhood of a hyperbolic fixed point whose separatrices have a homoclinical transversal intersection. My guess was that the positiveness of the measure of the "Smale's horseshoe" type Cantor set on which the dynamics is chaotical should be a physically generic event. As far as I know, this conjecture is still neither proved nor disproved. Its topological version (not referring to measure) was proved by V.M.Alexeev in a very general situation.

I was still an undergraduate student. Once Gelfand invited me to talk on the circle rotations and when I explained him my theorems, he said that they could be applied to what he was working on. He was working with M. L. Zeitlin on the mathematical model of the heart beat. In the heart, you have the resonance between the ventricles and the atria.

There is an atria-ventricular node and then there is an electric system synchronizing the ventricles and the atria. In the model of Gelfand and Zeitlin this system was described by a mapping of the circle into itself. My theorems were applicable, and I added several pages to my paper of 1959 on the applications of the theory of resonances and structural stability of the mappings of a circle into itself and on small denominators, to the heart beat problem. The paper was sent to Vinogradov's journal "Izvestiya of the Russian Academy of Sciences" for publication, but was rejected. Kolmogorov told me: "you should delete the Gelfand theory part". I was puzzled because I liked it, but Kolmogorov's reaction was that the heart beat theory, although very interesting, is not of the kind mathematicians should work on. "You should better concentrate on the three body problem", he told me. This was the only mathematical advice I ever got from Kolmogorov.[7] When I deleted the part on the Gelfand theory from the paper, it was accepted by Vinogradov and the shortened text appeared in the "Izvestiya" in 1961. Together with the heart beat theory I have deleted a paragraph about the influence of a small noise on the circle diffeomorphism invariant measure. Today these problems are included into the general Morse-Witten theory (but the discrete time case I was studying seems to remain still unsettled in the modern theory). Kolmogorov did not approve my naïve approach to the theory of asymptotics of solutions of the (discrete time) Focker-Planck equation in the small diffusion limit — which was of course his kingdom.

---

[7]Later, when I was his graduate student (in 1961), Kolmogorov learned about the existence of differential topology from Milnor's talk in Leningrad. He immediately suggested that I should include it in my graduate curriculum (thinking on the relations to the superposition problem). As a result, I started to study differential topology from Novikov, Fuchs and Rokhlin — and even served as an opponent for the candidate thesis of S. P. Novikov, the supergenial topologist and the glory of Russian mathematics, on the differential structures on the products of spheres.

The deleted heart beat part of the paper lied on my shelf for 25 years. Then two events happened.

The Canadian physiologist Leon Glass discovered that the mathematical theorems on resonances proved in my published paper have applications to heart beat. He published them in a paper and later in a book titled "From Clocks to Chaos".

About the same time Gelfand told me that he was preparing his collected works. "My congratulations, I said, I am very glad". "Yes, he answered, but I want your paper to be published in it". I was puzzled, but, since this was not the dangerous genuine collaboration, I gave him the old paper. And the paper was published almost simultaneously with the paper by Glass. The results were practically identical!

This was the story of how my works in what is now called KAM theory have started. Later I worked on the many body problem, following Kolmogorov's suggestion. Reading the "Méthodes nouvelles de la Mécanique Céleste" of Poincaré and having discussions with V. M. Alekseev during our weekly common "windows" (breaks between two classes) at Moscow University, I realized that that the problem of celestial mechanics has several difficulties which one might tackle separately. The first difficulty ("the limit degeneration") is already present in the simplest problem on the plane area-preserving diffeomorphisms near a fixed point, the so called Birkhoff problem. Suppose that the mapping linearized at a fixed point is a plane rotation. A rotation is *resonant* if the rotation angle is commensurable with 2π. If the linearized mapping is a non-resonant rotation, Birkhoff was able to reduce the mapping to a rotation (through a variable angle) using some symplectic (area preserving) formal coordinate change. The celebrated problem, formulated by Birkhoff, was to decide whether the fixed point was stable in this case. The difficulty is that Birkhoff's series (reducing the mapping to the Birkhoff normal form, which is a rotation through an angle depending on the distance to the fixed point) is generically divergent, due to the isolated periodic orbits born at the places where the rotation angle is commensurable with 2π — these periodic orbits form the "materialization of resonances" in this problem. I had solved this Birkhoff's problem and the paper was presented to Doklady by Kolmogorov in 1960.

At the Stockholm Congress of 1962, Moser, speaking about his recent results on the Birkhoff problem, explained how to replace the analyticity assumption by the continuity of the 333rd derivative. His method was not too far from Kolmogorov's 1954 paper, but the details were different. His result was even better than the solution of Birkhoff's problem: he had proved the stability provided that the rotation angles of the linearized mappings were not of the form $k\pi/2$ or $k\pi/3$. Rational numbers with denominators higher than 4 behave in this problem like irrational numbers! The resonances of order smaller than 5 are now called *strong* resonances, those of higher order — *weak* resonances. Moser discovered that the stability holds even in the presence of resonances, provided that they are weak.

Listening to Moser, I immediately understood that my 1960 stability proof was applicable (for the analytical mappings) to the case of weak resonances, while I had formulated the result only in the non-resonant case. Instead of studying the phenomenon, I was trying to solve a celebrated problem and was hypnotized by Birkhoff's formulation, which forbade all resonances. This was a good lesson: one should never be hypnotized by the authority of the predecessors.

The second main difficulty of the planetary motion problem was the so called "proper degeneration" (the terminology was introduced, I guess, by M. Born). The point is that some of the frequencies of the quasiperiodic motion of the perturbed system might be small together with the perturbation parameter. The simplest case is the adiabatic invariants theory. Consider, for instance, the motion of a charged particle along a surface under the influence of a strong magnetic field, orthogonal to the surface. Mathematically it is the problem of the description of the curves of prescribed large geodesic curvature on the surface. In the first approximation such a curve is a circle of small radius, the so called Larmor circle. But in the next approximation (provided by the adiabatic invariants theory) the center of the Larmor circle starts to move along the surface. The drift of the Larmor circle is described by the averaged system. In the adiabatic approximation, the center moves along the level line of the prescribed geodesic curvature (that is, the line where the intensity of the given magnetic field is constant). In the case of a constant magnetic field intensity the drift occurs in a higher order approximation. In this case the Larmor circle center follows the level line of the Gaussian curvature of the surface.

On a compact surface a typical approximate trajectory of the Larmor center is a closed curve and one may ask whether the genuine orbits of the charged particle remain close to these closed trajectories. The theory of the proper degeneration that I had constructed (the paper was presented to Doklady by Kolmogorov in 1960) gave a positive answer to this question, also providing many other physically important results on the infinite time behaviour of adiabatic invariants.

Just at that time these physical problems were formulated at Kolmogorov's seminar on dynamical systems by two well-known physicists, M. A. Leontovich and L. A. Artsimovich, who related them to the plasma confinement problem important for the controlled thermonuclear reaction project. Kolmogorov suggested that I send the resulting paper to ZETP, the main Russian physical journal.

A few weeks later M. A. Leontovich (who was, as far as I remember, the vice-chairman of the editorial board) invited me to his home, near the Atomic Energy (now Kurchatov) Institute to discuss the paper. Leontovich, heading the theoretical physics division of the thermonuclear controlled reaction project, was a friend of Kolmogorov and also of my father (he helped our family to survive when my father died and I was 11 years old). Treating me, as usually, with buckwheat porridge and, calling me, as usually, "Dimka" (he used this nickname until his death some 20 years later), Leontovich explained me the reasons why the paper cannot be published in ZETP:

i) the paper uses the forbidden words "theorem" and "proof",

ii) the paper claims "A implies B" while every physicist knows examples showing that B does not imply A,

iii) the paper uses non-physical notions like "Lebesgue measure", "invariant tori", "Diophantine conditions".

He proposed that I should rewrite the article.

Now I understand how right he was trying to defend a physical journal from the Bourbakist mathematical style.[8]

An author, claiming that A implies B, *must* say whether the inverse holds, otherwise the reader who is not spoiled by the mathematical slang would understand the claim as "A is equivalent to B". If mathematicians do not follow this rule, they are wrong.

Nowadays, every physicist, studying the Hamiltonian chaos or using KAM theory in plasma confinement or accelerator control problems, freely uses the Lebesgue

[8]Rumours later reached me that the paper had been reviewed by Landau, but I do not know if this was indeed the fact.

measure, the invariant tori and the Diophantine conditions. But in 1961 one of the first papers on what is now called KAM theory was, as we see, rejected by a leading physical journal for the use of these words.

I took the paper back from ZETP and it appeared a year later in Doklady. By that time, I had already combined the study of degenerations of both kinds and applied them to the planetary motion problem. The results were first presented at the conference on theoretical astronomy held in Moscow on 20–25 November 1961. The conference's main topic was the artificial satellite motion. I was delighted to meet there and make friends with M. L. Lidov whose students A. I. Neishtadt and M. L. Zieglin later made profound contributions to perturbation theory, averaging, adiabatic invariants, Hamiltonian chaos and materialization of resonances. The resulting theories are well known, and I shall only mention one small relevant detail.

Before I turn to this small detail, let me remark that now you have almost the whole picture of all my mathematical subjects. They all are starting from this problem of superpositions and you now see how they are connected. There is one more topic, hydrodynamics and hydrodynamic stability, but this is also related to the same origin.

When I finished the works on celestial mechanics and on other applications of what is called KAM, I tried to find some applications of the theory of dynamical systems to the continuous systems, in particular, to hydrodynamics. Kolmogorov, of course, was also a classic of hydrodynamics and he had a seminar at that time (1958-1959) called "Seminar on dynamical systems and hydrodynamics", where the celebrated work by Sinai and Meshalkin was done on Kolmogorov flows instability and continuous fractions, where the Kolmogorov–Sinai entropy was invented and so on.

In 1961 Smale came to Moscow. He was the first foreigner I met in my life. We were discussing a lot of interesting projects on the roof of Moscow University (he speaks of "the *steps* of Moscow University" in his reminiscences). Among other things we were discussing structural stability and he formulated the conjecture that torus diffeomorphisms and geodesic flows on negatively curved surfaces should be structurally stable. I have even written a paper with my friend Sinai, proving the first conjecture. Describing this proof at my Dynamical Systems seminar, I suggested that one might prove the second Smale's conjecture, identifying the perturbed geodesic with the nonperturbed one, connecting the same two points at the absolute. Next week Anosov reported his proof of it, but my proofs were wrong, since I was using too many derivatives of the invariant foliation.

And this is why I have never tried again to prove collective theorems. This happened in 1961–63, and since that time I was trying to find applications of this philosophy of structural stability. My first idea was to think on the hard balls model of statistical mechanics. I speculated that such systems might be considered as the limit case of geodesic flows on negatively curved manifolds (the curvature being concentrated on the collisions hypersurface). I had never proved anything in this direction, but I explained this idea to the greatest expert in dynamical systems and ergodic theory I knew, Sinai, and he started a long series of works continued by many people (let me mention only the recent works by D. Szász and V. Simányi — the project is still alive and not exhausted). My second project was to apply the new theories of dynamical systems to hydrodynamics. I started to discuss this project already in 1961–1962 with V. Yudovich and O. Ladyzhenskaya.

The idea was that because of the high sensitiveness of the flows on surfaces of negative curvature, the positive Lyapunov exponents are stable. The Euler and

Navier–Stokes equations contain many "parameters": the domains and the exterior forces. One might hope — had I conjectured — to find somewhere, at least numerically, an attractor of the Navier–Stokes equation on which the geodesic flow of a negatively curved surface is realized. It was of course very naïve, but I have tried, and in 1964 I made some numerical experiments (with the help of N. Vvedenskaya) on a model with 6 Fourier modes. Unfortunately I was unable to find the positive Lyapunov exponent numerically. At that time, computers produced very-very long tapes with numbers, kilometers of numbers. We were trying to imagine the orbit in 6-dimensional phase space looking at those numbers. I think that probably the Reynolds number was not sufficiently high, so what I have observed was a 3-dimensional torus in 6-dimensional space — a scenario predicted by Landau. But I was certain that with more work you might find the positive Lyapunov exponents, perhaps even the geodesic flow on a surface of negative curvature. This was the reason of my paper of 1966 on the differential geometry of infinite dimensional Lie groups, on the diffeomorphism groups which is the configuration space in hydrodynamics. I have calculated the curvature of this group[9] and I even used it to show that the weather prediction is impossible for periods longer than 2 weeks. In a month you lose 3 digits in the prediction, just because of the curvature. This instability is not the Euler instability, it's not describing a chaotic attractor of the Euler equations — but it comes from the same line of ideas. Thus all my hydrodynamical works were the byproduct of the works on dynamical systems which were the byproduct of the works on the Hilbert problem.

Trying to study the slow mixing in Hamilton dynamics, I have introduced the "interval exchange" model (as the simplest discrete time description of the events in the pseudo-periodic system whose Hamilton equations are defined by a closed but not exact 1-form on a surface of genus 2).

This intervals exchange model is so natural that I was always amazed not to find it in the works on ergodic theory prior to my 1963 paper where it was introduced. I have returned to pseudo-periodic topology many times since 1963. Pseudo-periodic functions are sums of linear functions and periodic functions, like $f = ax + by + \sin(x + y)$. Pseudo-periodic manifolds are those defined by pseudo-periodic equations, like the plane curve $f = 0$ (think of the Pacific coast of California and try to understand whether such a curve may have many unbounded components — a typical problem of the young pseudo-periodic topology, to which the interval exchange model also belongs).

The present state-of-the-art in pseudo-periodic topology might be understood from the forthcoming book by A. Zorich, S. Pajitnov and D. Panov (prepared in 1997 for the Advances in Soviet Mathematics AMS series). Studying the intervals exchange, Zorich discovered (by computer experimentation) astonishing new laws of correlation decay in such systems. In a recent joint work with Kontsevich they were able to explain most of these observations, relating them to the ergodic theory of geodesic flows on the Teichmüller spaces. The study of the intervals exchange model has thus returned to the non-exact Hamiltonians pseudo-periodic topology on higher genus surfaces, which was the initial motivation for the introduction of this model in 1963.

[9] A few years earlier I had translated into Russian Milnor's wonderful "Morse theory". My calculations in the 1966 paper were based on his short description of the Riemannian geometry. Milnor later (in 1972) proved the formulas for the curvature of a left invariant metric on a Lie group, which are essentially equivalent to my coordinate-free formulas of the hydrodynamical paper.

In the late sixties I have also explored some other areas related to dynamical systems with my undergraduate students:

— G. Margulis (in his first unpublished paper he started the theory of Diophantine approximations on submanifolds of Euclidean space, later continued by A. Pyartli, A. Neishtadt, V. Bakhtin),

— D. Kazhdan (who studied the ergodic properties of the Euclidean actions of free groups, continued later by R. Grigorchuk),

— N. Nekhoroshev (whom I have persuaded to apply the Diophantine net geometry to the problem of the action variables drift)

— A. Kushnirenko (slow mixing, structural stability of analytical semisimple groups actions, later — Newton polyhedra and fewnomials conjecture),

— A. Khovanskii (non-solvability of differential equations, later — Newton polyhedra and fewnomials theory).

I turn now to the KAM theory. This theory is called KAM, or Kolmogorov–Arnold–Moser, and people say that there is even a KAM theorem. I was never able to understand what theorem is it. In 1954 Kolmogorov proved his marvellous theorem on the preservation of the tori in Hamiltonian systems, when the Hamiltonian is almost integrable and all functions are analytical. What I have contributed was the study of some degenerate cases — when one of the frequencies is 0 in the non-perturbed system or when you have the vicinity of the fixed points or periodic points or tori of a smaller dimension — and then applications to celestial mechanics. All these facts are separate theorems. My main contribution was the discovery (in 1964) of the universal mechanism of instability in the systems with many degrees of freedom, close to integrable, — later called "Arnold diffusion" by the physicists.

In 1962 Moser has extended Kolmogorov's theorem to the case of smooth functions.[10] In the first papers of Moser the number of derivatives was enormous. Now we know that in the simplest case of plane rotation you only need 3 derivatives, and this is just the limit, the critical number of derivatives. But in the beginning the number was 333. For Kolmogorov, this was like a complete change of philosophy, he told me, because he was expecting and even claiming in his Amsterdam talk that the result should be wrong even in C^∞ and one needs analyticity or something close to it, like the Gervais condition.

Moser criticized that a proof of the theorem in the case of analytic Hamiltonians was never published by Kolmogorov. I think that Kolmogorov was reluctant to write the proof, because he had other things to do in the years still remaining of active work — which is a challenge, when you are 60. According to Moser, the first proof was published by Arnold. My opinion, however, is that Kolmogorov's theorem was proved by Kolmogorov.

Thanks for your attention.

Question (*J. Milnor*). You often told us about important mathematical work in Russia we did not know about and you gave another example today. I wonder if you can explain us how do you locate something interesting in the literature starting with zero information.

[10]It is interesting to note that, when Moser's papers appeared, some American mathematicians began to publish their papers that "extended the Moser theorem to the case of analytical functions". J. Moser himself has never supported these attempts to attribute Kolmogorov's results to him.

Answer. First of all (it is especially important for the Americans), do not forget that some mathematical results appear in Russian, in French, in German, in Japanese...

To learn the state-of-the-art in a domain new for me, I usually start with the German *Encyclopædia of Mathematical Sciences* edited by Klein and published around 1925. It contains an enormous amount of information. Then there are papers in the *Jahrbuch* which was published before the Mathematical Reviews and Zentralblatt had been organized — it is full of information. Then, I usually consult the collected works of Felix Klein and Poincaré. In Klein's *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert* there's a lot of information on whatever has happened in 19th century and before.

The other books by Klein are also extremely informative. For instance, in Klein you can find the article by Emil Artin on continuous fractions and on braid groups, and I think you can find there the 1918 article by Radon which has never been published and which contains the first draft of the Berry phase theory. It is the theory of the adiabatic pendulum, slowly moving along a surface. According to Radon, the Levi-Civita connection can be defined by the adiabatic invariants theory. You have the fast phase oscillations in a system, like a pendulum, located at a point of the surface. You slowly move the point along the surface, and the direction of the pendulum oscillations is parallel transported according to the Levi-Civita connection. I think this is the most physical way to define the Levi-Civita connection which otherwise is mathematically a rather complicated thing in higher dimensions. The adiabatic transportation defines it as a physically natural object. I think this can't be found in any textbook, I only find this in Klein.

Then from the 40's starts the Mathematical Reviews and the Zentralblatt and then later the Russian *Referativnyi Zhurnal Matematika* and then it's more or less OK. Of course, the MR and Zbl are not sufficient, because if you are trying to find a Russian paper and if in the Russian paper it was written that *A implies B*, then in the translation and hence in the MR you will usually find that *A is implied by B*. However, if you understand the topic, you can reconstruct the author's correct statements.

Also, in Russia, mathematics has never been completely separated from physics and mechanics. There were the same people doing mathematics, mechanics and physics. For example, in Kolmogorov's collected works there is a paper by Kolmogorov and Leontovich, who was a famous physicist, on the neighbourhood of a Brownian trajectory. This is a paper of a mathematician and a physicist which consists of two parts, the mathematical part containing evaluations of integrals, asymptotics, Riemannian surfaces, monodromies, Picard-Lefschetz theorem etc. and the physical part containing the background equations and so on. And, *of course*, the mathematical part was written by Leontovich, and the physical part by Kolmogorov. This is very typical for Russia.

Another useful rule is that you can usually learn a lot about the state-of-the-art in some domain from your neighbours. Many times I have used the opportunity to pose silly questions to Fuchs, Novikov, Sinai, Anosov, V. M. Alekseev, Rokhlin and later to my own students. Once I asked the greatest number-theorist I knew, whose works in many domains of mathematics I always admired, a question in number theory. His answer was, "Sorry, I have forgotten all of it, I am no longer a number theorist: several months ago, I have turned to another domain, logic". "Well, I said, can you recommend me a graduate student of yours still interested in number theory, to explain me what is known?" "How naïve you are, he replied, you think that my students may continue to be interested in number theory while I have turned to logic already three months ago!"

To facilitate the search of mathematical information, Russian mathematicians have tried to cover most of the present day mathematics in the more than one hundred volumes of the *Encyclopædia of mathematical sciences*, several dozens of which have already been translated into English. The idea of this collection was to represent the living mathematics as an experimental science, as a part of physics rather than the systematic study of corollaries of the arbitrary sets of axioms, as Hilbert and Bourbaki proposed. I hope that this Encyclopædia is useful as the source describing the real origins of mathematical ideas and methods (see, for instance, my paper on the catastrophe theory in vol. 5). Unfortunately, in the Library of Congress, and hence in all the USA libraries, the volumes of the Encyclopædia of Mathematical Sciences are scattered according to the author/subject alphabetical order, which makes its use as an encyclopædia extremely difficult. I have seen, however, all the collection arranged on one shelf in some European universities, for example in France.

Of course, in spite of all these precautions, you may discover too late that your result was known many years ago. It happened to me to rediscover the results of many mathematicians. And I am especially grateful to Prof. Milnor who explained me the relation to the works of G. Tyurina of my ADE paper of 1972 dedicated to her memory.