FORMULA AND SHAPE

NICOLAI VOROBJOV

1. Introduction

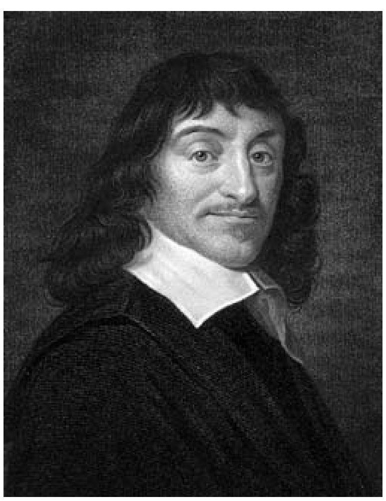
This lecture is about relations between formulae and shapes. Naturally, about shapes defined by formulae, but also about *shapes of formulae.*

Traditionally, since ancient times, mathematics consisted of two major parts: arithmetic (the study of numbers) and geometry (the study of figures). As time went on, new notions have been introduced and new subjects appeared within mathematics. Nevertheless, the fundamental duality between a *formula* and a *shape*, between algebraic manipulations and geometric imagination still is the heart of mathematics.

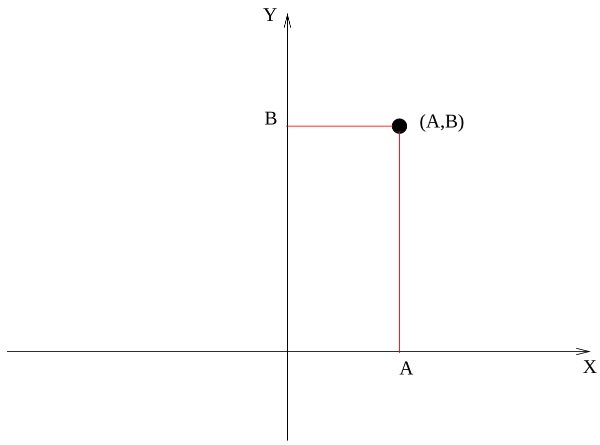
Mathematical thinking about spacial forms always includes rigorous (formal) reasoning. On the other hand, one can acquire intuition necessary to work with abstract algebraic or analytic objects only through visual concepts of some kind.

Probably the most important idea, linking geometry to algebra, is Cartesian coordinates.

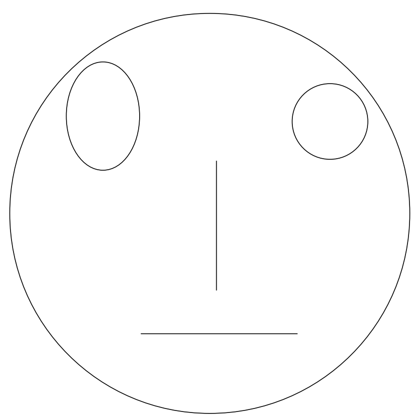
The idea of René Descartes (*Lat.* Cartesius) and Pierre Fermat



was to associate a point in the plane with a pair of numbers, its *coordinates*, according to the picture:



Note that this point is the only one in the plane satisfying the system of equations $X - A = Y - B = 0$. Generally, how can we represent a geometric object in the computer? Or just describe it accurately to a colleague? One very precise way of doing it would be to define the object by a *formula*. For example, the following curve

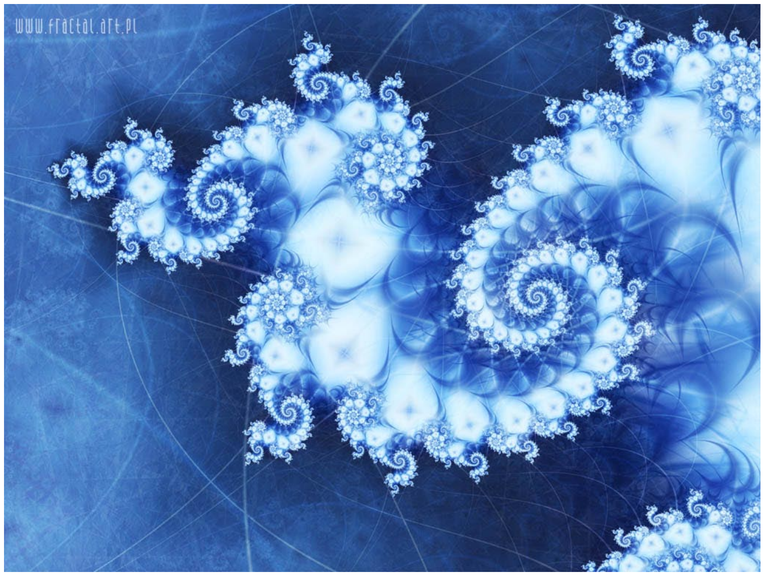


is defined by the equation

$$\big(X^2+Y^2-25\big)\big(|X-2|+|X+2|-4+(Y+3)^2\big)\big(X^2+|Y-1|+|Y+2|-3\big)$$

$$\left((X+3)^2+\frac{(Y-3/2)^2}{2}-1\right)\big((X-3)^2+(Y-2)^2-1\big)=0. \tag{1.1}$$

A bit later I'll specify more precisely what I mean by a *formula*. Formulae in our sense define only "tame" geometric objects, for example they can't define a fractal:



(these are defined by dynamical systems).

The method of coordinates allows to use the geometric intuition on the sets of solutions of equations. The result of this approach for *algebraic* equations, coordinates being *complex numbers*, is one of the central areas of modern mathematics – algebraic geometry.

In this lecture I want to discuss the following fundamental principle connecting algebra and topology:

> **A geometric object described by a "simple" formula should have a "simple shape'.**

Thus, we have to learn how to measure complexities of formulae and complexities of geometric objects.

2. Complexity of a formula

Let us start with formulae. We will consider the ones built from *multivariate polynomials*.

Let $X_1, X_2, \dots, X_n$ be some distinct letters (symbols). A polynomial in variables $X_1, \dots, X_n$ is *any* expression one can construct from variables and numbers using only additions, subtractions and multiplications.

For example, $X^2 + 2Y^2 - 7XY + X - 25$ is a polynomial in two variables, any polynomial in one variable can be written in the form

$$a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0, \tag{2.1}$$

where some of coefficients a_i can be equal to zero.

(When I was a student, I knew a girl, an archeologist, who said she was completely lost in maths because she could never understand what is X.)

Notice that the left-hand side of the expression (1.1) is not a polynomial since it contains, additionally, symbols $| \cdot |$ of absolute value.

What is natural to take as a complexity measure of a polynomial? Computer Science approach will be to take the number of symbols in the expression, but mathematically it does not lead to anything interesting.

Degree. A classical measure is the *degree* of a polynomial. Every polynomial, for example

$$X^3Y^5Z^{11} - 2X^2Z^8 + 3Y^7 + 100, \tag{2.2}$$

is the sum of terms with non-zero coefficients, called *monomials.* The degree of a polynomial is the maximal number of multiplications needed to compute a monomial.

In one variable, a polynomial of the kind (2.1) has degree d. Clearly, d is related to the "length" of the expression, if most of the summands have non-zero coefficients.

On the other hand,

$$X^{100} - 1$$

is obviously a "simple" expression. This prompts another complexity measure:

Number of monomials. Now, $X^{100} - 1$ has complexity two, and (2.2) also has a small complexity (four) comparing to its degree.

A polynomial considered with this complexity measure is called *fewnomial.*

On the other hand,

$$(X + 1)^{100}$$

is obviously a "simple" expression. It is equal to

$$X^{100} + 100X^{99} + 4950X^{98} + \cdots + 100X + 1$$

(Newton's binomial), so the number of monomials is large. This prompts yet another complexity measure:

Additive complexity. This is the minimal number of additions or subtractions needed to *compute* the polynomial, using any number of multiplications. Thus, the additive complexity of

$$(X + 1)^{100}$$

is just 1. A version of this measure is when we allow also an unlimited number of divisions. Then the additive complexity of the polynomial

$$X^{100} + X^{99} + X^{98} + \cdots + X + 1$$

is small since this expression is equal to

$$\frac{X^{101} - 1}{X - 1}$$

being the sum of a *geometric progression.*

3. Complexity of shape: Bezout and Khovanskii

In the case when a formula describes a finite set of points in an Euclidean space, the complexity of this set is easy to define: it's just the number of these points. Let us first consider the set defined by a single polynomial equation in one variable:

$$a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0 = 0.$$

In terms of the **degree** measure, by the Fundamental Theorem of Algebra, the number of distinct complex numbers satisfying this equation is at most d. It follows that the number of real numbers satisfying the equation (number of distinct points on the straight line) is also at most d.

Let the polynomial F have m monomials (of course, $m \leq d+1$). *Descartes rule* implies that the number of *positive* real solutions of the equation $F = 0$ is less than m (Descartes rule itself is slightly more complicated). Replacing X by $-X$ and adding 0, we see that the number of *all* solutions is at most $2m + 1$.

Now let us consider a finite set of points in the n-dimensional space, defined by a system of equations. If we measure the complexity of a formula in terms of the degree, then the *fundamental principle* above can be made quantitative due to *Bezout Theorem*.

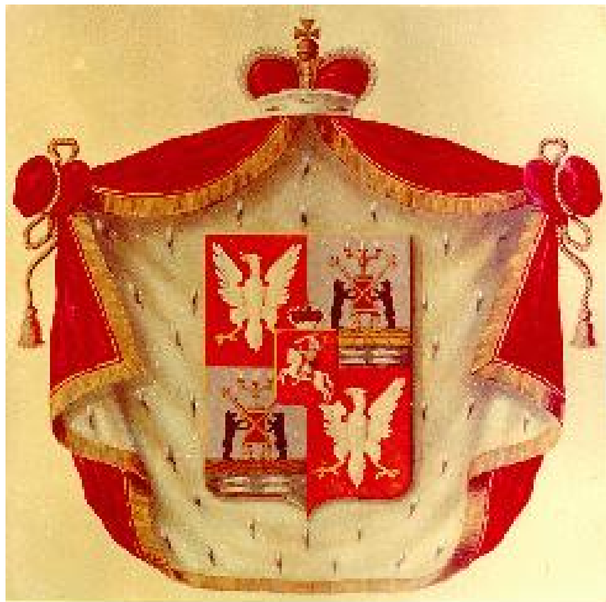
It implies that

> *the number of real solutions of a* generic *system (conjunction) of n polynomial equations of degrees $d_1, d_2, \ldots, d_n$ respectively, in n variables, does not exceed the product*
>
> $$D = d_1 d_2 \cdots d_n.$$

It is easy to find an example of such system with the number of real solutions exactly D, so the bound is *tight*. It is not difficult to deduce from Bezout that if the formula is now *any* system of k polynomial equations, and maybe inequalities, and if the number of solutions is finite, then this number does not exceed

$$d(2d-1)^{n-1}, \text{ where } d = \max\{d_1, \ldots, d_k\}.$$

Unlike the degree complexity measure, the analogy of the Bezout Theorem for *fewnomials* is a relatively recent result (1970s) and is due to Askold Khovanskii.



An unusual thing about Khovanskii is that he is a prince (knyaz') of the most ancient Russian noble family. More ancient and "noble" that Romanovy, the tzar family that ruled Russia for more than 300 years before 1917 revolution, and with whom Khovanskiis had fallen out during the Moscow Uprising in the second half of XVII century. These events are known in history as *Khovanshchina*, and are behind the famous opera of the same name by Modest Mussorgskii.

You would recall that the opera ends up with mass suicide of prince Khovanskii's followers. Askold's colleagues sometimes affectionately refer to the theory of fewnomials as to *Khovanshchina*.

Khovanskii's Theorem:

> *the number of real solutions* with all positive coordinates *of a generic system of n polynomial equations in n variables, having m different*

monomials in all polynomials does not exceed

$$2^{m(m-1)}(n+1)^m.$$

In particular, this estimate does not depend on degrees of polynomials. Unlike the bound from Bezout Theorem, it is not sharp. Some improvements were achieved recently by Bihan and Sottile.

Actually, Khovanskii proved much more than this: an upper bound on the number of solutions for real analytic functions satisfying triangular systems of partial differential equations with polynomial coefficients (Pfaffian functions). This class of functions includes iterations of exponentials, trigonometric functions in appropriate domains. Fewnomials are a very special case.

An upper bound in terms of the *additive complexity* can be easily obtained by introducing a new variable for every addition operation and thus reducing the problem to Khovanskii's Theorem. I leave it as an exercise :)

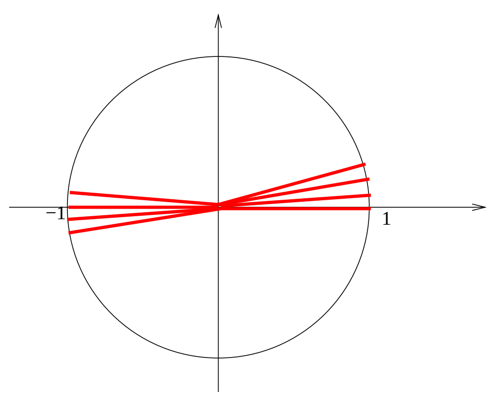
Before Khovanskii, the existence of a good upper bound in terms of the additive complexity was a famous open problem in computer science. That is because *upper bounds* in mathematics become *lower bounds* in computer science: if the number of solutions is bounded from above via the number of additions needed to compute a polynomial, then this number of additions is bounded from below by the number of solutions. If there is many solutions then the task of computing the polynomial is hard. Theoretical computer science is very much concerned by proving that various things are hard to compute.

4. Over complex numbers

You've probably noticed that the *fundamental principle* does not quite work for equations over *complex numbers*, for example

$$X^{100} - 1 = 0$$

has exactly 100 different complex numbers as solutions:

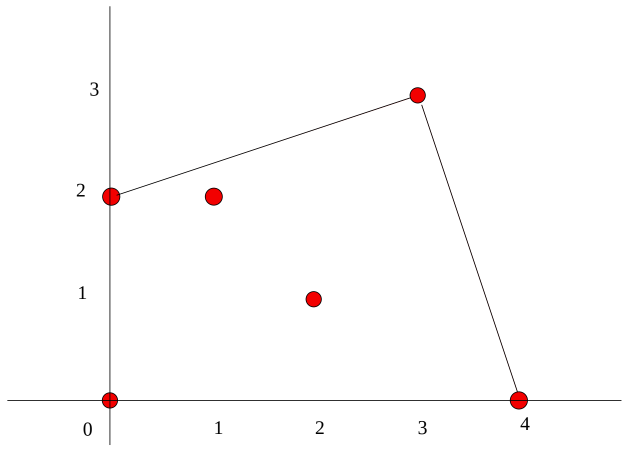


Over complex numbers we need to use another measure of the complexity of a polynomial: the volume of its Newton polyhedron. What is that?

Consider for example the polynomial

$$X^3Y^3 + 2X^2Y - XY^2 + 5X^4 - 3Y^2 + 1,$$

and for each monomial X^iY^j draw a point with coordinates (i, j) in the plane (red points in the picture):



Then take the *convex hull* of this set of points, i.e., the smallest convex set containing all these points. (You can imagine that red points are pegs sticking out of the screen, stretch a rubber band around the pegs, and the let it go. The band will become the boundary of the convex hull.) This convex hull is called *Newton polyhedron.*

Kushnirenko's theorem:

> *the number of complex* $\neq 0$ *solutions of a generic system of* n *polynomial equations in* n *variables, having the same Newton polyhedron, does not exceed the volume of this polyhedron multiplied by* $n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$.

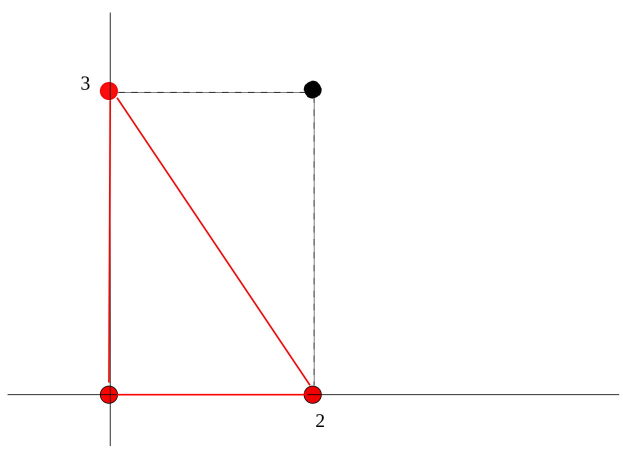
Example 1. A single equation $X^{100} - 1 = 0$. Newton polyhedron in this case is just a segment of a straight line, and its volume is its length, and $n = 1$. We get the same result as in Fundamental Theorem of Algebra.

Example 2. A system of equations

$$X^2 + Y^3 - 1 = X^2 - Y^3 + 2 = 0$$

obviously has six complex solutions. Exercise: prove it, and find them all! (Hint: introduce new unknowns $U = X^2$ and $V = Y^3$.)

The common Newton polygon here looks like this (in red):



The volume in 2D is called *area*. The area of the polygon is equal to 3. Hence, by Kushnirenko's Theorem the number of solutions is indeed $3 \times 2! = 3 \times 2 = 6$.

If polynomials have different Newton polyhedra, then in the theorem one should take their *mixed volume*.

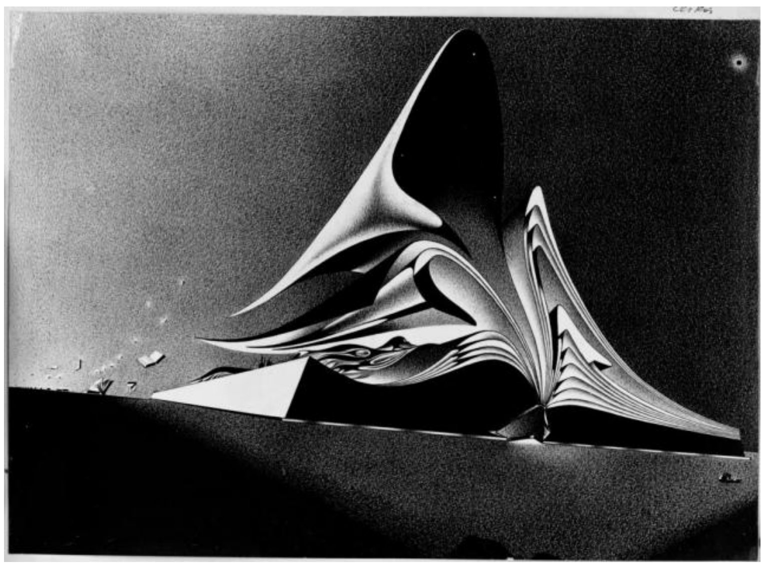
Khovanskii found a common generalization of his theorem on fewnomials and Kushnirenko's Theorem. To get a flavor of this generalization, observe that in the case of equation

$$X^d - 1 = 0,$$

for solutions X which satisfy the restriction $\alpha_0 \leq \arg X \leq \alpha_0 + \alpha$, where α is small, the Descartes bound takes place, while with the growth of d the solutions become uniformly distributed by arguments.

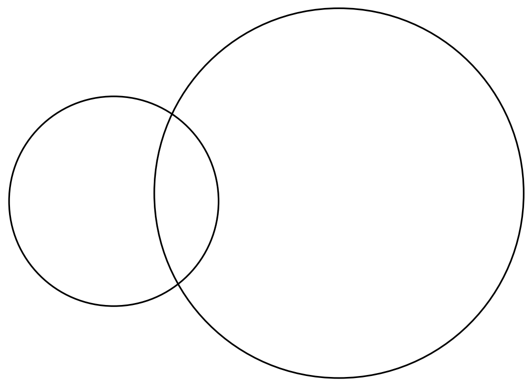
5. Complexity of higher-dimensional shapes

So far our geometric objects (sets) consisted of finite number of points, and this number is a natural measure of their complexity. But what is natural to take as complexity of the sets like this:



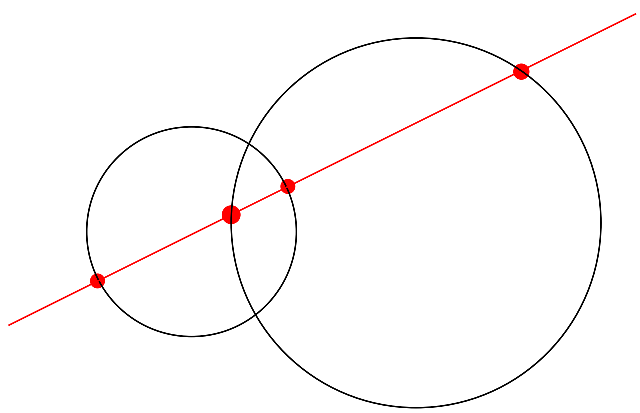
(This is a work of Anatolii Fomenko, a renown topologist, artist, and a highly controversial figure in modern Russian culture.)

Or something simpler, like this:



Various approaches are possible.

One way is to stick a straight line through the set in such a way that the number of intersection points is finite and maximal:



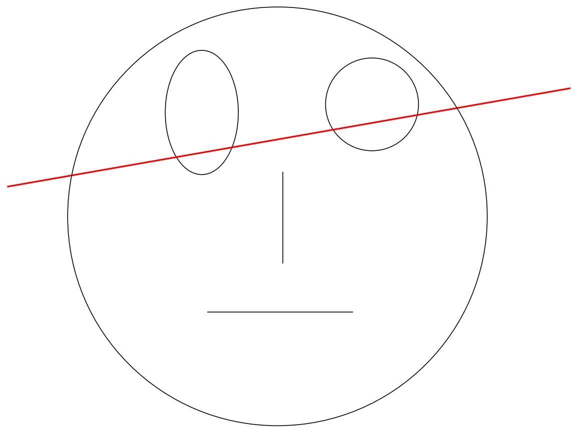
This number is called the *degree* of the set. In the similar way the degree can be defined for the Fomenko's picture above. Intuitively, the degree can serve as a complexity measure for a set.

Of course, in general, a set has to be intersected with the linear space of complementary to the set's dimension. For example, if the set is a curve in 3-dimensional space, it should be intersected with a plane.

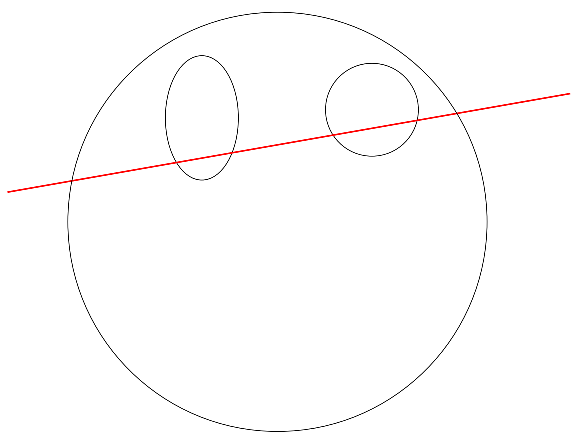
In any case, when the set is defined by a system of equations, the intersection points are also defined by a bit larger system of equations. But that is very convenient, because we already know how to estimate the number of isolated solutions of systems of equations.

Thus, by Bezout's Theorem, if the set consists of points satisfying a system of k polynomial equations of degrees $d_1, d_2, \dots, d_k$, in n variables ($k \leq n$) then the degree (i.e., the complexity) of this set is at most $D = d_1 \cdot d_2 \cdots d_k$.

We notice, however, that the degree may not capture the intuitive complexity in full:



no straight line through this picture is able to cross all components of the set. So the degree complexity of this set is the same as the complexity of



which is not right.

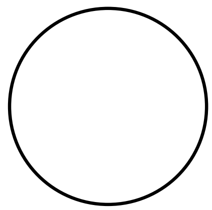
This happens because the degree is really an *algebraic* complexity measure and behaves awkwardly over real numbers. To make it adequate, we should take the degree of the *complexification* of the set, but this is a different story.

I am more interested in complexity measures that are topological invariants. In topology two geometric objects are considered equivalent if one can be obtained from another by continuous transformations, without cutting or pasting. This vague description can be formalized in essentially different ways, the one useful for us is *homotopy equivalence*.

Examples of homotopy equivalent sets are:

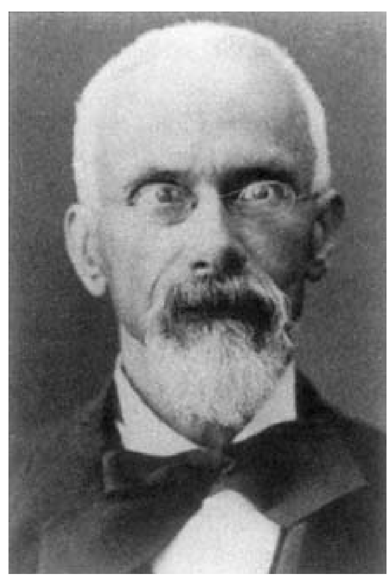
(Solid cup, solid ball, and a point.)

Or another triple:

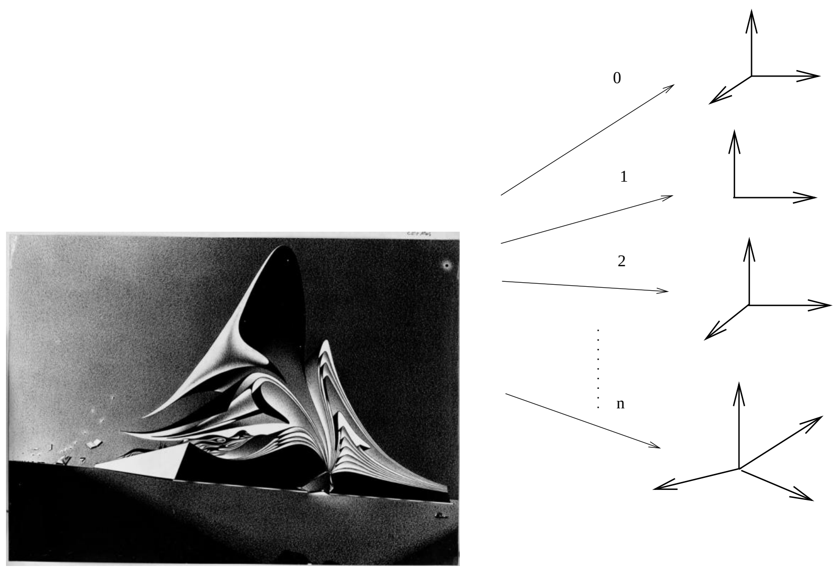


Complexities within both groups are the same. Clearly, the complexity of the second triple is larger.

This complexity is called *the sum of Betti numbers*. Betti numbers and the whole subject of *algebraic topology*, to which they belong, were invented by Henri Poincaré who referred to some ideas of Enrico Betti.



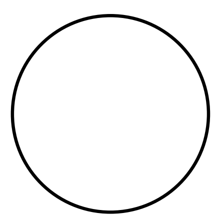
The exact definition of Betti numbers is quite complicated. As its consequence, with a given topological space, a sequence of vector spaces, called *homology groups*, is associated:



(Here numbered arrows are called *functors*.)

The dimension of nth vector space H_n is called nth Betti number. The sequence of Betti numbers is the same for homotopy equivalent spaces, for example, it is $\dim H_0 = 1,\ \dim H_1 = 1,\ \dim H_2 = 0$ for

and for

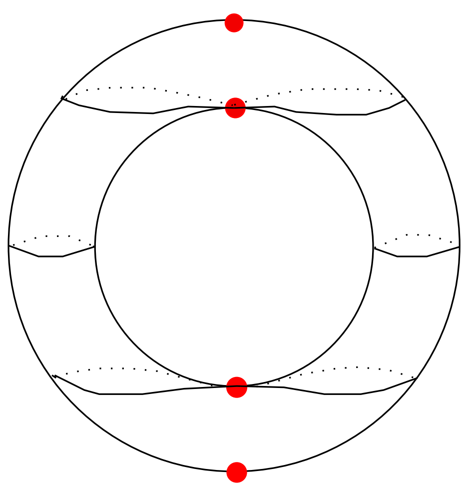


The sum of all Betti numbers can serve as a measure of complexity of a given set.

The following observation is crucial for estimating the sum of Betti numbers in terms of the complexity of the defining formula, and also may explain why this measure is natural. It is called *Morse Theory* by the name of its inventor Marston Morse.

Suppose that a surface is *smooth* (does not have sharp “angles”), and compact (does not stretch to infinity).

For example:



Let us move the horizontal plane from far above down. Indicate by red the points where the plane touches the surface, but not cuts through the surface (i.e., is *tangent*), while moving. These points are called *critical points*. In this example there is a finite number of them (four). We might have been unlucky if the surface was oriented symmetrically with respect to the vertical line, then the set of critical points would be infinite (the union of two circles). But clearly we can always rotate of surface slightly, so that the number of critical points is finite, they all lie on different levels, and moreover the surface has a non-zero curvature (whatever that means) at each of these points.

According to Morse Theory, the sum of Betti numbers does not exceed the number of critical points. (In our example these two numbers coincide.) Thus, the number of critical points can be considered as a complexity measure of the surface. It is consistent with geometric intuition: every connected component, every "hole", produces at least one critical point.

Notice that the set of critical points coincides with the set of all solutions of a system of equations. (If the surface in the example is defined by the equation $F = 0$, then this system is

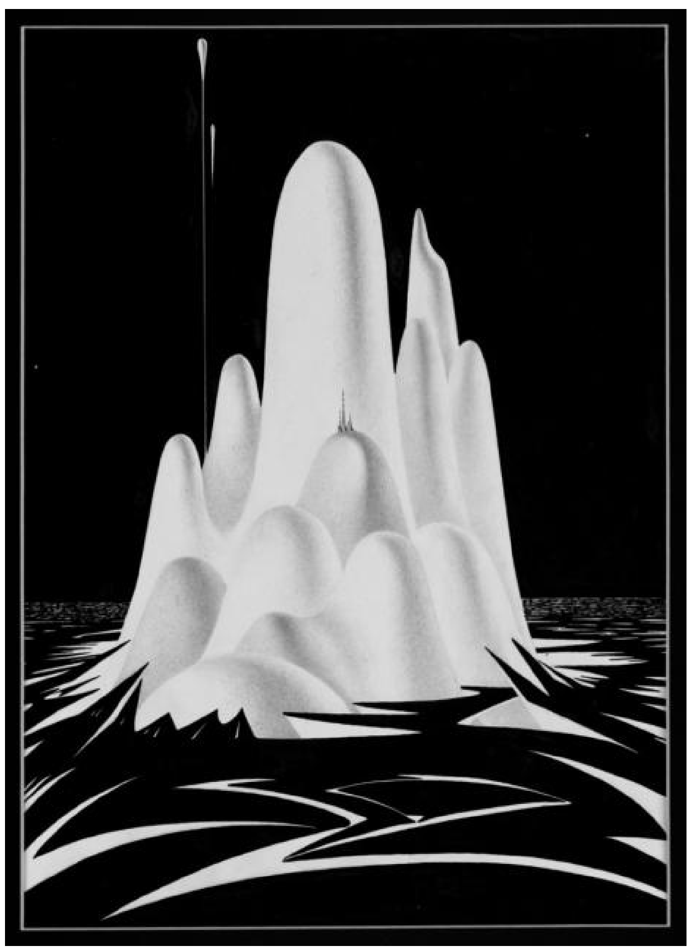
$$F = \frac{\partial F}{\partial X_1} = \frac{\partial F}{\partial X_2} = 0.)$$

But that is very convenient, because we already know how to estimate the number of isolated solutions of systems of equations.

Thus, if the set consists of points satisfying a polynomial equation of degree d in n variables, then the sum of Betti numbers (i.e., the complexity) of this set is at most

$$d(d-1)^{n-1}.$$

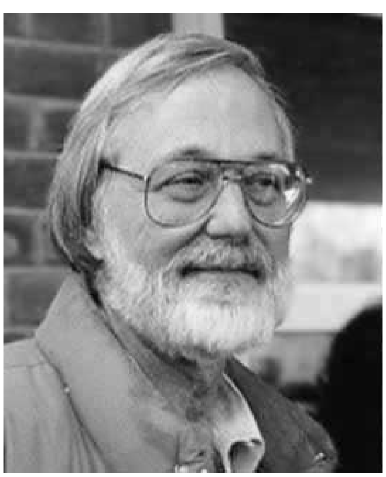
Fomenko's vision of a smooth surface:



Passing to general (not necessarily smooth) sets of arbitrary dimensions defined by systems of equations is a difficult problem, which is a natural extension of Hilbert's 16th problem. It was first solved in late 1940s by Ivan Petrovskii and his student at the time, Olga Oleinik.

Petrovskii was an absolutely remarkable man. He served as Rector (in our terminology – VC) of the Moscow University (a gigantic institution even then), member of the Central Committee, of the Supreme Soviet, at the same time being one of the wold's leading mathematicians. People say he was also a decent man, at those troubled times.

The problem was also independently solved in 1960s by John Milnor and by René Thom, the 20th century greatest topologists.



My contribution to the subject is a further generalization of Petrovskii-Oleinik-Thom-Milnor bounds to sets defined by more general formulae than just conjunction of equations and inequalities. One way of generalizing such formulae is to consider unions of these sets, images under maps, e.g., projections, and complements to images. In the language of logic it means that we consider sets defined by formulae with quantifiers. Another type of generalization appears when we pass from polynomials to more general functions, like above mentioned Pfaffian (which include fewnomials), definable in o-minimal structures. Some results in this direction were obtained by Basu, Pollack, Roy, and Zell. With my American colleague Andrei Gabrielov, we managed to advance further.

I don't have neither time nor popularization skills to discuss these results, instead I'll show you two last pictures by Fomenko which, I feel, capture the mood of the technique we invented.