

Modern Variational Calculus in Large

Anatoly T. Fomenko, Alexandr O. Ivanov, and Alexey A. Tuzhilin

January 31, 2000

The aim of the present paper is to illustrate the modern state of variational calculus in large with several physical and geometrical examples. In spite of the great progress in variational calculus during the last century, some classical variational problems are still far from its final solution. But, as it often happens in Science, the attempts to solve classical problems give rise to beautiful theories that give us an understanding of the problems and, sometimes, lead to unexpected discoveries. As bright examples we have chosen Minimal Surfaces Theory and Extreme Networks Theory, and here we explain our choice. Both theories deal with visual geometrical objects, namely, with two-dimensional surfaces in Euclidean space and one-dimensional continuums — networks, respectively. Both theories combine into themselves concepts and methods of many branches of mathematics, such as differential geometry, calculus of variations, functional analysis, topology, discrete mathematics, etc. But differences also take place. Minimal Surfaces Theory is very well developed. In the modern sense of the words, it was formed in 30th–40th by the pioneer works of T. Rado, J. Douglas, and Ch. Morry then it was developed by H. Federer, W. Fleming, and E. Reifenberg in 60th and A. Fomenko and Dao Trong Thi in 70th. It contains many deep and beautiful theorems, but also many open questions, of course, and it attracts the attention of many scientists all over the world. Quite the contrary, Extreme Networks Theory is now just in the making. It has begun to form in works of A. Ivanov and A. Tuzhilin in 80th–90th, who has understood that it could be useful to apply differential methods in investigations of Steiner problem that was usually considered as a problem of discrete geometry (the roots of these ideas can be found in works of E. Gilbert, F. Hwang, A. Fomenko, F. Morgan, H. Pollak and H. Rubinstein). Now, after 10–15 years of the investigations, it becomes clear that Extreme Networks Theory has all main features of a new branch of mathematics, namely, it contains its own deep and non-trivial results, it requires special techniques, and it seems to have a lot of relations with other branches of mathematics. One of the aims of this article is to attract the attention of scientists to Extreme Networks Theory and to present some bright results of the theory.

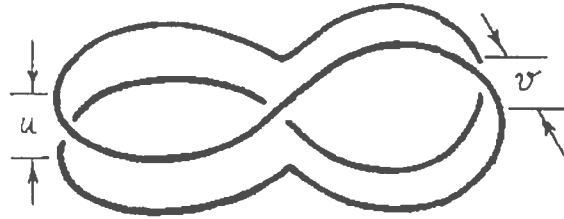


Figure 1: Douglas contour.

1 Interfaces between two media

The theory of minimal surfaces and surfaces of constant mean curvature is a branch of mathematics that has been intensively developed, particularly recently. On the basis of this theory we can investigate soap films and soap bubbles, interfaces between two media, which occur widely in chemistry and biology, for example, membranes in living cells, capillary phenomena, and so on. Minimal surfaces also turn out to be useful in architecture. Before giving exact definitions, let us consider some examples.

1.1 Soap films and soap bubbles

If we dip a wire contour into soapy water, and then carefully lift it out, a soap film is left on the contour. For many “young researchers” this is the time to obtain an iridescent bubble by blowing on the film. However, the soap films themselves conceal unexpected properties, which we can immediately see by making a simple experiment. Bend a wire contour, as shown in Figure 1. This is a so-called Douglas contour. Let u and v denote the distances between the extreme circles of the contour. It turns out that by lifting the contour out of the soapy water differently we can obtain different soap films. Figure 2 shows some of them (for small $u = u_0$ and $v = v_0$).

If the contour is not deformed in the process of obtaining the film, then as a rule films of type 1a and 1b are formed. A film of type 2 is obtained if at the time of lifting the contour out of the soapy water the left and right circles are kept joined ($u = 0, v = 0$), but after lifting the contour out they are let free. The elastic contour returns to its original position ($u = u_0, v = v_0$), and a film of type 2 remains on it. Films of type 3a and 3b can be obtained similarly, by combining only right (or only left) circles. To obtain a film of type 4 it is sufficient to puncture a film of type 3a by a disc D^2 .

We observe that films of type 2 and 3 are not smooth surfaces. They have singularities: a singular point A where four singular edges meet (type 2), or a singular edge S^1 (type 3). Moreover, smooth films of types 1 and 4 have different topological type: a film of type 1 is a two-dimensional disk D^2 , while a film of type 4 is a torus with a point deleted (see Figure 3).

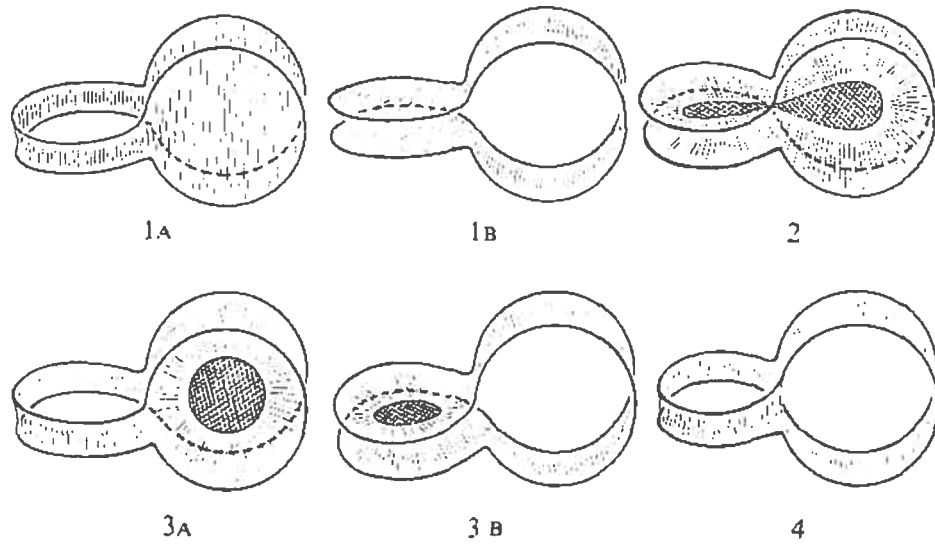


Figure 2: Soap films spanning Douglas contour.

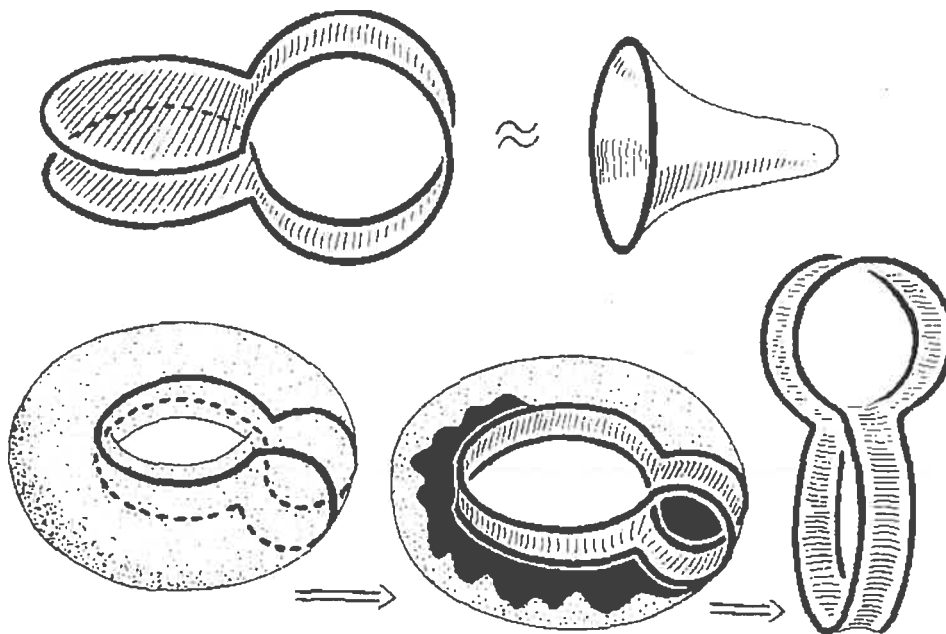


Figure 3: Soap films spanning Douglas contour have different topological types.

Thus, on a given contour we can, generally speaking, stretch many soap films, and not every film need be a smooth surface. Moreover, on a contour that is a bent circle we can stretch a minimal torus with a point deleted (a film of type 4). How many soap films can span a given contour? What topological types of films can occur? What singularities can be found on a soap film? These and other questions are considered in the theory of minimal surfaces (see below).

Soap bubbles and soapsuds, that is, a system of soap bubbles, are just as interesting to the researcher. Physical bubbles differ from soap films in that bubbles bound regions of space in which air is under greater pressure than the air outside. If we look closely at the soapsuds, we observe that the singularities at the junction of different bubbles are similar to the singularities of soap films. By carrying out many experiments with the interfaces between two media the Belgian physicist Joseph Plateau (1801–1883) formulated four principles, which describe the possible stable singularities on these surfaces. It turns out that the two types of singularities revealed above — a smooth singular edge at which three sheets meet, and a singular point at which four smooth singular edges meet, each pair of which is spanned by a smooth sheet — are the only possible singularities of stable soap films and soapsuds (Plateau's 1st, 2nd, and 3rd principles). Moreover, in the first case the sheets meet at a singular edge at an angle of 120° , while in the second case the singular edges meet at a point at an angle of about $109^\circ 28' 16''$ (more precisely, the cosine of this angle is $-1/3$), like the four line segments drawn from the center of a regular tetrahedron to its vertices (Plateau's 4th principle). We give some elementary justification of these principles below, first describing the variational principle that underlies soap films and soap bubbles.

1.2 The Poisson–Laplace theorem

Soap films and soap bubbles can be regarded as the interfaces between two homogeneous media in equilibrium. A soap film with boundary locally, that is, in a neighborhood of each of its points, separates two media, air–air, in each of which the pressure is the same. Therefore the total pressure on each small area of a soap film is zero. In a soap bubble the pressure inside is greater than the pressure outside, so the vector of the resultant pressure is directed outwards. This force must be compensated by the forces of surface tension. Since the pressure is always directed along the normal to the interface and is the same in absolute value at all points of this interface because of the homogeneity of the media, the interface is “curved in the mean” identically at all its points. To give a precise meaning to this statement, we need to define the geometric concept of “mean curvature” (for the details see [2], [5]).

Let M be a smooth two-dimensional surface in \mathbb{R}^3 , suppose that the point P lies on the surface M , and that $N(P)$ is one of the two unit normals to M at P (the vector $N(P)$ is orthogonal to the tangent plane to M passing through the point P ; we denote this tangent plane by $T_P M$).

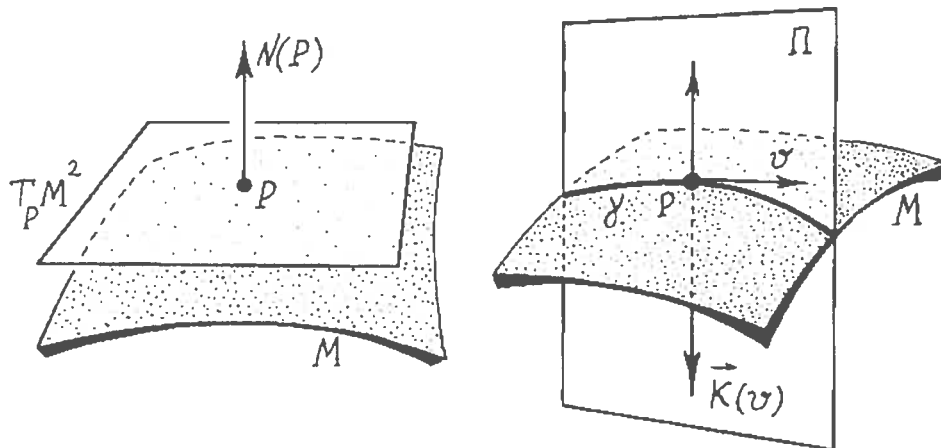


Figure 4: Tangent plane and normal section.

Through P we pass a plane Π containing the $N(P)$. The plane Π intersects M along a curve γ called a *normal section*. The unit vector v tangent to γ at P is called a *direction* of this normal section. Clearly, the vector $-v$ determines the opposite direction of the normal section γ , and the vectors v and $-v$ lie in the tangent plane $T_P M$ (Figure 4).

Let $\vec{k}(v)$ denote the curvature vector of γ in the direction v , that is, the acceleration vector at P under motion along γ with unit speed. We note that $\vec{k}(v) = \vec{k}(-v)$. It is easy to show that the curvature vector $\vec{k}(v)$ is collinear with the normal $N(P)$. We define the curvature $k(v)$ of a normal section γ in the direction v with respect to the normal $N(P)$ as the quantity $k(v) = \langle \vec{k}(v), N(P) \rangle$, where the brackets $\langle \cdot, \cdot \rangle$ denote the standard scalar product of vectors in \mathbb{R}^3 . Clearly, the continuous function $k(v)$ takes maximum and minimum values (since $k(v)$ is a function on the (compact) circle S^1 formed by all directions v). These values are called the *principal curvatures* k_1 and k_2 of the surface M at the point P , and the normal sections in which the values k_1 and k_2 are attained are called the *principal sections*.

Definition. The mean curvature H of a surface M at a point $P \in M$ with respect to the normal $N(P)$ is half the sum of the principal curvatures: $H = (k_1 + k_2)/2$.

Let φ denote the angle between the direction v of an arbitrary normal section γ at a point $P \in M$ and the direction of the principal section γ_1 (Figure 5).

If k_1 is the curvature of the principal section g_1 , and k_2 is the other principal curvature, then by Euler's well-known formula

$$k(v) = k(\varphi) = k_1 \cos^2 \varphi + k_2 \sin^2(\varphi).$$

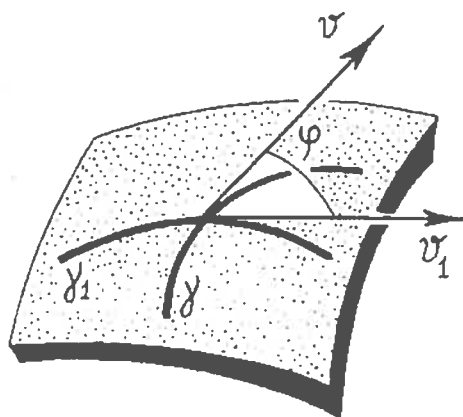


Figure 5: Principal sections.

To represent more clearly the distribution of the curvatures of the normal sections γ as the angle φ changes, we construct in the plane with polar coordinates (ρ, φ) the graph $\rho = |k(\varphi)|$. We can distinguish the following cases:

- a) k_1 and k_2 are nonzero and have the same sign. In this case the graph is an ellipse-like curve with “semi axes” $|k_1|$ and $|k_2|$ (where $|k(\varphi)|$ is maximal and minimal). If $k_1 = k_2 \neq 0$, this “ellipse” degenerates to a circle (Figure 6a);
- b) k_1 and k_2 are nonzero and of different signs. In this case the graph is a “four-leafed rose” (Figure 6b);
- c) one of the curvatures k_i is zero — the “four-leafed rose” degenerates into a “two-leafed rose” (Figure 6c);
- d) both the principal curvatures are zero. The graph is a point (the origin).

It is now clear that if k_1 is not equal to k_2 , then there are exactly two principal sections, which are in fact orthogonal to each other. If the principal curvatures are equal, then the curvature of all the normal sections is the same and equal to the mean curvature H .

Remark. It is easy to see that half the sum of the curvatures of any two mutually orthogonal normal sections is constant and equal to H .

Theorem 1.1 (Poisson–Laplace) *Suppose that a smooth two-dimensional surface M in \mathbb{R}^3 is the interface between two homogeneous media in equilibrium. Let P_1 and P_2 be the pressures in the media. Then the mean curvature H of the surface M is constant and equal to $H = h(P_1 - P_2)$, where the constant $\lambda = 1/h$ is called the coefficient of surface tension, and $P_1 - P_2$ is the difference between the pressures in the media (the resultant pressure).*

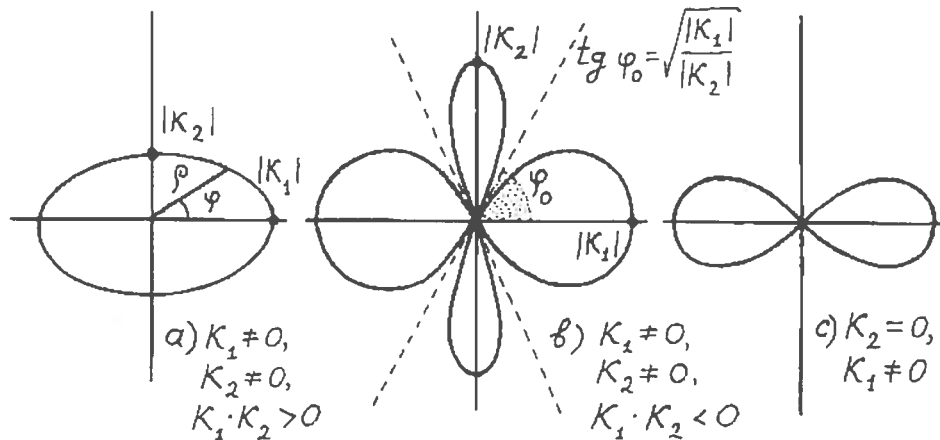


Figure 6: Distribution of curvatures of normal sections.

Thus, the expression “curved in the mean identically” implies that the mean curvature of the surface is constant. Taking account of what we said above, we can conclude that the mean curvature H of a soap film is zero, $H \equiv 0$, and the mean curvature H of a soap bubble is a constant $\neq 0$. In mathematics surfaces with $H = \text{const}$ are called *surfaces of constant mean curvature*. For the case $H = 0$ these surfaces have a special name — *minimal surfaces* (due to the fact that they locally minimize the area functional). Sometimes they are called *soap films*, and surfaces with $H = \text{const} \neq 0$ are called *soap bubbles*.

Surfaces of constant mean curvature are widespread in nature and play an important role in various research. Thus, for example, surface interactions on the interface between two media determine the character and rate of chemical reactions. Various membranes, such as the ear-drum and membranes that separate living cells, are minimal surfaces. One more example consists of microscopic marine animals — Radiolaria (see [1]).

2 The principle of economy in Nature

In this section we talk about an alternative approach to the description of minimal surfaces and surfaces of constant mean curvature, based on the variational principle (for a more detailed historical survey see [3]).

2.1 Optimality and Nature

In 1744 the French scientist Pierre-Louis-Moreau de Maupertuis put forward his famous principle, which has become known as the principle of least action. In 1746 Maupertuis published a paper “The laws of motion and rest deduced from a metaphysical principle”. This metaphysical principle is based on the assumption that Nature always acts with the greatest economy. Starting from

this position, Maupertuis draws the following conclusion: if certain changes occur in Nature, then the total action needed to carry out these changes must be as small as possible.

In parallel and independently Leonard Euler in 1744 obtained a strict proof that the principle of least action can be used to describe the motion of a material point in a field of conservative forces such as the motion of the planets around the Sun. Euler also put forward the conjecture that for any phenomenon in the Universe we can find a maximum or minimum rule to which it is subject. This remark appeared in the Appendix to his famous work of 1743 “Methods of finding curves that are subject to a maximum or minimum property”, the first textbook on the calculus of variations.

When in 1746 Maupertuis published his work on the principle of least action he was well aware of Euler’s achievement, since he briefly described it in the Preface. Then, however, he added: “This remark . . . is a beautiful application of my principle to the motion of the planets”, thus asserting his priority.

Euler reacted to this remark by giving up his right of priority, for which he was strongly criticized by certain historians of science. We shall not go into the details of the subsequent keen discussion that developed over priority in the discovery of the principle of least action. We shall only say that other people (König and apparently Leibniz) laid claim to authority and that the discussion was linked to the anthropomorphic understanding of the terms “living force” and “action” and with theology (see [3], [4]).

However, we observe that Maupertuis, who formulated his principle starting from the idea of the perfection of God, tested it on a few examples, but he did not investigate some of them thoroughly. It turns out that the principle of least action is not always true.

Let us consider one of the examples given by Maupertuis — the reflection of light. Here the law of least action leads to the conclusion that a ray of light “selects” from all possible routes from the source to the receiver the one that can be covered in the least time (this rule had already been formulated by Fermat). If light is propagated in a homogeneous medium, then this minimum principle leads to the simpler rule: a ray of light moves along the shortest path joining the source and the receiver.

Consider a spherical mirror situated in a homogeneous medium. Suppose that the source S and the receiver T are symmetrical about some line ℓ passing through the center of the sphere. What characterizes the trajectory SMT of the motion of light emitted from the source S , reflected in the mirror at a point M , and received by the receiver T ? In this situation we apply the famous law presented in the work *Catoptrica* attributed to Euclid — the so-called law of reflection. The following brief formulation of it is well known: the angle of the incidence is equal to the angle of reflection.

Figure 7 shows two situations: a convex mirror (Figure 7a) and a concave mirror (Figure 7b). In both cases the point M is the point of intersection of the mirror and the line ℓ (we consider the trajectories of the motion of light in the

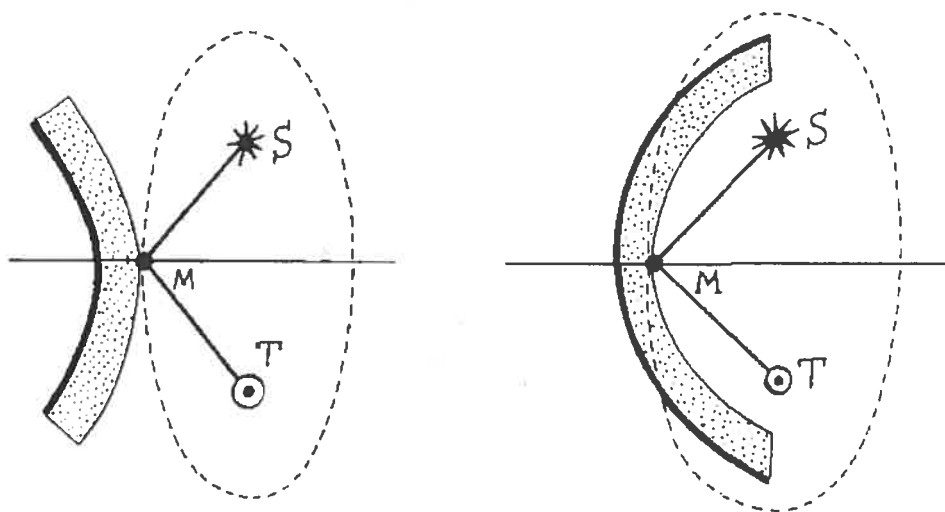


Figure 7: D'Arcy's example.

plane passing through the source, the receiver, and the line ℓ). Through M we draw the ellipse with foci at S and T . If M_1 is an arbitrary point lying outside the ellipse, M_2 is an arbitrary point inside it, and M is an arbitrary point on it, then $|M_1S| + |M_1T| > |MS| + |MT| > |M_2S| + |M_2T|$.

Hence it follows that in the case of a convex mirror the trajectory SMT actually has the shortest length, but for a concave mirror this is not always so. In Figure 7b the source S and the receiver T are symmetrical about the center of the sphere. It is easy to see that any path SM_2T , where $M_2 \neq M$, is shorter than the path SMT .

Thus, whether Nature is most economical or most wasteful depends on the form of the mirror. Citing similar arguments, d'Arcy showed in 1749 and 1752 that the principle of Maupertuis was not clearly formulated, and leads to incorrect assertions.

Nevertheless, the idea of optimality of the phenomena of Nature plays an important role in physics. The mathematical formulation of this idea has given birth to the calculus of variations, the founders of which are usually taken to be Lagrange and three Swiss mathematicians from Basel: the brothers Johann and Jakob Bernoulli and a student of Johann Bernoulli — Leonard Euler.

2.2 Minimal surfaces and optimality

It turns out that the forms of soap films are also optimal in a certain sense, namely the corresponding minimal surfaces are the extremals of the area functional. Let us explain this statement. For this we consider a soap film covering a given contour. Surface tension leads to the film tending to take up a form with the least possible surface energy (of course, this is an approximation, which

nevertheless works well in practice). Since the surface energy is directly proportional to the surface area, as a result of minimizing the surface energy the area of the soap film is least compared to the areas of all sufficiently neighboring surfaces covering the given contour.

Thus, soap films are local minima of the area functional. However, minimal surfaces, that is, surfaces of zero mean curvature, need not minimize the area even among all neighboring surfaces (with given boundary). To explain this we note that the concept of “neighboring surfaces” can be defined in different ways. We shall understand by a neighboring surface of a given surface M one obtained by a small deformation of M , leaving the boundary ∂M fixed. For a zero-dimensional surface M , that is, when M is a point, a small deformation is a small displacement of M . When the dimension of M is at least one, we can define two types of small deformations: a deformation with small amplitude, that is, when each point of M is displaced not far from its original position, and a deformation with small support, when only a sufficiently small region of M undergoes a deformation; the closure of such a deformed region is called the support of the deformation. A deformation with sufficiently small support always increases the area of the minimal surface, so for such deformations the minimal surfaces do minimize the area. For deformations with small amplitude this is not so. Nevertheless, for such deformations minimal surfaces are extremals (critical points) of the area functional.

It is easy to show that the (finite) area of any surface in \mathbb{R}^3 can always be increased by an arbitrarily small deformation of this surface, so no surface can be a local maximum for the area functional. It is well known that critical points other than a local minimum and a local maximum are called saddle points. Minimal surfaces corresponding to saddle points of the area functional (for deformations with small amplitude) are said to be unstable. If we have succeeded in creating a soap film that has the form of an unstable minimal surface, then fluctuations of this film that are small in amplitude, which always exist in the real world, would instantaneously lead to its collapse — the film constructed would turn out to be unstable.

Thus, minimal surfaces are critical points of the area functional. It turns out that the converse is true: a surface M that is a critical point of the area functional (considered on the space of all possible surfaces close (in amplitude) to M and having the same boundary ∂M) is minimal, that is, it has zero mean curvature.

Surfaces of constant mean curvature are also extremals of a certain functional. They can also be obtained as extremals of the area functional if we restrict the possible deformations. As an example we consider a soap bubble. If we blow on it, the film, sagging in one place, will swell at another place in such a way that the volume of the region inside the bubble is unchanged. This observation is the basis of the definition of a surface of constant mean curvature from the viewpoint of the variational principle. For a closed surface bounding a region in \mathbb{R}^3 , as admissible deformations we consider only those that preserve

the volume of the region bounded by this surface. The condition that the volume is preserved can be described in yet another way. For this we define the function of change of volume of a region V bounded by a surface M under a deformation M_t of this surface. For each $t = t_0$ we consider the totality of regions included between the surfaces M and M_{t_0} . From the total volume of those that lie outside V we subtract the total volume of those that lie inside V , and we call the resulting number the change in volume at the instant $t = t_0$. Varying t_0 , we obtain a function which is called the *change of volume function*. Clearly, the regions lying inside and outside V are on opposite sides of M , and so they can be defined without the use of V .

This observation enables us to define the change of volume function for a deformation M_t of an unclosed surface M (such as the soap film bounded by a wire contour or a hemisphere having the equator as its boundary), but in this case we must require that the deformation M_t is fixed ($= 0$) on the boundary ∂M of the surface M .

We say that a deformation M_t of a surface M (fixed on ∂M if $\partial M \neq 0$) preserves the volume if the change of volume function constructed from this deformation is identically zero. It turns out that *surfaces of constant mean curvature are critical points of the area functional restricted to the space of deformations that preserve the volume*.

Since surfaces of nonzero constant mean curvature are not minimal surfaces, they are also not critical points of the area functional considered on the space of all possible deformations (fixed on the boundary), not only those that preserve the volume. Thus, restriction of the space of admissible deformations naturally leads to an increase in the number of surfaces that are critical points of the area functional considered on this space.

A description of minimal surfaces as extremals of the area functional proves very useful. For example, this approach has given the possibility of solving the so-called Plateau problem, which consists in the following: for any contour, among all surfaces of given topological type that bound it, is there a surface of least area? A positive answer to this problem in the case when the contour is a simple rectifiable Jordan curve (it has finite length) and the surface has the topological type of the disk D^2 was obtained in 1928 by the young American mathematician Jesse Douglas. However, his proof turned out to be incomplete, and up to 1931 his paper had not appeared in print. At about the same time a solution of Plateau's problem, obtained by the Hungarian mathematician Tibor Rado, was published. In the following decades Jesse Douglas also solved a number of other problems that arose in the theory of minimal surfaces. In particular, the powerful mathematical technique that he developed enabled him to prove the existence of minimal surfaces of high genus spanning one or finitely many contours. For his achievement Douglas was awarded in 1936 the highest prize in mathematics — the Fields Medal.

In conclusion of Section 2 we give the one-dimensional version of Plateau's problem, the so-called Steiner problem, and show how from its solution there

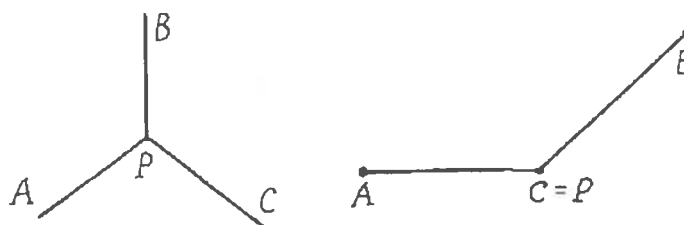


Figure 8: Solution of Steiner problem for three points.

follows a proof of Plateau's empirical principles, which describe all possible singularities of soap films of stable minimal surfaces. In Section 3 we discuss the one-dimensional case in more details.

2.3 Steiner's problem

Let us begin with the simplest case. Suppose we need to connect the towns A , B , and C by a system of roads, that there are no obstructions, and that we are free to construct the roads where we like. Suppose that the region in which the towns lie is flat. The problem is to find a system of roads of least length. In mathematical language this means the following: for three given points A , B , and C lying in a plane, to find a point P and paths joining P to A , B , and C so that the total length of these paths shall be as small as possible. Since a line segment is the shortest path between its ends, the required paths are line segments PA , PB , and PC . It remains to choose P in an optimal way.

It turns out that the solution depends on the relative positions of A , B , and C . If all the interior angles of the triangle ABC are less than 120° , then the required point P is uniquely determined from the condition that the angles \widehat{APB} , \widehat{BPC} , and \widehat{CPA} are equal (they are thus equal to 120°). But if one of the angles of the triangle ABC , say the angle at C , is at least 120° , then P coincides with C (see Figure 8).

The proof of this assertion is based on some elementary geometrical lemmas.

Lemma 2.1 (Heron's theorem) *Suppose that points A and B do not lie on a line a . Then among all the points P of the line a the point $P = P_0$ such that $|AP| + |BP|$ is as small as possible is uniquely determined from the following condition: the angle between AP_0 and the line a is equal to the angle between BP_0 and the line a (Figure 9).*

If A is a source of light, B is a receiver, and a is a mirror, then Heron's law can be regarded as a special case of the law of reflection (see above).

Lemma 2.2 *Suppose that the points A and B are the foci of an ellipse, and that the line a touches this ellipse at a point P . Then the angles between the line a and the segments AP and BP are equal (Figure 10).*

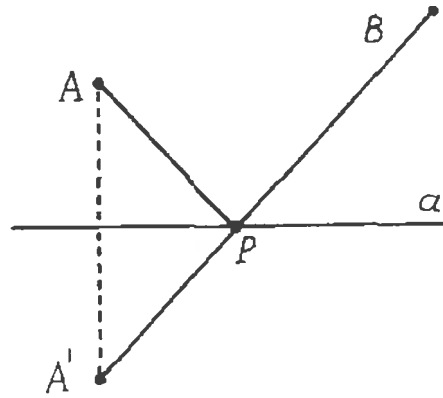


Figure 9: Heron's theorem

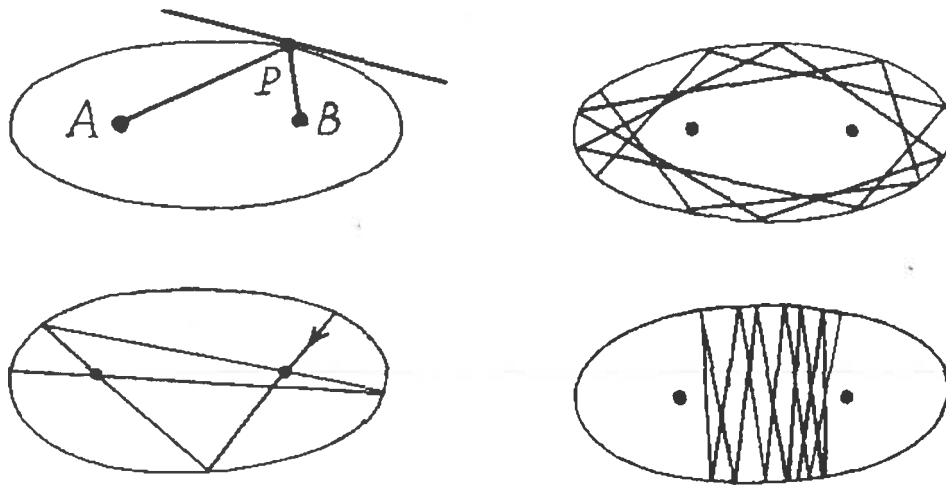


Figure 10: Ellipse properties

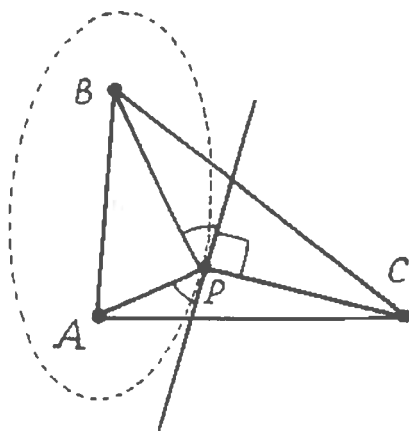


Figure 11: Solution of Steiner problem for three points.

It follows that if we put a source of light at one focus of an elliptical mirror, then all the rays collect at the other focus. Moreover, in elliptical billiards the ball always goes either outside the foci, or through the foci, or between the foci.

To prove Lemma 2.2 it is sufficient to observe that the sum of the distances from any point outside an ellipse to its foci is greater than the sum $|PA| + |PB|$ (since P lies on the ellipse).

Now suppose that P is a solution of Steiner's problem for a triangle ABC in which all the interior angles are less than 120° . Through P we draw the ellipse whose foci are A and B . Through P we draw the tangent a to this ellipse (see Figure 11). It is easy to see that CP is perpendicular to a . Therefore from Lemma 2.2 we see that $\widehat{APC} = \widehat{BPC}$. Similarly, $\widehat{BPC} = \widehat{APB}$. Thus, all the angles at P are equal to each other, and therefore equal to 120° . Now it is easy to construct the required point P and to see that the solution is unique.

If instead of three points A , B , and C we take any finite number of points, we obtain the generalized Steiner problem: it is required to join all these points by a finite system of curves of least length. This problem can be restated as follows: how do we join n towns by a network of roads with the least expense? From the combinatorial point of view the solution of this problem is a combination of two solutions, obtained above for the case $n = 3$. Here are some examples (Figure 12).

We observe that the solution of the generalized Steiner problem is not unique. For example, if four points are vertices of a square, we can obtain two symmetrical solutions (Figure 13). We note that a system of paths of least length joining n points of a plane is called a shortest network (in the plane), or Steiner minimal tree.

The generalized Steiner problem in the plane has still not been completely solved, so an experimental "solution" is of interest. Take two glass or transparent plastic sheets, place them in parallel planes, and join them by pieces of wire

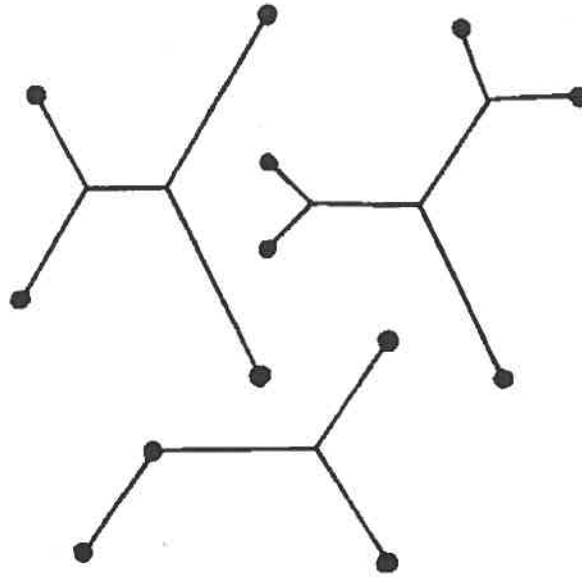


Figure 12: Examples of the shortest networks.

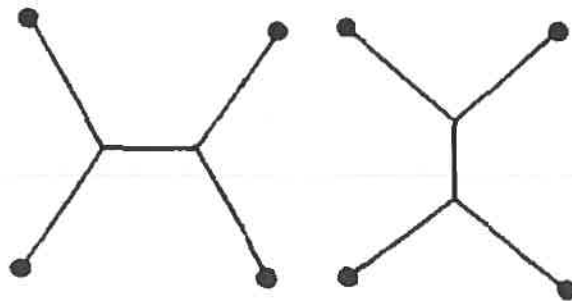


Figure 13: Two shortest networks spanning vertices of a square.

of the same length, equal to the distance between the sheets. Clearly, all the pieces of wire are parallel to one another and perpendicular to the sheets.

If we dip this configuration into soapy water, and carefully lift it out, then between the sheets there will be a soap film whose boundary consists of two parts: the set of joining wires and the set of "traces" which the film leaves on the sheets. We note that in accordance with the minimum principle the film is at an angle of 90° to each sheet. Moreover, this film consists of perpendicular sheets of planar rectangles that adjoin each other along the singular edges (see Figure 14). If a singular edge is not a joining wire, then in accordance with Plateau's principles exactly three rectangles meet on it at angles of 120° . We observe that the area of the resulting soap film is equal to the total length of the path joining the points where the wires are fastened (the "trace" of the soap film on a sheet) multiplied by the distance between the sheets. If the film has least area among all films with a partially free boundary consisting of the joining wires (rigid boundary) and the "traces" on the sheets (the hypothesis of the existence of such a film is called Plateau's problem with obstructions), then the corresponding "trace" on a sheet is a solution of the generalized Steiner problem for the configuration given by the fastening points.

Let us now turn to Plateau's principles, which describe all possible singularities of stable minimal surfaces. On a soap film we choose an arbitrary point P . We take a smaller and smaller neighborhood of P and blow it up to the same size. In the limit all the surfaces that join at P become planar, and the singular edges become segments of straight lines. Clearly, after such an enlargement the resulting fragment of stable film will also be stable. Now consider a sphere S^2 with center at P . Its intersection with the planar configuration we have constructed is a network on the sphere S^2 . Clearly the curves of intersection are parts of great circles of S^2 . Moreover, if l denotes the length of the network, and r is the radius of the sphere S^2 , then the area s of the part of the planar configuration inside the sphere is equal to $s = lr/2$. Therefore from the stability of the film it follows that at each node of the network only three arcs can meet at angles of 120° (otherwise by a small deformation we could lower the length of the network, and hence the area of the film).

The following natural problem arises (the so-called Steiner problem on the sphere): to describe all possible networks on the sphere consisting of arcs of great circles meeting at each vertex of the network three at a time at equal angles (of 120°). In contrast to the planar Steiner problem, the spherical problem has been completely solved. It turns out that there are exactly ten such networks, drawn in Figure 15. A more careful analysis shows that only three of these ten networks (the first three in Figure 15) correspond to configurations that minimize the area. Figure 15 shows soap films stretched on contours corresponding to minimal networks on a sphere. Only the first three of them are cones corresponding to the local arrangement of soap films described in Plateau's principles. This observation is a physical "proof" that only in the first three cases are the cones absolutely minimal, that is, they correspond to singularities occurring in stable

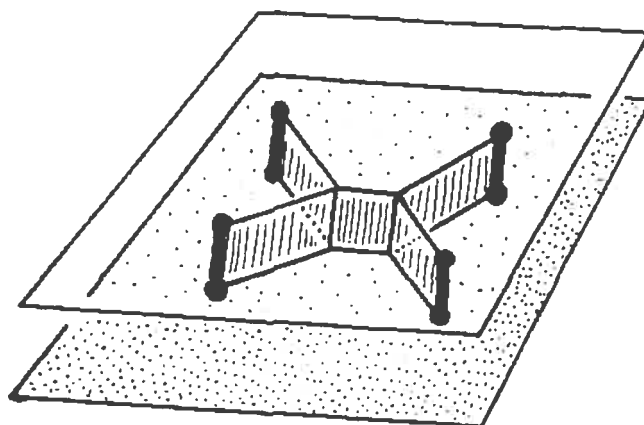


Figure 14: A soap film between two parallel sheets that simulates minimal network.

soap films.

3 Extreme Networks

The present Section is devoted to a new branch of mathematics studying branching solutions (*extremals*) of one-dimensional variational problems. This branch called *Extreme Networks Theory* appeared during attempts to understand how one can solve the generalized Steiner problem (see above).

We recall that Classical One-Dimensional Variational Problem can be stated as follows: to describe extremals of a variational functional defined on a space of curves joining a pair of fixed points in an ambient space. The theory of such extremals (so-called *extreme curves*) is well elaborated and presented in standard university courses such as Optimal Control, Differential Geometry, and Classical Mechanics. Let us assume now that we are given not with two but with three or more points of ambient space. The following natural questions arise: How is it possible to generalize the theory to this case? How is it possible to generalize the notion of an extreme curve? In such a way we come to an idea to introduce a new object, *networks*, which can be represented as unions of a finite number of curves into consideration.

Extreme networks for elementary variational functionals such as the length functional appeared in several ways in different theoretical and applied investigations. However, the problem was not systematically studied, in spite of its classical statement. Apparently, first problems of that type appeared in works of French mathematicians, Gergonne, Clapeyron and Lamé are among them. C. Gauss was also interested in such problems. In his letter to Schuhmacher, he brought up a question how to build the shortest system of roads joining

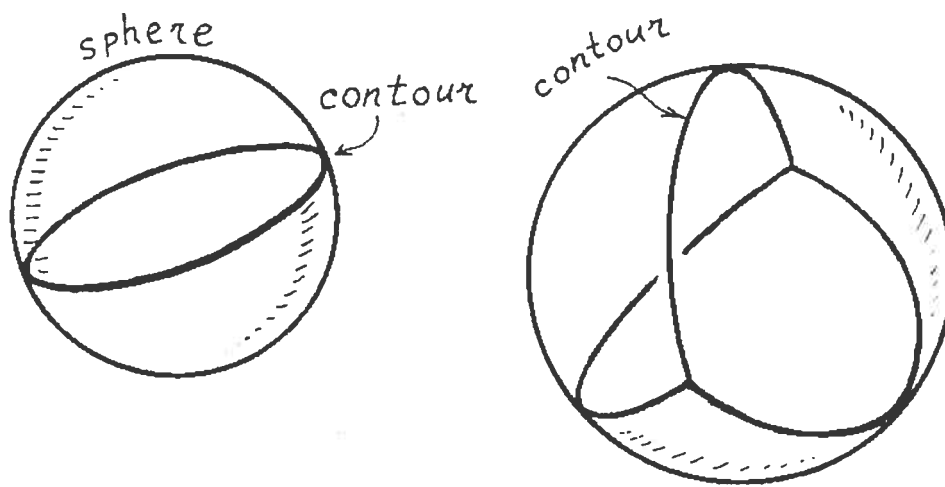


Figure 15: Minimal networks on sphere.

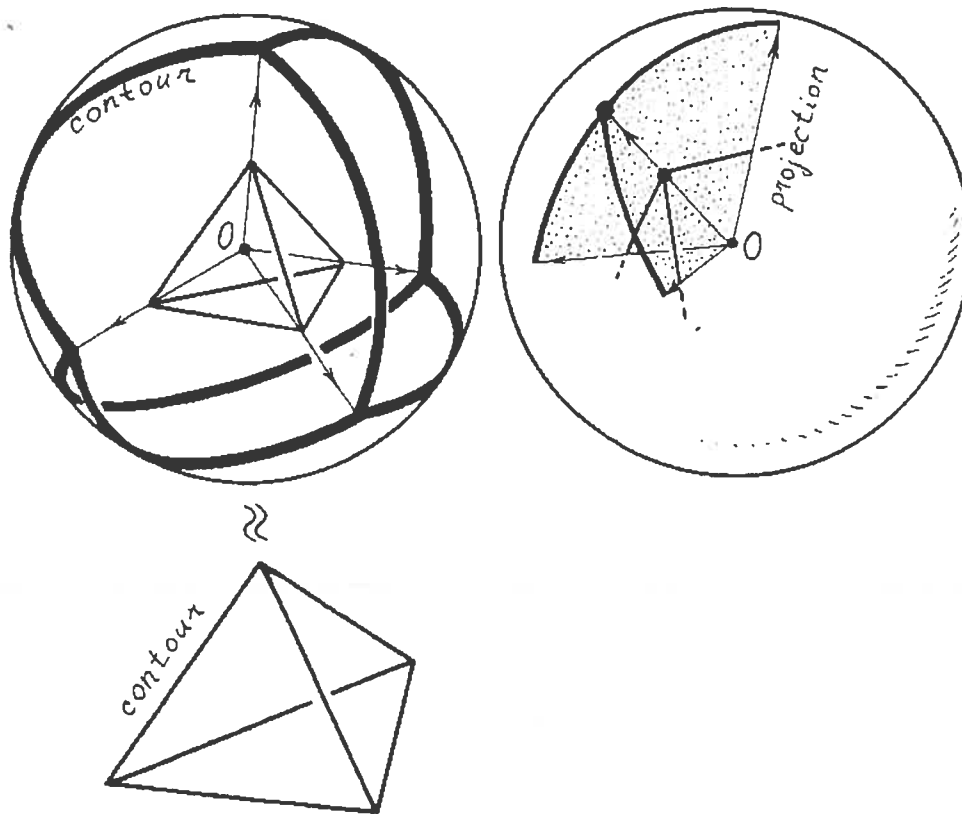


Figure 15: Minimal networks on sphere (continuation).

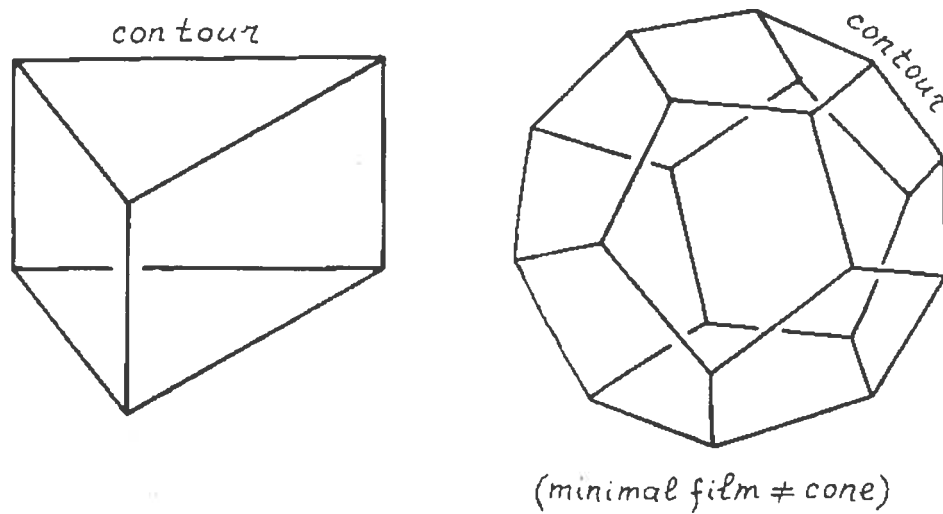


Figure 15: Minimal networks on sphere (continuation).

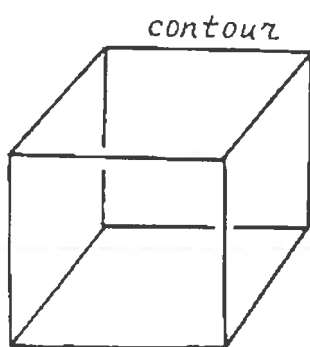
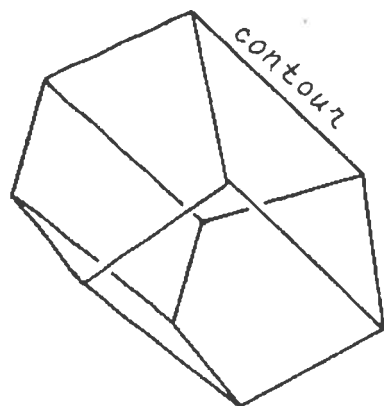
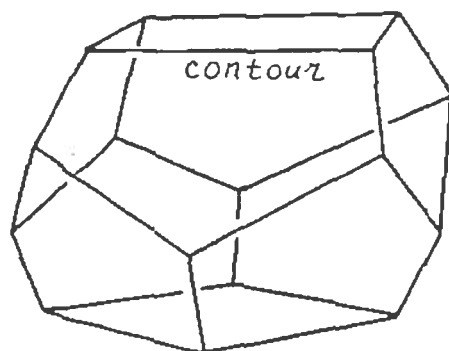


Figure 15: Minimal networks on sphere (continuation).



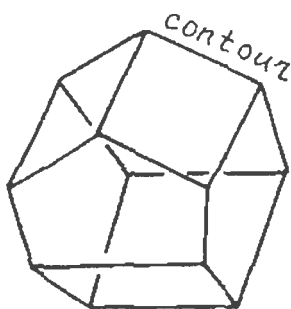
(minimal film \neq cone)

Figure 15: Minimal networks on sphere (continuation).



(minimal film \neq cone)

Figure 15: Minimal networks on sphere (continuation).



(minimal film \neq cone)

Figure 15: Minimal networks on sphere (continuation).

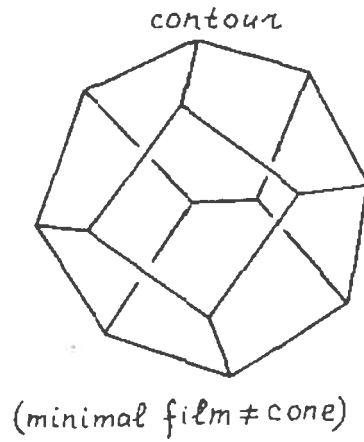


Figure 15: Minimal networks on sphere (continuation).

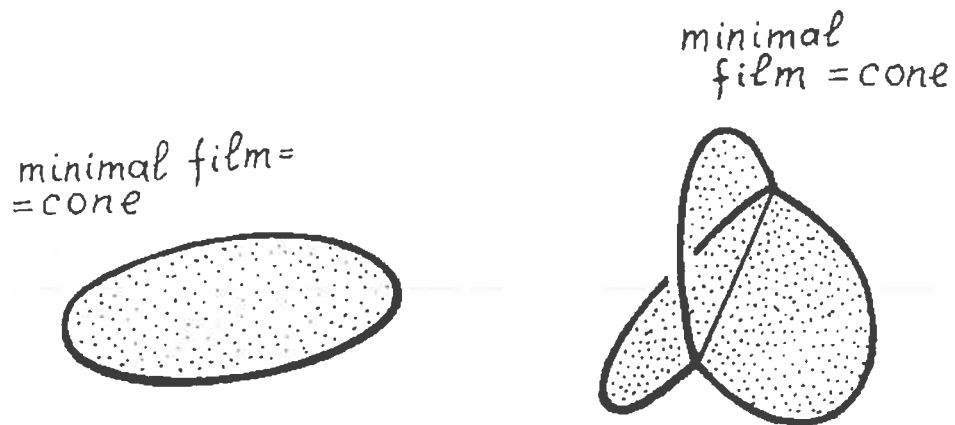


Figure 16: Minimal films corresponding to the networks on sphere.

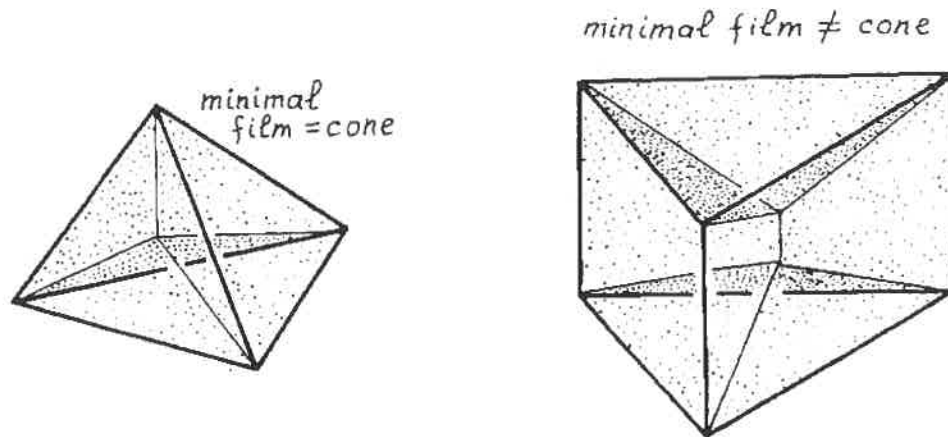


Figure 16: Minimal films corresponding to the networks on sphere (continuation).

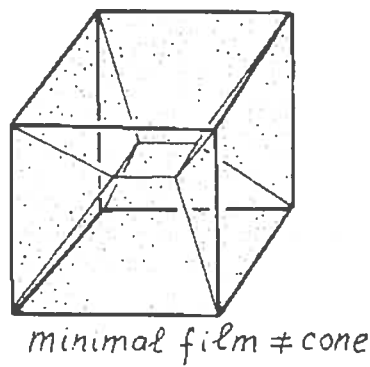


Figure 16: Minimal films corresponding to the networks on sphere (continuation).

four German towns: Hamburg, Bremen, Hannover, and Braunschweig. General problem of finding of the shortest networks joining a given finite set of points in the plane was stated by Jarnik and Kössler, see [7]. Later on, this problem became widely well-known as *Steiner problem* due to the famous book of Courant and Robbins (1941), see historical reviews in [8], [9], [10], or [11]. Notice that Steiner problem has several sorts, among them there is so-called *rectangular Steiner problem* consisting in description of the planar shortest networks whose edges are polygonal lines with parallel to coordinate axes links. The later problem appears naturally under arrangement of wires on printed circuit cards, see [3], and [9].

Thus, the idea of networks studying appeared several ages ago, and the interest to Steiner problem did not fade away as times goes by. Moreover, today a systematic studying of networks fulfilling some natural conditions, especially some extremality conditions, becomes even more important due to rapid development of networks of different kinds and levels. As an example we can speak about transportation networks, computer networks, networks prescribed by the structure of complex chemical molecules, such as DNA, etc.

What is the difference between the classical approach and the approach of extreme networks theory?

Firstly, main classical works concerning extreme networks are devoted to the studying of the shortest networks, the considered functional is usually either Euclidean length functional in a vector space (in the plane for the most part), or Manhattan length¹ functional (also in the plane mainly). But extreme networks theory supposed to give a common approach to investigate networks extreme with respect to an arbitrary functional in an arbitrary ambient space, say in a smooth manifold, or in a manifold with singularities such as a surface of polyhedron or an A. D. Alexandrov space, etc.

Secondly, the local structure of the shortest networks, that is, the structure of sufficiently small neighborhoods of the network points, is determined rigid enough. But if we consider extreme networks instead of the shortest ones then several possibilities appear. Notice that extremality is the network's property which characterizes the behavior of the functional under small deformations of the network. In the case of curves, the deformations are defined in more or less unique and natural way. In the case of networks one can consider several essentially different types of deformations. But the following two types are most important: deformations preserving the structure of the network, and deformations which can split some vertices of the network (see an example of such splitting of the vertex in Figure 17). Notice that the network Γ , spanning the vertices of the square and having exactly one interior vertex v located in the center of the square, is an extremal for the length functional with respect to the deformations preserving the network's structure. However, the deformation

¹ Recall that Manhattan length of vector equals to the sum of the length of its projections to coordinate axes.

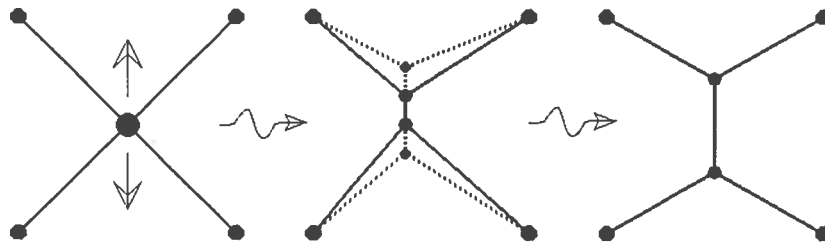


Figure 17: Splitting of the vertex.

splitting the vertex v as it is shown in Figure 17 decreases the length of the network Γ . The velocity of this decrease is not zero, thus the network Γ is no longer extreme with respect to such deformations.

The networks which do not change their structure under deformations are called *parametric networks*. The networks which allow the vertices splitting, that is, the networks whose structure can change under deformations are called *networks-traces*. (Formal definitions of these types of networks can be found in [11]).

Thirdly, by their own nature the networks possess both geometrical and combinatorial characteristics, they have both continuous and discrete properties. Therefore, creation of extreme networks theory initialized the development of new methods which do not have analogue in classical theory. The technique appeared combines methods of differential geometry and variational calculus on the one hand, and methods of discrete geometry and combinatorial analysis on the other hand. Besides, the authors hope that the continuous–discrete nature of networks could give a new way of thinking about problems dealing with geometrical quantization (in the most wide meaning of the words). Namely, properties of extreme networks (with respect to an appropriate variational functional) contain information about both continuous (geometrical) and discrete (quantum) characteristics of ambient space.

Below we present several spectacular geometrical results obtained in Extreme Networks Theory. An interested reader can find details in [10] or [?].

3.1 Possible structure of planar extreme network depends on geometry of its boundary

We consider planar binary trees (that is, planar connected graphs without cycles, whose vertices degrees equal either 1 or 3) that are extreme with respect to the standard Euclidean length functional. Such networks are usually called *local minimal 2-trees* or *full Steiner minimal trees*, and it is well-known, see for example [9], that each shortest tree in the plane can be naturally represented as a union of such networks. Moreover, the most present-day algorithms are

based on an enumeration of all such possible local minimal 2-trees. Therefore, each *a priori* elimination of possibilities can be applied to improve the existing algorithms solving Steiner problem. We give such an effective restriction in terms of geometry of the boundary set.

Recall the well-known partition of an arbitrary finite non-empty subset M of the plane into convexity levels. We assign all the points of M that lie on the boundary of the convex hull of M to the *first convexity level* M^1 of M . Notice that M^1 is not empty. We remove M^1 from M . If the resulting set is not empty we transform it by the same procedure. Namely, all the points of M that lie on the boundary of the convex hull of $M \setminus M^1$ are assigned to the *second convexity level* M^2 of M . We continue this process until all the points of M lie on some convexity level. The resulting partition $M = \sqcup_{i=1}^k M^i$ is called the *partition into convexity levels* and the set M^i is called the *i th convexity level* of M . If $M = \sqcup_{i=1}^k M^i$, then we say that M has k convexity levels. We denote by $\varkappa(M)$ the number of convexity levels of the set M . Note that M lies on the boundary of its own convex hull if and only if it has a single convexity level, which coincides with M itself.

Now we define a characteristic of a planar binary tree, that introduced in [10]. Let Γ be a planar 2-tree, and let a and b be some edges of Γ . Choose the path $\gamma \subset \Gamma$ connecting a and b , and let it be oriented from a to b .

Let the plane be oriented. Then the notions “left” and “right” are defined. Let us move along γ from a to b . During this motion, in each interior vertex of γ we turn either to the left or to the right (we omit rigorous definitions because they are obvious but cumbersome). We denote by $\text{tw}(a, b)$ the difference between the numbers of left and right turns. The $\text{tw}(a, b)$ is called *the twisting number* of an ordered distinct pair (a, b) . We put $\text{tw}(a, a) = 0$.

Definition. The *twisting number* $\text{tw}(\Gamma)$ of a planar 2-tree Γ is the maximum among the twisting numbers of all the ordered pairs of edges of Γ : $\text{tw}(\Gamma) = \max_{(a,b)} \text{tw}(a, b)$.

The following result demonstrates the connection between the twisting number of a local minimal 2-tree and the number of convexity levels of its boundary.

Theorem 3.1 *Let Γ be a planar local minimal 2-tree spanning a finite subset M of the plane. Then $\text{tw} \Gamma \leq 12(\varkappa(M) - 1) + 5$.*

If $k(M) = 1$, then the following inverse statement holds.

Theorem 3.2 *Let Γ be a planar binary tree with $\text{tw} \Gamma \leq 5$. Then there is a local minimal tree Γ_m spanning a finite subset M of the plane, such that Γ_m is planar equivalent to Γ and $\varkappa(M) = 1$, that is, M lies on the boundary of its own convex hull.*

Possible generalizations of these results for the case of so-called weighted length functional and the corresponding results in planar graph theory are discussed in [11].

3.2 Infinitely many different closed extreme networks on a flat torus

In Section 2 we discussed closed minimal networks in the standard sphere. We recall that up to isometry there are exactly 10 such networks. But it turns out, that on closed flat surfaces the situation is quite other. A. Ivanov and A. Tuzhilin in collaboration with I. Ptitsina obtained a complete description of closed local minimal networks on flat tori and flat Klein bottles. We omit the precise statements (the details can be found in [10]), and include only several non-trivial geometrical corollaries.

Theorem 3.3 *In each closed flat surface, that is, in each flat torus and in each flat Klein bottle, there are infinitely many non-equivalent closed local minimal networks. Each closed minimal network on a closed flat surface can be non-trivially deformed in the class of minimal networks.*

Theorem 3.4 *Any sufficiently small deformation of a closed flat surface through the class of homeomorphic flat surfaces does not destroy a closed local minimal network in it.*

Let us remark, that a classification of closed local minimal networks in closed surfaces of constant negative curvature is not obtained still. Some examples of such networks can be found in [10]. But the following general result was proved by M. Pronin who has elaborated an analogue of Morse index theory for minimal networks.

Theorem 3.5 *A closed local minimal network on a surface of strictly negative curvature is rigid, that is, it can not be deformed through the class of local minimal networks.*

3.3 Closed extreme networks on Platonic bodies' surfaces

Another interesting application of the extreme networks theory appears if we consider local minimal networks in the surfaces of polyhedra. Let us notice, that the surface of each polyhedron can be considered as a flat surface with point-wise singularities corresponding to the vertices of the polyhedron. Networks on polyhedra appears in applications, and first of all in computer science, due to the well-known fact that all surfaces are simulated as some close polyhedra. The detailed discussion of this topic can be found in [10]. Here we mention just one result concerning the most famous polyhedra — so-called Platonic bodies (this result was obtained by A. Ivanov and A. Tuzhilin in collaboration with T. Pavlyukevich (Anikeeva)).

Theorem 3.6 *In the surface of each Platonic body there are infinitely many non-equivalent closed local minimal networks, see 20.*

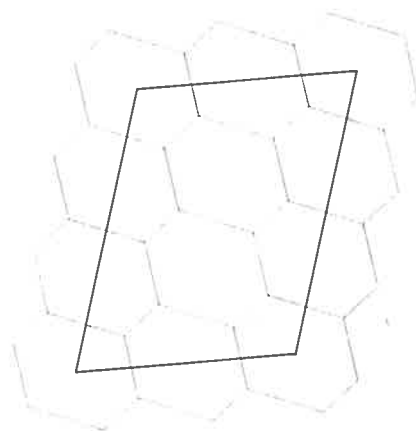


Figure 18: Minimal network on tori.

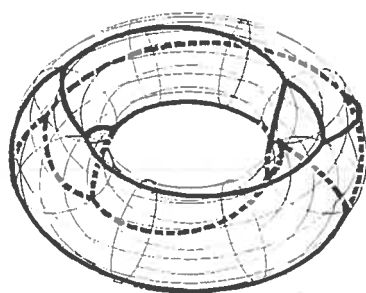


Figure 19: Topology of minimal network on torus.

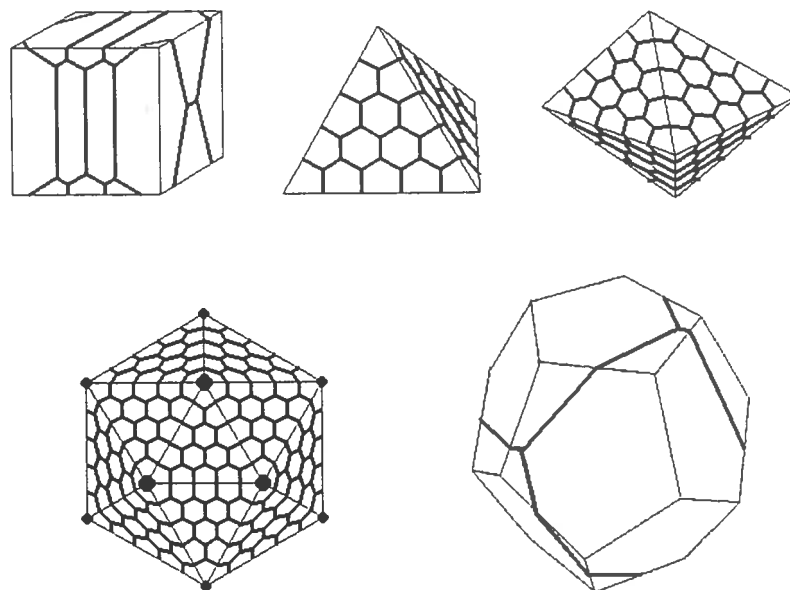


Figure 20: Closed local minimal networks in surfaces of Platonic bodies.

3.4 Extremality differs from local minimality in Manhattan plane

It is well-known that networks extremality criteria for Euclidean length functional have local nature, that is, extremality is equivalent to a special structure of small pieces of the network. It turns out that in general case it is not so. As an example we consider networks which are extreme with respect to so-called Manhattan length. We recall that the *Manhattan length of a vector in \mathbb{R}^2* is defined to be equal to the sum of the lengths of its projections onto the coordinate axes. We assume here that a standard basis is fixed in \mathbb{R}^2 .

Extremals of the Manhattan length functional form an important class of extreme networks. The first works devoted to the investigation of shortest networks with respect to Manhattan length appeared in 60th, see [12], because of rapid development of electronics and robotics. The interest to the Manhattan length emerged in view of that the conductors on printed circuit boards have, as a rule, the form of polygonal lines composed from horizontal and vertical straight segments, therefore their Manhattan length and their Euclidean length are the same. A similar situation takes place in robotics. Apparently, the first systematic investigation of shortest networks with respect to the Manhattan length (so-called *shortest rectangular trees*) was undertaken by Hanan, see [13]. Hanan described several important general geometrical properties of such networks. In particular, Hanan proved that there is a shortest rectangular tree, which is a subset of so-called *Hanan lattice*, that is, the set of all vertical and

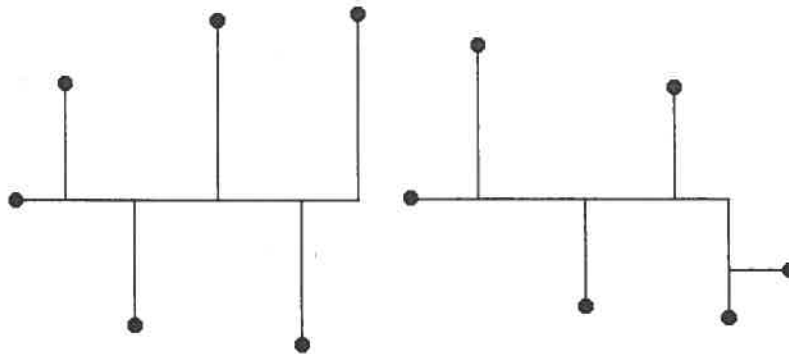


Figure 21: Two types of full shortest rectangular networks.

horizontal straight lines passing through the boundary points. We notice that, generally speaking, the edges of a shortest rectangular tree can be chosen in many ways without changing the length of the tree, of course. However, starting from Hanan, see [13] the tradition approach appeared to chose the edges of such trees as polygonal lines with links parallel to coordinate axes.

Ten years later Hwang, see [14], described possible structure of shortest rectangular trees under assumption that the given boundary set can be spanned by a *non-degenerate* shortest tree Γ_0 . The latter means that the degrees of all boundary vertices of the tree Γ_0 equal to 1. In particular, the tree Γ_0 has no vertices of degree 2. Hwang proved that in this case the shortest tree has one of two possible forms shown in Figure 21.

However, nobody succeeded in finding an effective algorithm constructing shortest rectangular trees. The reason of these failures was explained by Garey M.R. and Johnson D.S. (1977). Namely, they proved that the problem of constructing of a shortest rectangular tree is *NP*-complete, that is, most likely there is no polynomial algorithm to solve this problem. Due to this fact the studying of geometry of the shortest rectangular trees becomes even more topical.

As we have already mentioned above, the Manhattan length gives us an example of the functional for which networks extremality criteria is not local even under an assumption that all the edges of the network considered are shortest segments joining the corresponding vertices. Consider the following example: the edges of the network Γ depicted in Figure 22 are shortest curves, however, this network is not critical. In Figure 22 a deformation which linearly decreases the length is shown.

An interested reader can find a criterion of extremality of a tree in Manhattan plane in [11]. This criterion was obtained by A. Ivanov and A. Tuzhilin in collaboration with Hông Van Le. Let us mention that the Manhattan length is

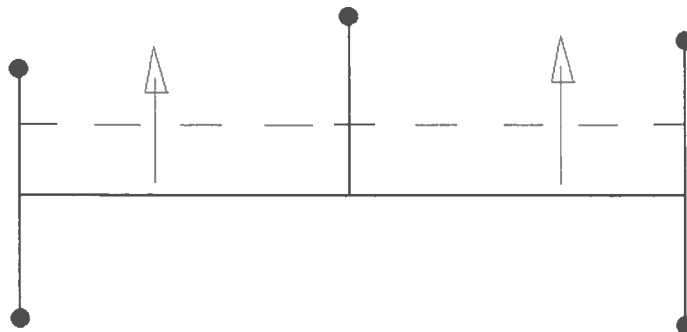


Figure 22: Local minimal non-critical network in the plane.

a particular case of the length in a normalized space. In [11] we give extremality criteria for networks—traces and for parametric networks in arbitrary normalized space.

References

- [1] *Dao Trong Thi, A. T. Fomenko*, Minimal Surfaces and Plateau Problem. — M., Nauka, 1987 (in Russian). (Engl. translation Amer. Math. soc., Providence, RI, 1991.)
- [2] *B. A. Dubrovin, A. T. Fomenko, S. P. Novikov*, Modern Geometry. — M.: Nauka, 1986. English translation: Part 1, GTM 93, 1984; Part 2, GTM 104, 1985. — New York, Springer-Verlag.
- [3] *S. Hildebrandt and A. Tromba*. The Parsimonious Universe. — Springer-Verlag, New York, 1996.
- [4] *L. S. Polak* (ed.), Variational principals in mechanics, Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow, 1959 (Russian).
- [5] *J. A. Thorpe*, Elementary topics in differential geometry, Springer, New York, 1979.
- [6] *A. A. Tuzhilin and A. T. Fomenko*, Elements of geometry and topology of minimal surfaces in three-dimensional space. — Translations of Math. Monographs, AMS, v. 93, 1992.
- [7] *Jarnik V. and Kössler M.* O minimalnich grafeth obeahujicich n danijch bodu. — Cas. Pest. Mat. a Fys., v. 63, pp 223–235, 1934.

- [8] *Cieslik D.*, Steiner Minimal Trees. — Kluwer Academic Publishers, 1998.
- [9] *Hwang F. K., Richards D. and Winter P.* The Steiners Tree Problem. — Elsevier Science Publishers, 1992.
- [10] *Ivanov A. O. and Tuzhilin A. A.* Minimal Networks. The Steiner Problem and Its Generalizations. — N.W., Boca Raton, Florida, CRC Press, 1994.
- [11] *A. O. Ivanov, A. A. Tuzhilin*, Branching Solutions of One-Dimensional Variational Problems. — World Publisher Press, 2000, in preparation.
- [12] *Francis R. L.*, A note on the optimum location of new machines in existing plant layouts. — J. Indust. Engrg., v. 14, pp. 57–59, 1963.
- [13] *Hanan M.*, On Steiner's Problem with Rectilinear Distance. — SIAM J. Appl. Math., v. 14, pp. 255–265, 1966.
- [14] *Hwang F. K.*, On Steiner minimal trees with rectilinear distance. — SIAM J. of Appl. Math., v. 30, pp. 104–114, 1976.