Topological Methods in Nonlinear Analysis
Journal of the Juliusz Schauder Center
Volume 16, 2000, 253–265

THE TOPOLOGICAL PROOF OF ABEL–RUFFINI THEOREM

Henryk Żołądek

Abstract. We present a proof of the non-solvability in radicals of a general algebraic equation of degree greater than four. This proof relies on the non-solvability of the monodromy group of a general algebraic function.

1. Introduction

In high school young people are learned how to solve the quadratic equation $x^2 + ax + b = 0$. Everybody knows the formula

$$x = -\frac{a}{2} + \sqrt{\frac{a^2}{4} - b}.$$

A general equation of third degree $x^3 + ax^2 + bx + c = 0$ is firstly reduced to the form $y^3 + py + q = 0$ (using the substitution $x = y - a/3$). The next substitution $y = z - p/3z$ leads to the equation $(z^3)^2 + q(z^3) - p^3/27$. From this we get $z = \sqrt[3]{-q/2 + \sqrt{q^2/4 + p^3/27}}$, what gives the *Cardano formula*

$$y = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

A general equation of fourth degree, which can be taken in the form $x^4 + px^2 + qx + r = 0$, is treated by means of the so called *Ferrari method.* Firstly we

2000 *Mathematics Subject Classification.* 14H30, 14H55, 12I10, 30F10.
Key words and phrases. Radicals, monodromy group, solvable group.
This work was supported by Polish KBN Grant No. 2 P03A 041 15.

rewrite it in the form

$$(x^2+\alpha)^2-[(2\alpha-p)x^2-qx+\alpha^2-r]=0, \tag{1}$$

where α is an additional parameter. We choose α such that the polynomial in the square brackets becomes a full square; we require that

$$q^2-4(2\alpha-p)(\alpha^2-r)=0. \tag{2}$$

Then the equation (1) is reduced to two quadratic equations. On the other hand, we already know how to solve the cubic equation (2). In this way one can solve the initial equation of fourth degree; (however the complete formula is so complicated than nobody tries to write it down).

We have proven that the roots of a general algebraic equation of degree ≤ 4 are expressed via the coefficients of the equation by means of the operations of addition, extraction, multiplication, division and extraction of root of natural degree. We say that the solution of an equation is *expressed in radicals.*

For long time mathematicians tried to find a method for solving a general equation of fifth degree in radicals. In 1799 P. Ruffini had presented a proof of non-existence of such solution. Unfortunately, the proof was too complicated to be accepted by the actual mathematical community. Social approval received a proof of the analogous statement done in 1824 by N. H. Abel.

Theorem (Abel and Ruffini). *A general algebraic equation of degree* ≥ 5 *cannot be solved in radicals. This means that there does not exist any formula which would express the roots of such equation as functions of the coefficients by means of the algebraic operations and roots of natural degrees.*

The Abel–Ruffini theorem constituted an essential step in the development of mathematics. Such notions like abelian group and solvable group take their origin just here.

Later E. Galois had started a general theory which associates with any algebraic equation certain invariant, known now under the name the *Galois group.* It is the group of those permutations of the roots of the equation which preserve all algebraic relations satisfied by these roots. Some properties of the equation (e.g. solvability in radicals) are translated to properties of its Galois group.

In such approach the main accent was shifted from analytic properties of solutions (dependence on the coefficients) to their algebraic character. One assumes that the coefficients belong to a given number field (e.g. of rational numbers) and investigates the extension of this field by means of the roots of the equation. But when one has to assume that the coefficients are variable, what is quite natural when solving general equations, then the algebraic theory ceases to be completely clear. In particular, in the proof of the Abel–Ruffini theorem people use some rather sophisticated tricks (transcendental extensions, algebraic independence of coefficients) in order to be compatible with the acquainted algebraic scheme.

Below we present another proof of the theorem of Abel and Ruffini. It relies upon topological properties of Riemann surfaces defined by algebraic equations with varying coefficients. The reader will see that it is natural and correct approach to the problem.

Unfortunately, none of the known to me books on number theory and algebra mentions about existence of such a proof. While working on this article I was using a short book of V. B. Alekseev [1]. The author attended a course of lectured on this subject delivered by V. I. Arnold for talented high-school pupils (in a school associated with Moscow University) and the book is based on these lectures. The lecturer had to begin with introduction of complex numbers, analytic functions and the notion of group.

B. A. Dubrovin, S. P. Novikov and A. T. Fomenko also write about the topological proof of Abel–Ruffini theorem in their monograph [3]. They even sketch the proof; unfortunately, most of the details are left to the readers.

Several people (from east and west) asked me where such a complete proof of this theorem can be found. I hope that this paper will serve as such a source.

2. Algebraic functions and their Riemann surfaces

A naive approach to algebraic functions may lead to misunderstandings. For example, it is known what is $\sqrt{x}$ (it takes two values). But how many values the function $\sqrt{x}+\sqrt{x}$ takes; two, four, or maybe three? The proper definition is following.

An *algebraic function* $y = f(x)$ is defined by the algebraic equation

$$g_n(x)y^n + g_{n-1}(x)y^{n-1} + \ldots + g_0(x) = 0 \tag{3}$$

(or shortly $F(x, y) = 0$), where g_j are polynomials. In what follows for simplicity we will assume that $g_n(x) \equiv 1$ (then the roots will not escape to infinity).

Let a point $a \in \mathbb{C}$ be such that the equation $F(a, y) = 0$ has n different roots $y = z_1, \ldots, z_n$. Then $F'_y(a, z_i) \neq 0$ and Implicit Function Theorem asserts that for any x from some neighbourhood U_a of the point a the equation $F(x, y) = 0$ (with respect to y) also has n different solutions. They define single-valued functions $f_{a,1}(x), \ldots, f_{a,n}(x)$ on the domain U_a. The functions $f_{a,i}(x)$ are expanded into convergent Taylor series in the point a; therefore we can choose U_a as a disc (with center at a) contained in the common convergence disc of these series.

The pairs $(f_{a,i}, U_a)$ constitute analytic elements of the function f. A general *analytic element* is denoted by (f_a, U_a), where U_a is a disc with center at a at which the Taylor series of the function f_a (at the point a) is convergent.

An analytic element can be prolonged. If the equation (3) has single-valued solutions, then they would be prolonged to the whole complex plane. For example, for the equation $F(x, y) = (y - x)(y - 1)$ we have two analytic elements which are prolonged to the functions $y = x$ and $y = 1$ on $\mathbb{C}$. It may turn out that several solutions (i.e. analytic elements) glue themselves and this constitutes an obstacle to prolongation. In the above example we have an illusory gluing at the point $x = 1$ (because each of the solutions is analytically prolonged there), but for the equation $y^3 - x = 0$ the singularity at $x = 0$ cannot be removed in this way.

Let $x_1, \ldots, x_m$ be the singular points of the function f. Beginning with the analytic element (f_a, U_a), $a \in \mathbb{C} \setminus \{x_1, \ldots, x_m\}$ we will construct the Riemann surface M of the function f. We prolong the element (f_a, U_a) along paths $\gamma \subset \mathbb{C} \setminus \{x_1, \ldots, x_m\}$ with beginning at a (and end at b). We cover γ by means of finite number of neighbourhoods U_{a_i}, $a_i \in \gamma$, which are domains of analytic elements (f_{a_i}, U_{a_i}) compatible at the intersections, $f_{a_i} \equiv f_{a_{i-1}}$ in $U_{a_i} \cap U_{a_{i-1}}$; we assume $U_{a_0} = U_a$. The final analytic element (f_b, U_b) constitutes a prolongation of the analytic element (f_a, U_a) along the path γ (see Figure 1).

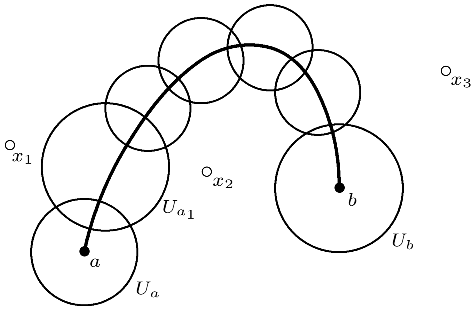

Figure 1

There arises the question about uniqueness of the analytic prolongation. It turns out that if two paths $\gamma^{(1)}$ and $\gamma^{(2)}$ (in $\mathbb{C} \setminus \{x_1, \ldots, x_m\}$, with beginning at a and end at b) can be deformed one to the other, with fixed ends and not touching the singularities, then the results of prolongations along these paths are the same, $f_b^{(1)} = f_b^{(2)}$. This is the theorem about monodromy. One can easily prove it by covering the domain swept by the deformed paths using domains U_c of analytic elements.

The union of all analytic elements obtained from the element (f_a, U_a) along all possible paths forms certain surface which we call the *Riemann surface* M of the algebraic function f. The surface M is equipped with the natural projection $\pi : M \to \mathbb{C} \setminus \{x_1, \ldots, x_m\}$ which associates to a value $f_c(x)$ (of a branch f_c) its argument x.

In fact, it is not yet the full Riemann surface. To be correct one should compactify it (in the topology induced by the analytic elements) and then smooth the cusps. This would give us a compact smooth analytic surface without self-intersections. Because we do not need it in this article, we will omit this part of the theory.

Example 1. $f(x) = \sqrt{x}$. The Riemann surface of this function is well known. We begin with the point $a = 1$ and the branch $f_a(x) = \sqrt{x}$ which is positive on the right real half-line. Prolonging this branch along the unit circle we arrive at the branch $-f_a(x)$. In order to imagine the Riemann surface of this root, we take two copies of the plane $\mathbb{C}$ cut along the negative real half-line, put one above another and glue the ridges of the cut of the upper sheet with the opposite ridges of the cut of the lower sheet. We cannot draw it in a planar picture without self-intersections (see Figure 2(a)). But when we turn the above sheet, then we can realize the gluings without self-intersections (Figure 2(b)). This is just the Riemann surface (over $\mathbb{C}^* = \mathbb{C} \setminus 0$). We see that it is homeomorphic with $\mathbb{C}^*$. This homeomorphism can be realized analytically: $t \to (x, y) = (t^2, t)$, $t \in \mathbb{C}^*$.

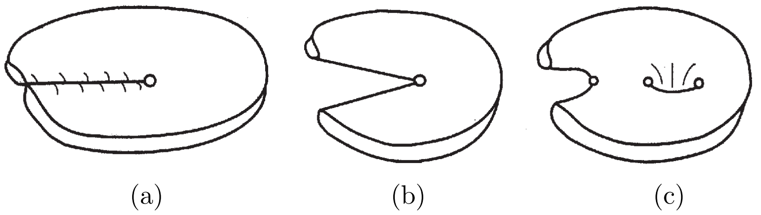

Figure 2

Example 2. $f(x) = \sqrt{x^3 - x}$. The subroot function has three zeroes $0, \pm 1$. We take two copies of the plane cut along the intervals $(-\infty, -1]$ and $[0, 1]$. We turn the upper sheet and glue. One can see that M is homeomorphic with the torus T^2 deprived of four points. One of these deleted points corresponds $x = y = \infty$ (Figure 2(c)).

The reader can prove himself that the Riemann surface of the function $\sqrt{x^2 - 1}$ is homeomorphic to $\mathbb{C} \setminus \{2 \text{ points}\}$.

Example 3. $y^3 - y = x$. Here the Riemann surface is isomorphic with $\mathbb{C} \setminus \{2 \text{ points}\}$ (Figure 3).

The general construction of the Riemann surface of an algebraic function $y = f(x)$ defined by an algebraic equation of degree n is following. Let $x_1, \ldots, x_m$ be the singular points. We cut the plane along straight radii starting at the points x_i, running to infinity and mutually disjoint; (one can do it). We take n copies of the plane cut in this way. Next we glue he ridges of cuts in a way

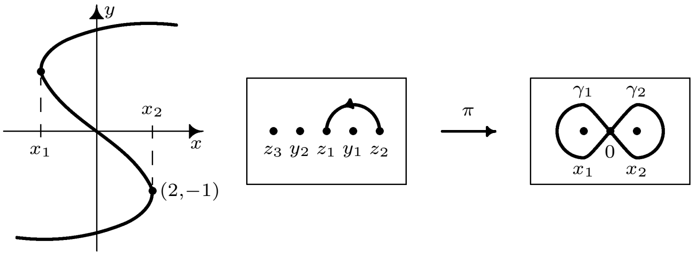

Figure 3

determined by the variations of values of the function $f(x)$ as the argument x varies around the singular points. It can be difficult to do it in some concrete (nontrivial) examples.

3. The monodromy group of an algebraic function

Consider the algebraic function $y = f(x)$ defined by the equation $F(x, y) = y^n + \ldots + g_0(x) = 0$, with the singular points $x_1, \ldots, x_m$. Let us choose the base point $a \in \mathbb{C} \setminus \{x_1, \ldots, x_m\}$. We have n analytic elements $(f_{a,i}, U_a)$, $i = 1, \ldots, n$ and the set $M_a = \{z_1, \ldots, z_n\}$ (identified with $\{1, \ldots, n\}$) of values of the function f in a). The monodromy group of the function f is a subgroup of the group $S(M_a) = S(n)$ of permutations of the set M_a defined as follows.

If γ is a loop in $\mathbb{C} \setminus \{x_1, \ldots, x_m\}$ with beginning and end at a, then the analytic prolongation of any analytic element $(f_{a,i}, U_a)$ along γ leads to a new element which coincides with one of the $(f_{a,j}, U_a)$. In particular, the point z_i is transformed to some point of the set M_a; we denote it by $\Delta_\gamma(z_i)$. On the surface M there exists a path δ_i with beginning at (a, z_i) and end at $(a, \Delta_\gamma(z_i))$ which is a lift of the path γ to M, $\pi(\delta_i) = \gamma$. The map $\Delta_\gamma : M_a \to M_a$ is the monodromy transformation defined by the loop γ.

The group generated by the maps Δ_γ, γ-loop, is called the *monodromy group* and is denoted by $\mathrm{Mon} = \mathrm{Mon}(f)$.

By the monodromy theorem the map Δ_γ is locally constant on the loop space; it does not change during a deformation of the loop. The equivalence classes of loops with respect to deformations forms the *fundamental group* of the set $\mathbb{C} \setminus \{x_1, \ldots, x_m\}$ with the base point a, $\pi_1(\mathbb{C} \setminus \{x_1, \ldots, x_m\}, a)$. The group operations rely on composition of loops and taking the inverse loop. We have then a homomorphism from $\pi_1(\mathbb{C} \setminus \{x_1, \ldots, x_m\}, a)$ to $S(M_a)$ whose image is $\mathrm{Mon}(f)$.

Examples. In Examples 1 and 2 we have $M_a = \{z_1, z_2)$ and the group $S(M_a) \simeq \mathbb{Z}/2\mathbb{Z}$ is generated by the transposition $(1, 2)$. If a loop γ surrounds even number of singular points (multiplicity counting), then $\Delta_\gamma = \mathrm{id} = e$. Otherwise $\Delta_\gamma = (1, 2)$. We have then $\mathrm{Mon}(f) = \mathbb{Z}/2\mathbb{Z}$.

Let us put $a = 0$ in Example 3; then $M_a = \{0, \pm 1\}$. With the loop γ_1 around $x_1 = -2$ the transposition of the values $z_1 = 0$ and $z_2 = 1$ is associated, i.e. $\Delta_{\gamma_1} = (1, 2)$. With the loop γ_2 around $x_2 = 2$ the transposition of the values $z_1 = 0$ and $z_3 = -1$ is associated, i.e. $\Delta_{\gamma_2} = (1, 3)$. Now it is easy to see that $\mathrm{Mon}(f) = S(3)$.

(We assume that the reader is acquainted with the description of a permutation by means of its decomposition into cycles. For example, the expression (142)(36) denotes the permutation 1 → 4, 4 → 2, 2 → 1, 3 → 6, 6 → 3, 5 → 5 in $S(6)$. We recall also that $\sigma \cdot (i_1, \dots, i_k) \cdot \sigma^{-1} = (\sigma(i_1), \dots, \sigma(i_k))$.)

Remark 1. The monodromy group $\mathrm{Mon} = \mathrm{Mon}(f)$ can be identified with the Galois group of certain extension of algebraic fields. As an initial field K we take the field $\mathbb{C}(x)$ of rational function of a variable x. Here we treat elements of K as functions on U_a. Next, we define an extension L as $K(f_{a,1}, \dots, f_{a,n})$, adjoining branches of the algebraic function. It turns out that the group of automorphisms of the extension $K \subset L$, i.e. its *Galois group* $\mathrm{Gal}_K L$, coincides with Mon. Indeed, because Mon permutes the branches, it induces an automorphism of the field L, and because functions from $\mathbb{C}(x)$ are single-valued, they are invariant with respect to the monodromy. This means that $\mathrm{Mon} \subset \mathrm{Gal}_K L$. Suppose that $\mathrm{Mon} \neq \mathrm{Gal}_K L$. By the fundamental theorem from the Galois theory (see [2]) the subgroup Mon is associated to a intermediary field $K \subset L_1 \subset L$, $L_1 \neq K$ such that $\mathrm{Gal}_{L_1} L = \mathrm{Mon}$ and $L_1 = L^{\mathrm{Mon}} = \{\varphi \in L : \mathrm{Mon}\,\varphi = \{\varphi\}\}$. The field L_1 consists of those functions which are invariant with respect to the monodromy. Therefore, they are single-valued functions. Their singularities are regular (of power type), also at infinity. From this it is easy to deduce that they are rational (we multiply them by $(x - x_i)^k$ and apply the Riemann's theorem about removable singularities) This means that $L_1 = K$ (a contradiction).

In some classical books on Riemann surfaces (like the Forster's book [4]) the Galois theory is used in a different way. Assume that a Riemann surface M is smooth, compact and equipped with a holomorphic map $\pi : M \to N = \mathbb{C}P^1$ (prolongation of the projection $(x, y) \to x$) called the *ramified covering.* One can achieve it after completing the construction from the point 2. The initial field is the field of rational functions on N, $K = C(N)$ which coincides with $C(x)$. However, the extension field L is the field of rational functions on M; here K is embedded into L by means of the induction $\pi^* : \varphi \to \varphi \circ \pi$. Essential for this theory is the group $\mathrm{Deck} = \mathrm{Deck}_N M$ of automorphisms of the covering $M \to N$ and consisting of homeomorphisms of M which preserve the fibers of the covering. In order that the group Deck be the Galois group of the extension $K \subset M$ one has to assume that it acts transitively on a typical fiber. The coverings which have this property are called the *Galois coverings.* The coverings from Examples 1 and 2 are Galois coverings, but Deck is trivial n Example 3. It is not observed in [4] that the class of such coverings is very thin in the class of finite coverings above the Riemann sphere.

4. The monodromy group of a typical algebraic function

By a *typical algebraic function* we shall mean a function given by an equation $F(x,y)=y^n+g_{n-1}(x)y^{n-1}+\ldots=0$ which satisfies the following conditions:

(i) The complex algebraic curve $\Gamma=\{F(x,y)=0\}\subset\mathbb{C}^2$ is smooth and restriction π of the projection $(x,y)\to x$ to the curve Γ has only the simplest singularities: non-degenerate critical points with different critical values.

(ii) The curve Γ is irreducible, i.e. the function F cannot be written in the form of product $F^{(1)}F^{(2)}$ of two polynomials.

The smoothness condition means that the (complex) gradient of the function F does not vanish; either $F'_x\neq 0$ or $F'_y\neq 0$. The critical points (x_i,y_i) of the projection π are the points where Γ is vertical, i.e. $F'_y=0$. Because Γ is nonsingular, we have $F'_x\neq 0$ and locally Γ is defined by the equation $x-x_i=\psi(y)$; moreover, $\psi(y_i)=\psi'(y_i)=0$. The non-degeneracy condition means that $\psi''(y_i)=-F''_{yy}/F'_x\neq 0$; only two branches of the algebraic function are glued. The critical values of the projection are equal to the numbers x_i; it is assumed that they are different.

Under the condition (i) the Riemann surface M can be identified with $\Gamma\setminus\{\text{critical points}\}$ and the singular points of the algebraic function are the critical values of the projection π.

Irreducibility of the complex algebraic curve Γ guarantees is topological connectivity, and also the connectivity of the Riemann surface $M=\Gamma\setminus\{\text{critical points}\}$. Indeed, suppose that Γ is not connected and that the assumption (i) holds. Then Γ consists of two disjoint curves $\Gamma^{(1)}$ and $\Gamma^{(2)}$. Let $f_i(x)$, $i=1,\ldots,k$ be the branches (suitably numerated) of the function $y=f(x)$ which lie in $\Gamma^{(1)}$ and let $f_i(x)$, $i=k+1,\ldots,n$ be the branches from the other curve. We define the functions $F^{(1)}(x,y)=(y-f_1(x))(y-f_2(x))\ldots(y-f_k(x))$ and $F^{(1)}(x,y)=(y-f_{k+1}(x))\ldots(y-f_n(x))$. Of course, we have $F=F^{(1)}F^{(2)}$. On the other hand, the curves $\Gamma^{(j)}$ are connected near the branching points, so permutations of branches from one group do not lead away of this group. This means that the coefficients (before powers of y) of the functions $F^{(j)}$ are analytic and single-valued functions of polynomial growth at infinity. Hence they are polynomials, then $F^{(j)}$ are also polynomials.

The irreducibility can be checked sometimes directly. For example, when Γ is an image of an irreducible algebraic curve under an algebraic mapping, then it is irreducible. If the algebraic closure of Γ in the complex projective plane $\mathbb{C}P^2$ is a smooth curve, then Γ is also connected. The latter property means that the highest degree homogeneous part of the polynomial F is factorized into different linear factors.

The above assumptions imply the following important properties of the monodromy group.

Lemma 1. *Let the algebraic function satisfies the conditions* (i) *and* (ii). *Then:*

(a) Mon *is generated by the transpositions* (k, l)*, corresponding to exchanges of the branches* $f_k(x)$*,* $f_l(x)$ *which glue themselves at critical points* (x_i, y_i).

(b) Mon *acts transitively on the set* $M_a = \{z_1, \dots, z_n\}$. *This means that for any two different values* z_k*,* z_l *there exists a* $\sigma \in$ Mon *such that* $\sigma(z_k) = z_l$.

Proof. The property (a) is obvious, because such transpositions are induced by the loops around x_i. The property (b) follows from the connectivity of $\Gamma \setminus \{\text{critical points}\}$. The points (a, z_k) and $(a, z_l))$ can be joined by means of a (real) curve δ in Γ. Moreover, we can assume that the projection $\gamma = \pi(\delta)$ does not pass through any of the points $x_i = \pi(x_i, y_i)$. γ is a loop and $\Delta_\gamma(z_k) = z_l$. We put $\sigma = \Delta_\gamma$. □

Lemma 2. *If a subgroup* $G \subset S(n)$ *is transitive and generated by transpositions, then it coincides with* $S(n)$.

Proof. We say that a subset $A \subset \{1, \dots, n\}$ is *complete* if any permutation from $S(A)$ can be prolonged to a permutation of the set $\{1, \dots, n\}$ which belongs to G. Any transposition (k, l) among the generators of G defines the complete subset $\{k, l\}$. Let A_0 be a maximal complete subset (with respect to the inclusion order). We claim that $A_0 = \{1, \dots, n\}$.

Suppose that A_0 is a proper subset. There exists a transposition $\tau = (k, l) \in G$ with $k \in A_0$ and $l \notin A_0$. Then the group generated by $S(A_0)$ and τ would be equal $S(A_0 \cup \{l\})$ and the set $A_0 \cup \{l\}$ would be complete. □

Corollary. *The monodromy group of a typical algebraic function is equal* $S(n)$.

Example 4 ([1]). The monodromy group of the algebraic function defined by the equation $F = 3y^5 - 25y^3 + 60y - x = 0$ equals $S(5)$.

Indeed, the condition for critical points of the projection π, $F = F'_y = 15(y^2 - 4)(y^2 - 1) = 0$, gives the four points $(x_i, y_i) = \pm(16, 2), \pm(38, 1)$ with different critical values. At these points the curve $F = 0$ is smooth ($F'_x \neq 0$) and the projection is non-degenerate ($F''_{yy} \neq 0$). On the other hand, the curve $F = 0$ is the image of the complex plane under an algebraic mapping (because x can be expressed by means of y). Therefore the typicality conditions (i) and (ii) are satisfied.

Remark 2. In the multidimensional case, when the coefficients of the algebraic equation depend on many parameters, we are dealing with multidimensional Riemann surfaces. Particular such case provides the so called *universal algebraic equation* $y^n + x_{n-1}y^{n-1} + \ldots + x_0 = 0$. The corresponding Riemann surface is n-dimensional and constitutes an n-fold covering above the discriminant locus $\Sigma = \{\Delta(x_0, \ldots, x_{n-1}) = 0\}$. The fundamental group of this complement $\pi_1(C^n \setminus \Sigma)$ is the same as the braid group $B(n)$ and the monodromy homomorphism turns out to be the same as the natural homomorphism of the *braid group* to the symmetric group $S(n)$. Of course, we also have $\mathrm{Mon} = S(n)$.

5. Solvable and nonsolvable groups

The *commutator* of a group G is its subgroup $G^{(1)} = [G, G]$ generated by the elements $[a, b] = aba^{-1}b^{-1}$, $a, b \in G$. In particular, if G is *abelian* (i.e. is commutative, $ab = ba$) then $G^{(1)} = \{e\}$. The group $G^{(1)}$ is a *normal subgroup*; if $a \in G$, $b \in G^{(1)}$, then $aba^{-1} \in G^{(1)}$. The set of cosets $G/G^{(1)}$ is an abelian group. We define by induction the subgroups (*derivative groups*) $G^{(k+1)} = (G^{(k)})^{(1)}$. Therefore we have a sequence of normal subgroups (*central derivative series*) $\ldots \subset G^{(2)} \subset G^{(1)} \subset G^{(0)} = G$ with abelian quotient subgroups $G^{(k)}/G^{(k+1)}$.

We say that G is *solvable* if its central derivative series is finite, i.e. $G^{(r)} = \{e\}$ for some r. The equivalent definition says that there is a finite series of groups $\{e\} = G_r \subset G_{r-1} \subset \ldots \subset G_0 = G$ such that the subgroups $G_{k+1} \subset G_k$ are normal and the quotients groups G_k/G_{k+1} are abelian.

It is useful to imagine the notion of a normal subgroup and the quotient group in the situation when the group G acts on some set A, in a way that some subset $B \subset A$ is invariant with respect to this action (images of elements from B do not leave B). Then the set of those maps which are identity on B constitutes a subgroup $H \subset G$. It is normal subgroup and the quotient group is treated as the restriction of the action of G to the subset B.

We shall use the following simple lemma.

Lemma 3.

(a) *A subgroup of a solvable group is solvable.*

(b) *The product $G \times H$ of solvable groups is a solvable group.*

(c) *If a group H is solvable and there exists a surjective homomorphism $G \to H$ with abelian kernel, then the group G is solvable.*

(d) *If G is solvable and a homomorphism $G \to H$ is onto, then H is solvable.*

Proof. Only the points (c) and (d) need explanation. In the point (c) we have the submersion of the derivative groups $G^{(1)} \to H^{(1)}$ with trivial kernel. Hence $G^{(1)} = H^{(1)}$ and $G^{(r)} = H^{(r)} = \{e\}$ for some r. In the case (d) we have the surjective homomorphisms $G^{(r)} \to H^{(r)}$. □

Example 5. The group $S(2)$ is abelian and hence solvable.

Example 6. The group $S(3)$ can be identified with the group of symmetries of a regular triangle (permutation of its vertices). It contains the *alternating subgroup* $A(3)$ consisting of permutations which are compositions of even number of transpositions (reflections of the triangle). The latter group consists of rotations of the triangle; it is normal subgroup with two-element quotient and is cyclic. This shows the solvability of $S(3)$.

Example 7. The group $S(4)$ has the following central derivative series

$$\{e\} \subset V \subset A(4) \subset S(4)$$

where the so-called *Vierergruppe* $V = \{e;\ (1,2)(3,4);\ (1,3)(2,4);\ (1,4)(2,3)\}$. The group $S(4)$ is isomorphic with the group of rotations of a cube (by permutations of the diagonals).

The next property is not as obvious as the previous ones.

Theorem 1. *The groups $S(n)$, $n \geq 5$, are not solvable.*

Proof (We follow the book of J. Browkin [2]). Because the alternating group $A(n)$ is normal subgroup of $S(n)$ with two-element quotient group, it is enough to show that $A(n)$ is not solvable. But this follows from the following observation.

If the cycles $\sigma = (123)$ and $\tau = (345)$ (with one common element) belong to a subgroup $H \subset A(n)$, then the elements $[\sigma, \tau] = (\sigma(3)\sigma(4)\sigma(5)) \cdot \tau^{-1} = (145) \cdot (354) = (143)$ and $[\sigma^{-1}, \tau^{-1}] = (253)$ belong to the commutator $H^{(1)}$. The latter are also cycles with one common element.

Repeating this argument we see that all the derivative groups $A(n)^{(j)}$ contain two cycles with one common element. Therefore none of them can be trivial. □

6. The monodromy groups of functions expressed in radicals

If $f(x)$ and $g(x)$ are algebraic functions with the branches $f_1, \dots, f_n$, $g_1, \dots, g_k$, then the sum $h(x) = f(x) + g(x)$ is also an algebraic function. Its Riemann surface is constructed as follows. We take $n \cdot k$ copies of the complex plane, cut along radii running from all singular points of the functions f and g. We label these sheets by $h_{i,j}$. Next we glue the ridges of cuts using the schemes of gluings for the functions f and g; it means that if after overrunning a singular point a sheet f_{i_1} passes to f_{i_2} and a sheet g_{j_1} passes to g_{j_2}, then the sheet h_{i_1,j_1} passes to the sheet h_{i_2,j_2}. Finally we have to identify (glue) those sheets for which the values of the functions $h_{i,j} = f_i + g_j$ are the same. For example, the function $y = \sqrt{x} + \sqrt{x}$ takes three values and satisfies the equation $y(y^2 - 4x) = 0$.

Analogously we define the algebraic functions and the Riemann surfaces for $f(x) - g(x)$, $f(x) \cdot g(x)$, $f(x)/g(x)$.

The function $h(x) = \sqrt[k]{f(x)}$ has kn branches $h_{j,l}(x) = e^{2\pi ij/k}h_{0,l}(x)$ for $j = 0, \ldots, k-1$, $l = 1, \ldots, n$, where $h_{0,l}(x)$ is a distinguished branch of the root $\sqrt[k]{f_l(x)}$. In the construction of its Riemann surface, besides the singular points of the initial function, we get additional branching points of the root, the zeroes and the poles of $f_l(x)$. So, we take n files, each with k copies of cut planes. The gluings of the ridges of cuts are analogous as in the case of the sum: if after overrunning a singularity x_i, f_{l_1} passes to f_{l_2}, then the cuts of sheets from the l_1th file are glued with cuts of sheets from the l_2th file, moreover the numbers of sheets in the files undergo a cyclic shift (which is trivial when $f_{l_1}(x_i) \neq 0, \infty$).

We say that an algebraic function of one variable is *represented in radicals* if it can be obtained from the constant functions $x \to c$ and the identity function x by means of the above operations.

Theorem 2. *The monodromy group of an algebraic function represented in radicals is solvable.*

This completes the proof of the Abel–Ruffini theorem. Therefore there exist algebraic equations which cannot be solved by means of the radicals.

Example. The equation $F = 3y^5 - 25y^3 + 60y - x = 0$ (from Example 4) cannot be solved in radicals.

Proof of Theorem 2. It is enough to show that if the groups $\mathrm{Mon}(f)$ and $G = \mathrm{Mon}(g)$ are solvable, then the groups $\mathrm{Mon}(f \pm g)$, $\mathrm{Mon}(fg)$, $\mathrm{Mon}(f/g)$ and $\mathrm{Mon}(\sqrt[k]{f})$ are also solvable. We consider only the cases $f + g$ and $\sqrt[k]{f}$.

Recall the construction of the Riemann surface of the function $f + g$. Firstly we have taken nk copies of the cut plane and glued the ridges of cuts and next we have glued the whole sheets with the same values of the branches $h_{i,j} = f_i + g_j$. Therefore, in the first step we have got certain surface M' whose monodromy group is isomorphic with a subgroup I of the group $F \times G$. (It can be a proper subgroup when some singularities of f and g coincide; e.g. $\mathrm{Mon}(\sqrt{x}) = \mathbb{Z}/2\mathbb{Z}$, $\mathrm{Mon}(\sqrt[4]{x}) = \mathbb{Z}/4\mathbb{Z}$, but $\mathrm{Mon}(\sqrt{x} + \sqrt[4]{x})$ is cyclic of order 4).

When we glue some sheets in the second step, some elements of the group I, those which permute the glued sheets, are sent to trivial transformations from the monodromy group H. However, any element from H (induced by a loop in the $x-$plane) is an image of an element from I, it is image of the permutation of the fiber M'_a induced by the same loop. Therefore, we have a surjective homomorphism $I \to H$. Now it is enough to use the points (a), (b), (d) of Lemma 3.

In the case of the function $h = \sqrt{[k]}f$ we are dealing with a process reverse to the process of gluings of sheets. We multiply sheets into files (of sheets). Therefore we have a surjective homomorphism $H \to G$. In order to be able to use the point (c) of Lemma 3, we have to show that the kernel of this homomorphism is an abelian group. But from the construction it follows that it is a subgroup of the cyclic group $\mathbb{Z}/k\mathbb{Z}$. □

References

[1] W. B. Alekseev, *Abel Theorem in Problems and Solutions*, Nauka, Moscow, 1974. (in Russian)

[2] J. Browkin, *Teoria Ciał*, PWN, Warsaw, 1978. (in Polish)

[3] W. A. Dubrovin, S. P. Novikov and A. T. Fomenko, *Modern Geometry*, Nauka, Moscow, 1986. (in Russian)

[4] O. Forster, *Riemannsche Flächen*, Springer–Verlag, Berlin, 1977.

Manuscript received January 24, 2000

Henryk Żołądek
Institute of Mathematics
University of Warsaw
Banacha 2
02-097 Warsaw, POLAND

E-mail address: zoladek@mimuw.edu.pl