

ON THE ANCIENT PROBLEM OF DUPLICATION OF A CUBE IN HIGH SCHOOL TEACHING

Andrea Grozdanić and Gradimir Vojvodić

Abstract. The paper is devoted to exposition of constructions with straightedge and compass, constructible numbers and their position with respect to all algebraic numbers. Although the large number of constructions may be accomplished with straightedge and compass, one of the known problems of this kind dating from Greek era is duplication of a cube. The given proof in this paper is elementary and self-contained. It is suitable for teachers, as well as for high school students.

ZDM Subject Classification: G45, H45; *AMS Subject Classification:* 97G40, 97H40.

Key words and phrases: Duplication of a cube; construction with compass and straightedge; constructible number; algebraic number.

1. Introduction

Constructive problems have always been the favorite subject of geometry. The traditional limitation of tools used for solving geometric constructions to just the compass and the straightedge reaches far into the past, although the Greeks had also been using certain other instruments. The well known Euclidean geometry (III B.C.) was based on geometric constructions performed only by compass and straightedge, treated as equal instruments in constructions. In addition, the straightedge may be used only as an instrument for the construction of a straight line, but not for measuring the lengths. Although the large number of constructions may be accomplished this way, we know of three problems dating from Greek era that cannot be solved in that way: *duplication (doubling) of a cube* – to find a side of a cube whose volume is twice that of a given cube; *trisection of an angle* – to find one third of a given angle; *squaring a circle* – to construct the square that has the same area as a given circle.

Unsolved problems of that kind initiated a completely new way of thinking – *how would it be possible to prove that certain problems could not be solved?* The answer is in modern algebra and group theory. The problem of solving algebraic equations dates far back in the past and for a long time it was the central content of algebra. Descriptions of solving certain simple algebraic equations had appeared as early as 2000 years B.C, for example in Egypt, during the Middle Dynasty, in the London papyrus known as Ahmess calculation, and on Babylonian tiles, approximately at the same time. The Babylonians were able to solve quadratic equations, while in the XVI century Girolamo Cardano, Nicolo Tartaglia, Lodoviko

Ferari, Scipione del Ferro and many others were dealing with solving cubic and quadratic equations.

For a long time, the question concerning the possibility to solve algebraic equations by radicals remained open in algebra. For an algebraic equation we shall say that it is solvable by radicals if its solutions may be obtained by using rational operations (addition, subtraction, multiplication, and division) and the operation of taking n^{th} roots, under the assumption that those operations are applied a finite number of times onto coefficients or onto functions of coefficients in which only the aforementioned operations appear. This way, the quadratic, cubic and biquadratic equations are solvable by radicals. It was to be expected that the equations of the fifth degree and of higher degrees would be solvable the same way, but it turned out to be impossible.

The initial foundations of solvability of algebraic equations were established by the French mathematician E. Galois, by connecting the solvability of algebraic equations by radicals with group theory. The demand that the roots of the algebraic equation $f(x) = 0$ may be expressed by coefficients of that equation, by using rational operations and taking n^{th} roots is expressed as a demand that the field F has to be a component of a radical extension field of K . When this demand is fulfilled one can say that the given algebraic equation is solvable by radicals. Galois has determined the criterion of solvability of algebraic equations that may be solved only by radicals, and such criterion is based on a fact that the corresponding group of that equation is solvable.

The general algebraic equation

$$\sum_{k=0}^n a_k x^k = 0 \iff a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = 0, \quad (n > 4)$$

of a degree higher than four with independent real coefficients a_k ($k = 0, 1, 2, \dots, n$) is not solvable by radicals. Many great mathematicians, such as for example L. Euler, thought that it was possible, but Ruffini and Abel disputed that at the beginning of the XIX century. This does not concern the issue of the existence of a solution of an algebraic equation of the n^{th} degree. That was proved by Gauss in 1799 in his PhD thesis. Abel's and Ruffini's problem was *whether that equation could be solved by radicals and taking n^{th} roots?* The path to the solution of that problem led to the development of modern algebra and group theory.

2. Solving the problem of duplication of a cube with compass and straightedge only

According to a legend, the problem of duplication of a cube arose when the Greeks of Athens sought assistance from the Oracle at Delphi in order to appease the Gods to grant relief from a devastating plague epidemic. The Oracle told them that to do so they had to double the size of the altar of Apollo which was in the shape of a cube. Their first attempt at doing that was a misunderstanding of the problem: they doubled the length of the sides of the cube. That, however, gave

them eight times the original volume since $(2x)^3 = 8x^3$. In modern notation, in order to fulfill the instructions of the Oracle, one must go from a cube of side x units to one of y units where $y^3 = 2x^3$, so that $y = x\sqrt[3]{2}$. Thus, essentially, given a unit length, they needed to construct a line segment of length $y = x\sqrt[3]{2}$. Now there are ways of doing this but not by using only the compass and an unmarked straightedge – which were the only tools allowed in classical Greek geometry. Constructive path of solving this problem was known to ancient Greeks unless we would not demand limitation of construction on use of only compass and straightedge. By using hard, 90 degree angle and movable cross-shaped rectangle, it is possible to construct a side of a cube whose volume is twice larger than the volume of a cube with unit side. For detailed description see the book [1].

For the sake of simplicity, let us assume that the given cube has the side length equal to the unit of measurement of length. Now the problem of duplication of a cube may be expressed in the following way:

For a given cube with a side of the unit of measurement find the side of a cube whose volume is twice that of the given cube.

The problem is reduced to the solving of the cubic equation $x^3 = 2$. We shall show that this problem *cannot be solved by using compass and straightedge only*, i.e. that the roots of polynomial $p(x) = x^3 - 2$ are not constructible.

To show this, we shall define the term of constructible and algebraic numbers. Then, we shall observe the algebraic equation $x^2 - 2 = 0$ for whose root $\sqrt{2}$ we shall show that is an irrational, algebraic and constructible number. Afterward we shall show that there is no analogy for the cubic equation, i.e. the equation $x^3 - 2 = 0$ has one root that is a real number and two conjugate complex roots, where the real root $\sqrt[3]{2}$ is irrational, algebraic, but not a constructible number.

2.1. On constructible numbers

As already mentioned, the quest for the answer concerning the possibility to solve algebraic equations by radicals has led us to the answer for the question of solving particular geometric problems by using the compass and straightedge only. In order to derive the proof on the duplication of a cube by applying algebra, it is necessary to convert that geometric problem to the language of algebra. Each geometric construction may be reduced to the following form: given a certain numbers of line segments a, b, c, \dots and looking for one or more line segments x, y, z, \dots . Geometric construction is than reduced to solving an algebraic problem:

- Determining the connection, i.e. equation between the wanted measure x and the given measures a, b, c, \dots ;
- Determining the unknown measure x by solving that equation;
- Determining whether that solution is arrived at through a procedure that corresponds to the construction performed by compass and straightedge.

Let us define the term of a constructible number. We shall say that a real number b is *constructible*, if it is possible, in a definite number of steps, to construct, with compass and straightedge, a segment of the length $|b|$.

Let us notice the connection between some of the simplest algebraic operations and elementary geometric constructions, where we shall assume that the given lengths a and b are measured according to the given “unit” measure, and that r represents any rational number.

1. Construction of a line segment that has the length $a + b$ or $a - b$

Let us spot an arbitrary point O on an arbitrary line. Construct the line segment OA that has the length a . Construct point B on that line, so that the line segment AB has the length b . Then, $OB = a + b$ (Figure 1).

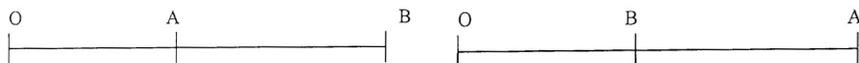


Fig. 1

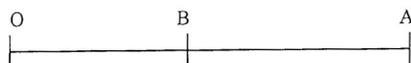


Fig. 2

The line segment $a - b$ ($a > b$) is constructed in a similar manner. On an arbitrary line, spot a point O . Construct the line segment OA that has the length a . Construct point B in the opposite direction on that line, so that the line segment AB has the length b . Then, $OB = a - b$ (Figure 2).

2. Construction of a line segment that has the length ra

In order to construct ra we simply apply r times $a + a + \dots + a$, where r is a natural number.

3. Construction of a line segment that has the length a/b

In order to construct a/b , we mark $OB = b$ and $OA = a$ on the arms of any angle with the vertex in point O , and on line OB we mark the segment $OD = 1$. Through D , we construct a straight line parallel to line AB that intersects OA at point C . Then OC will have the length a/b (Figure 3). Indeed, from the similarity of triangles OAB and OCD it follows that $OB : OD = OA : OC$, i.e. $b : 1 = a : OC$, wherefrom we can see that $OC = a/b$.

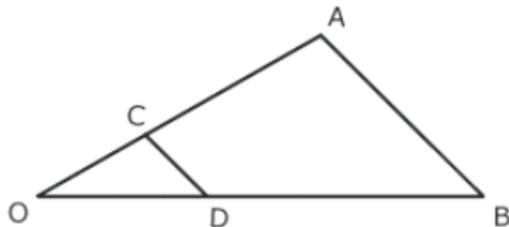


Fig. 3

4. Construction of a line segment that has the length ab

As we did thus far, we mark the line segments $OA = a$ and $OB = b$ on the arms of an angle whose vertex is in point O . On line OA , we mark the unit segment OC . Draw a straight line through points C and B , and after that a straight line through A , that is parallel to the straight line which will intersect the second arm of the angle in some point D . Then $OD = ab$ (Figure 4). The proof of validity of this construction follows from the similarity of triangles OAD and OCB (Figure 4). Indeed, from the mentioned similarity it follows that $OC : OA = OB : OD$, i.e. $OD = OA \cdot OB$. So, it is indeed true that $OD = a \cdot b$.

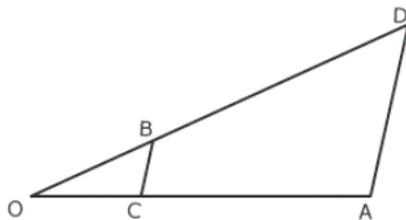


Fig. 4

From the aforementioned constructions it appears that “rational” algebraic operations of adding, subtracting, multiplying and dividing of known measures may be conducted by geometric constructions. A set of measures that may be calculated in that way form the so-called numerical field, i.e. a set of numbers such that the application of rational operations on two or more members of that set, results in a number that again belongs to that set. Rational, real and complex numbers create such fields.

Introduction of the construction of a square root takes us *out* the fields arrived at in that manner.

5. Construction of a line segment that has the length \sqrt{a}

We claim that if the given line segment is of length a , then the line of length \sqrt{a} may be constructed by compass and straightedge. Apply the line segments $OA = a$ and $AB = 1$ on a straight line. Draw a circle whose diameter equals to the line segment OB , i.e. a circle whose center is in the midpoint of line segment OB and whose radius equals $OB/2$, and construct a line perpendicular to line OB from the point A that intersects the circle in point C . Line segment $AC = \sqrt{a}$. The proof follows from the similarity of triangles OAC and ABC . Indeed, on the basis of this similarity it follows that $OA : AC = AC : AB$, i.e. $OA = AC^2$, wherefrom it follows that $AC = \sqrt{OA}$. Therefore, $AC = \sqrt{a}$.

Note that each geometric construction with a compass and straightedge in a Euclid’s plane always boils down to solving the following basic simpler tasks, not necessarily in following order:

- Construct a straight line through two points;

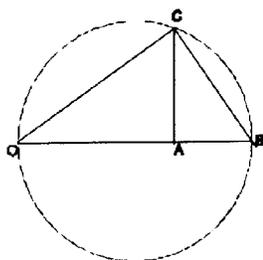


Fig. 5

- From a given point, as a center, construct a circle;
- Find points representing the intersection of two given circles;
- Find points representing the intersection of a given circle and a straight line defined by two points;
- Find points representing the intersection of two straight lines, each defined by two points,

where the basic elements – point, straight line and a circle are treated as known if they are specified at the beginning or if they were constructed within a previous step.

Let us assume that there is only one given element, unit segment 1. As the set of rational numbers is closed with regard to all rational operations, i.e. operations of adding, subtracting, multiplying and dividing two rational numbers (excluding division by zero), again result in a rational number, it turns out that all rational numbers may be constructed by compass and straightedge. Any set of numbers with this characteristic of being closed in regard to four rational operations is referred to as a numerical field. As we have shown that it is possible to construct \sqrt{k} with compass and straightedge, the only thing that remains is to check whether the extension of the set of rational numbers $Q(\sqrt{k})$ for any rational number k , will include constructible numbers only.

Let us observe any numerical field F of constructible numbers. Let us check whether it would be also possible to construct numbers in the form of $p + q\sqrt{k}$, where p , q and k come from the field of constructible numbers F .

Let us pick a number k from field F , let us find its square root and construct the field F' comprising of numbers in the form of $p + q\sqrt{k}$, where p and q are from F . It is easy to show that adding, subtracting, multiplying and dividing two numbers from the field F' again results in a number in the form $p + q\sqrt{k}$ of where p and q are from F . Field F is a subfield of the field F' . If we were to take that in the form $p + q\sqrt{k}$, $q = 0$, we conclude that all numbers from F are included in F' , assuming that \sqrt{k} is a number that does not belong to field F .

Following all those considerations we are ready to describe the set of *all constructible numbers*. Let us start from the field F_0 , e.g. the field of rational numbers,

which is defined if the unit line segment is given. Add $\sqrt{k_0}$ where k_0 is from F_0 , but $\sqrt{k_0}$ is not. By that, we construct the extended field F_1 of constructible numbers, comprising of numbers in the form of $p_0 + q_0\sqrt{k_0}$, where p_0 and q_0 are arbitrary numbers from F_0 . We can define a new extension of the field F_1 with numbers $p_1 + q_1\sqrt{k_1}$, where p_1 and q_1 are arbitrary numbers from F_1 , where k_1 is from F_1 , but $\sqrt{k_1}$ is not. By repeating this procedure, we arrive to the field F_n , after n additions of square roots. All constructible numbers are those and only those that may be arrived at with such a sequence of extended fields, in fact those ones that are located in the extended field F_n .

It would also be interesting to mention that if we assume that we can construct all numbers of a certain numerical field F , by using *only the straightedge* (in fact connecting two points within that field or finding the crossing point of two lines within that field) *we shall not exit the field in question*. It is also a fact, that with only one use of compass (the point acquired as the intersection of two circles or a circle and a line) *we can expand the field F* . The expansion of the field F would be the field F' composed of numbers in the form of $p + q\sqrt{k}$, where p and q are from F .

2.2. Duplication of a cube

As we have already mentioned, we shall start from the algebraic equation $x^2 - 2 = 0$. We shall show that the root of this equation, $\sqrt{2}$, is an algebraic, constructible and irrational number. Afterward we shall observe the equation $x^3 - 2 = 0$ and we shall show that this equation has one root that is a real number and two conjugate complex roots, where the real root $\sqrt[3]{2}$ is an algebraic, irrational number, but not a constructible number.

Let us define an algebraic number. We shall say that a certain real or complex number x is *algebraic* if it fulfills a certain algebraic equation of the following form

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0 \quad (n \geq 1, a_n \neq 0),$$

where a_k are whole numbers.

As $\sqrt{2}$ is a root of the algebraic equation $x^2 - 2 = 0$, we may conclude that $\sqrt{2}$ is an algebraic number. Based on the construction No. 5 above we may conclude that $\sqrt{2}$ is also a constructible number. It remains to be shown that $\sqrt{2}$ is an *irrational number*. We shall present an interesting geometric way of proving the irrationality of the number $\sqrt{2}$. Let us suppose the contrary, that $\sqrt{2}$ is a rational number, i.e. it may be written in the form of a real fraction $\sqrt{2} = \frac{m}{n}$ ($m, n \in \mathbf{Q}$, $\text{lcd}(m, n) = 1$). By squaring this equation it follows that $m^2 = 2n^2$. Note that the inequality $n < m < 2n$ is valid.

Let quadrangle $ABCD$ be a square with sides equal to m (Figure 6) and let A_1, A_2, C_1, C_2 represent the points on the sides of that square, so that it is valid that $AA_1 = AA_2 = CC_1 = CC_2 = n$. Let us notice points E, F, G and H so that the quadrangles $\square AA_1 EA_2$ and $\square CC_1 FC_2$ are squares with sides equal to n , and quadrangles $\square A_1 BC_1 G$ and $\square C_2 DA_2 H$ are squares with sides equal to $m - n$.

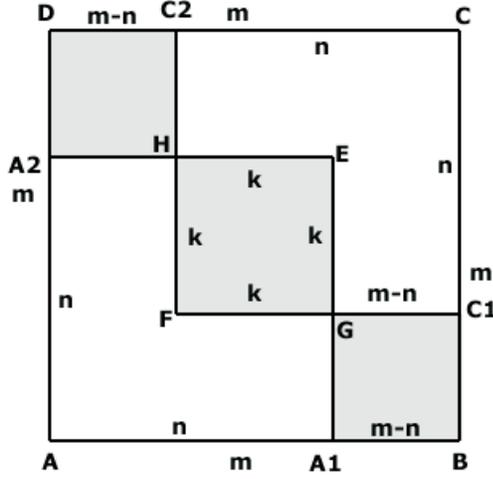


Fig. 6

Note that the quadrangle $\square FGEH$ is also a square. Let us mark the side of that square with k , where $k \in \mathbf{N}$, $k < m$. The area of the square $\square ABCD$ may be presented as

$$P_{ABCD} = P_{AA_1EA_2} + P_{A_1BC_1G} + P_{CC_1FC_2} + P_{A_2HC_2D} - P_{FGEH}$$

on the basis of which it follows that $m^2 = n^2 + (m-n)^2 + n^2 + (m-n)^2 - k^2$, i.e. $m^2 = 2n^2 + 2(m-n)^2 = k^2$. In view of the fact that $m^2 = 2n^2$, it follows that $k^2 = 2(m-n)^2$. That, however, is in contradiction with the assumption that m is the smallest non-negative whole number for which a natural number n exists so that $m^2 = 2n^2$. Therefore $\sqrt{2}$ is not a rational number.

With this we did show that $\sqrt{2}$ is an algebraic, constructible and irrational number.

Let us now observe the cubic equation $x^3 - 2 = 0$ to which the problem of cube duplication is reduced. Let us notice that this equation has one real solution and two complex solutions. Indeed, we can write the equation $x^3 - 2 = 0$ in the form of $(x - \sqrt[3]{2})(x^2 + x\sqrt[3]{2} + \sqrt[3]{2^2}) = 0$, wherefrom we can see that the solutions are $x_1 = \sqrt[3]{2}$ and $x_{2,3} = \frac{-\sqrt[3]{2} \pm \sqrt{\sqrt[3]{2^2} - 4\sqrt[3]{2^2}}}{2}$. By sorting out the expression further, we shall get $x_1 = \sqrt[3]{2}$ and $x_{2,3} = \frac{-\sqrt[3]{2} \pm \sqrt{-3\sqrt[3]{4}}}{2}$. Finally, we see that $x_1 = \sqrt[3]{2}$ and $x_{2,3} = -\sqrt[3]{2} \frac{1 \pm i\sqrt{3}}{2}$.

As $\sqrt[3]{2}$ is a root of the algebraic equation $x^3 - 2 = 0$, it follows that $\sqrt[3]{2}$ is an algebraic number. Let us show that is not a rational number. Let us suppose the

contrary, that $\sqrt[3]{2}$ is a rational number. It would follow that it could be written in the form of real fraction $\sqrt[3]{2} = \frac{p}{q}$, ($p, q \in \mathbf{Q}$, $\text{lcd}(p, q) = 1$). It would follow from this that $2 = \frac{p^3}{q^3}$, i.e. $2q^3 = p^3$. Since the left-hand side of the equation is an even number, it follows that also right-hand side of equation is also an even number, i.e. that $p = 2s$, $s \in \mathbf{Q}$. By including that in the preceding equation, equation $2q^3 = p^3$ becomes $2q^3 = 8s^3$, i.e. $q^3 = 4s^3$. From here, it turns out that q is also an even number, which is in contradiction with the assumption that $\frac{p}{q}$ is a real fraction, i.e. that the smallest common divisor of p and q is 1.

It remains to show that $\sqrt[3]{2}$ is not a constructible number.

Let us assume the opposite i.e. that the aforementioned construction is possible. As we saw in the previous paragraph, in that case x has to belong to a certain field F_k which is an extension of a set of rational numbers, arrived at by consecutive adding of square roots to the set of rational numbers.

Therefore, as we have shown that $\sqrt[3]{2}$ is not a rational number, we may conclude that x is not an element of the rational field F_0 , but instead belongs to another extended field F_k , where k is a natural number. Let us assume that k is the smallest natural number such that x belongs to the extended field F_k . As we saw that all elements of an extension of a set of rational numbers by a certain square root may be presented in the form of $p + q\sqrt{k}$, it follows that x may be written in the form of $x = p + q\sqrt{w}$, where p, q and w belong to certain field F_{k-1} , while \sqrt{w} does not.

Let us show now that if $x = p + q\sqrt{w}$ is a solution of the cubic equation $x^3 - 2 = 0$, then $y = p - q\sqrt{w}$ also is its solution. We saw that by applying the basic rational operations (adding, subtracting, multiplying, dividing), as well as applying the square root, we shall not exit a given field, then we may conclude that as x belongs to the extended field F_k , then $x^3 - 2$ also belongs to that field, and from there it follows that $x^3 - 2 = a + b\sqrt{w}$ where a and b are from the field F_{k-1} . By introducing $x = p + q\sqrt{w}$ into the preceding equation, we get $(p + q\sqrt{w})^3 - 2 = a + b\sqrt{w}$, where, after cubing the equation, it follows that $p^3 + 3p^2q\sqrt{w} + 3pq^2w + q^3w\sqrt{w} - 2 = a + b\sqrt{w}$. By grouping the corresponding factors, it follows that $(p^3 + 3pq^2w - 2) + (3p^2q + q^3w)\sqrt{w} = a + b\sqrt{w}$. Here it may be noticed that $a = p^3 + 3pq^2w - 2$ and $b = 3p^2q + q^3w$. In order to show that $y = p - q\sqrt{w}$ is a solution of cubic equation $y^3 - 2 = 0$, we switch values of q with $-q$ in expressions for a and b . It follows that $a = p^3 + 3pq^2w - 2$ and $b = -(3p^2q + q^3w)$ wherefrom it may be noticed that $y^3 - 2 = a - b\sqrt{w}$.

As we have assumed that x is a solution of cubic equation $x^3 - 2 = 0$, it follows that $a + b\sqrt{w} = 0$. From here it follows that $a = b = 0$. Let us assume the opposite, i.e. for example that $b \neq 0$. Then, from the equation $a + b\sqrt{w} = 0$ it would follow that $\sqrt{w} = -a/b$, wherefrom we may conclude that \sqrt{w} belongs to the field F_{k-1} , to which both a and b belong, which is contrary to the assumption. Therefore, it must be valid that $b = 0$, and since $a + b\sqrt{w} = 0$ it follows that $a = 0$. So, with this, we have proved that $a = b = 0$.

However, that conclusion takes us to the claim that $y = p - q\sqrt{w}$ also is a

solution of cubic equation $y^3 - 2 = 0$, keeping in mind that $y^3 - 2 = a - b\sqrt{w}$. Let us notice that it is valid that $x \neq y$, i.e. $x - y \neq 0$, since $x - y = 2q\sqrt{w}$, and $2q\sqrt{w} = 0$ only if $q = 0$, which consequentially leads to the conclusion that $x = p$, i.e. x would belong to the field F_{k-1} , which is in contradiction with the assumption.

By this, we have proved the claim that if $x = p + q\sqrt{w}$ is a solution of cubic equation $x^3 - 2 = 0$, then $y = p - q\sqrt{w}$ is also its solution, where $x \neq y$. But, we are thus facing a contradiction, because since p , q and \sqrt{w} are real numbers, it follows that x and y are real numbers, which is in contradiction with the fact that the equation $x^3 - 2 = 0$ has only one real solution.

Thus, the initial assumption has led us to a contradiction, i.e. the solution of the equation $x^3 - 2 = 0$ cannot belong to the field F_k , therefore the duplication of a cube using a compass and straightedge is impossible.

2.3. The connection between constructible and algebraic numbers

Let us now notice some of the links between constructible and algebraic numbers. Let us show that *all constructible numbers are algebraic*.

From the definition of algebraic numbers it follows that we claim that *each constructible number is a root of an n^{th} degree polynomial*.

Let start from the field F_0 , the field of rational numbers generated by one line segment. Numbers in the field F_1 are roots of a square equation, numbers from F_2 are roots of equation of the fourth degree, and, generally, numbers of the field F_k are roots of an equation of 2^k degree, with rational coefficients. Let us observe any number from F_2 field, in the form of $x = p + q\sqrt{w}$, where p , q , w are from F_1 , i.e. of the form $p = a + b\sqrt{s}$, $q = c + d\sqrt{s}$, $w = e + f\sqrt{s}$ where a, b, c, d, e, f, s are rational numbers. By squaring the equation $x = p + q\sqrt{w}$ we arrive to the equation $x^2 - 2px + p^2 = q^2w$, where all the coefficients are in field F , which is the field generated by \sqrt{s} . Furthermore, this equation may be written in the following form:

$$\begin{aligned} x^2 - 2(a + b\sqrt{s})x + (a + b\sqrt{s})^2 &= (c + d\sqrt{s})^2(e + f\sqrt{s}), \\ &\dots\dots \\ x^2 - 2ax + (a^2 - c^2e + b^2s - d^2es - 2cdf s) &= \\ &= \sqrt{s}(2bx + (-2ab + 2cde + c^2f + d^2fs)). \end{aligned}$$

Keeping in mind that a, b, c, d, e, f, s are rational numbers, we come to the conclusion that this equation may be presented in the form of $x^2 + ux + v = \sqrt{s}(rx + t)$ where u, v, s, r, t are all rational numbers. By squaring this equation, we get an equation of the fourth degree $(x^2 + ux + v)^2 = s(rx + t)^2$ with rational coefficients, as it was claimed.

It is worth noting that the opposite is not valid, i.e. that not all algebraic numbers are constructible! We have shown that the roots of the third degree equation $x^3 - 2 = 0$, i.e. algebraic numbers, are not constructible.

F is the algebraic extension of the field \mathbf{Q} , if each and every element from F is algebraic over \mathbf{Q} . If that polynomial does not exist, than the number α is

