STRAIGHTEDGE AND COMPASS CONSTRUCTIONS

JIHONG CAI AND PARTH DESHMUKH

1. A Bit of History

The straightedge and compass problem is one of the oldest problems in mathematics. Straightedges and compasses are simple tools to construct and use, making them easily accessible to ancient mathematicians. Additionally, the first applications of mathematics were in mapmaking and surveying. Using taut ropes and markers to survey land lets one draw out circles and lines, the same as drawing lines and circles with a straightedge and compass.

The Ancient Greek mathematicians were fascinated by the straightedge and compass, and sought to learn what constructions they could and could not draw with the two tools. Traditionally, their allowed tools was an unmarked straightedge and a compass that closes when lifted from the paper, so the compass couldn't "remember" a distance. The restriction on compasses to not be able to remember distances faded since they found a construction that let them duplicate any length, so the restriction simply added extra steps. Thus the modern version, with a compass that could remember distances, emerged. Their goal was to ultimately prove exactly they could and could not construct with the two tools, one that would not be realized until the advent of Galois theory millenia later.

2. Rules of Construction

We start from a line segment of unit length. In practice, imagine drawing a random line segment, and letting the length of that segment determine the units of measurement. Three constructions are allowed:

- Given two points, draw a line between them.
- Extend an existing line past its endpoints. (The endpoints are not lost.)
- Draw a circle given a center point and some distance.

The first two are how we use a ruler: drawing lines between points and extending them as necessary. The last one is how we use a compass: from a center, set the pencil either a random distance away or at a given distance corresponding to the distance between two points. Importantly, note that the compass cannot have marks on it, e.g. it cannot measure distances, only replicate them or pick at random.

3. Known Examples and Non-examples

A few examples of possible constructions:

- Construct perpendicular and parallel lines
- Duplicate an angle (using parallel lines)
- Split an angle into two
- Split a line segment into two

This note is for the MATRIX event on 10/03/2023.

- Construct triangles, squares, and pentagons
- Given a regular polygon, make a new one with double the sides

This doesn't mean we can make everything. For example, we can make regular polygons with $3 \cdot 2^k$, $4 \cdot 2^k$, and $5 \cdot 2^k$ sides where $k \in \mathbb{N}$, but we cannot make a seven-sided regular polygon. There are three classical examples of impossible constructions:

- Duplicating the cube: create a cube twice the volume of a given cube
- Squaring the circle: given a circle, draw a square of the same area
- Trisecting an angle: split an arbitrary angle into three

For the first two examples, we can rephrase them in which segment is the given segment of unit length, and ultimately, what length segment we want to construct off it. In Section 7, we show why these two are impossible. We will show why the last example is impossible in Section 9.

For duplicating the cube, the segment of unit length is the side length of a cube. Said cube would have volume 1, so we want to construct a cube of volume 2. Thus the side length of the new cube we would want to construct is $\sqrt[3]{2}$.

For squaring the circle, the segment of unit length is the radius of the circle. Since the circle then has area π, we need to construct a segment of length $\sqrt{\pi}$ to be the side length of the new square.

4. Some Basic Constructions

We can add and subtract lengths a and b very easily:

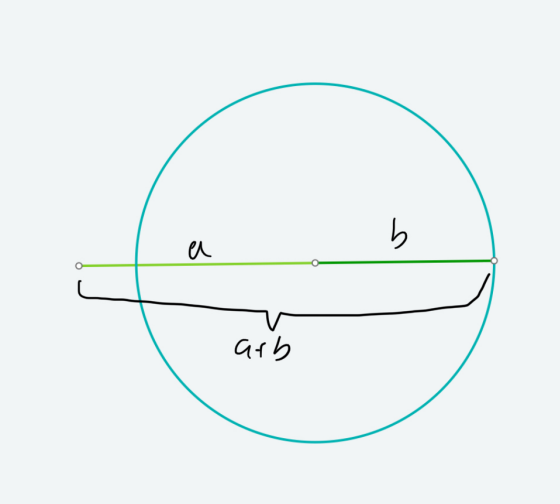


FIGURE 1. Adding two lengths a and b.

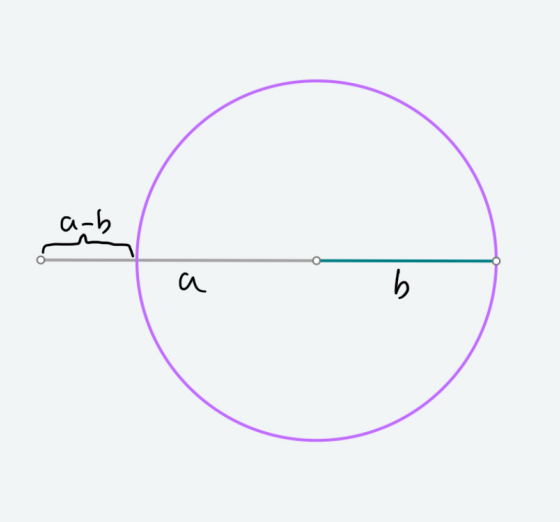


FIGURE 2. Subtracting two lengths a and b.

To add and subtract two lengths, we start with a line segment $\overline{A}$ with length a, and extend it continuously. We then take the length b with our compass, and draw a circle of radius b centered at the right endpoint of $\overline{A}$. The places where the circle intersects with the line through $\overline{A}$ correspond with the distances $a - b$ and $a + b$ from the left endpoint from $\overline{A}$.

Next, by drawing similar triangles, we can multiply and divide a and b:

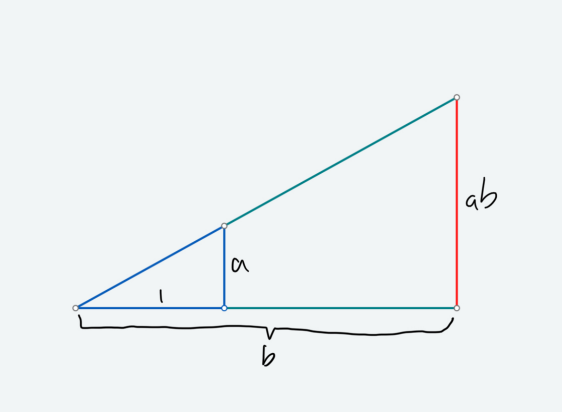


FIGURE 3. Multiplying two lengths a and b.

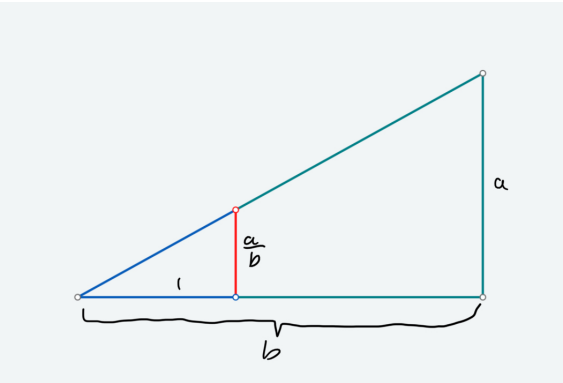


FIGURE 4. Dividing two lengths a and b.

The core idea is constructing two similar right triangles.[1] For multiplying a and b, we construct a right triangle with legs a and 1, and extend one leg so it has length b. We do this by extending the leg continuously, then using our compass to mark a distance of b from the endpoint. Thus the new leg of the larger right triangle has length ab. When dividing, we start from the larger triangle with legs a and b, and mark off a smaller leg for the new triangle of length 1 with our compass. Thus the other leg of the new triangle is length $\frac{a}{b}$.

Finally, we can take the square root of a length a, and additionally, explicitly find $\sqrt{n}$ for all $n \in \mathbb{N}$:

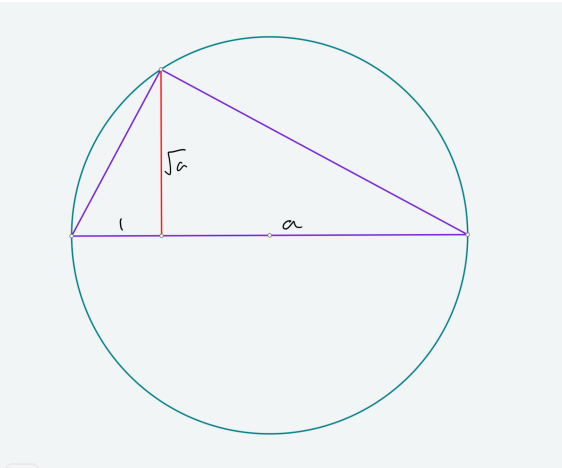


FIGURE 5. Taking $\sqrt{a}$.

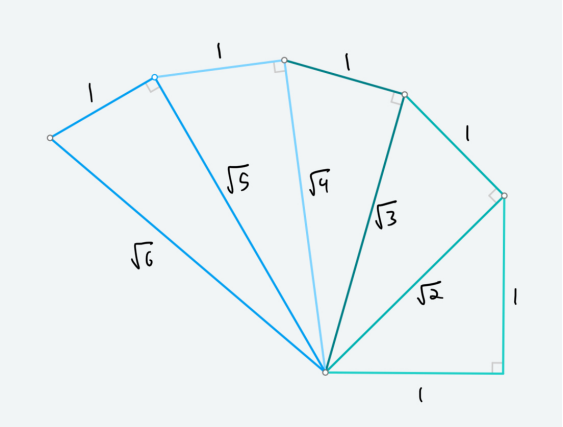


FIGURE 6. Repeating a process to get $\sqrt{n}$ for any whole number n.

The process for taking the square root of a is somewhat longer. First, extend the line segment of length a with a line segment of length 1 on one side. We want to use the new line segment of length $a + 1$, denoted $\overline{D}$, as the diameter of a circle; to do that, mark off two circles with equal radii and centers at either endpoint of $\overline{D}$, then join their intersection points to find the perpendicular bisector. This gives us the midpoint, which we can draw a circle from. Next, draw a perpendicular line up from the point splitting the two segments in

[1]We make right triangles because we can construct perpendicular lines using the process described in the square-root construction.

$\overline{D}$, indicated in red in the figure. Denote said line $\overline{A}$. Take the point it intersects with the circle and draw lines between it and both endpoints of $\overline{D}$ to complete our construction.

We have created two similar triangles, which can be seen because both are right triangles and because the top two angles of the two triangles add up to 90 degrees. One of these triangles has side lengths 1 and $|\overline{A}|$, and the other has side lengths $|\overline{A}|$ and A. So $\frac{1}{|\overline{A}|} = \frac{|\overline{A}|}{a}$, which solving gets us $|\overline{A}| = \sqrt{a}$.

The other construction is simpler. We start with a right triangle with legs of length 1. Its hypotenuse has length $\sqrt{1^2 + 1^2} = \sqrt{2}$. We draw a new leg from the hypotenuse by drawing a perpendicular line at the edge, and mark off its length as 1. We have thus drawn a new right triangle, whose hypotenuse is length $\sqrt{(\sqrt{2})^2 + 1^2} = \sqrt{3}$. We can repeat this over and over; at each step, we go from a hypotenuse of length $\sqrt{n}$ to $\sqrt{(\sqrt{n})^2 + 1^2} = \sqrt{n+1}$. By induction, we can construct any $\sqrt{n}, n \in \mathbb{N}$.

5. An Algebraic Interpretation

From the constructions, we learned that there are five allowed operations in the straightline and compass constructions: $+, -, \times, \div, \sqrt{\ }$. The first four operations can be summarized as the operations for a *field* and the square root represents field extensions.

Formally, field is a set F with two binary operations $+$ and $\times$ satisfying the following properties:

(1) Addition is associative: $(a + b) + c = a + (b + c)$ for all $a, b, c \in F$
(2) Addition is commutative: $a + b = b + a$ for all $a, b \in F$
(3) Additive identity exists: there exists an element $0 \in F$ so that $0 + a = a = a + 0$ for all $a \in F$
(4) Additive inverse exists: for all $a \in F$, there exists $b \in F$ so that $a + b = 0 = b + a$
(5) Multiplication is associative: $(ab)c = a(bc)$
(6) Multiplication is commutative: $ab = ba$ for all $a, b \in F$
(7) Additive identity exists: there exists an element $1 \in F$ so that $1 \cdot a = a = a \cdot 1$ for all $0 \neq a \in F$
(8) Additive inverse exists: for all $0 \neq a \in F$, there exists $b \in F$ so that $ab = 1 = ba$
(9) Distributive law: $a(b + c) = ab + ac$ for all $a, b, c \in F$.

Example. The rational numbers $\mathbb{Q} = \left\{ \frac{p}{q} : p, q \in \mathbb{Z} \right\}$ is a field.

Example. The real number $\mathbb{R}$ is a field.

Example. The set of complex numbers $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$ is a field.

From the basic constructions, we know that given the unit interval, i.e. the interval with length 1, we can construct all integers (by doing addition and subtraction). We can construct quotients to get all rational numbers. Therefore, the field that we are interested in are the field of rational numbers $\mathbb{Q}$.

6. Field Extension

Notice that we have one more operation not mentioned yet: the square-root $\sqrt{\ }$. This can be realized via so-called field extension, i.e. extending the field $\mathbb{Q}$ so that $\sqrt{r}$ is included. We further require that the resulting set is still a field.

Therefore, we define

$$\mathbb{Q}(\sqrt{r}) = \left\{a + b\sqrt{r} : a, b \in \mathbb{Q}\right\}.$$

Check that this is in fact a field. We can represent this extension visually as

$$\begin{array}{c} \mathbb{Q}(\sqrt{r}) \\ | \\ \mathbb{Q} \end{array}$$

Example. Consider $\mathbb{Q}(\sqrt{2})$. This is defined as $\left\{a + b\sqrt{2} : a, b \in \mathbb{Q}\right\}$. Numbers like $1 + 5\sqrt{2}$, $\frac{2}{3} - \sqrt{2}$, $-18\sqrt{2}$ are elements of $\mathbb{Q}(\sqrt{2})$.

For simplicity, we will only consider the extension of $\mathbb{Q}$. Everything we discuss here can be formulated in a more general sense, but it is not necessary for solving the straightedge and compass problem.[2] We will denote K/F if K is an extension of F.

Let us begin by comparing the number of variables we have for $\mathbb{Q}$ and $\mathbb{Q}(\sqrt{2})$. $\mathbb{Q} = \{a : a \in \mathbb{Q}\}$. Hence it can be represented by one letter. On the other hand, $\mathbb{Q}(\sqrt{2}) = \left\{a + b\sqrt{2} : a, b \in \mathbb{Q}\right\}$, so we need two variables we describe this field. Hence, we will define the degree of the extension

$$[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$$

as the quotient of the numbers of variables needed in the extended field compared with the original one. Similarly, we can conclude that $[\mathbb{Q}(\sqrt{3}) : \mathbb{Q}] = 2$, $[\mathbb{Q}(\sqrt{-5}) : \mathbb{Q}] = 2$, and $[\mathbb{Q}(i) : \mathbb{Q}] = 2$.

Now consider the extension $\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}(\sqrt{2})$. We want to solve for the degree of this extension

$$[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})].$$

This question will be easy if we know how to represent $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. We know that $1, \sqrt{2}, \sqrt{3}$ will need to be in this field, but is this enough? Recall that a field is closed under multiplication. So $\sqrt{2}\sqrt{3} = \sqrt{6}$ should be in the field. One can check that any other product will be some variations of a, $b\sqrt{2}$, $c\sqrt{3}$, or $d\sqrt{6}$. Hence,

$$\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \left\{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} : a, b, c, d \in \mathbb{Q}\right\}.$$

By counting, $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ has four variables, whence $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})] = 4/2 = 2$.

It is not hard to see that $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$. But more generally, we can generalize this to the tower law.

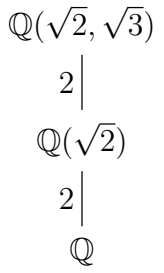
Proposition (Tower law)**.** *Suppose we have an inclusion of fields $F \subseteq K \subseteq L$. Then*

$$[L : F] = [L : K][K : F].$$

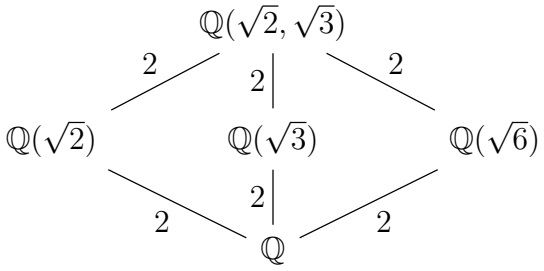
As a sanity check,

$$4 = [\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2 \cdot 2.$$

[2]If we talk about advanced materials, such as the unsolvability of quintics, we will state the more abstract version and reference this note for concrete examples.

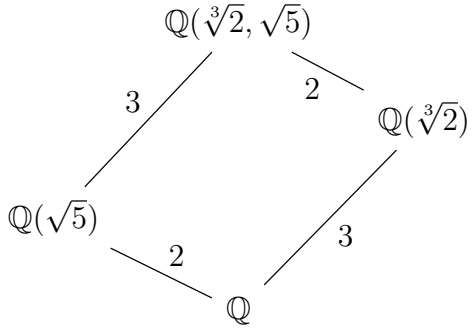


In fact, there are many different extensions we can get as an intermediate step:



This is also implied from the tower law that it is possible to do the extension via different ladder.

Example. Consider the extension $\mathbb{Q}(\sqrt[3]{2},\sqrt{5})/\mathbb{Q}$. There are two route we can take with different degree of extension: $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{5}) \subseteq \mathbb{Q}(\sqrt[3]{2},\sqrt{5})$ and $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2}) \subseteq \mathbb{Q}(\sqrt[3]{2},\sqrt{5})$. The product of the degree has to end by tower law, though.



We can derive a more general formulation by applying tower law repetitively.

Corollary. *If $F = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_r = L$, then*

$$[L : F] = [K_r : K_0] = [K_r : K_{r-1}] \dots [K_1 : K_0].$$

Note that the tower rule is only useful when each extension is finite, i.e. $[K_i, K_{i-1}] < \infty$. Otherwise, we are multiplying infinity, which does not tell us any useful information.

We say an extension is finite is K/F is a field extension and $[K : F] < \infty$.

Example. $[\mathbb{C}, \mathbb{R}] = [\mathbb{R}^2 : \mathbb{R}] = [\mathbb{R}(i) : \mathbb{R}] = 2$

Example. $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$

Example. $[\mathbb{Q}(\pi) : \mathbb{Q}] = \infty$

Example. $[\mathbb{R} : \mathbb{Q}] = \infty$

A technical note here. There is a difference between finite extension and finitely generated extension. Finite extension means given K/F an extension, $[K : F] < \infty$. A finitely generated extension is when $K = F(a_1, \ldots, a_n)$. Every finite extension is finitely generated, but the converse is false. Consider $\mathbb{Q}(\pi)$. This is finitely generated infinite extension, i.e. the only generator for this extension is π, but the extension $\mathbb{Q}(\pi)/\mathbb{Q}$ has infinite degree.

7. Constructible Numbers

Going back to straightedge and compass construction. Since the only allowed construction besides the field operations is square-root. Hence, we can only do extension of $\mathbb{Q}$ by square-roots. That means, all extension will be degree two from the previous field.

More explicitly, given a finite list of numbers we want to adjoint to $\mathbb{Q}$, say $\sqrt{a_1}, \sqrt{a_2}, \ldots, \sqrt{a_n}$, the extension $\mathbb{Q}(\sqrt{a_1}, \sqrt{a_2}, \ldots, \sqrt{a_n})/\mathbb{Q}$ gives me the set of all possible points I can construct by ruler and compass by choosing arbitrary a_i. Of course, we want the a_i to be non-trivial and non-repetitive. For example, $\sqrt{4} = 2 \in \mathbb{Q}$ and $\mathbb{Q}(\sqrt{4}) = \mathbb{Q}(2) = \mathbb{Q}$. Also consider the case where $\sqrt{12} = 2\sqrt{3}$ and $\mathbb{Q}(\sqrt{12}) = \mathbb{Q}(\sqrt{3})$, and $\mathbb{Q}(\sqrt{3}, \sqrt{12}) = \mathbb{Q}(\sqrt{3})$. Therefore, these extension will have degree 1, which is trivial, and does not add much to the conversation. This is the same as saying I want to draw a point that I already have, which is not helpful.

The construction we are interested in are so-called 2-radical extensions.[3] Explicitly, K/F is called 2-radical if there exists a finite tower of subsets of the form

$$F = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_r = K$$

where

$$[K_j : K_{j-1}] = 2$$

for all $j = 1, \ldots, r$.

Example. For example,

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt[4]{2}) \subseteq \mathbb{Q}(\sqrt[8]{2}) \subseteq \mathbb{Q}(\sqrt[16]{2}) \subseteq \cdots \subseteq \mathbb{Q}(\sqrt[2^n]{2})$$

is a 2-radical extension of degree $[\mathbb{Q}(\sqrt[2^n]{2}) : \mathbb{Q}] = 2^n$.

Example. A more complicated 2-radical extension is given by

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt{1+\sqrt{2}}) \subseteq \mathbb{Q}(\sqrt{1+\sqrt{1+\sqrt{2}}}) \subseteq \mathbb{Q}(\sqrt{1+\sqrt{1+\sqrt{1+\sqrt{2}}}}) \subseteq \ldots.$$

As an optional challenge, check that this is a 2-radical extension.

I claim that all these numbers are constructible since they are all 2-radical extensions.

Theorem. *Given a initial set $\mathcal{P}$ where we start the construction form. A point α is constructible in $\mathcal{P}$ if and only if $\alpha \in F^{2rad}$. In particular, when $\mathcal{P} = \{0, 1\}$ (or more generally $\mathcal{P} \subseteq \mathbb{N} \cup \{0\}$), α is constructible if and only if $\alpha \in \mathbb{Q}^{2rad}$.*

We can then show that why the two constructions from Section 3 are impossible.

First, duplicating the cube requires constructing a segment of length $\sqrt[3]{2}$. Extending $\mathbb{Q}$ to include it requires a degree 3 extension

$$\mathbb{Q}(\sqrt[3]{2}) = \left\{ a + b\sqrt[3]{2} + c\sqrt[3]{2^2} : a, b, c \in \mathbb{Q} \right\}$$

[3]This terminology is introduced by Charles Rezk in his lecture note, and is not a standard terminology.

Second, squaring the circle requires constructing a segment of length $\sqrt{\pi}$. We know we can take arbitrary square roots, so we just need the length π. For this to be a field, we need

$$\mathbb{Q}(\pi) = \left\{a_0 + a_1\pi + a_2\pi^2 + a_3\pi^3 + \cdots + a_n\pi^n + \cdots : a_i \in \mathbb{Q}\right\}.$$

This means that we need (countably) infinite number of variables for this extension to be closed under multiplication, as π^2 cannot be expressed in the form of $a + b\pi$. This means that $[\mathbb{Q}(\pi) : \mathbb{Q}] = \infty$. It follows that $[\mathbb{Q}(\sqrt{\pi}) : \mathbb{Q}] = \infty$, certainly not 2-radical. A more direct way to reach this conclusion is to say that π is transcendental, i.e. not algebraic, by Lindemann's theorem.

8. Fermat Prime and Constructible Regular p-gons

There is an equivalent statement of degree 2 (or 2-radical) extension in terms of minimal polynomial of degree 2^r. For example, we want to construct a regular n-gon, what we are really asking is if we can construct each internal angle. Recall that for a regular n-gon, the internal angle is of degree $\frac{360}{n}^\circ$. Equivalently, we can use the exponential form to write $\zeta_n = e^{2\pi i/n}$. ζ_n is the root for the polynomial. We can factorize this polynomial into irreducible, and to see if the solutions are 2-radical, or is it a solution to some 2^r degree polynomial.

Consider a special case when $n = p$ a prime, then $x^p - 1 = (x-1)(x^{p+1} + x^{p-2} + \cdots + 1)$. Its solutions are $\zeta_p = e^{2\pi i/p}$, satisfying

$$\mathbb{Q}(\zeta_p) = p - 1.$$

Hence, ζ_p is constructible if and only if $p - 1 = 2^r$ for some r, i.e. $p = 2^r + 1$. This is exactly what is known as Fermat number. There are only five known examples of Fermat primes:

$$3 = 2^1 + 1, \quad 5 = 2^2 + 1, \quad 17 = 2^4 + 1, \quad 257 = 2^8 + 1, \quad 65537 = 2^{16} + 1.$$

9. Impossibility of Trisecting General Angle

To show it is impossible to trisect any angle, we just need to show one example. We will show that it is impossible to construct $\theta = \frac{2\pi}{3}$. This is the same showing $\zeta = e^{2\pi i/9}$ is not constructible, i.e. not 2-radical. Note that $\zeta^9 = 1$ but $\zeta^3 \neq 1$. Hence $\zeta^9 - 1 = (\zeta^3-1)(\zeta^6+\zeta^3+1) = 0$ if and only if $\zeta^6+\zeta^3+1 = 0$. This means that $f = x^6+x^3+1 \in \mathbb{Q}[x]$, so $[\mathbb{Q}[\zeta] : \mathbb{Q}] \leq 3$. Now let $\alpha = \zeta + \zeta^{-1} \in \mathbb{Q}(\zeta)$. Using the fact that $f(\zeta) = 0$, we can show that

$$\alpha^3 = \zeta^3 + 3\zeta + 3\zeta^{-1} + \zeta^{-3} = 3\alpha - 1.$$

Thus α is a root of $g = x^3 - 3x + 1 \in \mathbb{Q}[x]$. By the rational roots test, this has no root in $\mathbb{Q}$ so is irreducible over $\mathbb{Q}$. Thus $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$. Since $[\mathbb{Q}(\zeta) : \mathbb{Q}] = [\mathbb{Q}(\zeta) : \mathbb{Q}(\alpha)][\mathbb{Q}(\alpha) : \mathbb{Q}]$, we conclude that 3 divides $[\mathbb{Q}(\zeta) : \mathbb{Q}]$, whence $[\mathbb{Q}(\zeta) : \mathbb{Q}] = 3$.

Email address: jihongc2@illinois.edu

Email address: parthd2@illinois.edu

Mathematical Advancement Through Research Idea eXchange (MATRIX)

Department of Mathematics, University of Illinois Urbana-Champaign