UNDERSTANDING RULER AND COMPASS CONSTRUCTIONS WITH FIELD THEORY

ISAAC M. DAVIS

Abstract. By associating a subfield of $\mathbb{R}$ to a set of points $P_0 \subseteq \mathbb{R}^2$, geometric properties of ruler and compass constructions on P_0 can be understood algebraically, creating a powerful tool for proving the possibility or impossibility of certain constructions. In this paper, field theory will be used to prove the impossibility of doubling the cube and squaring the circle, and will be used in studying the constructibility of regular n-gons.

1. Background

There are several conventions for defining a ruler and compass construction, all of which are for the most part equivalent. Ian Stewart's book *Galois Theory* is the source of the definitions and conventions used in this paper [1].

Definition 1.1 (Ruler and Compass Construction)**.** A **ruler and compass construction** on a set of points $P_0 \subseteq \mathbb{R}^2$ is defined by two operations:

(1) draw a straight line between two points in P_0
(2) draw a circle centered at some point $p \in P_0$ with radius equal to the distance between two points in P_0.

Definition 1.2 (Constructibility)**.** A point $p \in \mathbb{R}^2$ is said to be **constructible in one step from** P_0 if p is the intersection of two circles, two lines, or a line and a circle constructible on P_0. If $r_i = (x_i, y_i)$ is constructible in one step from P_{i-1}, then we denote $P_i = P_{i-1} \cup \{r_i\}$. A point r_n is said to be **constructible from** P_0 if there exists a finite sequence of points $r_1, \ldots, r_n \in \mathbb{R}^2$ such that for all $i \in \{1, ..., n\}$, r_i is constructible in one step from the set $P_0 \cup \{r_1, \ldots, r_{i-1}\}$.

Given these definitions for a ruler and compass construction, we can now associate a subfield of $\mathbb{R}$ with a set of points in $\mathbb{R}^2$.

Definition 1.3 (Point field)**.** Let P_0 be a set of points in $\mathbb{R}^2$. Then the **point field** of P_0, denoted K_0, is the smallest subfield of $\mathbb{R}$ containing every coordinate of every point in P_0. If some point $r_i = (x_i, y_i)$ is constructible in one step from P_{i-1}, then the point field K_i of $P_i = P_{i-1} \cup \{r_i\}$ is the smallest subfield of $\mathbb{R}$ containing K_{i-1}, x_i, and y_i.

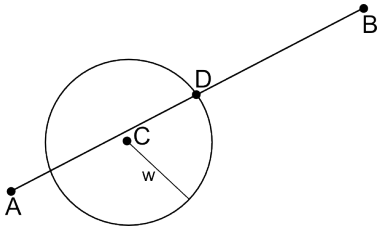
A very simple result follows from our definition of a point field:

Lemma 1.4. *Let $r_i = (x_i, y_i)$ be a point constructible from P_0, and let K_i be the point field of P_i. Then both x_i and y_i are roots of quadratic polynomials over the field K_{i-1}.*

Date: August 1, 2008.

Proof. Note that if r_i is constructible from P_0, it must be constructible in one step from some set of points P_{i-1}. Therefore r_i must be the intersection of two lines, two circles, or a circle and a line constructible on P_{i-1}, and we need only prove the lemma for these three cases. These can be proven with simple coordinate geometry. We will only prove the line-meets-circle case, but proofs of the other two cases are very similar.

Consider a line from $A = (p, q)$ to $B = (r, s)$, and a circle centered at $C = (t, u)$ with radius w, where w is the distance between two points in P_{i-1}, and let $D = (x_0, y_0)$ be a point of intersection.



The equation of $\overline{AB}$ is

$$\frac{x-p}{r-p} = \frac{y-q}{s-q}.$$

The equation of the circle is

$$(x-t)^2 + (y-u)^2 = w^2.$$

Solving for x_0 gives

$$(x_0-t)^2 + \left[\frac{(s-q)}{(r-p)}(x_0-p) + q - u\right]^2 - w^2 = 0. \tag{1.5}$$

The coordinates of A,B,C, and D are all elements of K_{i-1}, as are w and w^2. Therefore Equation 1.5 is a quadratic equation over K_{i-1}, and x_0 is a root of a quadratic equation over K_{i-1}. Solving for y_0 gives a similar quadratic equation. □

We are almost ready to prove the key theorem for ruler and compass constructions, but first we will need the following lemma from basic field theory, which we will not prove here.

Lemma 1.6. *Let K, L, and M be fields such that $K \subseteq L \subseteq M$. Then $[M : K] = [M : L][L : K]$.*

Corollary 1.7. *If $K_0 \subseteq \ldots \subseteq K_n$, then $[K_n : K_0] = [K_n : K_{n-1}] \cdots [K_1 : K_0]$.*

Proof. This is a simple proof by induction using Lemma 1.6. □

Theorem 1.8. *Let $P_0 \subseteq \mathbb{R}^2$, and let K_0 be its point field. Then for all constructible points $r = (x, y)$, the degrees $[K_0(x) : K_0]$ and $[K_0(y) : K_0]$ are powers of 2.*

Proof. The point $r_n = (x, y)$ is constructible from P_0, so there exists a finite sequence of points $r_1, \ldots, r_n$ such that for all $i \in \{1, ..., n\}$, r_i is constructible in one step from $P_0 \cup \{r_1, ..., r_{i-1}\}$. By Lemma 1.6, if $r_i = (x_i, y_i)$, then $[K_{i-1}(x_i) : K_{i-1}]$ is 1 if the quadratic polynomial over K_{i-1} of which x_i is a root is reducible over

K_{i-1}, or 2 if it is irreducible. The same holds for $[K_{i-1}(y_i) : K_{i-1}]$. Therefore by Corollary 1.7, we have

$$[K_{i-1}(x_i, y_i) : K_{i-1}] = [K_{i-1}(x_i, y_i) : K_{i-1}(x_i)][K_{i-1}(x_i) : K_{i-1}] = 2^n$$

where n = 0, 1, or 2, and $K_i = K_{i-1}(x_i, y_i)$, so $[K_i : K_{i-1}]$ is a power of 2. If K_n is the point field of P_n, then

$$[K_n : K_0] = [K_n : K_{n-1}] \cdots [K_1 : K_0]$$

and by Corollary 1.7 $[K_n : K_0]$ is a power of 2. But

$$[K_n : K_0] = [K_n : K_0(x)][K_0(x) : K_0]$$

so $[K_0(x) : K_0]$ is a power of 2. Similar steps show the same for $[K_0(y) : K_0]$. □

Now we have the tools necessary to prove the impossibility of two significant constructions: doubling the cube and squaring the circle.

2. Impossible Constructions

Theorem 2.1. *Given a cube of volume V, a cube of volume $2V$ is impossible to construct using rulers and compasses.*

Proof. If we have a cube, then we have a side of the cube, and we may, with no loss of generality, assume that one side of the cube is the line between points $(0, 0)$ and $(1, 0)$. The volume of such a cube is 1, so constructing a cube of volume 2 would be equivalent to constructing some point $(\alpha, 0)$ such that $\alpha^3 = 2$. However, the smallest field containing 0 and 1 is $\mathbb{Q}$, and the minimum polynomial of α over $\mathbb{Q}$ is $\alpha^3 - 2$. This polynomial has degree 3, so we have

$$[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$$

But by Theorem 1.8, if $(\alpha, 0)$ is constructible from $\{(0, 0), (1, 0)\}$, then $[\mathbb{Q}(\alpha) : \mathbb{Q}]$ must be a power of 2. This is contradictory; therefore, such a point $(\alpha, 0)$ cannot be constructed, and we cannot construct a cube of volume 2. □

Theorem 2.2. *Given a circle of area A, a square of area A is impossible to construct using rulers and compasses.*

In the proof of this theorem, we will use without proof the fact that π and $\sqrt{\pi}$ are transcendental over $\mathbb{Q}$.

Proof. We can, with no loss of generality, assume that our circle is the unit circle centered at $(0, 0)$. The area of this circle is π, and constructing a square with area π is equivalent to constructing a point $(\sqrt{\pi}, 0)$. The smallest field containing 0 and 1 is $\mathbb{Q}$, so the point field obtained from adjoining $(\sqrt{\pi}, 0)$ to $\mathbb{Q}$ is $\mathbb{Q}(\sqrt{\pi})$. However, by Theorem 1.8, $[\mathbb{Q}(\sqrt{\pi}) : \mathbb{Q}]$ must be a power of 2, and $\sqrt{\pi}$ must be algebraic over $\mathbb{Q}$, which is clearly not true. Therefore such a point $(\sqrt{\pi}, 0)$ cannot be constructed; hence we cannot construct a square of area π. □

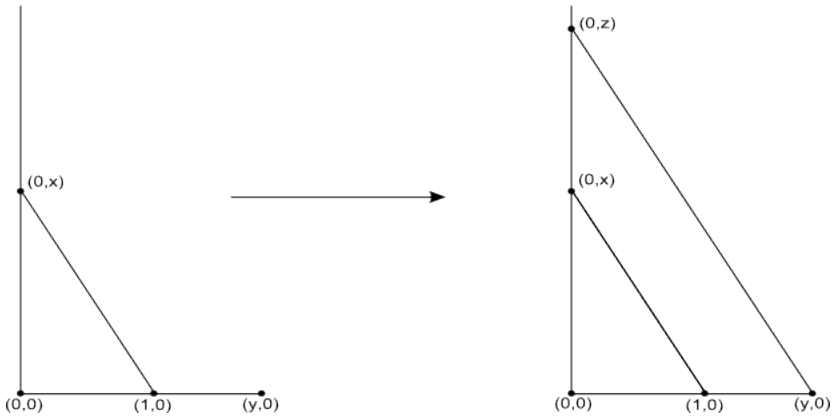
3. Field Properties of the Constructible Points

Theorem 3.1. *Let $P \subseteq \mathbb{R}^2$ contain $(0,0)$ and $(1,0)$, and let K be the point field of P. Then the point $r = (x,y)$ is constructible from P if $x, y \in K$.*

The reader should note that, because P contains $(0,0)$ and $(1,0)$, we can construct the coordinate axes. We will use this fact extensively in the proof.

Proof. First, we must prove that (x,y) is constructible from P if and only if $(x,0)$ and $(y,0)$ are constructible. Suppose (x,y) is constructible from P. We can construct lines that project this point onto the coordinate axes. The projection onto the x-axis is the point $(x,0)$. The projection onto the y-axis is the point $(0,y)$. A circle centered at $(0,0)$ with radius y will intersect the x-axis at $(y,0)$. Doing these steps in reverse order will construct the point (x,y) from $(x,0)$ and $(y,0)$.

Second, we must prove that the set of all numbers z such that $(z,0)$ is constructible from P forms a field containing K, which we will denote F_P. To do so, it will suffice to prove that if $(x,0)$ and $(y,0)$ are constructible from P, we can construct the points $(x+y,0)$, $(x-y,0)$, $(xy,0)$ and $(\frac{x}{y},0)$. Suppose we have points $(x,0)$ and $(y,0)$. A circle centered at $(x,0)$ with radius y will intersect the x-axis at points $(x+y,0)$ and $(x-y,0)$. For $(xy,0)$, consider the following construction:



The two right triangles constructed are similar triangles, so the ratios of corresponding sides are equal, which gives us the following equivalent equations:

$$\begin{aligned} \frac{x}{1} &= \frac{z}{y} \\ z &= xy. \end{aligned}$$

A similar construction follows for $(\frac{1}{y},0)$, and from this we can construct $(\frac{x}{y},0)$. So the set of all z such that $(z,0)$ is constructible from P forms a field F_P. And K is the field spanned by the coordinates of all points in P, so clearly F_P must contain K.

Now, suppose x and y are in K. Then x and y must also be in F_P. Therefore the points $(x,0)$ and $(y,0)$ are constructible from P. And from the first part of the proof we know that (x,y) is constructible if and only if $(x,0)$ and $(y,0)$ are constructible, so the point (x,y) must also be constructible from P. □

Theorem 3.2. *Let $P_0 \subseteq \mathbb{R}^2$, and let K_0 be the point field of P_0. Suppose $K_0(\alpha)$ is an extension of K_0 with degree 2 such that $K_0(\alpha) \subseteq \mathbb{R}$. Then for all $x, y \in K_0(\alpha)$, the point (x, y) is constructible from P_0.*

Proof. Since $[K_0(\alpha) : K_0] = 2$, the minimum polynomial of α over K_0 is quadratic, so for some $p, q \in K_0$, we have

$$\alpha^2 + p\alpha + q = 0.$$

From this, we get

$$\alpha = \frac{-p \pm \sqrt{p^2 - 4q}}{2}.$$

However, $K_0(\alpha) \subseteq \mathbb{R}$, so α must be real, and we have $p^2 - 4q \geq 0$. By Theorem 3.1, the desired result will follow if for all $k \in K_0$, we can construct the point $(0, \sqrt{k})$ from P_0. This is provable with simple coordinate geometry. □

Using induction on the result of Theorem 3.2 gives us the following corollary, which is stated without proof:

Corollary 3.3. *Suppose $P_0 \subseteq \mathbb{R}^2$, and K_0 is the point field of P_0. Let $x, y \in L$, where L is an extension of K_0, and $L \subseteq \mathbb{R}$. If there exists a finite series of subfields $K_0 \subseteq \ldots \subseteq K_n = L$ such that for all $i = 1, ..., n$, $[K_i : K_{i-1}] = 2$, then (x, y) is constructible from P_0.*

We can now use these facts to study for what values of n a regular n-gon is constructible.

4. Constructibility of Regular N-Gons

The proofs in this section will be less rigorous than those in the the first three sections, and we will use algebra to better understand the properties of constructible n-gons. In particular, note that:

(1) In the complex plane, the n^{th} roots of unity are the vertices of a regular n-gon of unit radius.
(2) Each n^{th} root of unity is a zero in $\mathbb{C}$ of the polynomial

$$t^n - 1 = (t - 1)(t^{n-1} + ... + t + 1)$$

First, we shall see how the constructibility of n-gons implies the constructibility of other m-gons.

Theorem 4.1. *(1) If a regular n-gon is constructible, and m divides n, then a regular m-gon is also constructible.*
(2) If regular n-gons and m-gons are constructible and $(m, n) = 1$, then a regular mn-gon is constructible.

Proof. (1) Suppose a regular n-gon is constructible, and $n = md$, where $d \in \mathbb{N}$. Drawing a line through every dth vertex will result in a regular m-gon.
(2) Suppose regular n-gons and m-gons are constructible, and $(m, n) = 1$. Then there exist $a, b \in \mathbb{Z}$ such that

$$am + bn = 1$$

which implies that

$$\frac{1}{mn} = a\frac{1}{n} + b\frac{1}{m}$$

From this equation and Theorem 3.1, we can, given angles of $\frac{2\pi}{m}$ and $\frac{2\pi}{n}$, construct an angle of $\frac{2\pi}{mn}$, and we can then replicate this angle mn times to create the vertices of an mn-gon.

□

Already, we have the following obvious corollary for constructing regular n-gons:

Corollary 4.2. *Let $n = p_1^{a_1} \cdots p_k^{a_k}$ where $p_1, \ldots, p_k$ are distinct primes. Then a regular n-gon is constructible if and only if, for each $p_i^{a_i}$, a regular $p_i^{a_i}$-gon is constructible.*

We will need three more lemmas before coming to the final theorem on constructible n-gons. These lemmas are proven mostly with algebraic manipulations of polynomials, so their proofs shall not be given here.

Lemma 4.3. *Suppose p is a prime number and a regular p^n-gon is constructible for some $n \in \mathbb{N}$. Let γ be a $(p^n)^{th}$ root of unity in $\mathbb{C}$. Then the degree of the minimum polynomial of γ over $\mathbb{Q}$ is a power of 2.*

Lemma 4.4. *Let p be prime and γ be a primitive p^{th} root of unity in $\mathbb{C}$. Then the minimum polynomial of γ over $\mathbb{Q}$ is*

$$f(x) = 1 + x + \ldots + x^{(p-1)}$$

Lemma 4.5. *Let p be prime and γ be a primitive $(p^2)^{th}$ root of unity in $\mathbb{C}$. Then the minimum polynomial of γ over $\mathbb{Q}$ is*

$$g(x) = 1 + x^p + \ldots + x^{p(p-1)}$$

Several facts should be obvious from these three lemmas. First of all, if a regular p-gon is constructible, then the minimum polynomial of a pth primitive root of unity in $\mathbb{C}$ will be

$$f(x) = 1 + x + \ldots + x^{(p-1)}$$

But then by Lemma 4.3, $p-1$ must be a power of 2. Therefore we already know that $p = 2^r + 1$. And from Lemma 4.2, it follows that if a regular n-gon is constructible, and $n = p_1^{a_1} \cdots p_k^{a_k}$ for distinct primes p_i, then each p_i must be of the form $2^s + 1$ for some $s \in \mathbb{N}$. The final theorem in this paper determines explicitly those values of n for which a regular n-gon is constructible:

Theorem 4.6. *A regular n-gon is constructible using rulers and compasses if and only if*

$$n = 2^r p_1 \cdots p_s$$

where r and s are non-negative integers, and each p_i is a distinct prime of the form $p = 2^{2^{s_i}} + 1$, a Fermat prime.

To prove that constructibility of the regular n-gon implies n is a product of distinct Fermat primes, we must simply combine the conclusions of the previous three lemmas. To prove the other direction of implication, however, requires knowledge of Galois groups, which is beyond the scope of this paper, so we will only prove the first statement.

Proof. Suppose a regular n-gon is constructible. Every positive integer n can be written as $n = 2^r {p_1}^{a_1} \cdots {p_s}^{a_s}$, where $p_1, \ldots, p_s$ are distinct odd primes. From

Lemma 4.2, it follows that for each p_i, a regular ${p_i}^{a_i}$-gon must constructible. Suppose that $a_i \geq 2$. Then ${p_i}^2$ divides ${p_i}^{a_i}$, so by Theorem 4.1, a regular ${p_i}^2$-gon must be constructible. However, by Lemma 4.5, the degree of the minimum polynomial over $\mathbb{Q}$ of the $({p_i}^2)^{th}$ root of unity is $p_i(p_i - 1)$. And by Lemma 4.3, this degree must be a power of 2. But p_i is odd, so $p_i(p_i - 1)$ cannot be a power of 2. This is a contradiction. Therefore, for all p_i we must have

$$a_i = 1.$$

So for each p_i, a regular p_i-gon is constructible. From Lemmas 4.3 and 4.4, it follows that $p_i - 1$ must be a power of 2, so we have

$$p_i - 1 = 2^{s_i}. \tag{4.7}$$

Suppose s_i has some odd divisor a greater than 1. Then equation 4.7 becomes

$$p_i = (2^b)^a + 1 = (2^b + 1)\left((2^b)^{(a-1)} + ... + (2^b) + 1\right). \tag{4.8}$$

But this implies that p_i is not prime, so s_i can only have even divisors, and must therefore be a power of 2. This gives us

$$s_i = 2^{r_i}.$$

Plugging this back into equation 4.7 now gives us

$$p_i = 2^{2^{r_i}} + 1.$$

Therefore if a regular n-gon is constructible, n must be the product of a power of 2 and distinct Fermat primes. □

References

[1] Ian Stewart. Galois Theory. Chapman and Hall Ltd. 1973.