CLASSICAL, CONIC, AND CONSTRAINED CONSTRUCTIONS

TAE KYU KIM

Abstract. We review the classical results on compass and straightedge constructions and use field theory to characterize the constructible points. Then, we introduce three conic drawing tools and prove that they are equivalent. We determine the set of conic constructible points. Finally, we prove that certain restrictions of our tools does not affect the set of points that we can construct. In particular, we discuss the Mohr-Mascheroni theorem and the Poncelet-Steiner theorem and prove that all conic constructible points can be drawn using only compass and straightedge constructions if we start with a non-degenerate conic different from a circle.

1. Introduction

We define straightedge and compass constructions in the following manner. Initially, we are given the points $(0, 0)$ and $(1, 0)$ on the Cartesian plane, and we can make the following moves:

- Draw a line through two already constructed points.
- Given two non-parallel lines, find their point of intersection.
- Given a point P and a segment of length a, draw the circle centered at P with radius a.
- Given a line and a circle, find their point(s) of intersection, if any.
- Given two circles, find their point(s) of intersection, if any.

Compass and straightedge constructions were first defined and attempted by Greek mathematicians. They were able to add, subtract, multiply, and divide lengths as well as take square roots. Moreover, they could bisect any arbitrary angle. However, the following three challenges remained unsolved:

(1) Given an angle, is it possible to trisect it?
(2) Given a segment of length 1, is it possible to construct a cube of volume 2?
(3) Given a circle, is it possible to construct a square with the same area?

Now, with the techniques of field theory and field extensions, these challenges can be shown to be impossible (See Corollary 2.4). In fact, we have a nice characterization for when a length is constructible from a compass and straightedge (See Theorems 2.3 and 2.6).

We can ask about which lengths are constructible given different sets of tools. The following sets of tools will be addressed in this paper.

(1) What if we only have a compass? or only a straightedge?
(2) What if we can draw ellipses? or parabolas? or a general conic?
(3) What if we are given a single conic and can only draw using a compass and a straightedge?

Date: June 8, 2020.

The rest of the paper is organized as follows. Section 2 gives an overview of the process used to determine the classical constructible numbers (only using a compass and straightedge). Section 3 answers which lengths can be constructed if we have a conic drawing tool. Section 4 gives some ways in which we can restrict our usage of the compass and/or straightedge but still construct the classically constructible lengths. Section 5 answers which points can be obtained with compass and straightedge constructions if we are initially given a fixed, non-degenerate conic.

If the reader took Simon's class on Abstract Algebra, then we suggest skipping much of Section 2, but taking the time to read Definition 2.5 and Theorem 2.6 as those were not covered in the class.

This paper assumes a basic understanding of field theory and field extensions.

2. Constructible Numbers

In this section, we will find a criterion for constructible points in terms of towers of field extensions of degree 2. This allows us to answer the three Greek challenges and identity the constructible regular n-gons. Then, we introduce the notion of a normal closure of a field extension and use it to find a criterion for constructible points only dependent on the degree of the normal closure over $\mathbb{Q}$.

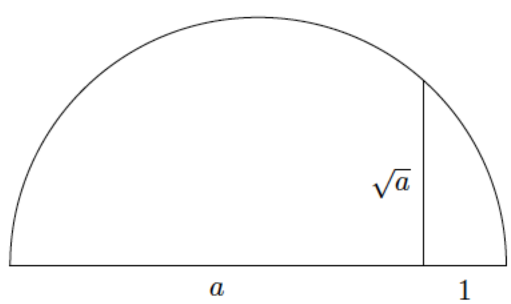
The Greeks showed that given segments of lengths a and b, segments of lengths $a+b$, $a-b$, a/b (assuming $b \neq 0$), ab, $\sqrt{a}$ can be constructed. They also figured out how to bisect any given angle. These constructions give us field operations on the set of constructible lengths. Thus, we can define the fields of constructible numbers and constructible points.

Definition 2.1. A real number r is a *constructible number* if a line segment of length $|r|$ can be constructed with a compass and a straightedge in a finite number of steps. These numbers form a subfield of $\mathbb{R}$.

Definition 2.2. A point $P = (x, y)$ is a *constructible point* if it can be constructed in a finite number of straightedge and compass construction steps. Equivalently, P is constructible if x and y are constructible numbers. We identity this point as the complex number $x + iy$ and recognize the field of constructible points as a subfield of $\mathbb{C}$.

It turns out that all three Greek challenges are impossible using straightedge and compass constructions. We will show this in the rest of this section using field theory and the fact that π is transcendental.

First we should check that the field of constructible points is the smallest subfield of $\mathbb{C}$ that is closed under square roots. We leave it as an exercise to prove that we can find $\sqrt{a}$ given a using the following diagram.



Thus, we know how to take square roots, so we now need to check that every possible operation only requires field operations and taking square roots of already constructed lengths.

A line through two constructed points is expressible as $y = ax + b$ where a, b are constructible numbers. A circle with a constructed center and radius is expressible as $(x - x_0)^2 + (y - y_0)^2 = r^2$ where x_0, y_0, r are all constructible numbers. Finding the intersection of any two lines, line and circle, or two circles requires solving a polynomial of degree at most 2 in either of the coordinates. Thus, the coordinate will be expressible in terms of square roots of constructible numbers. Hence, the field of constructible points is the smallest subfield of $\mathbb{C}$ that is closed under square roots.

This proves the following theorem.

Theorem 2.3. *A point $z \in \mathbb{C}$ is constructible if and only if there is a sequence of fields $F_0 = \mathbb{Q} \subseteq F_1 \subseteq \cdots \subseteq F_n \subseteq \mathbb{C}$ with $[F_i : F_{i-1}] = 2$ for all i, and $z \in F_n$.*

We will refer to a tower of fields with consecutive degrees 2 as a (2)-*tower*. This notion will be more useful when we look conics in Section 3. If we allow the consecutive fields to have degree 2 or 3, then we call the tower of fields a $(2, 3)$-*tower*.

Corollary 2.4. *The three Greek problems are impossible.*

Proof. Trisecting an angle is equivalent to producing $\cos\theta$ given $\cos 3\theta$, or solving for $\cos\theta$ in

$$\cos 3\theta = 4\cos^3\theta - 3\cos\theta.$$

Setting $\theta = 60°$, we can show that the resulting polynomial in $\cos\theta$ is irreducible over $\mathbb{Q}$. This implies that $[\mathbb{Q}(\cos 20°) : \mathbb{Q}] = 3$. Thus, Theorem 2.3 implies that $\cos 20°$ is not constructible, and we cannot trisect a $60°$ angle.

Duplicating the cube is equivalent to producing a length of $\sqrt[3]{2}$, which has the minimal polynomial $x^3 - 2$. Thus, $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$. So $\sqrt[3]{2}$ is not constructible.

Squaring the circle requires producing $\sqrt{\pi}$, or equivalently π. However, π is transcendental, so $\mathbb{Q}(\pi)$ would have infinite degree over $\mathbb{Q}$. Thus, π is not constructible. ■

It follows from Theorem 2.3 that if z is a constructible number, then $\mathbb{Q}(z)$ has degree 2^n over $\mathbb{Q}$ for some $n \in \mathbb{N}$. The converse turns out to be false; there exists a $z \in \mathbb{C}$ such that $[\mathbb{Q}(z) : \mathbb{Q}] = 4$ but z is not constructible. To understand why, we need to introduce the concept of a normal closure.

Definition 2.5. If K is a field and L is an algebraic extension of K, then there is some algebraic extension M/L such that M is a normal extension of K. There is only one such extension that is minimal, up to isomorphism. This extension is called the *normal closure* of the extension L/K.

Example. The field extension $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is not normal since the minimal polynomial of $\sqrt[4]{2}$, $x^4 - 2$, has non-real zeros. The normal closure of $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is $\mathbb{Q}(i, \sqrt[4]{2})/\mathbb{Q}$, obtained by adjoining all of the zeros of $x^4 - 2$.

We are now ready to state and prove a stronger criterion for constructible numbers.

Theorem 2.6. *A complex number z is constructible if and only if z is algebraic over $\mathbb{Q}$ and the normal closure $K/\mathbb{Q}$ of $\mathbb{Q}(z)/\mathbb{Q}$ has dimension 2^n over $\mathbb{Q}$ where $n \in \mathbb{Z}_{\geq 0}$.*

Proof. Suppose that z is constructible. By Theorem 2.3, it is contained in an (2)-tower $\mathbb{Q}(\alpha_1, \dots, \alpha_l)$. We may assume that this (2)-tower is a Galois extension over $\mathbb{Q}$: the argument that we used to show that a radical extension can be turned into a Galois extension holds (See Chapter 9, Section 4 of the Abstract Algebra book). Now, the normal closure $K/\mathbb{Q}$ of $\mathbb{Q}(z)/\mathbb{Q}$ is a subfield of $\mathbb{Q}(\alpha_1, \dots, \alpha_l)$, which has dimension 2^n over $\mathbb{Q}$. Thus the normal closure has order dividing 2^n, so it has order 2^s for some $s \in \mathbb{N}$.

Conversely, suppose the normal closure $K/\mathbb{Q}$ of $\mathbb{Q}(z)/\mathbb{Q}$ has dimension 2^s over $\mathbb{Q}$. The Galois group $G = \text{Gal}(K/\mathbb{Q})$ has order 2^s. By group theory, G is solvable and G has a decomposition series

$$G = G_1 \rhd G_2 \rhd G_3 \rhd \cdots \rhd G_k = \{e\}$$

such that G_i/G_{i+1} is of order 2. By the Galois correspondence, we obtain a sequence of subfields of K such that

$$\mathbb{Q} = F_1 \subset F_2 \subset F_3 \subset \cdots \subset F_k = K$$

where F_i is the subfield of K fixed by G_i for $1 \leq i \leq k$. In particular, $[F_{i+1} : F_i] = 2$. Write $F_{i+1} = F_i(\alpha_i)$ for some element $\alpha_i \in F_{i+1}$.

Any degree 2 extension is obtained by adjoining a square root: any $\alpha \in K_{i+1} \setminus K_i$ has a minimal polynomial of degree 2 over K_i; adjoin the square root of the discriminant to K_i to obtain K_{i+1}. Hence, z belongs to a (2)-tower over $\mathbb{Q}$ and is constructible by Theorem 2.3. ∎

We provide an example that demonstrates the utility of Theorem 2.6. Consider the polynomial $p(x) = x^4 - x - 1$, which is irreducible over $\mathbb{Q}$. Thus, if $z \in \mathbb{C}$ is a zero of $p(x)$, then $[\mathbb{Q}(z) : \mathbb{Q}] = 4$. Let $F/\mathbb{Q}$ be the normal closure of $\mathbb{Q}(z)/\mathbb{Q}$. The cubic resolvent of $p(x)$ is a cubic polynomial irreducible over $\mathbb{Q}$, and its roots are elements of F. This implies that 3 divides $[F : \mathbb{Q}]$. Thus, z is not constructible even though $[\mathbb{Q}(z) : \mathbb{Q}] = 4$.

Another problem of interest for mathematicians was determining which regular n-gons are constructible. Remember that a regular n-gon is a regular polygon with n sides. The Greeks knew how to construct regular 3, 4, 5-gons as well as construct a regular $2n$-gon given a regular n-gon. In 1796, Carl Fredrich Gauss proved the constructibility of the regular 17-gon and, five years later, found a sufficient condition for a regular n-gon being constructible. He claimed that the condition was also necessary, but he did not give a proof of it. A proof of the condition's necessity was given by Pierre Wantzel in 1837. They proved the following theorem:

Theorem 2.7 (Gauss-Wantzel). *A regular n-gon can be constructed with compass and straightedge if and only if n is the product of a power of 2 and any number of distinct Fermat primes (including none).*

A Fermat prime is a prime in the form $2^{2^k} + 1$. It turns out that if $2^k + 1$ is a prime, then k must be a power of 2. Assuming this fact, we can use Theorem 2.6 to prove Theorem 2.7.

Proof. An exterior angle of a regular n-gon is $\frac{2\pi}{n}$ radians. Thus, if we can construct a $\frac{2\pi}{n}$ radian angle, we can construct consecutive sides of the regular n-gon and eventually construct the entire polygon. Conversely, given a regular n-gon, we can extend one of the sides to find a $\frac{2\pi}{n}$ radian angle. We can construct a $\frac{2\pi}{n}$ angle if and only if we can construct the point $(\cos \frac{2\pi}{n}, \sin \frac{2\pi}{n})$. This point is represented by the nth of unity $\zeta_n = e^{2\pi i/n}$. Thus, a regular n-gon is constructible if and only if ζ_n is constructible.

Note that the normal closure of $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is just $\mathbb{Q}(\zeta_n)/\mathbb{Q}$, so it has degree $\phi(n)$ over $\mathbb{Q}$ where $\phi(n)$ is the Euler Totient function. Writing the prime factorization of n as $2^{e_1}p_2^{e_2}\cdots p_k^{e_k}$, we obtain $\phi(n) = 2^{e_1-1}p_2^{e_2-1}(p_2-1)\cdots p_k^{e_k-1}(p_k-1)$. Theorem 2.6 tells us that ζ_n is constructible if and only if $\phi(n)$ is a power of 2. Thus, $e_i = 1$ for $i > 1$, and $p_i - 1$ must be a power of 2 for $i > 1$. It is well-known that if $2^k + 1$ is prime, then must have k be a power of 2. Thus, for $i > 1$, p_i must be a Fermat prime. ■

3. Drawing with Conics

Definition 3.1. A point is *conic constructible* (or is said to have *solid construction*) if it can be constructed using a straightedge, compass, and a hypothetical tool that can draw any conic with an already constructed focus, directrix, and eccentricity.

Example. Consider the conic given by the focus $(0, 1/4)$, the directrix $y = -1/4$, and eccentricity $e = 1$. A point (x, y) is on this conic if the distance to the focus is equal to the distance to the directrix. Using the distance formula gives

$$x^2 + \left(y - \frac{1}{4}\right)^2 = \left(y + \frac{1}{4}\right)^2.$$

Some algebraic manipulation gives an equation for a parabola:

$$\begin{aligned}
x^2 &= \left(y + \frac{1}{4}\right)^2 - \left(y - \frac{1}{4}\right)^2, \\
x^2 &= \left(y + \frac{1}{4} + y - \frac{1}{4}\right)\left(y + \frac{1}{4} - y + \frac{1}{4}\right), \\
x^2 &= 2y \cdot \frac{1}{2}, \\
x^2 &= y.
\end{aligned}$$

In the following two propositions, we discuss two alternative methods of drawing conics that turn out to be equivalent to the method using the directrix, focus, and eccentricity. To simplify notation, let K be a subfield of $\mathbb{R}$ such that every positive number $x \in K$ has a square root in K. A line passing through two points of K^2 is called a *line in* K. A circle is called a *circle in* K if its center is in K^2 and it passes through a point in K^2. Similarly, a conic is called a *conic in* K if its foci are in K^2, its directrix line is in K, and its eccentricity is in K. Note that our conic drawing tool allows us to draw conics in K.

We can define an ellipse by its two foci F_1, F_2 and its semi-major axis a. The ellipse is the set of points P for which the $F_1P + F_2P = 2a$. The next proposition shows that this construction method for ellipses is equivalent to our conic drawing tool for eccentricity $e < 1$. For the next proposition, the variables a, b, c represent the semi-major axis, semi-minor axis, and half of the distance between the two foci of an ellipse.

Proposition 3.2. *Let E be a non-degenerate ellipse. Then E is in K if and only if its foci are in K^2 and its semi-major axis is in K.*

Proof. Suppose that an ellipse E is in K. We know that the directrix l, the focus F, and the eccentricity $e < 1$ are in K. We can find the perpendicular distance m between the directrix and the focus using a compass and straightedge, so $m \in K$. The original ellipse is congruent to the ellipse given by directrix $x = 0$, focus $(m, 0)$, and eccentricity e. Thus, it suffices to prove that a, b, c of the new ellipse are in K. Using the distance formula gives

$$\begin{aligned}
\frac{\sqrt{(x-m)^2+y^2}}{x} &= e, \\
(x-m)^2+y^2 &= e^2x^2, \\
(1-e^2)x^2-2mx+m^2+y^2 &= 0, \\
(1-e^2)x^2-2mx+\frac{m^2}{1-e^2}+y^2 &= \frac{m^2}{1-e^2}-m^2, \\
(1-e^2)\left(x-\frac{m}{1-e^2}\right)^2+y^2 &= \frac{m^2e^2}{1-e^2}, \\
\frac{\left(x-\frac{m}{1-e^2}\right)^2}{\left(\frac{me}{1-e^2}\right)^2}+\frac{y^2}{\left(\frac{me}{\sqrt{1-e^2}}\right)^2} &= 1.
\end{aligned}$$

Thus, $a = \frac{me}{1-e^2}$, $b = \frac{me}{\sqrt{1-e^2}}$, and $c = \sqrt{a^2-b^2}$. This proves the forward direction.

For the other direction, we can find c because it is half of the distance between the two foci. Moreover, $b = \sqrt{a^2-c^2}$ so b and c are in K. Then we explicitly solve for e in terms of a and b: $e = \sqrt{1-\frac{b^2}{a^2}}$. This allows us to also solve for m using the above formulas. We can find the slope of the directrix as it is perpendicular to the line connecting the two foci. Finally, we can find a single point on the directrix by noticing that the point of intersection of the directrix and the line connecting the foci is a distance $(a-c)\frac{e+1}{e}$ away from one of the foci. Thus, the directrix is in K, both foci are in K, and e is in K, so the ellipse is in K as desired. ■

Another way that we can define a conic is by the equation

$$ax^2+bxy+cy^2+dx+ey+f = 0. \tag{3.1}$$

where a, b, c, d, e, and f are in K. This construction is equivalent to the directrix, focus, eccentricity construction. First, we need a lemma and proposition to help us rotate conics so that the bxy term vanishes.

Lemma 3.3. *Let E be a conic in K. If $\cos\theta \in K$, then the conic obtained by rotating E about the origin by an angle θ can be expressed by an equation with coefficients in K.*

Proof. Let θ be the angle of rotation. The transformed conic is described by the equation $a'x^2+b'xy+c'y^2+d'x+e'y+f' = 0$ where the coefficients are:

$$\begin{aligned}
a' &= a\cos^2(\theta)+b\sin(\theta)\cos(\theta)+c\sin^2(\theta), \\
b' &= 2(c-a)\sin(\theta)\cos(\theta)+b(\cos^2(\theta)-\sin^2(\theta)), \\
c' &= a\sin^2(\theta)-b\sin(\theta)\cos(\theta)+c\cos^2(\theta), \\
d' &= d\cos(\theta)+e\sin(\theta), \\
e' &= e\cos(\theta)-d\sin(\theta), \\
f' &= f,
\end{aligned}$$

which all only require field operations on the elements of K. Therefore, the coefficients of the transformed conic are in K. ■

Proposition 3.4. *Let E be a non-degenerate conic (different from a circle), defined by the equation $ax^2 + bxy + cy^2 + dx + ey + f = 0$. Then E is in K if and only if the coefficients a, b, c, d, e, and f are in K. If E is in K, then it can also be rotated and translated into a conic in K in standard position i.e. it can be expressed by the equation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ for $a, b \in K$.*

Proof. The forward direction involves explicitly writing out the equation for the conic defined by the directrix, focus, and eccentricity. Then, one can collect like-term to show that the coefficients lie in K. We leave this as an exercise to the reader.

For the converse, let the conic E be expressed by the equation $ax^2+bxy+cy^2+dx+ey+f = 0$. Let θ denote the angle of rotation that is required to vanish the bxy term. For $b = 0$, θ is 0 or $\frac{\pi}{2}$. Thus assume that $b \neq 0$. In this case, $\cot(2\theta) = \frac{a-c}{b} \in K$. Moreover, $\sin(2\theta) = \frac{1}{\sqrt{1+\cot^2(2\theta)}} \in K$. Then, from $\cos(2\theta) = \sin(2\theta)\cot(2\theta) \in K$ and the half-angle formulas $\cos(\theta) = \pm\sqrt{\frac{1+\cos(2\theta)}{2}}$ and $\sin(\theta) = \pm\sqrt{\frac{1-\cos(2\theta)}{2}}$, we find that $\sin(\theta)$ and $\cos(\theta)$ are in K. Thus, Lemma 3.3 enables us to rotate E by θ to get a conic expressed by the form

$$\frac{(x-x_0)^2}{(a')^2} + \frac{(y-y_0)^2}{(b')^2} = 1.$$

Translation by $(-x_0, -y_0)$ gives the conic in standard form.

From the coefficients of the standard form equation, we can find its directrix, its focus, and its eccentricity. Then, we can translate the focus and the points on the directrix by (x_0, y_0) and apply a rotation of $-\theta$ about the origin to obtain the directrix and the focus of E. As the directrix line and the eccentricity are in K and the focus is in K^2, E is in K. ■

Note that by choosing some of the constants in Equation 3.1 to be 0, we can reduce the equation of a conic to that of a line or a circle. Thus, when we are considering the possible intersections of lines, circles, and conics, we only need to look at the intersection of two conics.

This problem reduces to solving two conic equations simultaneously, which turns out to be equivalent to solving a quartic polynomial in x or y.

Proposition 3.5. *The intersection points of two conics in K have x, y-coordinates that are roots of a quartic equation with coefficients in K.*

Proof. We provide a sketch of the proof. Assume that the coefficients of the y^2 terms are nonzero. By multiplying the equations for the conics by an appropriate factor, we can force the coefficient of the y^2 term to be 1 for both conics. Subtract the equations to obtain an equation linear in y. Solve for y in terms of x: we get a rational function in x—degree 2 in the numerator and degree 1 in the denominator. Now, this expression can be substituted back into one of the original equations to obtain a quartic equation in x.

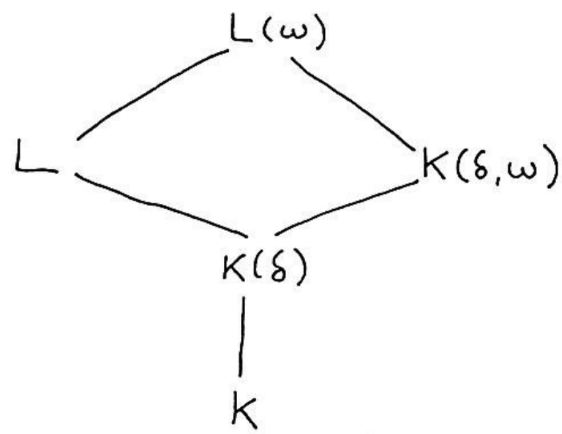
The case when one or both of the coefficients of the y^2 terms are zero can be treated in a similar manner by skipping the first few steps of the above algorithm. ■

We have already seen how to solve quartic polynomials; roots of degree 4 polynomials can always be expressed in terms of square roots and cube roots of the coefficients. However, we describe the proof in an alternative way that enables us to decompose a degree 3 or 4 field extension into a tower of field extensions so that at each step, a single cube root or square root is adjoined.

Theorem 3.6. *The roots of any cubic or quartic polynomial can be expressed in terms of square roots and cube roots of the coefficients.*

Proof. Let K be a field of characteristic 0, and let $f = x^3 + a_2x^2 + a_1x + a_0$ be an irreducible cubic over K. Let $y = x = \frac{1}{3}a_0$ so that $f = y^3 + py + q$ where $p = q_1 - \frac{1}{3}a_2^2$ and $q = a_0 + \frac{2}{27}a_2^3 - \frac{1}{3}a_2a_1$. Let L be the splitting field of $x^3 + px + q$ over K. There is an element $\delta \in L$ such that $\delta^2 = \Delta = 4p^3 - 27q^2$, the discriminant of $x^3 = px + q$ (This was an exercise in Chapter 7 of the Abstract Algebra book). Let $\omega \neq 1$ be a cube root of unity. Its minimal polynomial is $x^2 + x + 1$ over K, unless $\omega \in K$.

We consider the following diagram of fields:



In the field $L(\omega)$, set $\beta = \alpha_1 + \omega\alpha_2 + \omega^2\alpha_3$ and $\gamma = \alpha_1 + \omega^2\alpha_2 + \omega\alpha_3$ where $\alpha_1, \alpha_2, \alpha_3$ are the three roots of $x^3 + px + q = 0$.

We can be compute that β^3 and γ^3 are the elements $\frac{27}{2}q \pm \frac{3}{2}(2\omega + 1)\delta$. Thus, β^3 and γ^3 belong to $K(\delta, \omega)$. The field extension $K(\delta, \omega)/K$ has degree at most 4. If it has degree 4, then it has a subfield $K(\delta)$ of degree 2 over K. Notice that $\gamma = -3p/\beta$, and that $\alpha_1 = \frac{1}{3}(\beta + \alpha)$, $\alpha_2 = \frac{1}{3}(\omega^2\beta + \omega\gamma)$, and $\alpha_3 = \frac{1}{3}(\omega\beta + \omega^2\gamma)$. Hence, $L(\omega) = K(\delta, \omega, \beta)$, and $L(\omega) = K(\delta, \omega)(\beta)$ with $\beta^3 = K(\delta, \omega)$.

Finally, we have shown that if α is a root of $x^3 + px + q = 0$, then there exist fields $K \subset K(\delta) \subset K(\delta, \omega) \subset K(\delta, \omega, \beta)$ such that $\gamma^2 \in K$, $\omega^3 \in K(\delta)$, $\beta^3 \in K(\delta, \omega)$, and $\alpha \in K(\delta, \omega, \beta)$. Thus, at each step we need to adjoin a square root or cube root. This shows that α can be expressed in terms of square roots and cube roots of elements in K.

The case for quartic polynomials is treated in a similar way. We look at the quartic formula that we have derived before, at look at which square roots or cube roots we need at each step. Then, we can construct the subfields one at a time to obtain a sequence of fields $K \subset K(\alpha_1) \subset K(\alpha_1, \alpha_2) \subset K(\alpha_1, \alpha_2, \alpha_3) \subset K(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ where $\alpha_1^n \in K$ and $\alpha_i^n \in K(\alpha_1, \ldots, \alpha_{i-1})$ for $i = 2, 3, 4$ and $n \in \{2, 3\}$; the roots of the quartic equation will belong to $K(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$. This shows that the roots can be expressed as square and cube roots of elements in K. ■

In reality, we would first derive the cubic or quartic formula in a more primitive manner, then translate that into a proof about forming a tower of field extensions only involving adjoining square roots and cube roots.

Now we know that the field of conic constructible numbers has to be closed under square roots and possibly cube roots. We don't know quite yet if we require closure under cube roots as it could turn out that every quartic polynomial we want to solve only requires taking square roots (we don't have full control of the coefficients of the quartic polynomial). Proving that we require cube roots requires explicit construction. The following theorems give us some different ways to construct the cube root.

The following construction is due to Aliska Gibbins and Lawrence Smolinsky [GS09].

Theorem 3.7. *If a is conic constructible, then $\sqrt[3]{a}$ can be constructed using two ellipses.*

Proof. Consider the following equations of ellipses with coefficients in F:

$$2x^2 + y^2 - ax + 2\sqrt{2}y + 1 = 0$$
$$3x^2 + y^2 - ax + (1 + 2\sqrt{2})y + \sqrt{2} = 0.$$

It can be shown that the x-coordinates of the points of intersection satisfy

$$x^4 - ax = 0.$$

The real roots are 0 and $\sqrt[3]{a}$. ■

The following construction is given in [Vid97], and the author accredits it to Menachmus from 350 B.C.

Theorem 3.8. *If a is conic constructible, then $\sqrt[3]{a}$ can be constructed using two parabolas.*

Proof. Draw a parabola P_1 with focus $(0, \frac{1}{4})$ and directrix $y = -\frac{1}{4}$. Its equation is $y = x^2$. The parabola P_2 has its focus at $(\frac{a}{4}, 0)$ and its directrix at $x = -\frac{a}{4}$. Its equation is $x = \frac{y^2}{a}$. Thus, the point of intersection is obtained by solving

$$y = x^2 = \left(\frac{y^2}{a}\right)^2,$$

so

$$y(y^3 - a^2) = 0.$$

The real roots are $y = 0$ and $\sqrt[3]{a^2}$. The x-coordinates are 0 and $\sqrt[3]{a}$, respectively. ■

Theorems 3.7 and 3.8 actually tell us something stronger about conic constructible numbers: we only need to be able to draw ellipses (or parabolas) to construct all conic constructible numbers.

Let us complete this section by describing the field of conic constructible numbers using field theory.

Theorem 3.9. *Let $z \in \mathbb{C}$. Then z is conic constructible if and only if z is contained in a subfield of $\mathbb{C}$ of the form $\mathbb{Q}(\alpha_1, \alpha_2, \ldots, \alpha_l)$ where $\alpha_1^n \in \mathbb{Q}$ and $\alpha_i^n \in \mathbb{Q}(\alpha_1, \alpha_2, \ldots, \alpha_{i-1})$ for $2 \le i \le l - 1$ and $n \in \{2, 3\}$ (n can be different for each α_i). We will call such a field a $(2,3)$-tower over $\mathbb{Q}$.*

Proof. This is automatic by the fact that conic constructible numbers form the smallest subfield of $\mathbb{C}$ closed under square roots and cube roots. ■

The following theorem due to Carlos R. Videla [Vid97] characterizes conic constructible numbers.

Theorem 3.10. *A complex number z is conic constructible if and only if z is algebraic over $\mathbb{Q}$ and the normal closure $K/\mathbb{Q}$ of $\mathbb{Q}(z)/\mathbb{Q}$ has dimension 2^n3^m over $\mathbb{Q}$ where $m, n \in \mathbb{Z}_{\geq 0}$.*

Proof. Suppose that z is conic constructible. By Theorem 3.9, it is contained in an $(2,3)$-tower $\mathbb{Q}(\alpha_1, \dots, \alpha_l)$. We may assume that this $(2,3)$-tower is a Galois extension over $\mathbb{Q}$: the same proof that we used to show that a radical extension can be made into a Galois extension holds here (See Chapter 9, Section 4 of the Abstract Algebra book). Now, the normal closure $K/\mathbb{Q}$ of $\mathbb{Q}(z)/\mathbb{Q}$ is a subfield of $\mathbb{Q}(\alpha_1, \dots, \alpha_l)$, which has dimension 2^n3^m over $\mathbb{Q}$. Thus, $[K : \mathbb{Q}]$ divides $[\mathbb{Q}(\alpha_1, \dots, \alpha_l) : \mathbb{Q}] = 2^n3^m$, so $[K : \mathbb{Q}] = 2^s3^t$ for some $s, t \in \mathbb{Z}_{\geq 0}$.

Conversely, suppose the normal closure $K/\mathbb{Q}$ of $\mathbb{Q}(z)/\mathbb{Q}$ has dimension 2^s3^t over $\mathbb{Q}$. The Galois group $G = \mathrm{Gal}(K/\mathbb{Q})$ has order 2^s3^t. By Burnside's p–q theorem, G is solvable. Thus, G has a decomposition series

$$G = G_1 \rhd G_2 \rhd G_3 \rhd \cdots \rhd G_k = \{e\}$$

such that G_i/G_{i+1} is of order 2 or 3. By the Galois correspondence, we obtain a sequence of subfields of K such that

$$\mathbb{Q} = F_1 \subset F_2 \subset F_3 \subset \cdots \subset F_k = K$$

where F_i is the subfield of K fixed by G_i for $1 \leq i \leq k$. In particular, $[F_{i+1} : F_i] = 2$ or 3.

Any degree 2 extension is obtained by adjoining a square root: any $\alpha \in K_{i+1} \setminus K_i$ has a minimal polynomial of degree 2; adjoin the square root of the discriminant to K_i to obtain K_{i+1}. The extension F_{i+1}/F_i may be of degree 3 but may not be obtained by adjoining a cube root. In this case, we can use the field decomposition as found in Theorem 3.6 to replace the extension $F_i \subset F_{i+1}$ with the sequence $F_i \subset L_i' \subset L_i'' \subset L_i'''$ with $\alpha_i \in L_i'''$ such that the sequence forms a $(2,3)$-tower. Hence, z belongs to a $(2,3)$-tower over $\mathbb{Q}$ and so is constructible by Theorem 3.9. ■

With this theorem, we can characterize the conic constructible regular n-gons, again due to Videla [Vid97].

Theorem 3.11. *The regular n-gon is conic constructible if and only if $n = 2^s3^tp_3p_4 \dots p_k$ with $s, t \geq 0$ where p_i are distinct Pierpont primes: primes in the form $2^u3^v + 1$ for integers $u, v \geq 0$.*

Proof. Constructing a regular n-gon is equivalent to constructing $\zeta_n = e^{2\pi i/n}$. Let F_n be the cyclotomic field of nth roots of unity. $F_n/\mathbb{Q}$ is a Galois extension and has degree $\phi(n)$. If $n = 2^s3^tp_3^{e_3} \cdots p_k^{e_k}$ with $s, t \geq 1$, then $\phi(n) = 2^s3^{t-1}p_3^{e_3-1}(p_3 - 1) \cdots p_k^{e_k-1}(p_k - 1)$. A similar formula is obtained for $s = 0$ or $t = 0$. Theorem 3.10 tells us that ζ_n is conic constructible if and only if $F_n/\mathbb{Q}$ (which is the normal closure of $\mathbb{Q}(\zeta_n)/\mathbb{Q}$) has dimension 2^a3^b over $\mathbb{Q}$, which is just $\phi(n)$. Thus, we require that for $i \geq 3$, $e_i = 1$ and p_i is in the form $2^u3^v + 1$ for integers $u, v \geq 0$. ■

4. Restricted Construction

What kinds of points can we construct if we are only allowed to use a compass or only a straightedge? In this section, we apply restrictions to our straightedge and compass and prove that we can still construct classically constructible points with the restricted tool set.

Theorem 4.1 (Mohr–Mascheroni)**.** *Any point constructible by a compass and ruler can be constructed with just a compass.*

Proof. The proof is quite lengthy, so we leave the avid reader to read the following course handout from the Illinois Math and Science Academy (link). ■

What points can we construct using only a straightedge? Well, we can only form linear equations using a straightedge, so we can never obtain any square roots. Thus, there is no hope for getting constructible numbers from just a straightedge. However, it turns out that once we have a single circle and its center, we can get all constructible numbers using only a straightedge.

Theorem 4.2 (Poncelet–Steiner). *Any line or point constructible by a compass and ruler can be constructed with just a ruler, provided that a single circle and its center are given.*

Proof. While we provide an outline of the proof here, for the complete proof, we direct readers to the following course handout from the Illinois Math and Science Academy (link).

We prove the following results in order.

(1) Given a line $\overleftrightarrow{AB}$, the midpoint C of the segment $\overline{AB}$ and a point P, we can draw the unique line that passes through P and is parallel to $\overleftrightarrow{AB}$. This requires the usage of Ceva's Theorem.
(2) We can use the given circle to construct three equidistant points on any given line.
(3) It directly follows that given a line $\overleftrightarrow{AB}$ and a point P, we can draw the unique line that passes through P and is parallel to $\overleftrightarrow{AB}$.
(4) We can translate line segments i.e., given a segment $\overline{AB}$ and point P, we can construct a point Q such that $\overleftrightarrow{AB}$ is parallel to $\overleftrightarrow{PQ}$ and the lengths of the segments $\overline{AB}$ and $\overline{PQ}$ are equal. This follows from our ability to draw the parallelogram defined by the points A, B, and P to obtain the fourth vertex Q.
(5) Given a line $\overleftrightarrow{AB}$ and a point P, we can construct a line perpendicular to $\overleftrightarrow{AB}$ that passes through P. We use the fact that an inscribed angle that subtends half of a circle must be a right angle. We can force one of the rays of the angle to be parallel to $\overleftrightarrow{AB}$, which makes the other ray perpendicular to $\overleftrightarrow{AB}$. We can now translate this segment to the point P.
(6) Given a line $\overleftrightarrow{AB}$ and two points P and Q, we can construct a point R on the ray $\overrightarrow{PQ}$ (pointing out from P) such that the length of the segments $\overline{PR}$ and $\overline{AB}$ are equal. The proof utilizes the fact that any two radii of a circle are of the some length. Similar triangles obtained by three sets of parallel lines gives us our desired point R.
(7) Given segments of lengths a, b, and s, we can construct a segment of length $\frac{a}{b}s$. For an arbitrary pair of lines intersecting at point P, we can construct points A, B, and S on the lines so that their distances to P are a, b, and s, respectively. Drawing the line through A parallel to line $\overleftrightarrow{BS}$ and noting similar triangles gives the result.
(8) Given a segment of length a, we can construct a segment of length $\sqrt{a}$. With appropriate re-scaling, this can be obtained using a similar diagram as covered in the Abstract Algebra class.

The last two results prove that we can construct all classically constructible numbers. ■

5. Restricted Conic Constructions

What if we are initially given a fixed conic, but we can only make straightedge and compass constructions? We cannot draw new conics, but we may be able to still take cube roots of constructed numbers, thereby getting all conic constructible numbers.

A discovery in 2013 shows that this is true [BCJ+13].

Theorem 5.1. *Any point constructible from conics can be constructed using a ruler and a compass, together with a single fixed nondegenerate conic different from a circle.*

We must first define two new notions of conic constructible points:

Definition 5.2. Let C be a non-degenerate conic in the field of constructible numbers. Suppose that instead of starting of with just $(0,0)$ and $(1,0)$, we start with a set of points $P \subset \mathbb{C}$. We say that a point is *C-constructible from P* if it is obtained from the same process as conic-constructible points except that the conics are confined to the fixed one, C.

Definition 5.3. We call an ellipse or a hyperbola of eccentricity $e > 0$ *regular* if it is given by the equation

$$(1-e^2)(x-a)^2 + (y-b)^2 = \lambda^2$$

for some $a, b, \lambda \in \mathbb{R}$. A parabola is *regular* if it is of the form

$$x = \lambda(y-a)^2 + b$$

for some $a, b, \lambda \in \mathbb{R}$. Essentially, conics are regular if their directrix is parallel to the coordinate axes.

Definition 5.4. Let $e > 0$ be a constructible number. A point is *e-constructible from P* if it is obtained by the same process as conic constructible points except that:

(1) the conics are confined to the the regular ones of eccentricity e, and
(2) the intersection of two conics, neither of which is a circle, are not adjoined.

When $P = \{0, 1\}$, we say that a point is *C-constructible* and *e-constructible*, respectively. We have the following lemma:

Lemma 5.5. *Every conic-constructible point is e-constructible for every constructible number $e > 0$.*

Proof. Let $z \in \mathbb{C}$ be a conic-constructible point. From Theorem 3.9, z lies in a $(2,3)$-tower $\mathbb{Q}(\alpha_1, \ldots, \alpha_n)$ where either α_i^2 or α_i^3 is contained in $\mathbb{Q}(\alpha_1, \ldots, \alpha_{i-1})$ for each i. Let $e > 0$ be a fixed constructible number. We will show that z is e-constructible by induction on n. The base case is $n = 0$, but it is clear that every number in $\mathbb{Q}$ is e-constructible as it is classically constructible. Now suppose that all the points in $H = \mathbb{Q}(\alpha_1, \ldots, \alpha_{n-1})$ is e-constructible. if $\alpha_n^2 \in H$, then α_n is constructible from H from just using compass and straightedge constructions.

Now suppose that $\alpha_n^3 = re^{i\theta} \in H$. Let $q = \cos\theta$ and let K be the field of constructible numbers derived from $0, 1, r$, and q. Our strategy is to find curves over K such that their intersections give us $\sqrt[3]{r}$ and $\cos(\theta/3)$ from which we can construct α_n.

Consider the intersection of the following circle and conic:

$$x^2 + y^2 - rx - y = 0 \quad \text{and} \quad (1-e^2)x^2 + y^2 - rx - (1-e^2)y = 0.$$

Subtracting the second equation from the first gives $x^2 = y$. Substituting y with x^2 in the first equation then gives $x^4 - rx = 0$. Thus, the x-coordinate of the intersection points other than the origin corresponds to the cube root of r.

By the triple-angle formula, $\cos(\theta/3)$ is a real solution to the equation

$$4x^3 - 3x - q = 0.$$

To obtain this equation, we intersect the following circle and conic:

$$x^2 + y^2 - \frac{q}{4}x - \frac{7}{4}y = 0 \quad \text{and} \quad (1-e^2)x^2 + y^2 - \frac{q}{4}x - \left(\frac{7}{4} - e^2\right) y = 0.$$

Notice that the two conics we used are regular and have eccentricity e. We have obtained $\cos(\theta/3)$ and $\sqrt[3]{r}$, and ultimately the point α_n from those conics. We assumed that all the points in H were e-constructible, in particular α_n^3, r, and q. This shows that α_n is an e-constructible point. By induction, z is e-constructible. ■

We prove Theorem 5.1 by showing that every e-constructible point is C-constructible. Note that we can choose the value of e.

Proof. Let C be a nondegenerate conic with eccentricity $e > 0$. First, we establish a hierarchy on the e-constructible points starting from the field of constructible points $\mathbb{F}_0$:

$$\mathbb{F}_0 \subset \mathbb{F}_1 \subset \mathbb{F}_2 \subset \cdots .$$

Draw all of the regular conics of eccentricity e in $\mathbb{F}_0$. Let Q_1^e be the set of points by adjoining to $\mathbb{F}_0$ all the intersection of lines and circles with any of the drawn regular conics. Let $\mathbb{F}_1$ be the field of constructible points derived from Q_1^e (i.e. obtainable from compass and straightedge constructions with the starting set Q_1^e).

Define the rest of the fields in the sequence similarly, drawing regular conics of eccentricity e from $\mathbb{F}_k$ and intersecting them with lines and circles to form Q_{k+1}^e, then defining $\mathbb{F}_{k+1}$ as the field of constructible points derived from Q_{k+1}^e. The set of e-constructible numbers is $\bigcup_{k=0}^{\infty} \mathbb{F}_k$.

Now, we prove that every e-constructible point is C-constructible by induction on k. Clearly, $\mathbb{F}_0$ is C-constructible. Now, assume that all of the points in $\mathbb{F}_k$ are C-constructible. Take any point $z \in \mathbb{F}_{k+1} \setminus \mathbb{F}_k$ obtained by intersecting a circle R_k and a regular conic C_k of eccentricity e, both in $\mathbb{F}_k$.

Note that C_k and C both have eccentricity e. By Lemma 3.4, C can be rotated an angle θ so that C is regular. Moreover, we can scale and translate the rotated C to send it to C_k. The scaling factor $\lambda > 0$ and translation factor $a + bi$ both lie in $\mathbb{F}_k$. Thus, we can explicitly define a bijective function $f : \mathbb{C} \to \mathbb{C}$ that sends C to C_k, solely using the numbers in $\mathbb{F}_k$. In other words, a conic in $\mathbb{F}_k$ remains as a conic in $\mathbb{F}_k$ after f is applied.

The intersection point $z \in R_k \cap C_k$ can be obtained as follows:

(1) Apply f^{-1} to R_k to get R'_k.
(2) Intersect R'_k with C to get the intersection point z'
(3) Apply f to z' to obtain z.

We can verify that

$$f(C \cap R'_k) = f(f^{-1}(C_k) \cap f^{-1}(R_k)) = C_k \cap R_k = z.$$

We assumed that all of the points in $\mathbb{F}_k$ are C-constructible, so the the point $z' \in R'_k \cap C$ is a C-constructible point because R'_k is a circle in $\mathbb{F}_k$. Finally, z is C-constructible.

Every point in $\mathbb{F}_{k+1}$ is C-constructible from $\mathbb{F}_k$. This completes the proof. ■

References

[BCJ+13] Seungjin Baek, Insong Choe, Yoonho Jung, Dongwook Lee, and Junggyo Seo. Constructions by ruler and compass, together with a fixed conic. *Bulletin of the Australian Mathematical Society*, 88(3):473–478, 2013.

[GS09] Aliska Gibbins and Lawrence Smolinsky. Geometric constructions with ellipses. *The Mathematical Intelligencer*, 31(1):57–62, 2009.

[Vid97] Carlos R Videla. On points constructible from conics. *Mathematical Intelligencer*, 19(2):53–57, 1997.