FIELD EXTENSIONS AND THE CLASSICAL COMPASS AND STRAIGHT-EDGE CONSTRUCTIONS

WINSTON GAO

ABSTRACT. This paper will introduce the reader to field extensions at a rudimentary level and then pursue the subject further by looking to its applications in a discussion of some constructibility issues in the classical straight-edge and compass problems. Field extensions, especially their degrees are explored at an introductory level. Properties of minimal polynomials are discussed to this end. The paper ends with geometric problems and the construction of polygons which have their proofs in the roots of field theory.

CONTENTS

1. INTRODUCTION TO THE CLASSICAL GEOMETRIC PROBLEMS

One very important and interesting set of problems within classical Euclidean geometry is the set of compass and straight-edge questions. Basically, these questions deal with what is and is not constructible with only an idealized ruler and compass. The ruler has no markings (hence technically a straight-edge) has infinite length, and zero width. The compass can be extended to infinite distance and is assumed to collapse when lifted from the paper (a restriction that we shall see is irrelevant). Given these, we then study the set of constructible elements. However, while it is interesting to note what kinds objects we can create, it is far less straight forward to show that certain objects are impossible to create with these tools. Three famous problems that we will investigate will be the squaring the circle, doubling the cube, and trisecting an angle.

The study of what is not actually constructible is the motivation for studying field theory. Ultimately, it turns out, field theory, especially those theorems concerning the multiplicativity of degrees, is a very useful tool when studying classical geometric problems and in proving that certain constructions are impossible. The three major problems that we will attempt to show are not constructible in this paper are described below.

Date: August 5th.

Question 1.1. *Squaring the Circle: Given a square of length l, is it possible to construct a circle with the same area as the square, that is, l^2?*

Question 1.2. *Doubling the Cube: Given a cube of length l, and hence volume l^3, is it possible to create a cube with volume exactly $(2l)^3$?*

Question 1.3. *Trisecting the Angle: Is it possible with only compass and straight edge to trisect any given angle θ, that is, construct $\theta/3$?*

With these questions in mind, we progress onwards to the field theory.

2. Fields, Field Extensions, and Preliminaries

Definition 2.1. Let F and E be fields, E is a *field extension* of F if F is a subfield of E, $F \leq E$. We will write this as E/F. It is important to note that the extension E of F can be treated as a vector space in that it is abelian under addition and any "vector" $x \in$ E can be multiplied by a "scalar" $\lambda \in F$.

Definition 2.2. The dimension of this vector space is called the *degree* of the extension, which we will write as $[E : F]$.

Definition 2.3. If $[E : F] = n < \infty$ then we say the degree is finite, that is, *E has degree n*, which means that E is a *finite extension* of F. Otherwise, E is an infinite extension.

At this point, we already almost have the terminology necessary to talk about the classical geometric constructions. However, we need to first prove two major theorems regarding the degree of field extensions. The first is that the degrees of field extensions are multiplicative, meaning that if E is an extension of F and K is an extension of E; $F \leq E \leq K$ then $[K : F] = [K : E]\ [E : F]$. The other major theorem is that if E is a finite extension of F, then E is an algebraic extension of F. (The exact meaning of this statement will be explained later). To get to these results, however, we must first talk about the roots of polynomials that have coefficients in our field.

Lemma 2.4. *If $f : F \to E$ is a homomorphism of fields, then it is a monomorphism.*

Proof. First note that a field F has no ideals except 0 and F. But if $a \in F$ is a nonzero member of ideal I, then $\exists\ b \in F$ such that $ab = 1$, since we have multiplicative inverses in a field. But then $1 \in I$ which means that $\forall a \in F$, $a \in I$. We now take I to be the kernel of f, I must either be 0 or all of F. However, if I were all of F then $f(1) = 0$, but f is a homomorphism, so we have a contradiction. Hence, I must be 0 which means that the kernel of f is 0 so f is injective, a monomorphism. □

Theorem 2.5. *If f is a nonconstant polynomial over field F, there exists an extension E/F such that $\alpha \in E$ such that $f(\alpha) = 0$.*

Proof. We lose no generality in assuming that f is irreducible (since if not, we can factor f into irreducibles). We let the ideal I be the one generated by $f(X)$ so $I = \langle f(X) \rangle$. I in $F[X]$ is prime, because f is irreducible. For further reference see Robert Ash's *Abstract Algebra* section 2.6.1 for details.[1] Since I is a prime ideal, it must be maximal, which means that $E = F[X]/I$ is a field. (Ash) We place a isomorphic copy of F in E by taking the homomorphism $h : a \to a + I$, which is a monomorphism by above. Letting $\alpha = X + I$, then image of $f(\alpha)$ in $F[X]/I$ satisfies $f(\alpha) = f(X) + I = 0$ □

Definition 2.6. Let E be an extension of F, the element $\alpha \in E$ is said to be *algebraic over* F if $\exists$ a nonconstant polynomial $f \in F[X]$ such that $f(\alpha) = 0$. Otherwise, α is said to be *transcendental.* Likewise, if a field extension E/F is such that $\forall\ \alpha \in E$, α is algebraic, then E is said to be an *algebraic extension* of F.

Lemma 2.7. *Let f and g be polynomials over the field F. Then f and g are relatively prime $\Leftrightarrow$ f and g have no common root in any extension of F.*

Proof. Assume first that f and g are relatively prime, this means that their greatest common divisor is 1. This implies that $\exists$ polynomials a and b over F such that $a(X)f(X) + b(X)g(X) = 1$. This means that they must have no common root in any extension of F, if they did, we would substitute that common root α for X and we would get $1 = 0$. Conversely, assuming that f and g are not relatively prime, then they would have a non-constant common divisor, call it $d(X)$. We can show from theorem 2.5 that d has a root, α in some extension E of F, which means that both f and g have that same root α since d divides f and g. □

Note to Tom and Katie, I'm still not labeling correctly, can you look at what's wrong with my tex?

Corollary 2.8. *If f and g are distinct, monic, irreducible polynomials over F, then f and g have no common roots in any extension of F.*

Proof. If d is a nonconstant divisor of the irreducible polynomials f and g, then d coincides with both f and g, making f a multiple of g. This contradicts f and g being monic and distinct, hence relatively prime. □

Definition 2.9. Here, we define the *minimal polynomial* of α over F, written as $\min(\alpha, F)$. Let $\alpha \in E$ be algebraic over F, we can construct an ideal I by taking the set of all polynomials g such that $g(\alpha) = 0$. Since F is a principal ideal domain(see Robert Ash), I consists of all multiples of some $m(X) \in F[X]$. If we require that $m(X)$ to be monic then $m(X)$ is unique. We call this polynomial $m(X)$ the minimal polynomial of α over F.

Proposition 2.10. *The minimal polynomial $m(X)$, as defined above, has the following properties:*

1. *If $g \in F[X]$, then $g(\alpha) = 0$ if and only if $m(X)$ divides $g(X)$.*
2. *$m(X)$ is the monic polynomial of least degree such that $m(\alpha) = 0$.*
3. *$m(X)$ is the unique monic irreducible polynomial such that $m(\alpha) = 0$.*

Proof. Property (1) holds because $g(\alpha) = 0 \Longleftrightarrow g(X) \in I$. I is generated by our minimal polynomial, m, so m must divide g. Property (2) holds because of property (1). Specifically, we know that $m(X)$ is monic of least degree because other wise there could exist a $g(X) \in F[X]$ such that $g(\alpha) = 0$ but $g(X)$ is not divided by $m(X)$. Property (3) is obvious from the previous lemma as soon as we show that m is irreducible. If m is not irreducible then $m(X) = h(X)k(X)$ with deg h, deg k less than deg m. But this means that either $h(\alpha)$ or $k(\alpha) = 0$, which means that either h or k is a multiple of m, which contradicts our previous statement to show that m is irreducible. □

Definition 2.11. If E is the extension of F that contains α a root of some polynomial $f \in F$ then it is often interesting to study $F(\alpha)$, defined as the field generated by F and α. It is basically the smallest field extension that contains all the elements of F as well as α. Intuitively, the field $F(\alpha)$ would be all rational functions with polynomials in α for both the numerator and denominator. However, with the basic tools given to us by our study of the minimal polynomial, the field $F(\alpha)$ which is the smallest subfield of the extension E that has both F and α develops a much simpler representation, one that is related to $F[\alpha]$ which is the set of polynomials in α with coefficients in F.

Theorem 2.12. *If $\alpha \in E$ is algebraic over F and the minimal polynomial $m(X)$ of α over F has degree n, then $F(\alpha) = F[\alpha]$ is the set of all polynomials in α with coefficients in F. And, in fact $F(\alpha) = F_{n-1}[\alpha]$, the set of polynomials of degree at most n - 1, and furthermore, 1, $\alpha, \alpha^2, ..., \alpha^{n-1}$ form a basis for the the vector space $F[\alpha]$ over the field F.*

Proof. To begin our proof, must first show that $F_{n-1}[\alpha]$ is a field. To do so is simple given a trick we learned with the relative primeness of polynomials. Let $f \in F_{n-1}[\alpha]$, we know that deg f must be less than deg m since m is our minimal polynomial, irreducible and of degree n. Since these two polynomials are relatively prime, we know that $a(X)f(X) + b(X)m(X) = 1$ for some polynomials $a(X)$ and $b(X)$ over F. However, substituting α for X, we get $a(\alpha)f(\alpha) = 1)$ which shows us that f has an inverse and hence, $F_{n-1}[\alpha]$ is a field.
Next we show that $F_{n-1}[\alpha] \subseteq F[\alpha] \subseteq F(\alpha)$. This is true because it is obvious that any field containing F and α must also contain all of the polynomials in α as well as all of the polynomials of degree at most $n - 1$. However, $F(\alpha)$ is the *smallest* field containing α and F so it must be a subset of the field $F_{n-1}[\alpha]$. Hence, equality is established.
To prove that a basis of this field is 1, $\alpha, \alpha^2, ..., \alpha^{n-1}$ we just need to show that they span $F_{n-1}[\alpha]$ and are linearly independent. To show they span is a trivial matter because any polynomial in α that has degree less than $n - 1$ can certainly be generated by those elements. To show that they are linearly independent, we just note that if they were not then there would be a polynomial with non-zero coefficients with α as its root. However, this polynomial is of degree less than n which means it is a polynomial with degree less than our minimal polynomial m, which is a contradiction. □

Importantly for the next theorem (the first of the two major ones we want to cover), this shows that $[F(\alpha) : F] = n$.

Theorem 2.13. *The degree of field extensions is multiplicative. In other words, Let $F \leq E \leq K$, then [K : F] = [K : E] [E : F]. If either side is infinite, then both sides of the inequality are infinite.*

Proof. For this proof, we first look at a multiplicative property of bases. We note that if $\alpha_1, ..., \alpha_i$ form a basis for K over E and $\beta_1, ..., \beta j$ form a basis for E over F, then the set of $\alpha_m\beta_n, 0 < m < i, 0 < n < j$ form a basis for K over F. To see this note that for a vector $v \in K$, v is a linear combination of coefficients $a_i \in E$ and the basis α_i. However, each of the $a_i \in E$ is a linear combination of β_j and $b_j \in F$. This shows that the $\alpha_i\beta_j$ span K over F. To show linear independence, we note that for λ_{ij} constants, $\sum_{i,j} \lambda_{ij}\alpha_i\beta_j = 0$ then $\sum_i \lambda_{ij}\alpha_i = 0$ for all j, which means

that $\lambda_{ij} = 0$ for all i, j, which proves that $\alpha_i\beta_j$ are linearly independent, hence a basis of K over F.
Then for $[E : F] = j$, and $[K : E] = i$, we have $[K : F] = i * j$ □

Theorem 2.14. *If E is a finite extension of F, then E is an algebraic extension of F.*

Proof. Let $\alpha \in E$, and let the degree $[E : F] = n$. We have that $1, \alpha, \alpha^2, ...\alpha^n$ are $n + 1$ vectors in an n-dimensional vector space, making them linearly dependent. Hence, we have

$$\beta_0 + \beta_1\alpha + \beta_2\alpha^2 + ... + \beta_n\alpha^n = 0$$

for $\beta \in F$ not all 0. Hence α is a root of a nonzero polynomial with coefficients in F, so α is algebraic over F. □

3. Geometric Problems

At this point, we now have a firm enough grasp of field extensions to take a stab at some of the more interesting questions in classical geometry.

As stated before, we will look at specifically the compass and ruler constructions. The most elementary actions one can make with these two tools are: drawing a line between two points, drawing a circle centered a point with radius going out to another point, drawing a point at the intersection of two lines, two circles, or a line and a circle.

As a general rule, we begin the our constructions with only two points (and will often designate the distance between them as 1) and with the ability to only make the most elementary actions, but as we progress, we can assume more complicated starting conditions and shortcut a lot of the more complicated actions.

Lemma 3.1. *Given two points on a two dimensional plane A and B, with just compass and straight-edge, we can find the point that is equidistant from A and B, and we can find the perpendicular line that runs through this point. In other words, we can bisect that line segment A-B with a perpendicular line.*

Proof. Draw a circle centered at A through passing through B. Draw another circle centered at B passing through A. These two circles intersect at two points, C and D. Draw a line through C and D, this line will be perpendicular to line A-B, and the intersection between these two lines will be the midpoint of the segment A-B. □

Proposition 3.2. *Given two points, call them $(0, 0)$ and $(1, 0)$ on a two dimensional plane, with compass and straight-edge, we can construct the integer lattice.*

Proof. Let our two points be A and B. We can assign the value $(0, 0)$ to A and $(1, 0)$ to B. By drawing a line through the two points, we have generated the x-axis. Drawing a circle with A as the center and B as the length of the radius, then finding where that circle intersects the x-axis will grant us $(-1, 0)$. Doing this similarly with B as the center grants us $(2, 0)$, and we can recursively construct the entire set of points $(n, 0)$ where n is an integer. To find the values of y that are integers, we can use the previous lemma to construct the y-axis as well as all the other lines perpendicular to the x-axis through it's integer co-ordinates, which will give us the entire two dimensional plane with points (x, y) where $x, y \in \mathbb{Z}$. □

Next we will construct the rational numbers, but first it is necessary to illustrate two simple constructions that we can make.

Lemma 3.3. *Given a line that goes through A and B, we can construct a line parallel to it through a third given point, C.*

Proof. Begin by drawing a circle centered at C through the point A — this will give us a point D, at the intersection of the line and the circle. By the previous lemma we can draw a perpendicular line through the midpoint of D and A, which consequently passes through C. Now, take the previously drawn circle and find the two points that intersect with this new line E and F. Finding the perpendicular line that runs through the midpoint of these two points will give us a line that passes through C, which is parallel to the line made by A-B. □

Lemma 3.4. *Given a line-segment A-B, and a point C, we can construct a line-segment C-D that is parallel to A-B and has the same length.*

Proof. Begin by drawing a line, x, through A and C. Then construct a line, y, parallel to this one through B. Construct a third line z parallel to A-B through C. Where z intersects y is the point D that we are looking for. □

Note, with this lemma, we can then draw a circle through D centered at C to rotate our newly constructed segment in whatever direction we want, giving us the ability to "move" line segments from place to place.

Theorem 3.5. *Given two line-segments a, and b (as well as a third length, 1) in the two dimensional plane, we can add, subtract, multiply, and divide their lengths.*

Before we begin the proof of this theorem, something ought to be said about the concept of adding, subtracting, multiplying, dividing, and "length". What we mean by adding a length is that given two line segments x and y , we can construct a third line segment such that the length is equal to the length of x added to the length of y. This concept is analogous with subtracting, multiplying and dividing.

Proof. With lemma 3.4 , addition and subtraction become entirely trivial. For multiplication and division we use similar triangles to get our desired values. For multiplication, we can move segment a to have the end points $(0, 0)$ and (x_a, y_a) to share an endpoint with segment b which we place at $(0, 0)$ (x_b, y_b), rotate a about the origin so that it is over the positive x-axis and rotate b about the origin so that it is over the positive y-axis (note, we do this so the lines are non-parallel, although any angle that keeps them non-parallel will do). Marking the length 1 on the segment a $(1, 0)$, we then draw a line x through (1,0) and the endpoint of $b, (x_b, y_b)$. We then make a line y, parallel to x that goes through the end point of $a, (x_a, y_a)$. Now draw a line z by extending the segment b infinitely in both directions, this line will intersect y at a point D. The length of the segment created by D and the origin will be a line segment with the length of the desired product. The setup to division is the same as multiplication, only the first line we draw is one from the endpoint of $a, (x_a, y_a)$ through the endpoint of $b, (x_b, y_b)$. A parallel line to this one that goes through $(0, 1)$ will intersect b at a point E. The segment from E to the origin is one with length of the desired quotient b/a. □

Theorem 3.6. *Given a line segment a in the two dimensional plane we can construct a line segment whose length is equal to the square root of our original length.*

Proof. Let a have endpoints A and B. Add 1 to this length by adjoining a length 1 onto the endpoint B making a segment of length $a+1$. Draw a circle from the midpoint of this segment through A The perpendicular of the segment a that goes through the point B will intersect the circle at two points, C and D. The length of C-B is equal to $\sqrt{a}$. □

Hence, at this point we were able to construct the integer lattice, and given the ability to add, subtract, multiply, and divide these co-ordinates, we are immediately given the rational lattice. However, we are also capable of constructing the square roots of the rationals as well (as well as the square roots of those numbers). In short, we have constructed the rationals as well as the ability to square root these numbers, giving us numbers such as $\sqrt{2}$ and $\sqrt{1+\sqrt{2}}$. The resulting set of numbers for the coordinates is a field. To see this, we note that any sum of two constructible numbers is also constructible. Furthermore, we can construct the multiplicative inverse of any constructible number by dividing that number from 1. We know that this field is strictly bigger than the rationals because of the ability to construct square roots. It is now of interest to find what field this extension of the rationals actually is.

Theorem 3.7. *The field consisting of all constructible points using only compass and straight-edge is the quadratic closure of* $\mathbb{Q}$

Proof. Let F be a field that we have generated through compass and straight-edge constructions. Given our basic methods of constructing points, there are basically only three different types of actions we can make to construct a new point, α. Note here that the act of constructing new points is equivalent to constructing a field extension E that is F adjoined α. The first action is to construct a point α by intersecting two lines. In other words, we can construct a point by solving for the equations of two lines. However, these lines are of the form $ax - by = c$ where $a, b, c \in F$ so solving for two such equations will result with another point in F. We see that intersecting lines will not give us points that are not in F. On the other hand, intersecting a circle with a line will give us points that are not in F, to see this, note that solving for x and y given

$$(x-d)^2 + (y-e)^2 = f^2$$

and

$$ax - by = c$$

with $a, b, c, d, e, f \in F$. Solving for these two equations gives us a point in at most a quadratic extension of F (of degree 2). Finally, solving for a point constructed by the intersection of two circles is the same as solving for a point constructed by the intersection between a circle and a line (specifically the line that is created, in fact, by the two points).

Hence we end up with a field extension K of F where $[K : F] = \infty$ and any E such that $K \geq E \geq F$ and $[E : F] = 2^m$ where m is a power of two. □

Theorem 3.8. *Given only a straight-edge and a compass. It is impossible to (1) square the circle (2) double the cube (3) trisect an angle.*

Proof. _

(1) To square the circle, as stated above, we are trying to construct the circle that has the same area as a square of given length. Hence we solve the equation, $l^2 = 2\pi r$ for r in terms of l. This eventually simplifies to whether or not we can construct π, which we say without proof, is transcendental. Hence the field extension that is required is $\mathbb{Q}(\pi)$. Noting that the field extension is not algebraic, because π is transcendental, we know that the degree $[\mathbb{Q}(\pi) : \mathbb{Q}]$ is not finite from theorem 2.14, and hence not a power of 2.

(2) To double the cube, we need to construct $\sqrt[3]{2}$ from 1. But $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$, which is not a power of two.

(3) To trisect a certain angle is equivalent to constructing $\cos(\theta/3)$ from $\cos(\theta)$. This is not always possible (note, it is sometimes possible, for example, $\theta = 180\,^\circ$). Consider the triple angle formula for cosine:

$$\cos(\theta) = 4\cos^3(\theta/3) - 3\cos(\theta/3),$$

Letting $\theta = 60\,^\circ$ then $\cos(\theta) = 1/2$. Subsituting x for $\cos(20\,^\circ)$, we get

$$4x^3 - 3x - 1/2 = 0.$$

Letting $u = 2x$, we then get

$$u^3 - 3u - 1 = 0.$$

This is the minimal polynomial of $\mathbb{Q}(u)$. We previously showed that the degree of the minimal polynomial is the degree of the field extension, but that means that $[\mathbb{Q}(u) : \mathbb{Q}] = 3$ which is not a power of 2, so u is not constructible. Hence $10\,^\circ$ is not constructible.

□

4. Constructing Regular Polygons

Finally, to conclude our discussion of field extensions and classical geometry, we look to some problems on the construction of regular polygons.

Examples 4.1. Now, given our prior knowledge, we can easily construct the regular 3-gon or the equilateral triangle, and the 4-gon a square. To construct a square, we merely need to construct the integer lattice and connect the points $(1,0), (0,1), (-1,0)$, and $(0,-1)$. An equilateral triangle is constructed by taking two given points, A and B and drawing a circle centered at each going through the other. We now have two circles of equal radius that go through the center of the other. This means that if we take one of the points of intersection between the two circles, call this point C, C is the same distance away from B as it is from A, since it lies on both circles and is exactly the radius away. Hence $C - B - A$ would be a regular triangle.

Since we know how to bisect angles, we immediately obtain the 6-gon and 8-gon. The 5-gon is constructible, but the proof is complicated and does not add much to our investigation of fields, for a proof of the 5-gon reference Dummit and Foote's *Abstract Algebra*.[2] That leaves us with the 7-gon and the 9-gon.

Theorem 4.2. *Given only a straight-edge and compass, it is impossible to construct a 9-gon.*

Proof. The proof of this is given to us directly from the previous theorem. Attempting to construct any polygon is equivalent to finding the internal angles. Given the internal angles, we can just draw a circle centered at the vertices of our angles and connect the points. Given a polygon, we can find the angles by drawing lines to the center of the polygon. However, we have already proven that it is impossible to trisect certain angles. Within that proof, we specifically showed that it was impossible to construct the 40° angle, which is the internal angle of the regular 9-gon. □

Acknowledgments. It is a pleasure to thank my mentors, Tom Church and Katie Mann, for their unending patience when instructing and teaching and helping me through this entire process. [**?**]

References

[1] Robert Ash. Abstract Algebra: The Basic Graduate Year.

[2] David S. Dummit, Richard M. Foote. Abstract Algebra. John Wiley and Sons, Inc. 2004.