NUMBERS AND SHAPES

NOTES AND PROBLEMS TO TOPICS IN GEOMETRY, MATH 231, SPRING 2016

Contents

These notes are meant to supplement our reading, in particular fill in the details of what might not have been clear. Along the way you should complete all the exercises that are contained below, they will serve to build your intuition about what is going on. Enjoy the ride!

1. Double a cube

Definition 1.1. Through F. Klein's book we will study the constructibility of certain numbers, for example $\sqrt[3]{2}$ and π. We say a number is *constructible* if it can be constructed through finite number of rational operations and square roots, for example

$$\sqrt{3+\sqrt[4]{2}}$$

which is to say, we may construct the number through (1) drawing lines between points, (2) circles with given radius and center, and (3) intersecting lines and circles.

In 420 B.C. the inhabitants of Athens appealed to the oracle at Delos how to stop the plague that afflicted their city, which replied that they must double size of the altar of Appollo, a cube. This translates to the problem of solving the equation $x^2 = 2$, given that the edge of the cube has unit length.

Definition 1.2. Throughout when we refer to polynomials we mean a polynomial with rational coefficients, i.e.,

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x_1 + a_0$$

where all the a_i are rational numbers. Moreover, we call a polynomial *irreducible* (over the rationals) if it cannot be factored as a product of two non-constant polynomials. For example, $x^2 - 4$ is reducible but $x^2 + 4$ is not. (Why?)

Remark 1.3. If you recall the Fundamental Theorem of Algebra, it tells us that every polynomial is reducible over the complex numbers, that is to say, any polynomial will have solutions in the field of complex numbers.

Theorem 1.4. *Let x be a constructible number made up out of m square roots, then it is a root of an irreducible equation $f(x) = 0$ of degree 2^m.*

Proof. The way we shall go through this proof of Klein's is by illustrating key steps with exercises and examples. First, here is an outline of the steps:

(1) First express x as having a term of given order m appearing only in the numerator and linearly. So without loss of generality assume x is in this form. For example, one can rationalize the denominator of

$$\frac{5}{2 - \sqrt[4]{3}}$$

using the identity $a^4 - b^4 = (a + b)(a - b)(a^2 + b^2)$. (Can you show this?)

(2) The polynomial $F(x) = (x - x_1) \dots (x - x_{2^m})$ containing all the conjugate values of x has rational coefficients. This is because all the conjugates appear, so expanding the product yields a rational expression. For example, see Exercise 1.

(3) If x_i satisfies $f(x_i) = 0$ for any $1 \leq i \leq 2^m$, then so does any other x_j. For example, see Exercise 2.

(4) Define the equation $\phi(x)$ of lowest degree satisfying all x_i. It is irreducible, without multiple roots, has no other roots than x_i, and is the only irreducible equation with rational coefficients satified by the x_is. If M is the number of distinct conjugates of x_i, then

$$\phi(x) = C(x - x_1) \dots (x - x_M)$$

For some nonzero constant C. All of these claims follow from an argument by contradiction, i.e., if it were not so then $\phi(x)$ would not be of lowest degree.

(5) $F(x) = C\phi(x)^n$ for some integer n, since the irreducible polynomial $\phi(x)$ must divide $F(x)$, and if $F(x)/\phi(x)$ is a non-constant function, then we may factor $F(x) = F_1(x)\phi(x)$ where $F_1(x)$ is again a polynomial with some root x_i, and by (3) above it must have all the other conjugates as roots.

(6) The degree of $F(x)$ is 2^m, while the degree of $\phi(x) = 2^m$. Conversely, an irreducible equation not of degree 2^m cannot be solved by square roots.

□

Exercise 1. Consider $\sqrt{3+\sqrt[4]{2}}$, which has order 3 because it involves three square roots. Show that it is a root of an irreducible polynomial of degree 2^3. Hint: we can easily make this polynomial out of the conjugates, but why is it irreducible?

Exercise 2. (Easy.) Suppose $a+b\sqrt{c}$ is a root of $f(x) = x^3+px^2+qx+r$. Namely,

$$\begin{aligned} f(a+b\sqrt{c}) &= (a+b\sqrt{c})^3 + p(a+b\sqrt{c})^2 + q(a+b\sqrt{c}) + r \\ &= (a^3+3ab^2c+pa^2+pb^2c+1a+r) + (3a^2b+b^3c+2pab+qb)\sqrt{c} \\ &= A + B\sqrt{c} = 0 \end{aligned}$$

So $A = B = 0$, otherwise $\sqrt{c}$ is a rational number, which cannot be. By a similar computation, show that the conjugate $a - b\sqrt{c}$ is also a root, since $f(a-b\sqrt{c}) = A - B\sqrt{c}$, which is zero from before.

Corollary 1.5 (The Delian problem)**.** *$\sqrt[3]{2}$ is not a root of an equation of degree 2^m for some integer m, so it is not a constructible number.*

Exercise 3. (Easy.) In class we showed by contradiction that $\sqrt{2}$ is not a rational number, i.e., it cannot be written as a fraction p/q. By a similar argument show that $\sqrt[3]{2}$ is also irrational. Note: some irrational numbers like $\sqrt{2}$ are constructible, while, as we see from above, others are not.

Remark 1.6. This proof appears in Euclid's Elements, as proposition 117 of Book X. Its discovery is attributed to Pythagoras, and the Pythagoreans treated this as an official secret.

2. Trisect an angle

Let's start with a refresher on complex numbers. We will return to the *discovery* of complex numbers later in the semester; for now we will use them as they are: a complex number is of the form

$$z = a + bi = a + b\sqrt{-1}$$

where a and b are any real numbers. Using the power series representations of e^x, $\sin x$, and $\cos x$, Leonhard Euler showed the remarkable identity for any real x

$$e^{ix} = \cos x + i \sin x$$

Exercise 4. Prove this.

From which we have, according to Richard Feynman, the most beautiful equation $e^{i\pi} + 1 = 0$ (because it relates all the important constants.) This allows us to write complex numbers in the polar form

$$z = re^{i\phi}$$

(see Fig 1. p. 14 in Klein.) Notice that this quantity has length one, hence parametrizes the unit circle. (Do you see why?)

Theorem 2.1 (De Moivre)**.** *Given any real number x and integer n, we have:*

$$(\cos\phi + i\sin\phi)^n = \cos(n\phi) + i\sin(n\phi)$$

Proof. (For positive integers n) The proof of this is by induction. (1) Base case: clearly this holds for $n = 1$. (2) Induction step: now assume the formula holds for n, we want to show that the formula is then true for $n + 1$, namely

$$\begin{aligned}(\cos\phi + i\sin\phi)^{n+1} &= (\cos(n\phi) + i\sin(n\phi))(\cos\phi + i\sin\phi) \\ &= \cos((n+1)\phi) + i\sin((n+1)\phi)\end{aligned}$$

using basic trigonometric identities. □

Remark 2.2. If you're not familiar with proof by induction, the idea is the following. Suppose you want to prove a statement P is true for all $n =, 1, 2, 3 \ldots$ First prove the *base case* $n = 1$, which often simplest. Then the *induction step* goes like this: assume now that P is true for n, then using this assumption show that P in fact holds for $n + 1$. Proving these two amounts to a kind of 'domino effect' that proves P for all n. A nice example of this is the statement that the sum of first n integers $1 + 2 \cdots + n$ is equal to $n(n + 1)/2$.

Definition 2.3. A *root of unity* is a solution to the equation $x^n - 1 = 0$ for some n. For example, for $n = 2$, the solutions are $\{\pm 1\}$, for $n = 4$, the solutions are $\{\pm 1, \pm i\}$.

The case $n = 3$ is the one we want to consider, in relation to the problem of trisecting an angle. For this, we will need the formula for geometric series:

$$1 + r + r^2 + \cdots + r^{n-1} = \frac{r^n - 1}{r - 1}$$

(Prove this!) This shows that $x^3 - 1$ is a *reducible* equation, and has one rational root. What are the other two roots? They are roots of unity, moreover, complex numbers.

Theorem 2.4 (Rational Root Theorem)**.** *If p/q is a root of the polynomial equation*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

where $a_0, \ldots, a_n$ are integers, then p divides a_0 and q divides a_n.

Exercise 5. (Easy.) Prove this. To start off, write $f(p/q)$ and clear denominators by multiplying with q^n.

Now here is the main theorem of this section:

Theorem 2.5. *There exists an angle ϕ such that $\phi/3$ cannot be constructed.*

We present two proofs, the first is Kein's, and the second by Wantzel (1837) which is easier to understand. First,

Proof. Given any angle ϕ, we may represent it on the unit circle as $e^{i\phi}$. By De Moivre, if we set $x^3 = e^{i\phi}$, then for any integer n,

$$(\cos(\frac{\phi + 2\pi k}{3}) + i\sin(\frac{\phi + 2\pi k}{3}))^3 = \cos\phi + i\sin\phi$$

solves the equation. In particular, with $k = 0$, we have the solution

$$x_1 = \cos\left(\frac{\phi}{3}\right) + i\sin\left(\frac{\phi}{3}\right).$$

which is a point on the unit circle trisecting ϕ.

We now show that

$$x^3 = \cos\phi + i\sin\phi \tag{2.1}$$

is irreducible. If it were reducible, then it would have a rational root, call it x_i. Observe that this root must be a rational function of $\cos\phi$ and $\sin\phi$, and we are able to construct it by trigonometry plus straightedge and compass. That is,

$$x_i = a_i\cos\phi + b_i\sin\phi + c_i$$

where a_i, b_i, c_i are rational numbers. Suppose this were true. Then there is some rational function f such that

$$f(\cos\phi, \sin\phi) = \cos\left(\frac{\phi}{3}\right) + i\sin\left(\frac{\phi}{3}\right)$$

but rotating by 2π does not change the value

$$f(\cos(\phi + 2\pi), \sin(\phi + 2\pi)) = \cos\left(\frac{\phi}{3}\right) + i\sin\left(\frac{\phi}{3}\right)$$

But from the previous discussion we see that the roots x_1, x_2, x_3 are permuted cyclically $1 \mapsto 2 \mapsto 3 \mapsto 1$, which contradicts what we have just shown, so (1) must be irreducible. And from the Delian problem we see that $\phi/3$ is not constructible in general. □

Now for the second proof.

Proof. This proof will use the triple angle formula

$$\cos\phi = 4\cos^3\frac{\phi}{3} - 3\cos\frac{\phi}{3}$$

to show that $\phi = 60°$ is impossible to trisect. Note $\cos 60° = \frac{1}{2}$ and let ω stand for $\cos 20°$. Conisder

$$\frac{1}{2} = 4\omega^3 - 3\omega$$

and replacing ω by $\frac{1}{2}\alpha$ we are reduced to

$$\alpha^3 - 3\alpha - 1.$$

But by the rational root theorem this equation is seen to be irreducible, i.e., has no rational roots. □

Exercise 6. Prove the triple angle formula, then apply the rational root theorem to show that $\alpha^3 - 3\alpha - 1$ is irreducible, and conclude that 20° is not constructible.

Remark 2.6. So far, we have shown that a cube may not be doubled because $x^3 - 2$ is irreducible, and that an angle may not be trisected because $x^3 - e^{i\phi}$ is irreducible in general. In both we have needed the theorem that a number is constructible only if it is a root of a polynomial of degree 2^m.

Definition 2.7. Before we proceed to the next section we make a comment on modular arithmetic. Given an integer expression $y = x + an$, we say y is congruent to x modulo a, meaning that x and y are equal up to some multiple of a. Denote this by $y \equiv x \mod a$. For example, if $11 \equiv 1 \mod 5$.

Conversely, if $y \equiv x \mod a$, then we may express y as x plus some integer multiple of a, say $y = x + an$.

Exercise 7. (If you are not familiar with modular arithmetic, you should do this exercise.) Write the number 11 through 15 modulo p, where p ranges from 2 to 5. Next, show that arithmetic works like we know it:

(1) If $a \equiv x \mod p$ and $b \equiv y \mod p$, then $a + b \equiv x + y \mod p$. Give a numerical example.
(2) If $a \equiv x \mod p$ and $b \equiv y \mod p$, then $ab \equiv xy \mod p$. Give a numerical example.

Note: this is needed for Fermat's Little Theorem below which, strictly speaking, we will not require, but we will encounter this crucially when constructing the 17-gon.

3. Divide a circle

For a long time it was known that a regular polygon could be constructed if it had $2^m, 3$, or 5 sides, or any product of these numbers. From the previous section, we see that if we can divide a circle into equal parts, then we may construct a regular polygon. This leads us to looking for the n-roots of unity, namely, solutions of the equation

$$x^n - 1 = 0$$

Clearly $x = 1$ is always a solution, and indeed we have the factorization

$$x^n - 1 = (x - 1)(x^{n-1} + \cdots + x + 1)$$

From Theorem 1.4, we know that for x to be constructible it must the root of an irreducible equation of degree 2^m, therefore $n - 1 = 2^m$.

Theorem 3.1 (Gauss)**.** *A circle can be divided into a prime number of equal parts if and only if* $p = 2^{2^m} + 1$.

Indeed, when $m = 0, 1$ this gives precisely the triangle and pentagon, and for the subsequent $m = 2$ we have the 17-gon, an achievement Gauss was particularly proud of, even though he went on to prove much more mathematics. We will prove this in a series of steps, beginning with a reduction to a simpler case:

Proof. Step 0. We first reduce to the case where n is a prime or a power of a prime. Given a composite number mn where m, n are prime to each other, then the Bézout's lemma implies that

$$1 = am + bn$$

for some integers a, b. Then since

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

we can divide a circle into mn parts if we can do so given m and n divisions, so we are reduced to only considering m, n prime.

Exercise 8. Look up a proof of Bézout's lemma and write it down. Make sure you understand it. This uses a result of Euclid.

Example 3.2. Notice that $\frac{1}{15} = \frac{2}{3} - \frac{3}{5}$, so if we are able to divide a circle into 3 and 5 equal parts, then from a combination of the two we may divide the circle into 15 equal parts.

Step 1. Next we show that if the circle can be divided into p parts, it must be of the form $2^{2^m} + 1$.

Lemma 3.3. *Let p be a prime. Then $p = 2^h + 1$ only if $h = 2^m$.*

There are two proofs of this in Klein. First, the simpler proof:

Proof. Suppose h is divisible by an odd number, so that $h = h'(2n + 1)$. Then substituting $x = 2^{h'}$ into the formula

$$x^{2n+1} = (x + 1)(x^{2m} - x^{2m-1} + \cdots - x + 1)$$

we see that $p = 2^{h'(2n+1)}$ is divisible by $2^{h'} + 1$, hence not prime. □

The second proof is more complicated:

Proof. We depend on the following

Theorem 3.4 (Fermat's Little Theorem)**.** *Let p be a prime number and a an integer not divisible by p. Then the following congruence holds true:*

$$a^{p-1} \equiv 1 \mod p$$

Remark 3.5. (Optional exercise) Prove this. Its proof may require Euclid's lemma, which requires Bezout's lemma, which you proved above.

The proof now goes as follows. Suppose we have a prime $p = 2^h + 1$ for some h, and let s be the smallest integer satisfying

$$2^s \equiv 1 \mod p$$

from these two facts we deduce that $s > h$. Also, we see that h is the smallest integer such that

$$2^h \equiv -1 \mod p$$

Then dividing the two equations,

$$2^{s-h} \equiv -1 \mod p$$

so $s - h$ must be greater than or equal to h, hence $s \geq 2h$.

Next, from squaring

$$2^{2h} \equiv 1 \mod p$$

we see that s, by minimality, cannot be more than $2h$, so we conclude that $s = 2h$. □

Step 2. Finally, we show that the cyclotomic polynomial $x^{2^{2^m}+1} + x^{2^{2^m}} + \cdots + x + 1$ is irreducible. We first need a lemma:

Lemma 3.6 (Gauss)**.** *Let $F(x)$ be a polynomial with integer coefficients that have GCD =1. Suppose further that it factors as $F(x) = f(x)g(x)$ where $f(x)$ and $g(x)$ have rational coefficients, then in fact $f(x), g(x)$ have integer coefficients.*

Proof. Let's prove this in the case we are considering, namely, suppose that

$$F(x) = x^n + x^{n-1} + \cdots + x + 1$$

is reducible, hence factors into a product of polynomials with rational coefficients,

$$f(x)g(x) = (x^m + a_1 x^{m-1} + \cdots + a_m)(x^{m'} + b_1 x^{m'-1} + \cdots + b_{m'})$$

where $m + m' = n$. Let a be the LCD of the a_i's and b of the b_i's. Consider now the product $abF(x)$; its coefficients are integral, and furthermore their GCD is exactly ab.

On the other hand, a and b being the LCD, $af(x)$ and $bg(x)$ have integral coefficients. Since the GCD of the coefficients of $f(x)$ and $g(x)$ is 1, the GCD of the coefficients of $af(x)$ and $bg(x)$ are again 1, and we conclude that $a = b = 1$, hence the coefficients are integral. □

We use this to prove that the cyclotomic polynomial $F(x)$ is irreducible. Consider first the case where $n = p$. Writing

$$F(x) = \frac{x^p - 1}{x - 1},$$

we make the substitution $x = z + 1$, we have

$$F(z+1) = \frac{(z+1)^p + 1}{z} = z^{p-1} + pz^{p-2} + \frac{p(p-1)}{1 \cdot 2} z^{p-3} + \cdots + \frac{p(p-1)}{1 \cdot 2} z + p$$

where the second equality follows from the Binomial Theorem (if you don't know this, look this up and convince yourself why the Binomial Theorem implies this.) In fact, all the coefficients except the first are divisible by p, being of the form

$$\frac{p!}{(p-k)!k!}$$

where k ranges from 1 to $p-1$. This expression is irreducible, for otherwise the factorization

$$f(x+1) = (x^m + a_1 x^{m-1} + \cdots + a_m)(x^{m'} + b_1 x^{m'-1} + \cdots + b_{m'})$$

implies $a_m b_{m'} = 1$, hence either $a_m = \pm p$ and $b_{m'} = \pm 1$ or vice versa. Next, the coefficient of x is

$$\frac{p(p-1)}{2} = a_{m-1} b_{m'} + a_m b_{m'-1}$$

and it follows that if $b_{m'} = \pm 1$ then a_{m-1} is divisible by p. Continuing thus we see that all the coefficients a_i are divisible by p, but this cannot be true of the coefficient of x^m, which is 1. Thus we conclude that $F(x)$ is irreducible for $n = p$.

Now we want to exclude the second case where n is a power of p if $p > 2$. Of course, it is enough to consider $n = p^2$ since if this is impossible to construct then so are higher powers of p. The cyclotomic equation is

$$\frac{z^{p^2} - 1}{z - 1} = 0$$

It has the roots of $(z^p - 1)/(z - 1)$, which we shall discard because we already showed that these are constructible. So we consider instead

$$F(x) = \frac{x^{p^2} - 1}{x^p - 1} = 0.$$

Using the formula for geometric series again we may write it as

$$x^{p(p-1)} + x^{p(p-2)} + \cdots + x^p + 1$$

and substituting again $x = z + 1$, we have

$$(z+1)^{p(p-1)} + (z+1)^{p(p-2)} + \cdots + (z+1)^p + 1.$$

Seeing that there are p number of terms in this sum, after expanding each term the constant terms 1 add up to p. Moreover, by the binomial expansion we see again that the p divides every term but the first, so we may express $F(z+1)$ as

$$z^{p(p-1)} + p \cdot (\text{lower order terms})$$

From the first case, we know that this is irreducible.

Finally, notice that the degree of this cyclotomic polynomial is $p(p-1)$, and that an irreducible equation is solvable by square roots only if is degree is a power of two. This forces $p = 2$, so a circle cannot be divided into p^2 equal parts if p is odd. □

Remark 3.7. Notice that from the criterion we proved, the polygons of sides 3, 4, 5, 6, 8, and 10 are able to be constructed. The lack of a construction for 7 (and also 9) was particularly frustrating for the Greeks and later mathematicians, until it was proven by Gauss to be impossible. Indeed, consider a root ζ of the equation

$$x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 = 0.$$

If we pair the roots as follows:

$$\begin{aligned} y_1 &= \zeta + \zeta^{-1} = \zeta + \zeta^6 \\ y_2 &= \zeta^2 + \zeta^{-2} = \zeta^2 + \zeta^5 \\ y_3 &= \zeta^3 + \zeta^{-3} = \zeta^3 + \zeta^4 \end{aligned}$$

(using the fact that $\zeta^{m+7n} = \zeta^m \zeta^{7n} = \zeta^m$) we obtain the relations

$$\begin{aligned} y_1 + y_2 + y_3 &= -1 \\ y_1 y_2 + y_2 y_3 + y_1 y_3 &= -2 \\ y_1 y_2 y_3 &= 1 \end{aligned}$$

which show that y_1, y_2, y_3 satisfy the irreducible cubic $Y^3 + Y^2 - 2Y - 1 = 0$.

Exercise 9. Justify the steps sketched above to show that the heptagon is not constructible.

4. The Heptadecagon

Now we shall construct the seventeen-sided polygon. Begin with the roots of

$$x^{16} + x^{15} + \cdots + x + 1 = 0$$

which, by De Moivre, are of the form

$$\zeta^k = \cos\frac{2\pi k}{17} + i\sin\frac{2\pi k}{17}$$

We want a root that whose successive powers will range over all 17 roots. This is called a *primitive* root, which is to say that $\zeta^k \neq 1$ for any $1 \leq k \leq 17$.

Notice that if $3^k = 17q + r \equiv r \mod 17$, then $\zeta^{3^k} = \zeta^r$, then the next remainder is the cube of the preceding: $\zeta^{3^{k+1}} = \zeta^{3^k+3} = (\zeta^r)^3$. In fact, we claim that ζ^3 is primitive. To see this, we produce a sequence of powers of ζ^3 that range through k, permuting the roots of unity in the order

$$k = 3, 9, 10, 13, 5, 15, 11, 16, 14, 8, 7, 4, 12, 2, 6, 1.$$

Exercise 10. (Easy) Verify this sequence.

This way, we will construct sums containing 8, 4, 2, and 1 roots, corresponding to the divisors of 16, as roots of quadratic equations. Note that these degrees depend on the prime factors of $p - 1$.

Taking alternate terms, we now form the sums

$$x_1 = 9 + 13 + 15 + 16 + 8 + 4 + 2 + 1$$
$$x_2 = 3 + 10 + 5 + 11 + 14 + 7 + 12 + 6$$

Where we have used the numbers to stand for the exponents of roots of unity, e.g., $x_1 = \zeta^9 + \zeta^1 3 + \ldots$ is written as $9 + 13 + \ldots$ (The relations $x_1 + x_2 = -1$ and $x_1 x_2 = -4$ shows that x_1, x_2 satisfy the quadratic $x^2 + x - 4 = 0$.)

Then taking again alternate terms,

$$y_1 = 13 + 16 + 4 + 1$$
$$y_2 = 9 + 15 + 8 + 2$$
$$y_3 = 10 + 11 + 7 + 6$$
$$y_4 = 3 + 5 + 14 + 12$$

(The relations $y_1 + y_2 = x_1, y_3 + y_4 = x_2$, and $y_1 y_2 = y_3 y_4 = -1$ shows that the y_1, y_2 satisfy $y^2 - x_1 y - 1 = 0$ and y_3, y_4 satisfy $y^2 - x_2 y - 1 = 0$.)

And finally

$$z_1 = 16 + 1 \qquad z_5 = 11 + 6$$
$$z_2 = 13 + 4 \qquad z_6 = 10 + 7$$
$$z_3 = 15 + 2 \qquad z_7 = 5 + 12$$
$$z_4 = 9 + 8 \qquad z_8 = 3 + 14.$$

(The relations $z_1 + z_2 = y_1$, $z_1 z_2 = y_4$ show that z_1, z_2 satisfy $z^2 - y_1 + y_4 = 0$, and finally, ζ, ζ^{16} satisfy $v^2 - z_1 v + 1 = 0$. Thus the roots can be obtained by extraction of square roots.)

Next, notice that the z_i's are real numbers, since

$$z_* = \zeta_k + \zeta_{17-k} = 2\cos\frac{2\pi k}{17}$$

Exercise 11. (Easy) Prove this last equation using the trigonometric identities.

Now we want to order the z_i's by length. To do this, start with a circle of radius 1, and consider a semicircle subdivided into 17 equal parts (cf. Klein p.28), enumerated $O, A_1, \ldots, A_{17}$. The the angle $A_k A_{17} O$ is half of the angle formed by A_1, A_{17} and the center of the circle by the formula

$$\text{arc length} \;=\; \text{radius} \;\times\; \text{angle}\,.$$

(Do you see why?) The arc length $A_k O$ is precisely $2\pi k/34$. Knowing this, the length of the chord S_k is given by

$$S_k = 2\sin\frac{\pi k}{34} = 2\cos\frac{(17-k)\pi}{34}$$

Exercise 12. (Easy) A *chord* is a straight line joining two points on a circle. If the two points are separated by an angle θ from the center of the circle, then show using the Pyhagorean theorem that the length of the chord in a circle of radius one is given by

$$2\sin\Big(\frac{\theta}{2}\Big).$$

Now we shall use this to examine the $z_i's$. Notice that the formula for z_i is related to that of the chord S_K. Indeed, we claim that eight out of these 34 chords are equal to the $z_i's$, since $z_* = 2\cos(2\pi k'/17)$, which is related to the previous formula by $k = 17 - 4k'$. As we let k' range over 1 through 8, we obtain an ordering of the z_i's:

$$4, 6, 5, 7, 2, 8, 3, 1.$$

Then noticing that the chords $OA_k, A_k A_{k+l}, OA_{k+l}$ form a triangle, the triangle inequality immediately gives

$$S_{k+l} < S_k + S_l$$

since the chord $A_k A_{k+l}$ has length S_l.

Exercise 13. Putting these together, prove that $y_3 < y_2 < y_4 < y_1$ and $x_2 < x_1$. (Hint: pp.28–29 of Klein.)

Now we are ready to construct these roots, and build the heptadecagon!

The method now is as follows: with one circle we will construct the roots of every quadratic equation involved (Klein p.34). Our circle will be the circle of radius one, centered at (0,1), $x^2 + (y-2)^2 = 1$, or equivalently,

$$x^2 + y(y-2) = 0. \tag{4.1}$$

Now given any quadratic equation $x^2 - px + q$ with distinct real roots x_1 and x_2, we will show how to construct the roots on the x-axis. We will consider the parallel lines $y = 0$ and $y = 2$ tangent to our circle. Connect the points $(q/p, 0)$ and $(4/p, 2)$ on the two lines. Since our roots are real, we must have $p^2 - 4q \geq 0$, and so

$$\frac{p}{4} \geq \frac{q}{p} \Rightarrow \frac{p}{4} \geq \frac{q}{p}$$

hence the picture in Klein. We want to show that this line intersects the two rays connecting x_1 and x_2 to $(0, 2)$. The equations of the rays are

$$2x + x_1(y-2) = 0, \quad 2x + x_2(y-2) = 0.$$

Multiply the two equations together and subtract the result from (4.1) to get

$$\frac{x_1+x_2}{2}x(y-2)+\frac{x_1x_2}{4}(y-2)^2-y(y-2)=0$$

This equation gives the intersections of the two lines with the circle. We can of course remove the factor $(y-2)$ (since we know that that point is). And from what's left we have

$$\frac{p}{2}x+\frac{q}{4}(y-2)-y=0$$

since $x_1+x_2=p$ and $x_1x_2=q$. Observe that this line intersects $y=0$ at $\frac{q}{p}$ and the line $y=2$ at $\frac{4}{p}$.

Now we use four quadratic equations to construct the length z_1, the ones with solutions $x_1>x_2, y_1>y_2, y_4>y_3$, and $z_1>z_2$. From the relevant equations (Klein p.36) we see that we only need to construct x_1, x_2, y_1, y_4 from these eight.

Considering the equations in Exercise 14 below, we will have to mark on $y=2$ the points

$$-4, \frac{4}{x_1}, \frac{4}{x_2}, \frac{4}{y_1}$$

and on $y=0$,

$$4, -\frac{1}{x_1}, -\frac{1}{x_2}, \frac{y_4}{y_1}$$

which will produce in the end for us z_1.

Remark 4.1. Okay! Let's finish this off. The following exercises should give us the remaining tools that we need to do out construction. Most of the answers should already be indicated in Klein, your job is to either understand his fuddy duddy language or solve it yourself!

Exercise 14. Show that the equations

$$x^2+x-4=0 \text{ has roots } x_1, x_2,$$
$$y^2-x_1y-1=0 \text{ has roots } y_1, y_2,$$
$$y^2-x_2y-1=0 \text{ has roots } y_3, y_4,$$
$$z^2-y_1z+y_4=0 \text{ has roots } z_1, z_2$$

Hint: The relations between the roots are as in p.10. Prove any statements you use.

Exercise 15. Prove that the two lines connecting (0,2) with the intersections of $4/p$ and q/p with the unit circle, will intersect the x-axis at the roots of $x^2-px+q=0$. (pp.34–35).

Exercise 16. Prove that $4/x_1$ is obtained, as in (Fig. 5, p.36), by extending the line connecting O and the intersection of $\overline{Ax_1}$ and the circle. Then prove that $-1/x_1$ is obtained by extending the line connecting $4/x_1$ and the intersection of $\overline{41}$ with $\overline{AO}$, as in (Fig. 6, p. 38). Conclude that you get y_2 from this.

Exercise 17. Use the same procedure to obtain y_4 (Fig. 7, p. 39).

Exercise 18. Show that the line connecting $\overline{4y_4}$ extended to the y-axis, then connecting back to $4/y_1$ intersects the x-axis at y_4/y_1 (Fig. 8, p. 40). Use this to get z_1.

Exercise 19. Finally, show that the line $\overline{Az_4}$ intersects $y = 1$ at exactly $x = \cos(2\pi/17)$, as in (Fig. 9, p. 41). Conclude that the perpendicular line meets the circle at the first and sixteenth (or second and seventeeth) side of the 17-gon.

Remark 4.2. You made it! Of course, as was remarked at the end of class, this process is a little mystifying. The point to keep in mind is that we have four quadratic equations to solve, and every construction of the 17-gon boils down to some way of getting your hands on these roots. In fact, I found a possibly easier method! But we've come this far; better to make our time worthwhile. Next week we will look at *both* methods.

Exercise 20. Draw the heptadecagon.

5. Transcendence

We have spent a lot of time so far talking about whether numbers are rational or not, that is, whether a number can be written in the form

$$\frac{a}{b}$$

for some pair of integers $a, b \in \mathbb{Z}$ (which is a fancy way of saying 'fractions'). But what do they look like on the so-called number line? Clearly the integers just look like an endless string of evenly spaced dots; the rationals numbers $\mathbb{Q}$, on the other hand, are referred to as being *dense* on the number line. Let's be a little technical for a moment:

Definition 5.1. Let A be a subset of B. We say A is *dense* in B if for every element b in B, every neighborhood of b also contains some element a of A. What do I mean by a neighborhood? In our case, if you pick any real number y in $\mathbb{R}$, then a neighborhood of y is the set of real numbers x within some fixed distance of y, i.e.,

$$\{x \in \mathbb{R} : |x - y| < C\}$$

where C is the fixed distance. So what we want to say is that the $A = \mathbb{Q}$ is dense in $B = \mathbb{R}$. This will be useful because it means that we can approximate irrational numbers rather well using only fractions.

Proposition 5.2. *$\mathbb{Q}$ is dense in $\mathbb{R}$.*

Proof. We'll let x stand for an arbitrary real number. What we want to show is that in every neighborhood there is a rational number in that neighborhood. (As you might guess, there's really infinitely many rationals in any neighborhood, but all we need to find is one.)

The idea is as saying the sequence of rational numbers 0.9, 0,99, 0.999, ... approximate 1. Since x is a real number it has a decimal expansion, which we'll write as

$$x_0.x_1x_2x_3\ldots$$

So we simply write down the sequence $x_0, x_0.x_1, x_0.x_1x_2, \ldots$. Then we can see that for any distance C, there is some n large enough that

$$|x - x_0.x_1 \ldots x_n| < C$$

and since every $x_0.x_1 \ldots x_n$ is a rational number, e.g., the fraction

$$\frac{x_0x_1 \ldots x_n}{10^n}$$

the same way $0.999 = 999/1000$, this proves the claim. □

So the rationals are dense in the set of real numbers. But obviously they are not all the real numbers, since $\sqrt{2}$ and the like are not rational. So along the number line, the rationals look like a very thick set of points, with lots of tiny holes. Question: how many tiny holes are there? The answer is not intuitive: there's more holes than there are rationals, even though the rationals are dense! In other words, the irrational numbers are also dense in $\mathbb{R}$, and even more so!

First, we separate irrational numbers into two kinds:

Definition 5.3. We say a real number x is *algebraic* if it is the solution to the polynomial equation

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$$

with coefficients $a_1, \ldots, a_n$ all integers. A real number that is not real algebraic is *transcendental*.

For example, the n-th roots $\sqrt[n]{2}$ is an algebraic number for every n (do you see why?). Note that i is algebraic in this sense too, but not *real algebraic*. We will only concern ourselves with real algebraic numbers for now.

We first need a notion of size, introduced by Cantor (in different terminology). Clearly, given two sets containing finitely many elements, we can say one is larger than the other if the number of elements, i.e., the *cardinality* of one set, is larger than the other.

How about infinite sets, like $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$? Let $\mathbb{N}$ denote the natural numbers $1, 2, 3, \ldots$ (or whole numbers, as we used to say).

Definition 5.4. Recall that a function f from a set A to B is *injective*, or one-to-one, if for every element a there is a unique element b such that $f(a) = b$. For example, $f(x) = 1/x$ is injective, but $f(x) = x^2$ is not (why?).

Recall also that a function is *surjective*, or onto, if for every b there is some a such that $f(a) = b$. For example, x^2 is not surjective but x^3 is. A function that is both injective and surjective is called *bijective*.

Definition 5.5. We say a set A is *countable* if there is an injective function f from A to $\mathbb{N}$, and hence they are in bijection with one another. (If moreover f is a bijection, then we say A is infinitely countable.)

Proposition 5.6. *The integers $\mathbb{Z}$ and the rationals $\mathbb{Q}$ are countable sets.*

Exercise 21. (Easy) Prove this. Fun: look up 'Hilbert's Grand Hotel'.

The way we prove this last statement gives us an arithmetic of infinity: countable + countable = countable, and countable $\times$ countable = countable. Of course, countable $-$ countable = countable, though it may be finite, or even empty (which is still countable, but kind of in a dumb way).

Proposition 5.7. *The set of algebraic numbers are also countable.*

Proof. Order the algebraic numbers by height, where the height of an algebraic number is N if it is the solution of a polynomial of degree n where

$$N =: n - 1 + |a_0| + \cdots + |a_n|,$$

so that if we order by height we have broken up the set of algebraic numbers into countably many sets. If we can show that each set is countable, then the sets altogether will be countable (this might not be true..). In fact, we can see that there are only finitely many equations given a height N, hence only finitely many algebraic numbers of height N. □

And here is the kicker:

Proposition 5.8. *The set of transcendental numbers is uncountable.*

Exercise 22. Prove this. This statement was shown by Cantor, using what we call Cantor's diagonal argument, which is quite simple, the idea being to exhibit a real number that does not lie in any countable subset of $\mathbb{R}$, by decimal expansions.

Definition 5.9. Now we have shown that the real numbers are a larger set than the natural numbers, where 'large' was defined using the notion of one-to-one functions. The size of a set, in this manner, we call the *cardinality* of the set. The cardinality $\mathbb{N}$ is denoted $\aleph_0$, and that of the real numbers is $2^{\aleph_0}$, referred to as the *continuum*.

Thus Cantor developed a method of distinguishing between different infinities, an idea that was controversial at his time (and still difficult to grasp) because as we emphasize in basic math courses, infinity is not a number. Indeed, Cantor went on to show that there were other cardinalities larger than $2^{\aleph_0}$, for example the power set of $\mathbb{R}$ (see below).

What is known as the continuum hypothesis states that there does not exist a cardinal number between $\aleph_0$ and $2^{\aleph_0}$. It turns out that this statement *cannot be proved nor disproved*, if we use the axioms for set theory that are standard today, called the Zermelo-Fraenkel axioms.

But let's take things down a notch:

Definition 5.10. Given a set A, the set of all subsets of A is called the *power set* of A, and denoted $P(A)$. For example, if $A = \{a, b, c\}$, then

$$P(A) = \{\{\varnothing\}, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$$

where $\varnothing$ denotes the empty set. Notice that the cardinality $|A| = 3$, while $|P(A)| = 2^{|A|} = 2^3 = 8$.

Exercise 23. Show that the power set $P(A)$ of any finite set A has cardinality $2^{|A|}$. Hint: suppose we enumerate the power set $\{A_1, \ldots, A_n\}$, how many possible choices of subsets can we make?

Remark 5.11. Here is something to think about: What is the set of all sets? Answer: such a thing cannot exist in set theory! But we can define the *category* of all sets, which leads to the wonderful world of category theory.

6. Square a circle

Now we want to answer an old problem: can a circle be constructed having the same area as a given square, or vice versa? Suppose it is the unit circle, then to make a square of equal area amounts to solving

$$\pi = x^2,$$

which leads to the question: is π constructible?

The answer, as we will see, is no. The proof will be a result of a much stronger statement, i.e., that π is transcendental. But first, we will need to know that e too is transcendental.

Theorem 6.1 (Hermite)**.** *e is transcendental.*

The proof will be by contradiction: we will pretend that e is algebraic, i.e., a solution to some polynomial

$$F(e) = c_n e^n + \cdots + c_1 e + c_0 = 0 \tag{6.1}$$

where the c_i's are integers and relatively prime (i.e., their gcd is 1), then arrive at a contradiction.

First consider the polynomial

$$\phi(x) = x^{p-1} \frac{((1-x)(2-x)\dots(n-x))^p}{(p-1)!}$$

where n is the degree of (6.1), and p is some prime number larger than n and $|c_0|$. Observe the following properties:

(1) For a fixed x, $\phi(x)$ tends to 0 as p tends to infinity. To see this, set $u = x(1-x)(2-x)\dots(n-x)$, so that

$$\phi(x) = \frac{u^{p-1}}{(p-1)!}\frac{u}{x}$$

Then as p tends to infinity, the expression tends to zero.

Exercise 24. Show that for any real number a,

$$\lim_{n\to\infty} \frac{a^n}{n!} = 0.$$

(There's many ways of doing this.)

Definition 6.2. Now we define a (strange) function as follows: our function $\phi(x)$ has a series expansion in terms of powers of x,

$$f(x) = \sum_k a_k x^k,$$

define a new series, defined on integers $m \geq 0$,

$$\Phi(m) = \sum_k a_k k!,$$

that behaves such that if we shift $\phi(m)$ by $\phi(m+n)$ where n is another integer, then the coefficients of f will change

$$\phi(m+n) = \sum_k a_k (m+n)^k = \sum_k a'_k m^k$$

as a power series in m, then we require

$$\Phi(m+n) = \sum_{k}^{n} a'_k k!$$

Note: It's not clear that this function is well-defined, e.g., is $\Phi(1+1) = \Phi(2)$?

Now here are the remaining two properties:

(2) **Claim A:** $\Phi(m)$ is an integer and is not divisible by p, hence non-zero.

To see this we examine the coefficients of $\phi(x)$ as a series in x. Expanding the products in $\phi(x)$, the exponent on x is $p-1$ and the highest is $np+p-1$. So when we replace the x^k by $k!$ we get the expression

$$\sum_{k=p-1}^{np+p-1} \frac{a_k}{(p-1)!} k!$$

where the coefficients a_k are integers (do you see why?). The first coefficient can be written as

$$\frac{(1\dot{2}\dots n)^p (p-1)!}{(p-1)!} = (n!)^p$$

and all subsequent coefficients are divisible by p (cf. binomial theorem).

(3) **Claim B:** If m is an integer and $1 \le k \le n$, $\Phi(m)$ is an integer divisible by p.

To see this, first note that in $\phi(m+k)$ one of the factors is $-m$, since

$$\phi(m+k) = (m+k)^{p-1} \frac{((1-m-k)(2-m-k)\dots(-m)\dots(n-m-k))^p}{(p-1)!},$$

so the term of lowest degree of m in the expansion is p (why?), and now we have

$$\sum_{k=p}^{k=np+p-1} c'_k m^k$$

Arguing as before, all the terms now are divisible by p (unlike before, there is no $p-1$ term). The first term c_p is

$$\frac{(1-)^{kp} k^{p-1} ((k-1)!(n-k)!)^p p!}{(p-1)!} = (-1)^{kp} k^{p-1} ((k-1)!(n-k!))^p p$$

which is divisible by p. You still have to convince yourself that it is an integer.

We now show that the equation (6.1) is impossible, by showing that

$$\Phi(m)F(e) = c_n \Phi(m) e^n + \dots + c_1 \Phi(m) e + c_0 \Phi(m)$$

can be written as an integer part that is nonzero and a fractional part that can be made to be zero, hence an equality between something that is zero and nonzero. We will decompose the coefficients $c_i \phi(h) e^i$ into an integer part and a fractional part.

Claim C:

$$e^k \Phi(m) = \Phi(m+k) + (\text{fractional part})$$

where the fractional part will tend to zero when we take p to be very large.

To see this, consider the terms

$$e^i\phi(m) = e^i \sum_k a_k x^k,$$

where each summand has a coefficient that involves

$$e^i x^k = (\sum_{n=0}^{\infty} \frac{i^n}{n!}) x^k,$$

so that for $n = 1, \ldots, k$, the coefficients are integers, and for $n > k$, we will have fractions. We'll use $\Phi(m)$ and $\phi(x)$ in a somewhat roundabout way: expand the power series to get

$$e^i x^k = x^k + \frac{i}{1}x^k + \frac{i^2}{2!}x^k + \cdots + \frac{i^k}{k!}x^k + \frac{i^{k+1}}{(k+1)!}x^k + \ldots,$$

then we'll take advantage of this substitution of x^k by $k!$, but only partially. We rearrange:

$$e^i x^k = x^k + \frac{i}{1}kx^{k-1} + \frac{i^2}{2!}k(k-1)x^{k-2} + \cdots + \frac{i^{k-1}}{(k-1)!}k!x + \frac{i^k}{k!}k! + \frac{i^{k+1}}{(k+1)!}k! + \ldots,$$

What the heck is going on with this? Here is the idea: For first k terms, since we are multiplying with x^k, thought of as $k!$, the denominator is cancelled and so the first k terms are *integers*. For the same reason, the terms after k will be *fractional*, because the denominator will not cancel entirely. More suggestively, the last line can be rewritten as

$$e^i x^k = x^k + \frac{k}{1}x^{k-1}i + \frac{k(k-1)}{2!}x^{k-2}i^2 + \cdots + \frac{k!}{(k-1)!}xi^{k-1} + i^k + \ldots$$

which is the integral part, plus

$$\cdots + \frac{i^{k+1}}{(k+1)} + \frac{i^{k+2}}{(k+1)(k+2)}\cdots,$$

which is the fractional part. (Note that here i is indexing integers, not the square root of -1.) Now observe that the integral part is in fact the binomial expansion of the product $(x+i)^k$ (this proves Claim C), while the fractional part being i^k times

$$0 + \frac{i}{(k+1)} + \frac{i^2}{(k+1)(k+2)}\cdots,$$

is strictly less than

$$e^i = 1 + \frac{i}{1} + \frac{i^2}{2\cdot 1} + \frac{i^3}{3\cdot 2\cdot 1} + \ldots$$

so the fractional part can be written as $q_{ik}e^i i^k$, where $0 < q_{ik} < 1$.

Claim D: The fractional part tends to zero as p tends to infinity.

[The rest is in p.67, follows quickly from here..]

Theorem 6.3 (Lindemann)**.** *e is not intertranscendental, that is, it is not a solution to a polynomial of the form*

$$c_0 + c_1(e^{k_1} + \cdots + e^{k_N}) + c_2((e^{l_1} + \cdots + e^{l_N})) + \cdots = 0 \tag{6.2}$$

where the c_i's are integers, and each family the exponents $k_1, \ldots, k_N$ are the complete set of roots of a algebraic equation of degree N.

Corollary 6.4. *e does not satisfy any equation of the form*

$$c_0 + c_1 e^{k_1} + c_2 e^{l_1} + \dots$$

where the $k_1, l_1, \dots$ are independent algebraic numbers, and so are the c_i's.

This immediately shows that π is not an algebraic number, seeing that we have the equation $1 + e^{i\pi} = 0$.

7. The parallel axiom

I'll only give the cliff notes here:

Axiom 7.1 (Euclid)**.**

Axiom 7.2 (Archimedes)**.**

Axiom 7.3 (Aristotle)**.**

Axiom 7.4 (Proclus)**.**

Axiom 7.5 (Clairaut)**.**

Axiom 7.6 (Clavius)**.**

Axiom 7.7 (Simson)**.**

Axiom 7.8 (Playfair)**.**

Exercise 25. Pick a statement that is equivalent to the Parallel axiom (not necessarily one listed above), and prove the equivalence.

8. Mid-Term: The story so far

Now for the big reveal: I will tell you what you have learned so far, in the language of modern algebra which is used today. This material would be typically covered in a first or second course in abstract algebra, and we will take the point of view of how the theory is used to study the famous problems in geometry we have looked at so far. This exercise will prove the following statement, stated informally:

Theorem 8.1. *The collection of extensions of $\mathbb{Q}$ generated by the roots of $x^3 - 2$ have the symmetries of an equilateral triangle.*

Note: A lot of what you will be asked to do is check that something you have studied satisfies a bunch of definitions. This will seem tedious, but it will tell you that there is structure to the objects you have studied, and you can abstract this structure and apply to seemingly unrelated situations.

8.1. **Group theory.** Let's begin with a simple idea:

Definition 8.2. A **group** $(G, \cdot)$ is a set G with a binary operation $\cdot$ which takes two elements in G and gives a third, satisfying the following axioms:

(1) (Closure) For any a and b in G, $a \cdot b$ is also in G.
(2) (Identity) There is an element e in G such that for any a in G, we have
$$e \cdot a = a \cdot e = a.$$
(3) (Associativity) For any a, b, and c in G, we have
$$(a \cdot b) \cdot c = a \cdot (b \cdot c).$$
(4) (Inverse) For any a in G, there is some b in G such that
$$a \cdot b = b \cdot a = e.$$
The inverse of a, if it is unique, will generally be denoted a^{-1}.

Moreover, if for any a and b in G, we have $a \cdot b = b \cdot a$, then we call $(G, \cdot)$ a *commutative* or *abelian* group. When the group operation is clear, we will simply write G instead of $(G, \cdot)$ for our group.

Finally, if H is a subset of G, and $(H, \cdot)$ and $(G, \cdot)$ are groups under the same operation, we say H is a *subgroup* of G.

Example 8.3. The set of real numbers under addition forms a commutative group $(\mathbb{R}, +)$, with e being 0, called the additive identity. The rational numbers $\mathbb{Q}$ form a subgroup of $\mathbb{R}$.

Exercise 26. The n-roots of unity form a group for any n. This group is typically denoted μ_n. Prove this for $n = 5$, that μ_5 is a group. (Hint: the operation $\cdot$ is usual multiplication and the identity e is 1, called the multiplicative identity.

A *finite* group is one that contains a finite number of elements, for example μ_n is a finite group for any n, while $(\mathbb{R}, +)$ is not. An important family of finite groups are the groups of symmetries of regular n-gons, called the dihedral group D_n. It works like this:

Consider an equilateral triangle (a regular 3-gon) with vertices 1,2, and 3, labelled clockwise. (This isn't important, you just have to remember how you label them.) There are two operations: (1) rotation by 120°, which we denote r, which sends
$$r : (1, 2, 3) \to (2, 3, 1)$$

and hence r^3 means rotating by 360°, which means not at all, hence it is the identity $r^3 = e$. (2) flipping across the perpendicular line through 1, which we denote s, sending

$$s : (1, 2, 3) \to (1, 3, 2)$$

and again we see $s^2 = 1$. These two operations generate a finite group of order $2n = 6$, this set can be expressed as

$$\{e, r, r^2, s, sr, sr^2\},$$

where the convention is to read from right to left, e.g., sr means rotate, then flip. To see that these six elements generate all the symmetries of the triangle, we have to show that every possible combination of r and s will result in some element of the set above. We will do this by filling out a multiplication table for s and r:

	e	r	r^2	s	sr	sr^2
e						
r						
r^2						
s						
sr						
sr^2						

Exercise 27. (Easy) Fill out the above multiplication table using only e, r, r^2, s, sr, and sr^2. (Hint: if you get stuck, cut out a triangle and do the operations physically!) Use this to infer that D_3 is not a commutative group.

Remark 8.4. As you can see, this representation of D_3 is not unique, i.e., we could have chosen s to be flipping along one of the other two axes of symmetry, or r to be rotation by 240°. But no matter which we choose, the multiplication table will still be a 6×6.

Remark 8.5. There are other basic examples of finite groups, for example the group of permutations: the set S_n of all possible permutations of n objects forms a group. For example $S_3 = \{(123), (132), (12), (23), (13), e\}$, where (123) means $1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 1$. Coincidentally, it is true that D_3 can be identified with S_3 (do you see why?), but in general the groups D_n an S_n are distinct.

8.2. **Field theory.** Great! Now you know about groups. Let's spice things up a little:

Definition 8.6. A **field** $(F, +, \cdot)$ is a commutative group with two operations, which we'll call addition $+$ and multiplication $\cdot$, satisfying the following axioms:

(1) (Closure) For any a and b in F, $a \cdot b$ and $a + b$ are also in F.
(2) (Identity) There is an element 1 in F such that

$$1 \cdot a = a \cdot 1 = a,$$

and an element 0 in F such that

$$a + 0 = 0 + a = a$$

for any a in F.
(3) (Associativity) For any a, b, and c in F, we have the relations

$$(a \cdot b) \cdot c = a \cdot (b \cdot c)$$

and

$$(a+b)+c = a = (b+c).$$

(4) (Inverse) For any a in G, there is some b, c in F such that

$$a \cdot b = b \cdot a = 1$$

if $a \neq 0$,[1] and

$$a + c = c + a = 0.$$

The multiplicative inverse of a, if unique, will generally be denoted a^{-1}, and the additive inverse $-a$.

(5) (Distributivity) For any a, b, c in F,

$$a \cdot (b+c) = a \cdot b + a \cdot c.$$

Finally, the commutativity requires that $a \cdot b = b \cdot a$ and $a + b = b + a$ for any a, b in F. But in all cases that we will consider this should be clear.

Remark 8.7. Fun fact: It can be shown that if the multiplicative and additive identities are equal, i.e., $0 = 1$, then the field F must consist only of that one element. Hence we will assume in our definition of a field that $0 \neq 1$.

Example 8.8. The familiar rational $\mathbb{Q}$, real $\mathbb{R}$, and complex $\mathbb{C}$ numbers all form fields under the usual arithmetic operations.

Exercise 28. Denote by $\mathbb{Q}(\sqrt{2})$ the field of rational numbers, adjoined with $\sqrt{2}$. A typical element in this field is written as

$$a + b\sqrt{2}, \quad a, b \in \mathbb{Q},$$

for example $-\frac{1}{2} + \frac{2}{3}\sqrt{2}$. Show that under the arithmetic operations $+, \times$, the set $\mathbb{Q}(\sqrt{2})$ forms a field.

The last exercise was tedious but straightforward. Now that you have done this, you can see that the proof works $\mathbb{Q}(\alpha)$, where α is any algebraic number. Moreover, if we add a finite number of algebraic numbers, $\mathbb{Q}(\alpha_1, \ldots, \alpha_n)$, the result is again a field. This process of adding elements to fields creates what we call (finite) *field extensions*.

Remark 8.9. Translated thus, the quest for constructibility was the search for particular extensions of $\mathbb{Q}$. The rational operations of compass and straightedge were the field operations in $\mathbb{Q}$, and the extraction of square roots corresponds to successively adjoining roots of quadratic equations. Note that once we have extended the field once to $\mathbb{Q}(\alpha)$, we can make a new number $a + b\alpha$, and then take the square root of this number. This was the process of getting nested square roots.

8.3. **Galois theory.** Now for the pretty stuff of how field extensions behave.

Definition 8.10. If a polynomial factors completely in a field F, we say that $p(x)$ splits in F, and that F is the splitting field of $p(x)$.

Now let's consider the familiar algebraic number $\sqrt[3]{2}$, which we'll denote by α. It satisfies the polynomial equation

$$p(x) = x^3 - 2.$$

[1] This prevents $0 = 1$.

Here is a subtlety passed over briefly in Klein's discussion of trisecting an angle: Even though α is a root of $p(x)$, the polynomial $p(x)$ will not factor completely in the field $\mathbb{Q}(\alpha)$, into

$$(x-\alpha)(x-\alpha')(x-\alpha'')$$

where the α's all lie in $\mathbb{Q}(\alpha)$.

Remark 8.11. If you have some linear algebra background, note that an element in $\mathbb{Q}(\alpha)$ can be expressed as $a+b\alpha+c\alpha^2$, a vector space over $\mathbb{Q}$ of dimension 3 with basis $1, \alpha, \alpha^2$. Do you see why α^2 must be an element of the field $\mathbb{Q}(\alpha)$?

To split $p(x)$, we must extend the field a little more. Over the field $\mathbb{Q}(\alpha)$, you can check immediately the polynomial $p(x)$ factors as

$$p(x) = x^3 - 2 = (x-\alpha)(x^2+\alpha x+\alpha^2).$$

Exercise 29. Use the quadratic formula to show that the remaining roots are

$$(\frac{-1\pm\sqrt{-3}}{2})\alpha,$$

then show that this factor $\dfrac{-1\pm\sqrt{-3}}{2}$ is a third root of unity, and conclude that the polynomial $p(x)$ factors completely in $\mathbb{Q}(\sqrt[3]{2}, \sqrt{-3})$.

Definition 8.12. Recall that the minimal polynomial of an algebraic number α is the lowest degree irreducible polynomial that α satisfies. In our case, the minimal polynomial of $\sqrt[3]{2}$ is $p_1(x) = x^3 - 2$. Then we define the **degree** of the extension $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$ to be the degree of the minimal polynomial satisfied by α.

Example 8.13. The extraction of square roots of integers, as in the classical problems we have considered, produces *quadratic extensions*, i.e., extensions of degree two, for example $\mathbb{Q}(\sqrt{2})$ and more generally, $\mathbb{Q}(\sqrt{b^2-4ac})$.

$\mathbb{Q}(\sqrt[3]{2})$ and $\mathbb{Q}(\sqrt{-3})$ have degree 3 and 2 over $\mathbb{Q}$ respectively. Since we know that the two complex roots of $p(x)$ do not lie in $\mathbb{Q}(\sqrt[3]{2})$, the polynomial

$$x^2+3$$

still does not split over $\mathbb{Q}(\sqrt[3]{2})$, thus generates an extension of degree 2 over $\mathbb{Q}(\sqrt[3]{2})$. So $\mathbb{Q}(\sqrt[3]{2}, \sqrt{-3})$ is a degree 6 extension. (Do you see why the minimal polynomial for $\sqrt{-3}$ is x^2+3?)

Remark 8.14. In our study of constructibility of real numbers, we needed to know the notion of an *irreducible* polynomial. Notice that $p(x)$ is irreducible over $\mathbb{Q}$ (and hence $\mathbb{Z}$, by Gauss' lemma) by the rational root theorem.

Now Galois theory is the study of symmetries possessed by roots of polynomials. We would like to talk about the Galois group of a field extension; to do this we will need some more technology:

Definition 8.15. Let G and H be groups. A **homomorphism** of groups $\sigma : G \to K$ is a map satisfying

$$\sigma(a\cdot b) = \sigma(a)\cdot\sigma(b)$$

for every a, b in G, where on the left $\cdot$ denotes the group operation in G and on the right that in H. If the map σ is one-to-one and onto, i.e., a bijection, we call σ an isomorphism. Finally, if $G = H$, then σ maps the group to itself and we call σ an automorphism. Now we have the language to say: D_3 and S_3 are *isomorphic*.

Exercise 30. Consider the field of complex numbers. Show that the operation of complex conjugation, sending

$$\sigma : a + ib \mapsto a - ib, \quad a, b \in \mathbb{R}$$

is an automorphism of $\mathbb{C}$. (Notice further that σ leaves real numbers fixed. We say $\mathbb{R}$ is a fixed field of the automorphism σ.)

I'll prove this next statement for you:

Proposition 8.16. *Let K be a finite extension of a field F. The set of automorphisms of K that fix F, denoted*

$$\mathrm{Aut}(K/F) = \{\sigma : \sigma(x) = x \text{ for every } x \in F\}$$

forms a group under composition.

Proof. If σ and τ are any two automorphisms of K fixing F, then the composition $\sigma \circ \tau$ acts by the identity on F, since given any x in F,

$$\sigma \circ \tau(x) = \sigma(\tau(x)) = \sigma(x) = x,$$

this verifies (1). The identity homomorphism (and it is a homomorphism)

$$\mathrm{Id} : x \mapsto x$$

clearly fixes F, this verifies (2). Associativity (3) follows immediately from composition of maps. Finally, the inverse τ^{-1} exists because it is bijection, and it fixes F, i.e., $\tau(x) = x$ for any x in F, then we see that

$$x = \tau^{-1}(\tau(x)) = \tau^{-1}(x).$$

This verifies (4). □

Definition 8.17. Let K be a finite extension of F. We say K is a Galois extension if the size of $\mathrm{Aut}(K/F)$ is equal to the degree of K over F, and moreover the define the **Galois group** of K over F to be the automorphism group $\mathrm{Aut}(K/F)$, denoted $\mathrm{Gal}(K/F)$.

Now we are ready to study the Galois group of $\mathbb{Q}(\sqrt[3]{2}, \sqrt{-3})$ over $\mathbb{Q}$. For convenience, let's write

$$K = \mathbb{Q}(\sqrt[3]{2}, \sqrt{-3}), \quad \alpha = \sqrt[3]{2}, \quad \zeta = \frac{-1 + \sqrt{-3}}{2}.$$

Then any automorphism of K fixing $\mathbb{Q}$ will send the roots α to $\alpha, \zeta\alpha$, or $\zeta^2\alpha$ and ζ to ζ or ζ^2. Since K is a Galois extension, we know that the number of automorphisms is 6, the degree of K over $\mathbb{Q}$. So our job is to find these six automorphisms. To that end, define

$$\sigma : \begin{cases} \alpha \mapsto \zeta\alpha \\ \zeta \mapsto \zeta \end{cases}, \qquad \tau : \begin{cases} \alpha \mapsto \alpha \\ \zeta \mapsto \zeta^2 \end{cases} \qquad 1 : \begin{cases} \alpha \mapsto \alpha \\ \zeta \mapsto \zeta \end{cases}$$

Notice that these are automorphisms of K that fix $\mathbb{Q}$.

Exercise 31. Show that $\sigma\tau = \tau\sigma^2$, and that $\sigma^3 = \tau^2 = 1$. Then conclude that by identifying σ with r and τ with s, we have

$$\mathrm{Gal}(K/\mathbb{Q}) \simeq D_3$$

and equivalently, S_3. Congratulations! You have proved Theorem 8.1 with some pretty sweet math.

Remark 8.18. Here is a foretaste of what is to come: We have considered finite extensions of $\mathbb{Q}$ by adding algebraic numbers, but what if you add *all* the algebraic numbers? We call this infinite field extension $\overline{\mathbb{Q}}$, the algebraic closure of $\mathbb{Q}$, the smallest field where every algebraic number can be found, i.e., where every polynomial with rational coefficients splits completely. It is a subfield of the complex numbers, $\mathbb{C}$ as there are still missing the transcendental numbers. So we can now add to the chain of inclusions

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \overline{\mathbb{Q}} \subset \mathbb{C}$$

where once we had $\mathbb{R}$ between $\mathbb{Q}$ and $\mathbb{C}$. The two fields $\mathbb{R}$ and $\overline{\mathbb{Q}}$ are not directly comparable, because $\mathbb{C}$ is degree 2 over $\mathbb{R}$, one only needs to add i (we say $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$), while $\overline{\mathbb{Q}}$ contains the algebraic number i, but is missing all the transcendental numbers, that is to say, $\mathbb{C}$ has infinite degree over $\overline{\mathbb{Q}}$.

Remark 8.19. Here is a historical tidbit: the eponymous Évariste Galois, lived in the early 1800s and died in a duel when he was 20. Nonetheless, by then he had already shown that in general polynomials of degree 5 and above cannot be solved using radicals. In contrast, for quadratic polynomials we have the quadratic formula, and similarly there are similar formulas for the cubic and quartic, the latter being related to the first two. Galois initiated the study of roots of formulas using their symmetries, encoded by what we now call the Galois group.

9. Make a triangle

Let's think about the equation

$$a^2 + b^2 = c^2.$$

Since we're thinking about triangles, we'll ask for a, b, c to be all positive. Suppose we wanted to know when a, b, c are moreover whole numbers, then we could view this as asking for when a square c^2 can be represented as a sum of two squares.

Definition 9.1. Any solution (a, b, c) to this problem is called a *pythagorean triple*, because of its connection to the Pythagorean theorem.

Indeed, the Greeks already had a way to produce all such solutions. The Babylonians before them found at least 15 different solutions, the largest being

$$(12709, 13500, 18541).$$

Of course, given any pythagorean triple (a, b, c), any integer multiple (na, nb, nc) will be another pythagorean triple. (Do you see why?) So we will only care about the case where a, b, c are relatively prime, i.e., $\gcd(a, b, c) = 1$.

Exercise 32. Here is a stronger statement: If (a, b, c) is a Pythagorean triple where any two of the numbers have a common divisor n, then the third number will also be divisible by n. Prove this.

Definition 9.2. As a consequence, we can further restrict ourselves to (a, b, c) where no two numbers have a common divisor. We'll call these *primitive* pythagorean triples.

Remark 9.3. Using what we know now, given a pythagorean triple we can divide the equation $a^2 + b^2 = c^2$ by c^2 to get an equivalent equation

$$\left(\frac{a}{c}\right)^2 + \left(\frac{b}{c}\right)^2 = 1.$$

The two numbers on the left are rational, and they satisfy the equation for the unit circle. So we call the pair $(\frac{a}{b}, \frac{a}{c})$ a rational points on the circle. Conversely, given any rational solution (x, y) to the equation $x^2 + y^2 = 1$, we can produce a pythagorean triple by clearing denominators. Thus the problem of finding pythagorean triples is equivalent to finding rational points on the unit circle.

Let's find all rational points on the the circle. We'll need the following construction: Start with the point (0,1) and another point P on the unit circle, and extend the line connecting the two down to the x-axis. If the line intersects the x-axis at the point $(0, r)$, then the equation for the line is

$$y = 1 - \frac{x}{r}.$$

Having the equation for the line and the circle, we can solve for the point P:

$$P = \Big(\frac{2r}{r^2 + 1}, \frac{r^2 - 1}{r^2 + 1}\Big).$$

Exercise 33. Prove this.

Remark 9.4. Now notice that P will be a rational point of r is a rational number, and conversely r will be rational if P is a rational points, for then the slope of the line must be rational. So finding a rational points on the circle is equivalent to finding rational points on the line.

Let now $r = \frac{p}{q}$, then the resulting Pythagorean triple is

$$(a, b, c) = (2pq, p^2 - q^2, p^2 + q^2)$$

(check this). Now we get all triples:

Theorem 9.5. *Any primitive pythagorean triple is of the form* $(a, b, c) = (2pq, p^2 - q^2, p^2 + q^2)$*, up to interchanging* a *and* b*, where* $p > q$*,* $\gcd(p, q) = 1$*, and have opposite parity, i.e, one is even and the other odd.*

Proof. In our statement we make the harmless assumptions that the fraction $\frac{p}{q}$ is reduced, hence p and q have no common factor, and that $p > q$.

Case 1: Suppose p and q have opposite parity. Then $p^2 - q^2$ and $p^2 + q^2$ are odd, so that if $2pq, p^2 - q^2, p^2 + q^2$ have a common divisor d, it must be odd. Further, if d divides $p^2 - q^2, p^2 + q^2$ then it also divides their sum $2p^2$ and difference $2q^2$. But since d is odd it must divide p and q, but by assumption p and q have no common factors, so $(2pq, p^2 - q^2, p^2 + q^2)$ is primitive.

Case 2: Suppose p and q have equal parity. Then they must be odd because if even then they have a common divisor. But then their sum and difference must be even, so write them as $2P$ and $2Q$. Any common factor will divide

$$P + Q = p, P - Q = q,$$

so P and Q must be relatively prime. Then we may express

$$(2pq, p^2 - q^2, p^2 + q^2) = 2(P^2 - Q^2, 2PQ, P^2 + Q^2),$$

but now we are back in Case 1 because P and Q can't both be odd for otherwise $p = P + Q$ and $q = P - Q$ would both be even, which is impossible because they have no common factor. □

Exercise 34. (Easy.) Since we already have some complex arithmetic under our belts, check this out: consider the complex conjugates $p+iq$ and $p-iq$. Computing

$$(p + iq)^2(p - qi)^2 = ((p + qi)(p - qi))^2$$

leads to the pythagorean identity

$$(p^2 - q^2)^2 + (2pq)^2 = (p^2 + q^2)^2,$$

which are precisely our a, b, c's.

10. Sum two squares

Now instead of asking when a square is a sum of two squares, we'll ask when any integer is a sum of two squares.

In 1640, Fermat announced that

Theorem 10.1. *An odd prime p is the sum of two squares*

$$p = x^2 + y^2$$

with $x, y \in \mathbb{Z}$ if and only if $p \equiv 1 \mod 4$.

but did not prove this. In 1749 Euler gave the first full proof of this statement. We will prove this in several steps; in the following we will always take our variables to be whole numbers.

Lemma 10.2 (Diophantus)**.** *The product of two sums of squares is again a sum of squares. That is,*

$$(a^2 + b^2)(p^2 + q^2) = (ap + bq)^2 + (aq - bp)^2 = (ap - bq)^2 + (aq + bp)^2. \tag{10.1}$$

Proof. Direct computation. □

Lemma 10.3. *If $a^2 + b^2$ is divisible by $p^2 + q^2$ and $p^2 + q^2$ is prime, then the quotient*

$$\frac{a^2 + b^2}{p^2 + q^2}$$

is a sum of two squares.

Proof. Since $p^2 + q^2$ divides $a^2 + b^2$, then $p^2 + q^2$ divides $p^2(a^2 + b^2) - a^2(p^2 + q^2)$ which can be expressed as

$$p^2(a^2 + b^2) - a^2(p^2 + q^2) = (pb - aq)(pb + aq).$$

Then since $p^2 + q^2$ is prime, it must divide one of the factors. If it divides the first factor, $(pb - aq)$ whic h is the second term on the right-hand side of (10.1), then $p^2 + q^2$ must also divide the first term, $ap + bq$. Then the square $(p^2 + q^2)^2$ must divide the equation (10.1), which is

$$\frac{a^2 + b^2}{p^2 + q^2} = \left(\frac{ap + bq}{p^2 + q^2}\right)^2 + \left(\frac{aq - bp}{p^2 + q^2}\right)^2.$$

But we see that the two terms on the right are integers, so we have written the quotient as a sum of two squares. □

Lemma 10.4. *If a $a^2 + b^2$ is divisible by a number x that is not a sum of two squares, then the quotient has a factor that is not a sum of two squares.*

Proof. Suppose the quotient has prime factorization $p_1 p_2 \dots p_n$. If all of the p_i can be written as a sum of two squares, then by the previous lemma we may divide

$$a^2 + b^2 = x p_1 p_2 \dots p_n$$

by all of the p_i and still have a sum of two squares on the left, which is now equal to x. But this contradicts our assumption that x was a not a sum of two squares, so there must be some p_i that is not a sum of two squares. □

Lemma 10.5. *If a and b are relatively prime, then every factor of $a^2 + b^2$ is a sum of two squares.*

Proof. Let x be a factor of $a^2 + b^2$. By the division algorithm write

$$a = mx \pm c, \quad b = nx \pm d,$$

where $c, d \leq x/2$. (For if we have $a = mx + c$ where $2c > x$, then can write $a = (m+1)x + (1-c)$ instead.) Then

$$a^2 + b^2 = (m^2x \pm 2mc + n^2x \pm 2nd)x + (c^2 + d^2),$$

so $c^2 + d^2$ must be divisible by x also, say $c^2 + d^2 = yx$. If c and d have a common factor, their gcd must be relatively prime to x, for otherwise the three terms on the right will have a common factor, but a^2 and b^2 are assumed to be relatively prime. So the gcd must divide y instead, we write $e^2 + f^2 = zx$ with e, f relatively prime, and $z <$, since

$$e^2 + f^2 \leq c^2 + d^2 \leq (\frac{x}{2})^2 + (\frac{x}{2})^2 = \frac{x^2}{2}.$$

If x is not the sum of two squares, then by the previous lemma there must be a factor of z that is not the sum of two squares, call it w. Then from x we have arrived at a smaller number w that is also not a sum of two squares but dividing a sum of two squares.

But this argument tells us that we there given any x we can produce such a w infinitely many times, but of course this is impossible because x is a finite positive number. This final step is called infinite descent. □

Lemma 10.6. *If $p \equiv 1 \bmod 4$, then p is a sum of two squares.*

Proof. Let $p = 4n + 1$. Then by Fermat's Little Theorem, the numbers

$$1^{4n}, 2^{4n}, \ldots, (4n)^{4n}$$

are congruent to 1 mod p. Label these as $a_1, \ldots, a_n$. The difference between any two *consecutive* terms

$$a_{i+1}^{4n} - a_i^{4n} = (a_{i+1}^{2n} + a_i^{2n})(a_{i+1}^{2n} - a_i^{2n})$$

is congruent to 0 mod p, hence divisible by p. p being prime, it divides one of the two factors on the right-hand side. If p divides the first factor, which is a sum of two squares, then p is a sum of two squares. (Since $a_{i+1} = a_i + 1$, they are indeed relatively prime and the previous lemma applies.)

To see that p does not divide the second factor, observe that if p divides all the $a_{i+1}^{2n} - a_i^{2n}$ then it also divides the $a_{i+2}^{2n} - a_i^{2n}$ and so on.

Exercise 35. (Hard.) At this point one has to show that p cannot divide all the differences $a_{i+1}^{2n} - a_i^{2n}$. One way to do this is by showing that these successive differences, which are called *finite differences*, which we are considering here of the sequence

$$1^{2n}, 2^{2n}, \ldots, (2n)^{2n}$$

are constant with value $(2n)!$. Then use the fact that $p = 4n + 1$ to show that p cannot divide $(2n!)$, that is, $(2n)! \not\equiv 0 \bmod p$.

□

Exercise 36. Look up an alternative proof of Lemma 10.6 that convinces you, and write it in your own words.

Remark 10.7. The Lemma of Diophantus has a cute expression: it shows that the product of two sums of squares can be written as a sum of two squares in two different ways, e.g.,

$$65 = 5 \cdot 13 = 8^2 + 1^2 = 4^2 + 7^2.$$

11. Imagine that

Before we continue our story, we'll build up some background on complex numbers and complex function theory.

Let's think back to the problem of trisecting an angle, which came down to the irreducibility of a cubic equation. The solution to a general quadratic was known to the Babylonians, but for a long time it was not known how to solve a general cubic.

We pick up the story in the 15th century, with the Italian mathematician Scipione del Ferro, who discovered the solution to the so-called depressed cubic

$$x^3 + px = q,$$

where in the coefficient of x^2 is zero. Note that the appearance of this equation avoids the appearance of negative numbers, because in the spirit of Euclid negative numbers had no meaning. The idea is to set $x = u + v$, leading to

$$u^3 + v^3 + (3uv + p)(u + v) = q$$

so that the solution reduces to

$$3uv + p = 0, \qquad u^3 + v^3 = q.$$

Solving for v in terms of p and q leads to a quadratic equation in u^3, thus

$$u^3 = \frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}.$$

Choosing the positive root, the solution to the depressed cubic is

$$x = \sqrt[3]{\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Exercise 37. Fill in the details.

As del Ferro took p and q to be positive, the resulting x is always a real number. Indeed, any cubic will have at least one real root.

Then given one root a_1 of the cubic, we can easily find the other two roots:

$$(x - a_1)(x - a_2)(x - a_3) = (x - a_1)(x^2 - (a_2 + a_3)x + a_2a_3)$$

by factoring, then the remaining a_2 and a_3 follow.

Example 11.1. Consider $x^3 + 6x = 20$, in which case del Ferro's formula gives

$$x = \sqrt[3]{10 - \sqrt{108}} + \sqrt[3]{10 + \sqrt{108}} = 2$$

But how does this come out to 2? You can check that 2 is a root of the cubic, then factoring you get

$$(x - 2)(x^2 + 2x + 10)$$

you find the other two roots are complex.

Exercise 38. Here's how to solve a general cubic from del Ferro's solution. Start with

$$ax^3 + bx^2 + cx + d = 0,$$

divide by a and substitute $x = t - b/3a$ to get the depressed cubic

$$t^3 + pt + q = 0$$

with

$$p = \frac{3ac - b^2}{3a^2}, \qquad q = \frac{2b^3 - 9abc + 27a^2d}{27a^3}.$$

FIll in the details. Now you see why it took literally thousands of years between solving the quadratic and cubic!

11.1. **Some complex function theory.** From this point on we'll see what comes out of the geometry of the upper-half plane. We'll do this by studying a special class of functions on $\mathbb{H}$, but first we'll need to know some complex analysis.

Definition 11.2. A function f from $U \subset \mathbb{C}$ to $\mathbb{C}$ is called holomorphic at a point z_0 if the complex derivative

$$f'(z_0) := \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists. If the limits exists for all $z \in \mathbb{C}$ then we say f is holomorphic (or complex analytic or complex differentiable).

This condition is much stronger than in calculus of a single real variable. Recall that if a function $f : \mathbb{R} \to \mathbb{R}$ is differentiable, its derivative might not be differentiable. If indeed it is infinitely differentiable, we call the function real analytic, and it is equal to its Taylor series expansion. In contrast, in the complex world we have the following:

Theorem 11.3. *If $f(z)$ is complex differentiable at z_0 then it is infinitely differentiable, moreover it is equal to its Taylor series at z_0.*

Proof. A first course in complex analysis. □

Definition 11.4. Let f be a function from $\mathbb{C}$ to $\mathbb{C}$. We call f meromorphic in $\mathbb{C}$ if it is holomorphic everywhere except on a set of isolated points. These points will be called the poles, or singularities of f.

An isolated point of a set X, roughly speaking, is a point x such that we can find an open set containing x but not any other element of X. For example, $\mathbb{Z}$ is a set of isolated points in $\mathbb{R}$, and

$$\mathbb{Z} + \mathbb{Z}i = \{x + iy \in \mathbb{C} : x, y \in \mathbb{Z}\}$$

is a set of isolated points in $\mathbb{C}$.

Later on we will need the following property of complex functions:

Definition 11.5. Let if be a holomorphic (resp. meromorphic) function on an open subset $U \subset \mathbb{C}$. If there exists a holomorphic (resp. meromorphic) function F on V another open subset containing U such that

$$F(z) = f(z)$$

for all z in U, then we call $F(z)$ the analytic continuation of $f(z)$.

Remark 11.6. Just in case you are worried: we won't do any difficult complex analysis here. You just need to know that differentiation rules in $\mathbb{C}$ behave the same as derivatives in $\mathbb{R}$, so you can essentially pretend $f(z)$ is $f(x)$ when you differentiate (most of the time).

Definition 11.7. Here are some quick facts: Any periodic, infinitely differentiable function f has a convergent Fourier series, i.e., if $f(z+\omega) = f(z)$ for all z and a fixed ω, then we have the Fourier expansion

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i(nz/\omega)} := \sum_{n=-\infty}^{\infty} a_n e\Big(\frac{nz}{\omega}\Big)$$

where the Fourier coefficients are

$$a_n = \frac{1}{\omega}\int_0^{\omega} f(z) e\Big(-\frac{nz}{\omega}\Big) dz.$$

Example 11.8. You know periodic functions: all the trig functions you have ever seen! Indeed, functions on $\mathbb{R}$ that are periodic can be viewed as functions on a circle, because they repeat themselves, like going around in circles. Indeed, the identity

$$e^{ix} = \cos x + i \sin x,$$

relates functions on the circle to trigonometric functions. This is the beginning of what is called harmonic analysis.

Definition 11.9. Notice that if $f(z+\omega_1) = f(z+\omega_2) = f(z)$ for all z and ω_1/ω_2 is rational, then $f(z)$ is again a (simply) periodic function. If irrational but real, then by complex analysis $f(z)$ reduces to a constant function. Finally, $f(z)$ is properly doubly periodic if $\mathrm{Im}(\omega_1/\omega_2) > 0$, and if moreover $f(z)$ is meromorphic (see below), we call $f(z)$ an *elliptic function*. Notice that any integer combination $n\omega_1 + m\omega_2$ with $n, m \in \mathbb{Z}$ is also a period, and the set of all these forms a lattice in the plane, denoted by

$$\Omega = \{n\omega_1 + m\omega_2 : n, m \in \mathbb{Z}\} = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2.$$

We can form the fundamental domain of $f(z)$, which is the parallelogram P with vertices $0, 1, \omega_2/\omega_1, 1 + \omega_1/\omega_2$. Identifying the edges on P, we can view f as a function on the torus.

Example 11.10. The Weierstraß $\mathcal{P}$ function

$$\mathcal{P}(z) = \frac{1}{z} + \sum_{\omega\in\Omega,\omega\neq 0} \Big(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\Big)$$

converges for all $z \notin \Omega$. It is an elliptic function, with a pole at the origin, and hence at every lattice point ω. It is one of the simplest(!) elliptic functions.

12. Sum three squares

So in 1640 Euler proved the claim of Fermat that an odd prime p can be written as a sum of two squares if and only if $p \equiv 1 \bmod 4$. As a corollary, any integer that can be written as a sum of two squares must have prime factorization such that the primes $p \equiv 3 \bmod 4$ occur only in even powers.

In 1770 Lagrange proved that every integer can be written as a sum of four squares.

In 1798 Legendre proved that every natural number can be written as a sum of three squares if and only if it is not congruent to 7 mod 8 nor divisible by 4.

In 1801 Gauss showed the number of ways an integer can be written as a sum of three squares. This is the result we will now discuss, as it will lead us to geometry. To describe his formulas, we introduce some notation:

Definition 12.1. Let $r_k(n)$ denote the number of distinct representations of the number n as a sum of k squares, i.e., solutions to

$$n = x_1^2 + \cdots + x_k^2$$

taking into account both sign and order.

Also define $\sigma(n)$ to be the number of divisors of n, and $d_1(n)$ and $d_3(n)$ the number of divisors of n that are of the form $4m + 1$ and $4m + 3$ respectively.

13. Sum four squares

In 1813 Cauchy proved that every integer can be written as a sum of n n-gonal numbers.

In 1834 Jacobi proved that the number of ways to represent n as a sum of 2, 4, 6, and 8 squares.

Theorem 13.1 (Jacobi)**.** *Let $n = 2^r n_1 n_3$ where n_1 and n_3 are the prime factors congruent to 1 and 3 mod 4 respectively. Then*

$$r_2(n) = \begin{cases} 4(d_1(n) - d_3(n)) & \text{if } n_3 \text{ has an odd exponent} \\ 0 & \text{otherwise} \end{cases}$$

and

$$r_4(n) = \begin{cases} 8\sigma(n) & \text{if } n \text{ is odd} \\ 24\sigma(k) & \text{if } n \text{ is even, } n = 2^r k, \ k \text{ odd} \end{cases}$$

The formulas for $r_6(n)$ and $r_8(n)$ are more complicated, which we omit here.

Example 13.2. To illustrate $r_2(n)$, notice that 3 cannot be written as a sum of two squares, but $90 = 9^2 + 3^2$. On the other hand, the formula for $r_4(n)$ tells us that every integer can be written as a sum of four squares.

Let's prove the sum of four squares. First, a warm up:

Exercise 39. Justify in words the following:

$$\sum_{n=1}^{\infty} \frac{x^n}{1-x^n} = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} x^{nk} = \sum_{n=1}^{\infty} x^n \sum_{k|n} 1 = \sum_{n=1}^{\infty} d(n) x^n,$$

where $k|n$ means the positive integers k dividing n, and $\sigma(n)$ is the number of divisors of n.

Definition 13.3. To describe the proof, we introduce the *theta function*:

$$\theta(x) = \sum_{n=-\infty}^{\infty} x^{n^2} = 1 + 2\sum_{n=1}^{\infty} x^{n^2}.$$

Using the definition one calculates the products:

$$\begin{aligned} (\theta(x))^2 = \Big(\sum_{n=-\infty}^{\infty} x^{n^2} \Big) &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} x^{n^2+m^2} \\ &= \sum_{k=0}^{\infty} x^k \sum_{n^2+m^2=k} 1 = 1 + \sum_{n=1}^{\infty} r_2(n) x^n. \end{aligned}$$

And by the same method,

$$(\theta(x))^k = 1 + \sum_{n=1}^{\infty} r_k(n) x^n.$$

So if we knew the Taylor expansion of $(\theta(x))^k$ and that the coefficients were positive, then that would imply that any integer could be written as a sum of k squares. But already Euler knew this. The difficulty is in finding an expression for these coefficients.

Proof. Let's assume for the moment that we know

$$(\theta(x))^4 = 1 + \sum_{n=1}^{\infty} r_4(n)x^n = 1 + 8\sum_{m=1}^{\infty} \frac{mx^m}{1+(-x)^m}. \tag{13.1}$$

Observe that for m odd,

$$\begin{aligned}\sum_{m \text{ odd}} \frac{mx^m}{1+(-x)^m} &= \sum_{m \text{ odd}} mx^m(1 + x^m + x^{2m} + \dots) \\ &= \sum_{n=1}^{\infty} x^n \sum_{\substack{m|n \\ m \text{ odd}}} m = \sum_{n=1}^{\infty} x^n(d_1(n) + d_3(n));\end{aligned}$$

whereas for m even,

$$\begin{aligned}\sum_{m \text{ even}} \frac{mx^m}{1+x^m} &= \sum_{m \text{ even}} mx^m(1 - x^m + x^{2m} \dots) \\ &= \sum_{n=1}^{\infty} x^n\Big(\sum_{\substack{m|n \\ m \text{ even} \\ n/m \text{ odd}}} m - \sum_{\substack{m|n \\ m \text{ even} \\ n/m \text{ even}}} m\Big).\end{aligned}$$

So we need to evaluate the coefficient of x^n in this last series. If $m|n$ and $n = 2^k v$ with v odd, then $m = 2^a d$ with $d|v$ and $1 \le a \le k$. If n/m is odd, then only $a = k$ occurs; if n/m is even, then only $1 \le a \le k-1$ occurs. Then for a fixed n, the expression in parentheses is

$$\begin{aligned}&\sum_{\substack{d|n \\ d \text{ odd}}} d(2^k - (2^{k-1} + 2^{k-2} + \dots + 2)) \\ &= \sum_{\substack{2d|n \\ 2d \equiv 2 \bmod 4}} 2d = \sum_{\substack{m|n \\ m \equiv 2 \bmod 4}} m = d_2(n).\end{aligned}$$

Then putting this back into (13.1) we have

$$1 + 8\sum_{n=1}^{\infty} x^n(d_1(n) + d_2(n) + d_3(n)) = 1 + 8\sum_{n=1}^{\infty} x^n \sum_{\substack{d|n \\ d \not\equiv 0 \bmod 4}} d,$$

then equating coefficients gives the result. □

Exercise 40. Convince yourself of the above.

It remains to prove the identity 13.1. This is the hard part. (Take a moment to look at the section on elliptic functions in the previous chapter.)

Define the more general theta function

$$\theta(v, \tau) = \sum_{n=-\infty}^{\infty} q^{n^2} e^{2\pi i n v}$$

where $q = e^{\pi i \tau}$. We see that this specializes to the $\theta(x)$ above by taking $q = x$ and $v = 0$. In general, the series converges for $|q| < 1$, or equivalently, for $\mathrm{Im}(\tau) > 0$ (do you see why?). We will assume this.

Exercise 41. From the definition, show that

(1) $\theta(v+1,\tau) = \theta(v,\tau)$
(2) $\theta(v+\tau,\tau) = q^{-1}e^{-2\pi iv}\theta(v,\tau)$
(3) $\theta(-v,\tau) = \theta(v,\tau)$.

where you get the last identity by rewriting $\theta(v,\tau) = 1 + 2\sum_{n=1}^{\infty} q^{n^2}\cos(2\pi nv)$ using Euler's identity for e^{ix}. Thus we see that theta functions are periodic, and very close to being elliptic functions.

Lemma 13.4 (Jacobi's triple product identity)**.** *Let $z = e^{2\pi iv}$. Then for $|q| < 1$ and $z \neq 0$,*

$$\theta(z,q) = \sum_{-\infty}^{\infty} q^{n^2} z^n = \prod_{n=1}^{\infty}(1-q^{2n})(1+q^{2n-1}z)(1+q^{2m-1}z^{-1}).$$

Proof. Consider the function

$$F(z) = \prod_{n=1}^{\infty}(1+q^{2n-1}z)(1+q^{2n-1}z^{-1}),$$

this converges for $|q| < 1$ and $z \neq 0$. Its zeroes are exactly those of $\theta(v,\tau)$ (take this for granted). From the definition, we see that $F(e^{2\pi i(v+1)}) = F(e^{2\pi iv})$. On the other hand,

$$F(e^{2\pi i(v+\tau)}) = \frac{1+q^{-1}z^{-1}}{1+qz}F(z) = \frac{1}{qz}F(z). \tag{13.2}$$

Exercise 42. Prove Equation 13.2. Hint: the first equality follows just by working through the definition, for the second equality you only need to show

$$1 + q^{-1}z^{-1}1 + qz = \frac{1}{qz}.$$

Then show that the ratio $\theta(z,q)/F(z)$ is invariant both under $v+1$ and $v+\tau$ (it takes some work), so we get an elliptic function!

Now, it turns out that the ratio is a constant with respect to v (again by complex analysis), but it could be a function of q, so we'll write

$$\theta(z,q) = T(q)\cdot F(z),$$

where $T(q)$ is determined as follows: We'll write the expression

$$G(q) = T(q)\prod_{n=1}^{\infty}(1-q^{2n})^{-1}$$

in two ways, by the expressions

$$\theta(-1,q) = T(q)\prod_{n=1}^{\infty}(1-q^{2n-1})^2$$

and

$$\theta(i,1) = T(q)\prod_{n=1}^{\infty}(1+q^{4m-2}).$$

Substituting these two into $T(q)$ respectively, we find that $G(q) = G(q^4)$. Moreover, one sees that $G(q) = G(q^{4^k})$ for all integers $k \geq 1$ (by induction, say). And since $|q| < 1$, we see that

$$G(q) = \lim_{k\to\infty} G(q^k) = G(0) = 1$$

and we have our expression for $T(q)$, which proves the identity. □

Exercise 43. (Hard.) Derive the formula for $T(q)$ using the steps outlined above.

Sketch of proof of Equation 13.1. Now we hand wave the rest of the proof, which relies on functions called Lambert series, which in our case, leads to the expression

$$f(v,\tau) = \frac{\pi}{\sin \pi v} + 4\pi \sum_{n=1}^{\infty} \frac{q^{2n-1}}{1-q^{2n-1}} \sin(2n-1)\pi v$$

Evaluated at $v = \frac{1}{2}$, gives the identity

$$\pi^2(\theta(1,q))^4 = f(\frac{1}{2},\tau)^2. \tag{13.3}$$

Then one expresses $f(v,\tau)^2$ as a difference of Weierstraß functions,

$$f(v,\tau)^2 = \mathcal{P}(v,1,\tau) - \mathcal{P}(\frac{\tau}{2},1,\tau) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \Big\{ \frac{1}{(v+m+n\tau)^2} - \frac{1}{(\tau/2+m+n\tau)^2} \Big\}.$$

Now using a technique called the Poisson summation formula, we arrive at

$$f(v,\tau)^2 = \frac{\pi_2}{\sin^2 \pi v} + 8\pi^2 \sum_{n=1}^{\infty} \frac{nq^n}{1-q^{2n}} - 8\pi^2 \sum_{n=1}^{\infty} \frac{q^{2n-1}}{1-q^{2n}} \cos(2\pi n v).$$

And then by the identity (13.3) we have upon setting $v = \frac{1}{2}$,

$$\theta^4(1,q) = 1 + 8\sum_{n=1}^{\infty} \frac{nq^n}{1-q^{2n}} - 8\sum_{n=1}^{\infty} (-1)^n \frac{nq^{2n}}{1-q^{2n}},$$

which, with $q = x$, gives us $(\theta(x))^4$.

Exercise 44. This last step you can do. To complete the proof of Equation 13.1, show that

$$\sum_{n=1}^{\infty} \frac{nq^n}{1-q^{2n}} - \sum_{n=1}^{\infty} (-1)^n \frac{nq^{2n}}{1-q^{2n}} = \sum_{n=1}^{\infty} \frac{nx^n}{1+(-x)^n}.$$

□

And we are done. What a work out! And we skipped a whole lot too!

14. A little hyperbolic

14.1. **Noneuclidean geometries.** Now we want to negate the parallel axiom. Recall that the parallel axiom can stated as:

(0) Given a line L in the plane and a point P not on L, then there exists a *unique* line L' passing through P and not intersecting L.

We can negate the parallel axiom in two ways:

(1) Given a line L in the plane and a point P not on L, then there *does not exist* a line L' passing through P and not intersecting L. This leads us to *ellipitic geometry*. In this world, the sum of the angles in a triangle add up to greater than π, and the typical example is the sphere. More on this later.
(2) Given a line L in the plane and a point P not on L, then there exists *more than one* line L' passing through P and not intersecting L. This leads us to *hyperbolic geometry*. In this world, the sum of the angles in a triangle add up to less than π, and we will spend much time here.

14.2. **The upper-half plane model.** For the moment, let's think about hyperbolic geometry. There are two well-known models of the hyperbolic plane, and we'll look at both of them. First, there is the *upper-half plane*:

$$\mathbb{H} = \{z = x + iy \in \mathbb{C} : y > 0\}$$

where we have written x and y for the real and imaginary part of z respectively.

Here's one way of thinking about the upper-half pane: we'll extend the group of transformations of the real projective line $\mathbb{RP}^1$ to the plane. That is, we will consider the action of the *special linear group*

$$SL_2(\mathbb{R}) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1; a, b, c, d \in \mathbb{R}\}$$

on $\mathbb{H}$ by the map

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z \mapsto \frac{az + b}{cz + d}.$$

We will call these fractional linear transformations, or Möbius transformations.

Remark 14.1. In class we discussed a larger group $GL_2(\mathbb{C})$, where we relax the determinant condition $ad-bc \neq 0$, and entries in $\mathbb{C}$. This is the general linear group, or the group of invertible matrices with entries in $\mathbb{C}$. (From linear algebra we know that a matrix having nonzero determinant is equivalent to having an inverse.) This set forms a group with the action given by matrix multiplication. It contains as subgroups

$$SL_2(\mathbb{Z}) \subset SL_2(\mathbb{R}) \subset GL_2(\mathbb{R}) \subset GL_2(\mathbb{C})$$

where we will focus on the first two.

Definition 14.2. If in a group G we look at the set of all elements z in G that commute with every element g in G, i.e.,

$$z \cdot g = g \cdot z,$$

these elements for a group called the *center* of G, which we will denote by Z. Observe that if G is a commutative group, then $Z = G$. Further, the center

$GL_2(\mathbb{R})$ is the set of diagonal elements,

$$\begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}$$

with $k \neq 0$.

Exercise 45. (Easy.) In the upper-half plane, we only consider matrices with determinant one because the property that for any nonzero scalar k,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d} = \frac{kaz+kb}{kcz+kd} = \begin{pmatrix} ka & kb \\ kc & kd \end{pmatrix} \cdot z = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z$$

shows that the action of $GL_2(\mathbb{R})$ on $\mathbb{H}$ reduces to the action of $SL_2(\mathbb{R})$. Argue this.

Exercise 46. (Easy.) In class we showed that matrices of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

leave the point $z = i$ fixed. Show the converse, i.e, if a matrix leaves i invariant,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot i = i$$

then it must be of the form above. We call these matrices the special orthogonal group SO_2, as they have determinant $a^2 + b^2 = 1$. These correspond to rotations in the plane, and the solutions to this equation are parametrized by $0 \leq \theta \leq 2\pi$. That is, we can write

$$SO_2 = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : 0 \leq \theta \leq 2\pi \right\}$$

The following claim will show the connection between $SL_2(\mathbb{R})$ and $\mathbb{H}$.

Proposition 14.3. *The action of $SL_2(\mathbb{R})$ on $\mathbb{H}$ is an automorphism. In fact, it is a many-to-one map.*

Proof. Proved in class. It is quite easy to see that the map is a group homormophism. □

14.3. **The Poincaré disk model.** Here is a second model for hyperbolic geometry. It is the disk

$$D := \{z \in \mathbb{C} : |z| < 1\}$$

or also the pairs satisfying $x^2 + y^2 < 1$. We will define *hyperbolic lines* in the disk to be Euclidean lines or circles which meet the boundary of the disk at right angles. (The boundary of the unit disk is the unit circle.)

Lemma 14.4. *The equation of a hyperbolic line is either*

$$ax + by = 0,$$

with a, b not both zero, or

$$x^2 + y^2 + cx + dy = 0$$

with $c^2 + d^2 > 4$.

Sketch of proof. The first equation is a line through the origin, thus a Euclidean line. The second equation, we must show that it gives precisely circles that intersect D at right angles. □

14.4. **Mobius transformations.** Now let's describe the hyperbolic lines in the upper-half plane. We make the following claims:

(1) Any point on a Euclidean line, which is a line of the form $x = a$ from some $a \in \mathbb{R}$, is defined by $z = a + iy$, and also $\bar{z} = a = iy$. Therefore we can describe the line uniquely by
$$z + \bar{z} = 2a.$$
(2) We also want lines that are semicircles with centers on the x-axis. A circle with center a and radius r satisfies
$$r^2 = |z - a|^2_{\mathbb{C}},$$
where remind the reader the complex absolute value is
$$|x + iy|^2 = (x + iy)(\overline{x + iy}) = x^2 + y^2,$$
that is, it measure the regular Euclidean distance. So in our case,
$$r^2 = (z - a)(\bar{z} - \bar{a}) = z\bar{z} - a\bar{z} - \bar{a}z + a\bar{a},$$
but since a is a real number, $a = \bar{a}$, whence
$$z\bar{z} - a(z + \bar{z}) + a^2 - r^2 = 0.$$

Exercise 47. We see that (1) and (2) satisfy the equation
$$A|z|^2 + B(z + \bar{z}) + C = 0$$
for some A, B, C with A, B not both zero. Prove the converse: show that given any such equation, it reduces to either (1) or (2).

We need to know another fact: The group of Mobius transformations is generated by the following operations:

(1) Horizontal translations $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$,
(2) Dilations $\begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}$,
(3) Reflections $z \mapsto -\bar{z}$. It is a property of Mobius transformations that the reflection of any hyperbolic line can be realized as a finite number of Mobius transformations.

Let's now describe the standard metric on $\mathbb{H}$, which is referred to as hyperbolic distance. To measure points vertically, we define
$$d(ai, bi) = |\log \frac{b}{a}|.$$
This distance function enjoys the following properties: (1) it is additive: given $a < b < c$, we have by the additive property of logs (check:)
$$d(ai, ci) = d(ai, bi) + d(bi, ci).$$
Then, for example, we see that the sequence of points $2^n i$ on the imaginary axis have the same distance from each other. You may think of this as a kind of 'shrinking ruler' as you vertically approach the boundary of the upper-half plane.

(2) It preserves Mobius transformations. This requires more work, and is related to the property of Mobius transformations that they preserve hyperbolic lines.

Exercise 48. (Optional) Prove that Mobius transformations preserve hyperbolic lines.

15. Modular forms

Now we look at complex-valued functions f on $\mathbb{H}$ rather $\mathbb{C}$ just by restricting the domain of our function.

Definition 15.1. Let k be an integer. We say a function f on $\mathbb{H}$ is weakly modular of weight $2k$ if f is meromorphic on the half-plane $\mathbb{H}$ and verifies the relation

$$f(z) = (cz+d)^{-2k} f\Big(\frac{az+b}{cz+d}\Big), \qquad \text{for any } \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $ad - bc = 1$.

Proposition 15.2. *Let f be a meromorphic function on $\mathbb{H}$. Then f is weakly modular of weight $2k$ if and only if it satisfies*

$$f(z+1) = f(z)$$
$$f(-1/z) = z^{2k} f(z).$$

Proof. (Sketch.) The forward implication is straightforward to check. The other direction we will not prove here. □

Given the first condition $f(z+1) = f(z)$, f can be expressed as a function of $q = e^{2\pi i z}$, since it is periodic with respect to $\mathbb{Z}$ and hence descends to a function on the circle. It is a meromorphic in the disc $0 < |q| < 1$ with the origin removed. If in fact f extends to a meromorphic (resp. holomorphic) function at the origin, then we say that f is meromorphic (resp. holomorphic) at infinity. That is, f admits a Laurent expansion in the neighborhood of the origin

$$f(q) = \sum_{-\infty}^{\infty} a_n q^n,$$

where the a_n are zero for n small enough (resp. $n < 0$).

Remark 15.3. This course doesn't assume any complex analysis, but you call to mind what you know about Taylor and Laurent series for functions on $\mathbb{R}$.

Definition 15.4. A weakly modular function is called a *modular function* if it is meromorphic at infinity. A modular function which is holomorphic everywhere (including infinity) will be called a *modular form*. If moreover the modular form is zero at infinity, it is called a *cusp* form.

Thus a modular form of weight $2k$ is given by a series

$$f(z) = \sum_{n=0}^{\infty} a_n q^n,$$

which converges for $|q| < 1$, i.e., for $\mathrm{Im}(z) > 0$ and verifies the identity $f(-1/z) = z^{2k} f(z)$. It is a cusp form if $a_0 = 0$.

Example 15.5. The first example is the theta series

$$\theta(z) = \sum_{n \in \mathbb{Z}} e^{2\pi i n^2 z}$$

We'd like to check that this is a modular form of weight $k = 1/2$, but we can't do this just yet.

Example 15.6. Let $k > 1$ be an integer, and Γ a lattice in $\mathbb{C}$. Consider the function

$$G_k(z) = \sum_{(m,n)\neq(0,0)} \frac{1}{(mz+n)^{2k}}.$$

It is a modular form of weight $2k$, called the *Eisenstein series*. Let's denote

$$g_2 = 60G_2, \qquad g_2 = 140G_3.$$

and define $\Delta = g_2^3 - 27g^3$. One uses the fact that $G_k(\infty) = \zeta(2k)$ where $\zeta(s)$ is the Riemann zeta function

$$\sum_{n=1^\infty} \frac{1}{n^s},$$

to show that $\Delta(\infty) = 0$, i.e., a cusp form of weight 12. Then we find Eisenstein series in the Laurent expansion of the Weierstraß $\mathcal{P}$ function,

$$\mathcal{P}(z) = \frac{1}{u^2} + \sum_{n=2}^{\infty}(2k-1)G_k u^{2k-2},$$

and if we set $x = \mathcal{P}(z)$ and $y = \mathcal{P}'(z)$, the following differential equation is satisfied:

$$y^2 = 4x^3 - g_2x - g_3.$$

This equation defines an *elliptic curve,* and the discriminant of the cubic $4x^3 - g_2x - g_3$ is Δ.

16. Instructions: Final project

Welcome to your final project! This assignment will make up a good chunk of your grade, and rightly so. You are to think of this as a final paper to a writing class, or a final assignment in an art class, or the like. It should be a exercise in the techniques and ideas that you have been exposed to in our class, but certainly not limited to them.

Your assignment is to explore and extend a topic that relates to the theme of our class, number and shapes. This is intentionally broad, because the focus is on a topic that interests you, that you would like to explore further. It could be a deeper study into a topic that we have or will cover, or some real-world application, or even a piece of art.

Definition 16.1. The final project will be judged based on the following criteria:

(1) Content: The mathematical content of your work must be substantial, there must be some mathematical foundations that guide your audience through your project. If it is an artwork, explain how the mathematics enters into it; if it is a paper, prove all assertions.
(2) Form: A mathematical paper without narrative is at best dry, and at worst unreadable. If mathematics were a bunch of equations we wouldn't know how to make sense of them, much less care about what they can tell us. But the beauty is that mathematics so often tells us something about the world we live in. Your project should communicate that not only the truth but the beauty of what your topic.

On that note, if your topic is not true but beautiful, then it's wrong. If it is true but not beautiful, you should prefer to spend your time elsewhere. Although a mathematician was once quoted saying if they had to choose between the true and the beautiful, they would choose the latter. (Go figure.)

Example 16.2. Here are some examples of topics you could explore:

(1) More on transcendental numbers, e.g., special values of the Riemann zeta function, the Milllenium Problem (Riemann hypothesis); other transcendental numbers
(2) Questions about infinity and cardinal numbers, e.g., the Continuum Hypothesis
(3) Galois theory of field extensions, e.g., Abel's insolvability of the quintic polyomial
(4) Hyperbolic geoemetry, e.g., connections to physical phenomena, space-time curvature
(5) Mathematical art, e.g., projective, affine, etc. Can mathematical art be beautiful?
(6) Introduction to algebraic geometry, i.e., studying geometry using polynomial equations over any field
(7) Introduction to arithmetic geometry, i.e., how do things change as we vary the field of definition?
(8) The shape of space, e.g., what is the shape of the universe?
(9) Topological ideas: gluing, cutting, and pasting space. e.g., orientations and manifolds.

Deadlines: Starting from now till next week please meet me with a short up to one page proposal of what you would like to do. It need not be overly detailed, but it should be well-thought out, with an outline of what you plan to do. If you can't meet during office hours we'll work out an alternative.

Monday May 2: First draft. As is usual, the more complete your draft is the better feedback I can give to you.

Wednesday May 9: Final project presentations. Your final work need be turned in at this point depending on the form of it.

Wednesday May 16: Tentative absolute due date.

Remark 16.3. A note on extensions: I will most likely be flexible with due dates, on the condition that the work that you turn in late is excellent. If not, it will only do you harm to miss a deadline.

17. Some solutions

(1) 1. The idea is the following: if $\sqrt{a}$ is a root, then so must $-\sqrt{a}$. Similarly for nested radicals, we have to flip all the signs to get all the roots. Thus we are looking at all (eight) combinations of

$$\pm\sqrt{3 \pm \sqrt{\pm\sqrt{2}}}.$$

Consider first the conjugates of the form $**+$,

$$\begin{aligned}&(x - \sqrt{3 + \sqrt[4]{2}})(x + \sqrt{3 + \sqrt[4]{2}})(x - \sqrt{3 - \sqrt[4]{2}})(x + \sqrt{3 - \sqrt[4]{2}})\\ &= (x^2 - (3 + \sqrt[4]{2}))(x^2 - (3 - \sqrt[4]{2}))\\ &= (x^2 - 3)^2 - \sqrt{2},\end{aligned}$$

while the remaining ones $**-$ give similarly:

$$\begin{aligned}&(x - \sqrt{3 + \sqrt{-\sqrt{2}}})(x + \sqrt{3 + \sqrt{-\sqrt{2}}})(x - \sqrt{3 - \sqrt{-\sqrt{2}}})(x + \sqrt{3 - \sqrt{-\sqrt{2}}})\\ &= (x^2 - (3 + \sqrt{-\sqrt{2}}))(x^2 - (3 - \sqrt{-\sqrt{2}}))\\ &= (x^2 - 3)^2 + \sqrt{2}.\end{aligned}$$

Then multiplying the two together gives by the binomial theorem

$$(x^2 - 3)^4 - 2$$

which opens up to our candidate polynomial $\phi(x)$

$$\begin{aligned}&(x^2)^4 + 4(x^2)^3(-3) + 6(x^2)^2(-3)^2 + 4(x^2)(-3)^3 + (-3)^4 - 2\\ &= x^8 - 12x^6 + 54x^4 - 108x^2 + 79\end{aligned}$$

Checking that $f(79)$ and $f(-79)$ are not zero, irreducibility follows from the rational root theorem.

18. Appendix A: Chebotarev density and other primes

First let's recall Euclid's result from antiquity:

Theorem 18.1 (Euclid)**.** *There are infinitely many primes.*

This is a very nice statement, and the proof is very short.

Exercise 49. Prove this. Bonus: find a proof of this statement different from Euclid's proof.

Then much, much later, we have the following result of Dirichlet on primes in arithmetic progressions:

Theorem 18.2 (Dirichlet)**.** *Let a and d be relatively prime, then there are infinitely many prime numbers $p \equiv a \mod d$.*

You don't see this immediately, but the theorem of Chebotarev below can be seen as a generalization of this. The proof of this theorem requires to show that a certain object $L(s, \chi)$ known as the Dirichlet L-function which generalizes the Riemann zeta function $\zeta(s)$, is nonzero at $s = 1$.

To give you a little more feeling for this, χ is a complex valued function on the integers that is periodic mod m, then we define the L-function to be

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s},$$

and you see immediately that this specializes to $\zeta(s)$ when χ is identically 1 for all n.

Dirichlet's theorem is a much stronger statement: given any arithmetic progression $a+nd$ where a and d are coprime, like $4, 7, 10, 13, 16, \ldots$, infinitely many primes will occur in this progression.

Example 18.3. We looked at the extension $\mathbb{Q}(i)$. To be historically accurate, Gauss consider the 'complex integers' $\mathbb{Z}[i]$ (this is a ring, but you don't know that), which we now sometimes refer to as the Gaussian integers. That is, these numbers $\mathbb{Z}[i]$ are to $\mathbb{Q}(i)$ what $\mathbb{Z}$ is to $\mathbb{Q}$. He showed that if

$$p \equiv \begin{cases} 1 \mod 4 & \text{then } p \text{ splits completely} \\ 3 \mod 4 & \text{then } p \text{ remains prime (called inert)} \\ 2 & \text{then 2 ramifies} \end{cases}$$

Splits completely means p can be written as a product of distinct Gaussian integers, like

$$5 = (2 - i)(2 + i)$$

and ramifies means that it can be written with some Gaussian integer raised to some power, like

$$2 = (1 - i)(1 + i) = -i(1 + i)^2$$

Note that $1 - i$ is not a prime in $\mathbb{Q}(i)$ because it factors as $-i(1 + i)$. From this, we see that the splitting behaviour in $\mathbb{Q}(i)$ of the primes in $\mathbb{Q}$ approach $1/2$ fairly quickly. This leads to the question, how do primes split in other extensions? This is answered in part by Chebotarev's density theorem.

I'll fill in details later, but a first version goes like this:

Theorem 18.4 (Chebotarev)**.** *Given a finite Galois extension K of $\mathbb{Q}$ of degree n, the number of primes that split completely in K has density $1/n$ among primes.*

19. Appendix B: Some projective geometry

Let's start from a different 'point of view': perspective drawings, in the style of the Italian renaissance, were drawn using a method called *construzione legittima'*, or legitimate construction, where parallel lines meet at a horizon. While at first this seems mathematically strange, this makes a lot of sense to us intuitively, if say, you were looking at a long, straight road, and the two ends of the road would meet at the horizon.

Exercise 50. Make a perspective drawing.

While it turns out that *construzione legittima'* is in fact a Euclidean construction because it involves lines that are really parallel, we can in fact make a perspective drawing of a tiled floor using *straightedge alone*, no compass. (See Stillwell p.92.)

Definition 19.1. A *projective plane*, as we have described above, can be defined by the following axioms:

(1) Any two points are contained in a unique line,
(2) Any two lines contain a unique point,
(3) There exists four points, no three of which are in a line.

Notice that these axioms do not involve any notion of length or angle. The third axiom is a little strange at first, but it tells us that there the projective plane has 'enough points' to be interesting. An example of four such points are the vertices of a quadrilateral.

Okay, so we have a bunch of axioms, but that's no use if we don't have a way to work with this. Our first *model* of the projective plane will be called the *real projective plane*, denoted $\mathbb{RP}^2$, defined to be the set of lines in $\mathbb{R}^3$ through the origin. Then the 'points' of $\mathbb{RP}^2$ will be lines through the origin, and the 'lines' will be the unique plane in $\mathbb{R}^3$ containing two such lines.

Exercise 51. Check that $\mathbb{RP}^2$ satisfies the axioms for the projective plane. For the third axiom, use the points

$$(1,0,0),(0,1,0),(0,0,1),(1,1,1)$$

Each point defines a line through the origin. To check that no three lines lie on the same plane, use the general equation of the plane

$$ax + by + cz = 0$$

(in the way $ax + by = 0$ defines a line) to argue this.

Definition 19.2. Notice that the equation of a line $ax+by+cz = 0$, can be viewed as the triple (a, b, c), and if we multiply by a nonzero scalar, (ta, tb, tc) still define the same plane, in this manner point and lines in $\mathbb{RP}^2$ are independent of scaling. (The point (a, b, c) and (ta, tb, tc) lie on the same line through the origin.)

The equation $ax+by+cz = 0$ is called a *homogenous equation*, because the sum of the exponents of all variables in each term are equal. For example,

$$3x^2y + 5xyz + 2yz^2 + x^3 = 0$$

is a homogeneous equation.

It turns out that there is a geometric object—in fact,a surface—that behaves like $\mathbb{RP}^2$, but first let's think of something else: the real projective line $\mathbb{RP}^1$ can be

thought of as the 'points' in a single 'line' in $\mathbb{RP}^2$, or really, the set of coplanar lines through the origin. There is a nice visualization of this as as the tangent line to a circle.

Exercise 52. Write down an equation for the projection of the circle on to a tangent line. Convince yourself that this gives you a model for $\mathbb{RP}^1$.

But why is this called projective geometry? Each plane in $\mathbb{R}^3$ *not* passing through the origin defines a perspective view of the projective plane, a view that contains all but one 'line' of $\mathbb{RP}^2$. Each point in this plane defines a line through the origin, and the set of these lines gives us all the lines through the origin except those parallel to the plane. These missing lines form the 'horizon', or the 'line at infinity'.

Remark 19.3. Digression: There are several transformations of $\mathbb{RP}^1$, which amount to combinations of the functions

(1) kx, projection from a finite point
(2) $x + k$, projection from the point at infinity
(3) $\frac{1}{x}$, projection of lines not parallel to each other.

These generate the group of *linear fractional transformations*:

$$\frac{ax + b}{cz + d},$$

where $ac - bd \neq 0$. We will study this in detail later.

Remark 19.4. As it turns out, our model of the projective plane using lines and planes through the origin does not depend on the fact that we have used coefficients in $\mathbb{R}$. In fact, we can take coefficients to be in any field, for example $\mathbb{C}, \mathbb{Q}$ or more strangely, a finite field $\mathbb{F}_p$.

The complex projective plane we shall visit later, the rational projective plane is basically the real one except we only keep the rational points. The smallest finite field $\mathbb{F}_2$ of two points gives us $\mathbb{F}_2\mathbb{P}^2$ having seven points, called the Fano plane. Its points are

$$(1,0,0), (0,1,0), (0,0,1), (0,1,1), (1,0,1), (1,1,0), (1,1,1)$$

This is the smallest projective plane. It has a nice visualization, and the lines satisfy

$$x = y = z = 0,$$
$$x + y = y + z = x + z = 0,$$
$$x + y + z = 0$$

Exercise 53. Convince yourself that this construction make sense.

20. Appendix C: The p-adic numbers

Let's formalize our little digression today. It in fact serves as the best introduction to nonarchimedean geometry. We first need some definitions:

Definition 20.1. A *metric space* is a set X with a metric d, which is a function $d : X \times X \to \mathbb{R}$ satisfying the following axioms for any $x, y, z \in X$:

(1) $d(x, y) \geq 0$
(2) $d(x, y) = 0$ if and only if $x = y$
(3) $d(x, y) = d(y, x)$
(4) $d(x, z) \leq d(x, y) + d(y, z)$.

Definition 20.2. We will call a metric space *complete* if for any sequence (x_n) in X such that

$$d(x_i, x_j) \to 0$$

as i and j tend to infinity, there exists some y in X such that

$$d(x_i, y) \to 0$$

as i tends to infinity.

Example 20.3. In this discussion we will start with X to be the field of rational numbers $\mathbb{Q}$. The *archimedean absolute value* is the usual absolute value defined as

$$|x|_\infty = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x \leq 0 \end{cases}.$$

Then $\mathbb{Q}$ is a metric space with the metric $d(x, y) =: |x - y|_\infty$. It is not a complete metric space because it is missing the the irrational numbers. For example, there is a the sequence

$$3, 3.1, 3.14, 3.141, 3.1415, \ldots$$

approaching π, which is not an element of $\mathbb{Q}$. This is what we mean when we say $\mathbb{Q}$ has 'holes'.

Definition 20.4. Given any metric space X, one can construct a complete metric space $\hat{X}$ that contains X as a dense subspace (in the sense of §5). We sometimes call this the metric closure of X.

Example 20.5. Here is an analogy with closed and open sets: consider the space $(0, 1) \subset \mathbb{R}$, endowed with the usual metric $|\cdot|_\infty$. The limit of the sequence $\frac{1}{n}$ does not belong to $(0, 1)$, so the space is not complete with this metric. Its closure is, as you would expect, the interval $[0, 1]$.

Example 20.6. The metric closure of $\mathbb{Q}$ with the absolute value $|\cdot|_\infty$ is $\mathbb{R}$.

Definition 20.7. You may know that any integer can written as a product of prime numbers, for example

$$63 = 3^2 7.$$

Now for any prime number p, we can express any rational number as

$$\frac{a}{b} = kp^{-n}$$

where n is some integer and k is prime to p, i.e., it does not contain p as a factor. Then we define the p-adic absolute value to be

$$\left|\frac{a}{b}\right|_p = p^n,$$

and define also $|0|_p = 0$.

Exercise 54. (Easy.) Given

$$x = \frac{63}{550} = 2^{-1} \cdot 3^2 \cdot 5^{-2} \cdot \cdot 7 \cdot 11^{-1}$$

find the p-adic absolute values $|x|_p$ for $p = 2, 3, 5, 7, 11$, and 13.

As you can see, for different p, the absolute values of x look very, very different. (Of course, when $p = \infty$ then we have the usual $|63/550|_\infty = 63/550$.) Here is your first indication that the p-adic number look very different from the real numbers: by definition of the p-adic absolute value, the p-adic absolute values of rational numbers are only integer powers of p. In other words,

$$\{|x|_p : x \in \mathbb{Q}\} = \{p^n : n \in \mathbb{Z}\}$$

where as

$$\{|x|_\infty : x \in \mathbb{Q}\} = \mathbb{Q}.$$

(This should be obvious once you've thought about it.)

Moreover, all multiples $5, 10, 15, 20, 30, 35, \ldots$ all have the same 5-adic absolute value, which is 1.

Exercise 55. (Easy.) Is the following sequence increasing or decreasing in 5-adic absolute values:

$$125, 25, 5, 1, \frac{1}{5}, \frac{1}{25}, \frac{1}{125}, \ldots$$

Write down the 5-adic absolute values of each term.

Proposition 20.8. *The p-adic absolute value defines a metric on $\mathbb{Q}$,*

$$d(x, y) = |x - y|_p$$

satisfies the axioms for a metric function. (Convince yourself of this.)

On the other hand, we can see that $\mathbb{Q}$ is not metrically closed under the p-adic metric: Consider the number represented by

$$a_n =: 1 + 5 + 5^2 + 5^3 + \cdots + 5^n.$$

Its 5-adic absolute value converges to a real number as n tends to infinity. For example,

$$|a_2 - a_1|_5 = |(1 + 5 + 5^2) - (1 + 5)|_5 = |5^2|_5 = 5^{-2} = \frac{1}{25}$$

and more generally

$$|a_{n+1} - a_n|_5 = |5^n|_5 = \frac{1}{5^n},$$

so that as n tends to infinity the absolute value tends to 0.

Definition 20.9. So there exists a metric completion of $\mathbb{Q}$ with respect to the p-adic metric for each prime number p, this we call $\mathbb{Q}_p$, the field of p-adic rational numbers.

Now that we have completed p-adically, let's ask what arithmetic looks like in this world: sticking to $\mathbb{Z}_p$, a p-adic integer has an expansion

$$a_0 + a_1 5 + a_2 5^2 + a_3 5^3 + \ldots.$$

so it is uniquely determined by the string

$$(\ldots, a_3, a_2, a_1, a_0)$$

written in ascending order by convention. This should remind you of how bunary, or more generally p-ary number are represented, for example, the number 7 has binary expansion

$$1 + 2 + 2^2$$

written as 111.

Addition in the p-adic world means that we work base p. Instead of carrying over 10 in our usual base 10 world we carry over in base p, for example, 5-adically we add two 5-adic integers:

$$\begin{aligned} &\ldots 3\ 1\ 2\ 4\ 0\\ +&\ldots 1\ 2\ 4\ 2\ 1\\ =&\ldots 4\ 4\ 2\ 1\ 1 \end{aligned}$$

Subtraction is performed the same way, but the usual borrowing you learned in second grade gives us negative integers:

$$\begin{aligned} &\ldots 0\ 0\ 0\ 0\ 0\\ -&\ldots 0\ 0\ 0\ 0\ 1\\ =&\ldots 4\ 4\ 4\ 4\ 4 \end{aligned}$$

by borrowing from the left. This gives the expression $-1 = \ldots 4, 4, 4$ in $\mathbb{Z}_5$. Convince yourself that this makes sense. Indeed, this shows that $\mathbb{Z}_p$ forms an abelian group under addition.

Multiplication is easier done than said; it works still the same as in grade school. We perform an example in 5-adics:

$$\begin{aligned} &\ldots 3\ 1\ 2\ 4\ 0\\ \times&\ldots 1\ 4\ 1\ 2\ 1\\ =&\ldots 3\ 1\ 2\ 4\ 0\\ +&\ldots 3\ 0\ 3\ 0\\ +&\ldots 2\ 4\ 0\\ +&\ldots 1\ 0\\ +&\ldots 0\\ =&\ldots 0\ 1\ 0\ 4\ 0 \end{aligned}$$

Exercise 56. Multiply the 5-adic integer $(\ldots, 3, 1, 2, 4, 0)$ by the number 5.

Division, on the other hand, does not work always in $\mathbb{Z}_p$. For example, one can't find a p-adic integer a such that $pa = 1$, because multiplying a p-adic integer by p always gives a p-adic integer whose expansion ends in 0. This should not be surprising because division doesn't work in $\mathbb{Z}$ either. You have to pass to the rational numbers.

On the other hand,

$$\begin{aligned} &\ldots 2\ 2\ 2\ 2\ 3\\ +&\ldots 2\ 2\ 2\ 2\ 3\\ =&\ldots 0\ 0\ 0\ 0\ 1 \end{aligned}$$

shows that the number $(\dots, 2, 2, 2, 2, 3)$ is 5-adically 'one half'. More generally, given some p-adic integer α ending in 0, the number

$$\beta = 1 + \alpha + \alpha^2 + \alpha^3 + \dots$$

ends in 1, and we can multiply by $(1 - \alpha)$ to obtain the inverse

$$(1-\alpha)\beta = (1-\alpha)(1 + \alpha + \alpha^2 + \alpha^3 + \dots) = 1 - \alpha + \alpha - \alpha^2 + \alpha^2 + \cdots = 1$$

by what we know about infinite series. For p-adic integers not ending in 0 or 1, there is a more complicated algorithm to obtain the inverses. Then once again, $\mathbb{Q}_p$ forms a field under these operations.

Exercise 57. (p-adic unit ball) Let's get a handle on what $\mathbb{Q}_p$ feels like. The unit ball in $\mathbb{R}$ is the set of numbers with distance less than one from the point 0, which is the closed interval $[-1, 1]$. Describe the unit ball in $\mathbb{Q}_p$

$$\{x \in \mathbb{Q}_p : |x|_p \leq 1\}.$$

Hint: it has a simple description, and coincides with what we call the p-adic integers, $\mathbb{Z}_p$.

Definition 20.10. In class we discussed another completion, called the algebraic completion of a field. Given any field F, we can construct the *algebraic closure* $\bar{F}$ of F as the set of solutions to all algebraic equations

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where all the coefficients a_i belong to F. A field F is such that all algebraic equations can be solved in F, then we say that F is algebraically closed.

Example 20.11. (Complex numbers) The algebraic closure of $\mathbb{Q}$ is denoted $\overline{\mathbb{Q}}$. The algebraic closure of the field of real numbers $\mathbb{R}$ is the field of complex numbers $\mathbb{C}$, in particular, $\mathbb{R}$ is only missing the algebraic number i. So we have the archimedean metric completion

$$\mathbb{Q} \subset \hat{\mathbb{Q}} = \mathbb{R} \subset \bar{\mathbb{R}} = \mathbb{C}$$

We will take it for granted the fact that $\mathbb{C}$ is metrically complete.

Example 20.12. On the other hand, $\mathbb{Q}$ can also be completed to $\mathbb{Q}_p$, which is metrically complete but not algebraically closed. We take algebraic closure to get

$$\overline{\mathbb{Q}}_p$$

which is no longer metrically closed! But as fields there exists an (unnatural) field isomorphism $\mathbb{C} \simeq \overline{\mathbb{Q}}_p$. Then we complete this space again with respect to the p-adic metric, to obtain the p-adic complex numbers

$$\hat{\overline{\mathbb{Q}}}_p = \mathbb{C}_p \supset \mathbb{C}$$

which strictly contains $\mathbb{C}$ as a subset.

So the p-adic numbers are quite wacky! The following theorem comforts us, saying that there nothing else you can get from $\mathbb{Q}$:

Theorem 20.13 (Ostrowski)**.** *Any absolute value on $\mathbb{Q}$ is equivalent to $|\cdot|_p$ where $p = \infty$ or any prime number.*

But here is the punchline: the p-adic numbers form a *nonarchimedean* field. Recall that the archimedean axiom essentially states that given two lengths, the shorter length can be made to exceed the longer one by adding to it enough small lengths.

Definition 20.14. A *nonarchimedean metric* is one where the triangle inequality, the fourth property of a metric function defined above, is replaced with the ultrametric inequality:

$$d(x, z) \leq \max(d(x, y), d(y, z))$$

The p-adic numbers form a complete ultrametric space. In such spaces, phenomena like the following occur:

(1) Every triangle is isoceles.
(2) Every point in a ball of a given radius is the center of the ball, where the ball of radius r around the point x is defined as
$$B(x, r) = \{y \in X : d(x, y) < r\}$$
(3) If two balls intersect then they are contained in each other.

Exercise 58. Prove one of the above properties in $\mathbb{Q}_p$.

Here is another interesting property:

Proposition 20.15. *$x^2 - 1 = 0$ has a solution in $\mathbb{Q}_p$ if and only if p is congruent to 1 mod 4. In particular, i belongs to $\mathbb{Q}_p$ so it is not directly comparable to the archimedean completion $\mathbb{R}$.*