This Talk is Under Construction
Berkeley Math Circle 2001–2002
Kiran Kedlaya

The straightedge and compass operations.

(a) Given two points, we may construct the line through them.

(b) Given two points, we may construct the circle centered at one point passing through the other point.

(c) Given two lines, two circles, or a line and a circle, we may construct their intersection points.

Recall some things you can construct from these steps:

- the perpendicular bisector of the segment joining two points;
- the angle bisectors of two given lines;
- the circle through three given points;
- a segment of length $\sqrt{a}$, given segments of length 1 and a;
- the regular hexagon, pentagon and 17-gon (!);

and some things you've probably heard cannot be constructed from these steps:

- the trisectors of an arbitrary angle;
- a regular heptagon (7-gon) or nonagon (9-gon);
- a segment of length $\sqrt[3]{2}$, given segments of length 1 and 2.

We call a real number r *(straightedge and compass) constructible* if, starting from two points at distance 1, we can produce two more points at distance $|r|$ by a finite sequence of straightedge and compass operations. We call a complex number constructible if its real and imaginary parts are constructible. Then:

1. the number 1 is constructible;
2. if r and s are constructible, then so are $r+s$, $r-s$, and rs;
3. if $r \neq 0$ is constructible, so is $1/r$;
4. if r is constructible, so are $\pm\sqrt{r}$;
5. a number z is constructible if and only if it can be written in terms of rational numbers using only addition, subtraction, multiplication, division and square roots.

In particular, each constructible number is *algebraic*: it is the root of some polynomial with integer coefficients. A famous theorem of Lindemann states that π is transcendental (not algebraic); that's why one cannot "square the circle". But there are algebraic numbers that are not constructible, such as $\sqrt[3]{2}$.

Warning. The next part of the discussion belongs to the subject of abstract algebra, so may involve some ideas you may not have seen before; but I hope the intuition will be clear. (A subset of that is what is called "linear algebra"; that's pretty much all I'm really using.)

Given a finite set S of algebraic numbers, the set of complex numbers obtained from S and the rational numbers by addition, subtraction, multiplication and division forms what is called a *number field.* If K is a number field, we can always find $z_1, \dots, z_n$ in K such that

$$K = \{a_1 z_1 + \cdots + a_n z_n : a_1, \dots, a_n \in \mathbb{Q}\}$$

and that representation is not redundant (there's only way to write any element of K this way). Such a set $\{z_1, \dots, z_n\}$ is called a *basis* of K, and the number n turns out not to depend on the choice of the basis (a fact from linear algebra that I'm not going to prove now). That number is called the *degree* of the number field. For example, the number field K generated by $\sqrt[3]{2}$ has degree 3, because we can take $z_1 = 1, z_2 = \sqrt[3]{2}, z_3 = \sqrt[3]{4}$. And if S consists of a single primitive n-th root of unity $e^{2\pi i/n}$, then the degree of K turns out to be the Euler phi function $\phi(n)$.

Theorem 1. *If S is a finite set of constructible numbers and K is the number field they generate, then the degree of K is a power of 2.*

Sketch. We prove this by induction on the size of S. It's obvious if S is of size zero, since then $S = \mathbb{Q}$. Suppose $S = S_1 \cup \{\sqrt{s}\}$, and assume that the degree of the number field K_1 generated by S_1 is a power of 2 and that $s \in K_1$. If $\sqrt{s}$ is already in K_1, then $K = K_1$ and we're done. Otherwise, pick a basis $z_1, \dots, z_n$ of K_1; then I claim that $z_1, \dots, z_n, z_1\sqrt{s}, \dots, z_n\sqrt{s}$ is a basis of K. Certainly every element of K can be written as

$$a_1 z_1 + \cdots + a_n z_n + b_1 z_1 \sqrt{s} + \cdots + b_n z_n \sqrt{s}$$

with the a_i and b_i all rational. So we only have to check that there aren't two ways to write some number; in fact, we only have to check that there is no way to write 0 besides using all zeroes. If we had a nontrivial way to write zero, we could write $a = a_1 z_1 + \cdots + a_n z_n$ and $b = b_1 z_1 + \cdots + b_n z_n$, which are both elements of K_1. Then $a + b\sqrt{s} = 0$; if b were nonzero, then $\sqrt{s} = -a/b$ would be in K_1, contradiction. So $b = 0$ and hence $a = 0$; but since $z_1, \dots, z_n$ is a basis of K_1, this can only happen if the a_i and b_i are all zero.

Conclusion: the degree of K is exactly twice that of K_1, so it's also a power of 2.

Why this is only a sketch: You can always arrange for this induction to work at the cost of making S larger. (Given whatever numbers you started with, you have to toss in the "intermediate" numbers you needed to construct them.) So you have to check that if the degree of a number field is a power of 2, then the degree of any number field it contains is also a power of 2, which requires a little more linear algebra. □

So for example, $\sqrt[3]{2}$ is not constructible, because the number field it generates has degree 3. Also, $e^{2\pi i/7}$ and $e^{2\pi i/9}$ both generate number fields of degree 6, so they're not constructible either; thus the regular 7-gon and 9-gon cannot be constructed with straightedge and compass!

It turns out the converse is also true: if an algebraic number generates a number field whose degree is a power of 2, then it is constructible. I won't prove this, but I'll give you an example. A 17-th root of unity $e^{2\pi i/17}$ generates a number field of degree 16, so this converse says that a regular 17-gon is constructible. That is, we can write $e^{2\pi i/17}$ in terms of rational numbers using square roots. We'll do that in the next section. One can do likewise for a 65537-gon, since 65537 is prime and $\phi(65537) = 2^{16}$; we won't do this here.

Example: constructing the 17-gon.

In case you've never seen the construction of the 17-gon (due to Gauss), it's worth a look, in part because it's a very simple example of a notion from abstract algebra called a "Galois group".

Here's the idea: let $z = e^{2\pi i/17}$ be a primitive 17th root of unity. Note that 3 is a primitive root modulo 17, so that $1, 3, \dots, 3^{15}$ cover all the nonzero residue classes modulo 17. Now write

$$\begin{aligned} -1 = x_0 &= z + z^3 + z^{3^2} + \cdots + z^{3^{15}} \\ x_1 &= z + z^{3^2} + z^{3^4} + \cdots + z^{3^{14}} \\ x_2 &= z + z^{3^4} + z^{3^8} + z^{3^{12}} \\ x_3 &= z + z^{3^8} \\ x_4 &= z. \end{aligned}$$

Gauss realized that each x_i can be written as the root of a quadratic polynomial in terms of the previous ones. I'll leave it to you to check my arithmetic:

$$\begin{aligned} x_1(-1 - x_1) &= -4 \\ x_2(x_1 - x_2) &= -1 \\ x_3(x_2 - x_3) &= \frac{x_1x_2 - x_1 + x_2 - 3}{2} \\ x_4(x_3 - x_4) &= 1. \end{aligned}$$

How do you guess this? I don't know how Gauss did it, but they are predicted by *Galois theory.* That's the branch of mathematics that tells us that you can't write the roots of a general degree 5 polynomial in terms of the coefficients using addition, subtraction, multiplication, division, and taking n-th roots (the way the quadratic formula does this in degree 2, and analogous formulas are known in degree 3 and 4).

Here's how the prediction is made in that language: the Galois group of the number field generated by z is a cyclic group of order 16, generated by the map g that sends z^i to z^{3i}. (The point is that this map preserves addition and multiplication, since all it's doing is changing from one 17-th root of unity to another.) The map g fixes only rational numbers; the maps g^2, g^4, g^8 fix number fields of degree 2, 4, 8, respectively. The image of x_1 under g is precisely $-1 - x_1$, so the product $x_1(-1 - x_1)$ is fixed by g and so must be some rational number. Likewise, the image of x_2 under g^2 is $x_1 - x_2$, so the product $x_2(x_1 - x_2)$ is fixed by g^2 and so must be in the number field generated by x_1, and so forth.

The marked straightedge operations.

In using the marked straightedge to trisect angles, Archimedes assumed an additional axiom: (here d is the length marked on the straightedge)

(d) Given a line ℓ, a circle C, and a point P, we can construct all pairs of points Q and R, with Q on ℓ and R on C, such that the line QR passes through P and the distance QR is equal to d.

But you might as well go a little further, replacing the line ℓ by another circle:

(e) Given circles C_1 and C_2 and a point P, we can construct all pairs of points Q and R, with Q on C_1 and R on C_2, such that the line QR passes through P and the distance QR is equal to d.

We can define *marked straightedge constructible* real and complex numbers as you might expect, but the definition will depend on the distance d between the marks on the straightedge. The "right" definition is to take $d = 1$, the same as the distance between the two starting points. (Or if you like, just take one starting point, and use the marked straightedge to produce the second one!)

The "solid" operations.

The ancient Greeks considered a geometric object to be "solid" constructible if if could be obtained using the ordinary straightedge and compass operations plus the ability to draw conic sections. One can phrase this as follows:

(d) Given five points, we may construct the conic section (parabola, hyperbola, ellipse, or degenerate conic) through them.

(e) Given a conic section and a line, circle or another conic, we may construct their intersection points.

Likewise, one can define solid constructible real and complex numbers. It turns out that a complex number r is constructible if and only if:

- r is algebraic, i.e., $P(r) = 0$ for some polynomial P with integer coefficients;
- for some such P, the number field generated by *all* of the roots of P has degree a power of 2 times a power of 3. (It's enough to check this for the polynomial P of smallest degree, called the *minimal polynomial* of r.)

Warning: the second condition is stronger than saying that the number field generated by r alone has degree a power of 2 times a power of 3. For example, if P is a "general" polynomial of degree 6 with integer coefficients, then the number field generated by r itself will have degree 6 but the number field generated by all of the roots of P will have degree $6! = 720$, which has the bad prime factor 5.

For a nice exposition of this, see the article "Constructions Using a Compass and Twice-Notched Straightedge" by Arthur Baragar, in the February 2002 *American Mathematical Monthly.* He also includes a discussion of the marked straightedge case, but a precise characterization of the constructible numbers in that case is still unknown.

The origami operations.

The Japanese art of paper folding gives an entirely different approach to constructing geometric figures. Here is one set of axioms for "origami geometry"; there may be others in the literature. (This one is from an "origami mathematics" web site.)

(a) Given two points, we may construct the line through them.

(b) Given two points, we may construct the perpendicular bisector of the segment joining the two points (i.e., we may fold one point onto another).

(c) Given two lines, we may form the angle bisector of either angle between them (i.e, we may fold one line onto another).

(d) Given a line and a point, we may construct the perpendicular to the line through the point.

(e) Given points P, Q and a line ℓ, we may construct the line ℓ_1 through P such that the reflection of Q across ℓ_1 lies on ℓ.

(f) Given points P_1 and P_2 and lines ℓ_1 and ℓ_2, we may construct the line(s) m such that the reflection of P_1 across m lies on ℓ_1, and the reflection of P_2 across m lies on ℓ_2.

Challenge problems.

Some of these are discussed in the article by Baragar mentioned above, which I again heartily recommend; he also includes many more challenge problems. Warning: the characterizations of constructible numbers involves the notion of a "Galois group", which we haven't talked about.

1. Find a construction of a regular 7-gon using marked straightedge.

2. You might wonder why I didn't state a version of the marked straightedge axiom (d) with two lines instead of a line and a circle. Prove that actually the resulting construction can already done with straightedge and compass.

3. Prove that origami constructions (a)-(e) can also be accomplished with straightedge and compass. (The only subtle one is (e).)

4. Find constructions for trisecting a given angle and doubling a cube using origami. (This will imply that origami construction (f) cannot be accomplished with straightedge and compass.)

5. Find a regular polygon that cannot be constructed using origami. Even better, find a characterization of origami constructible numbers. (I believe the latter is unsolved!)