LECTURE NOTES FOR
CONSTRUCTIONS WITH RULER AND COMPASS

K. N. RAGHAVAN
THE INSTITUTE OF MATHEMATICAL SCIENCES

Well known is the revolutionary idea of translating problems of geometry to algebra by means of the use of co-ordinates: we are all familiar with such terms as *Cartesian plane*, *Cartesian co-ordinates* in honour of René Descartes (1596–1650), to whom this idea is attributed. The manipulative power of algebra can thus be brought to bear upon geometry.[1] Not so well known is another equally revolutionary instance of the introduction of algebra into geometry, due to Carl Friedrich Gauss (1777–1855), dating from circa 1795, which lead to spectacular solutions of certain long standing problems of geometry (Can regular polygons be constructed? Can angles be trisected?), and which illuminated certain other long standing problems (Can the circle be squared?) thereby contributing to their eventual solution. This lecture is about this second instance of the application of algebra to geometry.

Formulation of the problem. To start with, we are given two (distinct) points on the *Euclidean plane*, a *ruler*, and a *compass*. We are allowed to *construct* points, (straight) lines, and circles (or arcs thereof) from our initial data consisting of the two given points, by means of the given tools (the ruler and compass), according to the rules specified below. Points, lines, and circles thus constructed are termed *constructible*.

- The two given points are by definition constructible.
- Using the ruler, we may construct the line through two constructible points.
- Using the compass, we may construct the circle (or arc thereof) with a constructible point as centre and radius the distance between this point and another constructible point.[2]
- Points obtained as intersections of either two lines, or a line and a circle, or two circles thus constructed are themselves in turn defined to be constructible.

It is convenient to allow application of the adjective *constructible* to other objects that are formed out of constructible points, lines, and circles: the line segment between two constructible points is constructible; a triangle whose vertices are constructible is constructible; an angle is constructible if it can be realized as the angle between two constructible lines that meet; etc.

The big question now is:

(1) Which geometric objects are constructible?

[1] And, conversely, algebraic equations can be interpreted as representing geometric shapes—the loci of their solutions—whose properties reflect those of the equations.

[2] You are NOT allowed to draw a circle of arbitrary radius.

Before addressing (some versions of) this question, let us warm up by first giving a more formal definition of the set of constructible points and then making a few simple constructions.

An alternative, formal definition of the set of constructible points. We define, inductively, an increasing sequence $\mathfrak{P}_0 \subseteq \mathfrak{P}_1 \subseteq \mathfrak{P}_2 \subseteq \mathfrak{P}_3 \subseteq \dots$ of sets of points in the plane. The initial set $\mathfrak{P}_0$ consists of just the original two points. Given $\mathfrak{P}_i$, the succeeding member $\mathfrak{P}_{i+1}$ in the sequence is defined as follows. For every pair of points in $\mathfrak{P}_i$, consider the line through these. For every ordered pair of points in $\mathfrak{P}_i$, consider the circle with centre the first point and passing through the second. We define $\mathfrak{P}_{i+1}$ as the set of all the points of intersection of these lines and circles with one another (of two lines, of a line with a cicle, or of two circles).

Clearly $\mathfrak{P}_{i+1}$ is finite and contains its predecessor $\mathfrak{P}_i$. The set of constructible points may be defined, alternatively, as the union:[3]

$$\mathfrak{P} := \mathfrak{P}_0 \cup \mathfrak{P}_1 \cup \mathfrak{P}_2 \cup \mathfrak{P}_3 \cup \cdots \tag{2}$$

Here is a curious looking observation, which follows from the inductive definition of $\mathfrak{P}_i$, and which we will actually use later:

(3) Each $\mathfrak{P}_i$ and so also $\mathfrak{P}$ is closed under reflection in the line through the original two points.

Some simple constructions. Suppose that we are given two distinct constructible points p and q. Let ℓ denote the line through these points. Then the following are constructible (the familiar constructions work even with the constraint on our compass that we cannot draw circles of arbitrary radii):

- ℓ
- the perpendicular to ℓ at p (or q)

Given also a constructible point r not on ℓ,

- the perpendicular to ℓ through r
- the parallel to ℓ through r
- the parallelogram of which pq and pr are two sides

Given a fourth constructible point s at a distance d from r,

- the circle with centre p and radius d

Note that this last item shows that we could exchange our compass for a more versatile one without changing the game. To clarify this remark, suppose that you place the vertices of our compass on points r and s. Can you now lift the compass off the plane, preserving the distance d between the vertices, keep the metallic vertex at p, and draw the resulting circle? While such a maneuvre is not allowed under our present rules, allowing it would not get us any further in the long run: if we exchanged our compass for a more versatile one that could perform this maneuvre, we would not get any more constructible objects! While it might take more steps for us with our poorer compass to get a point constructed than with the more versatile compass, constructibility or its lack thereof remains unchanged.

[3] Being a countable union of finite sets, the set $\mathfrak{P}$ is countable. This implies that there are lots of non-constructible points, there being uncountably many points in the Cartesian plane.

Given two constructible lines ℓ and ℓ' that meet, we can construct:

- the line that bisects the angle between ℓ and ℓ'.[4]

We can also *transfer* angles: that is, given further a constructible line ℓ'' and a constructible point u on it, we can construct:

- the lines through u that are at the same angle to ℓ'' as ℓ and ℓ' to each other.

In particular, the sum of two constructible angles is constructible.

Constructible numbers. Let us now translate the notion of constructibility from geometry to algebra. For this purpose, consider co-ordinates on the plane after the usual fashion: each point on the plane is represented by an ordered pair (x, y) of real numbers, the x- and the y-coordinates of the point. Let us suppose that the two original points given to us are in positions $(0, 0)$ and $(1, 0)$.[5] We may also identify the points on the plane with complex numbers.[6] We may then suppose that the original two given points correspond to 0 and 1.

Introducing Cartesian co-ordinates

A real number a is defined to be *constructible* if the point $(a, 0)$ on the real line (the x-axis) is constructible. More generally, a complex number is defined to be *constructible* if the corresponding point on the plane is constructible. Observe that the following conditions are equivalent for a complex number $z = x + iy = re^{i\theta}$:

- z is constructible;
- x and y are constructible numbers;
- the number r and the angle θ are constructible.

This seemingly innocuous notion of the constructibility of a number is in fact quite powerful. Witness, for starters, that it opens up an entirely new range of thought:[7]

(4) *Constructible numbers form a field closed under taking square roots.*

Proof: (i) Let a be a non-zero constructible complex number. Then a and $-a$ are the points at which the line through 0 and a meets the circle with centre 0 that passes through a. So $-a$ is constructible.

(ii) Let a and b be constructible numbers. Then the parallelogram with the line segment joining 0 to a and that joining 0 to b as two of its sides has $a + b$ as its fourth vertex, so $a + b$ is constructible.

(iii) Let $a = re^{i\theta}$ and $b = r'e^{i\theta'}$ be two constructible numbers. Then the angle $\theta+\theta'$ being the sum of two constructible angles is constructible. To show that ab is constructible, it is therefore enough to show that rr' is constructible. But the line through the origin and the point $(1, r)$ meets the line with equation $x = r'$ at the point (r', rr'), so rr' is constructible.

---

[4] For the lines to be constructible, we must in the first place have had two points apiece on each of them, so the usual construction is possible.

[5] There is no loss of generality in this supposition as the reader may later convince herself.

[6] Such an identification would have been inconceivable at the time of Descartes, for the notion of a complex number was not yet discovered—invented, if you prefer—at his time. In fact, Gauss was among the first to exploit this identification.

[7] A subset of the complex numbers is called a *field* if it contains 0, 1, and -1, is closed under addition, under multiplication, and under division by a non-zero element of the subset.

(iv) Let $a = re^{i\theta}$ be a non-zero constructible number. Then r is a positive real constructible number. The line joining the origin to the point $(r, 1)$ meets the line with equation $x = 1$ at the point $(1, 1/r)$, so $1/r$ is constructible too. Thus $a^{-1} = r^{-1}e^{-i\theta}$ is constructible.

(v) Let $a = re^{i\theta}$ be a constructible number. Its square roots are $\sqrt{r}e^{i\theta/2}$ and $\sqrt{r}e^{i(2\pi+\theta)/2}$. We've already seen that angles can be bisected, so, in order to show that the square roots are constructible, it is enough to show that $\sqrt{r}$ is constructible. The circle with centre origin and radius $r + 1$ meets the line $x = r - 1$ at the points $(r - 1, 2\sqrt{r})$ and $(r - 1, -2\sqrt{r})$, so $\sqrt{r}$ is constructible. □

As a corollary of the above result, we have:

(5) *If b and c are constructible numbers, then so are the roots of the quadratic equation $X^2 + bX + c = 0$.*

Constructibility or lack thereof of regular polygons. Let us now take up a particular version of our question (1), one which was a famous long standing open problem at the time of Gauss, namely:

(6) Which regular polygons are constructible?

What do we mean by the *constructibility* of a regular n-gon? It seems natural to take it to mean that we can find n constructible points that, in some cyclical order, form the vertices of a regular n-gon (never mind how big or small). But it is easy to see—given the constructions that we have made above—that the question can equally well be posed in any of the following equivalent ways:

- For which n can we construct the regular n-gon with centre origin and $(1, 0)$ as one of the vertices?
- For which n is the angle $2\pi/n$ constructible?
- For which n is the number $e^{2\pi i/n}$ constructible?

Since we can bisect angles, it is clear that:

(7) If the regular n-gon is constructible, so is the regular $2n$-gon.

If m and n are coprime integers, then we can find integers a and b such that $am + bn = 1$.[8] Dividing by mn, we get $a/n + b/m = 1/mn$, so that $(e^{2\pi i/m})^a(e^{2\pi i/n})^b = e^{2\pi i/mn}$. By (4),

[8] This follows from Euclid's division algorithm for finding the GCD of two positive integers m and n. Recall that the division algorithm constructs a finite sequence, defined inductively, of strictly decreasing positive integers as follows. If $n = m$, the sequence consists just of one element: n. If $n \neq m$, assuming $n > m$ without loss of generality, the first two terms in the sequence are n, m. Now suppose that we have found two or more terms of the sequence, the last two of which are r and s. To find the next term, find t, $0 \leq t < s$, such that $r = sq + t$ (there is a unique such t). If $t = 0$, then the sequence terminates at s. If $t > 0$, then we add it as the next element of the sequence: n, m, …, r, s, t. This sequence terminates since it is strictly decreasing. Since every third or later term of the sequence is expressed as an integer linear combination of the previous two—note $t = r - qs$ from the definition of t above—by successive back substitutions, we obtain the last term as an integral linear combination of the integers n and m. This proves that the GCD divides the last term. Moreover, since the last term divides the one before, by induction, it divides every preceding term, in particular m and n. So the last term is in fact the GCD.

if $e^{2\pi i/m}$ and $e^{2\pi i/n}$ are constructible, then so is $e^{2\pi i/mn}$. In other words:

(8) If for coprime integers m and n, the regular m-gon and the regular n-gon are constructible, then so is the regular mn-gon.

Let us now try to list the n for which the regular n-gon is constructible.

The equilateral triangle and square. Since $e^{2\pi i/3}$ is at a distance 1 from both the 0 and -1, it is constructible.[9] The unit circle—that with centre 0 and radius 1—and the y-axis intersect at the points $\pm e^{i\pi/2} = \pm i$, so the regular 4-gon is constructible.

The regular pentagon. The first non-trivial case therefore is $n = 5$. It was known for a long time before Gauss that the regular pentagon is constructible. Let us now give a proof of this fact following Gauss and using (5). Put $\zeta = e^{2\pi i/5}$. The constructibility of ζ follows from that of $\zeta + \zeta^4$, for ζ and ζ^4 are the roots of $X^2 - (\zeta + \zeta^4)X + 1 = 0$. Now consider $\zeta + \zeta^4$ and $\zeta^2 + \zeta^3$. Their sum is $\zeta + \zeta^2 + \zeta^3 + \zeta^4 = -1$ and their product is also $\zeta^3 + \zeta + \zeta^4 + \zeta^2 = -1$.[10] Thus they are the roots of $X^2 + X - 1 = 0$, and therefore constructible, which finishes our proof of the constructibility of the regular pentagon.

The state of the art at the time of Gauss. Putting together the constructibility of the regular 3-, 4-, and 5-gons with observations (7) and (8), we conclude that the regular n-gon is constructible if n has one of the following forms:

$$2^m, \quad 2^m 3, \quad 2^m 5, \quad 2^m 15 \qquad \text{for } m \text{ a non-negative integer}$$

Such was the state of knowledge at the time of Gauss. For any value of n other than those above, namely, 7, 9, 11, 13, 14, 17, 18, 19, 21, 22, 23, 25, ..., it was not known whether or not the regular n-gon is constructible.

Gauss's theorem. As a teenager, Gauss succeeded in constructing the regular 17-gon. In fact, he settled problem (6) for good, by proving:

(9) *The regular n-gon is constructible if and only if the number $\phi(n)$ of positive integers not greater than and coprime to n is a power of 2.*

The function $\phi(n)$ in the theorem above is called the *Euler totient function*. To unravel the condition that it be a power of 2, recall its following properties:

- it is multiplicative: that is $\phi(mn) = \phi(m)\phi(n)$ if m and n are coprime.
- for p a prime and r a positive integer, $\phi(p^r) = (p-1)p^{r-1}$.

Let $n = 2^r p_1^{r_1} \cdots p_k^{r_k}$ be the prime factorization of m where $r \geq 0$; $p_1, \ldots, p_k$ are distinct odd primes; and $r_1, \ldots, r_k$ positive integers. Then, by the properties above of ϕ, we get

$$\phi(n) = 2^{r-1} \cdot (p_1 - 1)p_1^{r_1 - 1} \cdot \ldots \cdot (p_k - 1)p_k^{r_k - 1}.$$

For $\phi(n)$ to be a power of 2, it is therefore necessary and sufficient that each of r_1, ..., r_k equal 1, each of $p_1 - 1$, ..., $p_k - 1$ be a power of 2. Gauss's theorem could thus equivalently be stated as follows:

[9] A point is constructible if it is at a constructible distance d from a constructible point p and a constructible distance d' from a second point $p' \neq p$: for it is then one of the points of the intersection of the constructible circles with centre p and radius d and with centre p' and radius d'.

[10] Since ζ is a root of $X^5 - 1$ which factors as $(X-1)(X^4 + X^3 + X^2 + X + 1)$, we obtain $(\zeta - 1)(\zeta^4 + \zeta^3 + \zeta^2 + \zeta + 1) = 0$. Evidently $\zeta - 1 \neq 0$, and so $\zeta^4 + \zeta^3 + \zeta^2 + \zeta + 1 = 0$.

The regular n-gon is constructible if and only if the factorization of n into primes has the following form

$$n = 2^r \cdot p_1 \cdot \ldots \cdot p_k$$

where r is a non-negative integer, and $p_1, \ldots, p_r$ are distinct odd primes such that $p_1 - 1, \ldots, p_r - 1$ are all powers of 2 (r could be 0, that is, n could be just a power of 2).[11]

For example, the full list of n up to 100 for which the regular n-gon is constructible reads as follows:[12]

$$1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 16, 17, 20, 24, 30, 32, 34, 40, 48, 51, 60, 64, 68, 80, 85, 96$$

Before turning to the proof of the theorem, let us note one important consequence of it.

Impossibility of trisecting angles. Gauss's theorem tells us that the regular 18-gon is not constructible, which means precisely that the angle $2\pi i/18$ ($= 20°$) is not constructible. So we cannot trisect the angle $2\pi i/6$ ($= 60°$) (not within the rules of our game), settling another long standing question of geometry from antiquity.

Proof of Gauss's theorem. The crucial albeit simple observation is the following complement to (4):[13]

(10) *The field of constructible numbers equals the smallest subfield of the complex numbers closed under taking square roots.*

Let K denote the latter subfield of the complex numbers. By (4), the field $\mathfrak{P}$ of constructible numbers contains K. To show $\mathfrak{P} \subseteq K$, it is enough, by the definition (2) of $\mathfrak{P}$, to show that each $\mathfrak{P}_i \subseteq K$, which we do my induction. Since $\mathfrak{P}_0 = \{0, 1\}$, it is clearly contained in K. Now suppose that $\mathfrak{P}_i \subseteq K$ for some i. For z in $\mathfrak{P}_i$, its conjugate $\bar{z}$ is also in $\mathfrak{P}_i$ by (3), so $(z + \bar{z})/2$ and $(z - \bar{z})/2i$, the real and imaginary parts of z belong to K (observe that i belongs to K, it being a square root of -1). One can thus write an equation with coefficients in K for any line through two points in $\mathfrak{P}_i$; and the same for any circle with centre at one of the points in $\mathfrak{P}_i$ and passing through another point of $\mathfrak{P}_i$. Solving for the co-ordinates of the points of intersection of such lines and circles with one another (of two lines, of a line with a circle, or of two circles) involves only the operation of extracting square root—while invoking the familiar formula for the roots of a quadratic equation—in addition to the field operations (of addition, subraction, multiplication, and division by non-zero elements). Since K is closed under taking square roots, the co-ordinates of the points of $\mathfrak{P}_{i+1}$ belong to K. Since i belongs to K, the points of $\mathfrak{P}_{i+1}$ (thought of as complex numbers) also belong to K, and (10) is proved.

---

[11] If $2^k + 1$ is a prime, then k itself is a power of 2: if h is an odd integer, then the polynomial $X^h + 1$ has -1 as a root and so $X + 1$ as a factor; this means that if $k = hj$ with h odd, then $2^{hj} + 1 = (2^j)^h + 1$ has $2^j + 1$ as a factor.

[12] If the presence of 1 and 2 on the list is causing discomfort, think of it as just asserting the constructibility of the numbers $e^{2\pi i/1} = 1$ and $e^{2\pi i/2} = -1$.

[13] Since the intersection of an arbitrary collection of subfields is a subfield, it makes sense to talk of such a smallest subfield as in the following statement.

Further preparation for the proof. Consider the inductively defined increasing sequence $K_0 \subseteq K_1 \subseteq K_2 \subseteq \ldots$ of subfields of the complex numbers, where K_0 is the field $\mathbb{Q}$, and, for each i, K_{i+1} is the smallest field containing K_i and the square roots of all elements of K_i. The smallest subfield K of the complex numbers closed under taking square roots is then the union $K_0 \cup K_1 \cup K_2 \cup \ldots$. Using this description of K, we now show:

(11) any finite extension F of $\mathbb{Q}$ contained in K has degree a power of 2 over $\mathbb{Q}$.

Since such an F is contained in some K_i, it is sufficient to prove the statement with K_i in place of K (i arbitrary). We do this by induction on i. Since $K_0 = \mathbb{Q}$, the case $i = 0$ is obvious. Now suppose we know the statement for K_i (induction hypothesis) and suppose that F is contained in K_{i+1}. Set $F' := F \cap K_i$. By enlarging F (and consequently also F') if necessary, we may assume that there exist a_1, a_2, $\ldots$, a_r, such that a_1^2, a_2^2, $\ldots$, a_r^2 belong to F' and $F = F'[a_1, \ldots, a_r]$.

By the induction hypothesis, $F' := F \cap K_i$ has degree a power of 2 over $\mathbb{Q}$. Since $[F : \mathbb{Q}] = [F : F'][F' : \mathbb{Q}]$, it is enough to show that $[F : F']$ is a power of 2. Consider

$$F' \subseteq F'[a_1] \subseteq F'[a_1, a_2] \subseteq \ldots \subseteq F'[a_1, \ldots, a_{r-1}] \subseteq F'[a_1, \ldots, a_r] = F$$

Since a_j^2 belongs to F' for $1 \leq j \leq r$, the degree of every extension in the above chain over the previous one is either 1 or 2. Thus F has degree a power of 2 over F' and we are done proving (11).

A complement. We will now show the following complement to (11):[14]

(12) Any finite Galois extension of $\mathbb{Q}$ of degree a power of 2 is contained in K.

Let F be a finite Galois extension of $\mathbb{Q}$ of degree 2^i. We will show, by induction, that $F \subseteq K_i$. For $i = 0$ we have $F = \mathbb{Q}$ and $K_0 = \mathbb{Q}$. Now let $i > 1$. Recall that a group of order p^i (where p is prime) has non-trivial centre. Let H be a subgroup of order 2 of the Galois group of F over $\mathbb{Q}$ that is contained in the centre. Then H is normal and its fixed field F' is a Galois extension of degree 2^{i-1} over $\mathbb{Q}$. By induction, we know that $F' \subseteq K_{i-1}$. To show that $F \subseteq K_i$ it thus suffices (since $[F : F'] = 2$) to observe that:

(13) Any degree 2 extension is obtained by adjoining a square root

Let α be an element of the extension field not in the base field. Let $\alpha^2 + b\alpha + c = 0$ with b and c in the base field. By the usual procedure of "completion of squares", we see that $(\alpha + b/2)^2 = b^2/4 - c$, so that the extension is obtained by adjoining a square root of $b^2/4 - c$.

Proof of (9). Put $\zeta := e^{2\pi i/n}$., Let $\mathbb{Q}[\zeta]$ be the smallest field containing ζ. As is well known, $\mathbb{Q}[\zeta]$ is a finite Galois extension of degree $\phi(n)$ over $\mathbb{Q}$.

Suppose first that ζ is constructible. Then, by (10), it belongs to K, so $\mathbb{Q}[\zeta]$ is a finite extension of $\mathbb{Q}$ contained in K. By (11), $\phi(n) = [\mathbb{Q}[\zeta] : \mathbb{Q}]$ is a power of 2.

Conversely, suppose $\phi(n)$ is a power of 2. Then, by (12), $\mathbb{Q}[\zeta]$ is contained in K. By (10), ζ is constructible. □

[14] Suggestion to the reader who is not familiar with Galois theory: if you accept as blackboxes the assertion (12) and the fact that $\mathbb{Q}[e^{2\pi i/n}]$ is a Galois extension of degree $\phi(n)$ of $\mathbb{Q}$, the proof becomes more accessible, for the rest of it uses only elementary field theory.

Impossibility of squaring the circle. Is a square with area equal to the unit circle constructible? This question too was for a long time outstanding at the time of Gauss. Since the area of the unit square is π, the side of such a square would be of length $\sqrt{\pi}$. So the question is equivalent to asking if $\sqrt{\pi}$ is constructible. As was shown by Lindemann in 1882, π is transcendental, so definitely not constructible (so also $\sqrt{\pi}$).[15]

EXERCISES

(1) Show that, for coprime positive integers m and n, the regular mn-gon is constructible if the regular m-gon and the regular n-gon are constructible.
(2) A yard stick has inch markings (one yard = 36 inches). It is also divided into hundred equal parts by another set of markings. What is the minimum positive distance that can be measured with such a stick?
(3) A prime p is called a ***Fermat prime*** if $p - 1$ is a power of 2. Show that any Fermat prime p is of the form $2^{2^k} + 1$.
(4) Show that the following two conditions are equivalent for a positive integer m:
    (a) $\phi(m)$ is a power of 2 (where $\phi(m)$ denotes the number of positive integers at most m that are coprime to m).
    (b) $m = 2^r \cdot p_1 \cdot \cdots \cdot p_k$, where $p_1, \ldots, p_k$ are distinct odd Fermat primes.
(5) Prove from first principles that the regular 17-gon is constructible.
(6) Prove or disprove: 3 is a generator of the multiplicative group of units modulo any Fermat prime.

[15] I do not know if it was known before Lindemann that π is not constructible: taken at face value, the non-constructibility seems much weaker than the transcendence.