12. Early on in the proof of the Lindemann–Weierstrauss theorem, we had an equation $\sum_{j=1}^{m} a_j e^{\alpha_j} = 0$, and we needed equations $\sum_{j=1}^{m} a_j e^{\sigma(\alpha_j)} = 0$, where σ is an automorphism of an appropriate field. If σ is continuous, then we can use infinite series to show that $\sigma(e^a) = e^{\sigma(a)}$. Show that if σ is an automorphism of a subfield F of $\mathbb{C}$, then σ is not continuous unless $\sigma = \mathrm{id}$ or σ is complex conjugation restricted to F.

15 Ruler and Compass Constructions

In the days of the ancient Greeks, some of the major mathematical questions involved constructions with ruler and compass. In spite of the ability of many gifted mathematicians, a number of questions were left unsolved. It was not until the advent of field theory that these questions could be answered. We consider in this section the idea of constructibility by ruler and compass, and we answer the following four classical questions:

1. Is it possible to trisect any angle?
2. Is it possible to double the cube? That is, given a cube of volume V, a side of which can be constructed, is it possible to construct a line segment whose length is that of the side of a cube of volume $2V$?
3. Is it possible to square the circle? That is, given a constructible circle of area A, is it possible to construct a square of area A?
4. For which n is it possible to construct a regular n-gon?

The notion of ruler and compass construction was a theoretical one to the Greeks. A ruler was taken to be an object that could draw perfect, infinitely long lines with no thickness but with no markings to measure distance. The only way to use a ruler was to draw the line passing through two points. Similarly, a compass was taken to be a device that could draw a perfect circle, and the only way it could be used was to draw the circle centered at one point and passing through another. The compass was sometimes referred to as a "collapsible compass"; that is, after drawing a circle, the compass could not be lifted to draw a circle centered at another point with the same radius as that of the previous circle. Likewise, given two points a distance d apart, the ruler cannot be used to mark a point on another line a distance d from a given point on the line.

The assumptions of constructibility are as follows. Two points are given and are taken to be the initial constructible points. Given any two constructible points, the line through these points can be constructed, as can the circle centered at one point passing through the other. A point is constructible if it is the intersection of constructible lines and circles.

The first thing we note is that the collapsibility of the compass is not a problem, nor is not being able to use the ruler to mark distances. Given two constructible points a distance d apart, and a line ℓ with a point P on ℓ, we can construct a point Q on ℓ a distance d from P. Also, if we can construct a circle of radius r, given any constructible point P, we can construct the circle of radius r centered at P. These facts are indicated in Figure 15.1. It is left as an exercise (Problem 4) to describe the construction indicated by the figure.

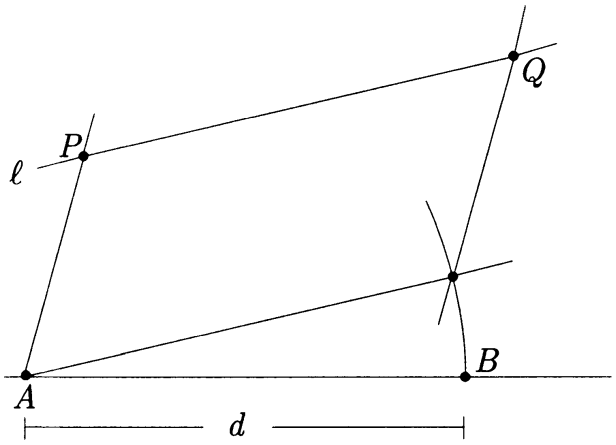

FIGURE 15.1. Construction of Q on ℓ a distance d from P.

There are some standard constructions from elementary geometry that we recall now. Given a line and a point on the line, it is possible to construct a second line through the point perpendicular to the original line. Also, given a line and a point not on the line, it is possible to construct a second line parallel to the original line and passing through the point. These facts are indicated in Figure 15.2.

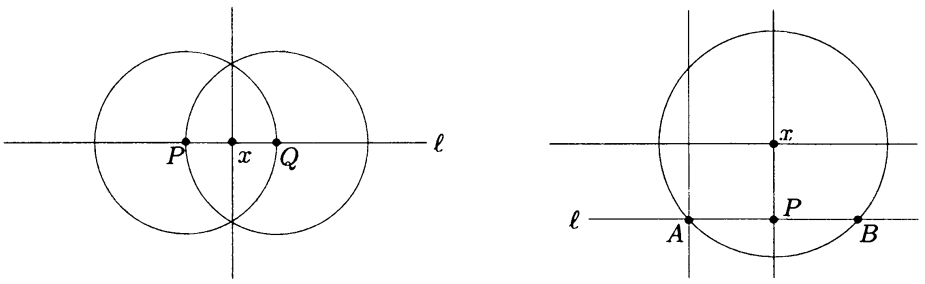

FIGURE 15.2. Construction of lines perpendicular and parallel to ℓ passing through x.

So far, our discussion has been purely geometric. We need to describe ruler and compass constructions algebraically in order to answer our four questions. To do this, we turn to the methods of analytic geometry. Given our original two points, we set up a coordinate system by defining the x-axis to be the line through the points, setting one point to be the origin

and the other to be the point $(1,0)$. We can draw the line perpendicular to the x-axis through the origin to obtain the y-axis.

Let $a \in \mathbb{R}$. We say that a is a constructible number if we can construct two points a distance $|a|$ apart. Equivalently, a is constructible if we can construct either of the points $(a,0)$ or $(0,a)$. If a and b are constructible numbers, elementary geometry tells us that $a+b$, $a-b$, ab, and a/b (if $b \neq 0$) are all constructible. Therefore, the set of all constructible numbers is a subfield of $\mathbb{R}$. Furthermore, if $a > 0$ is constructible, then so is $\sqrt{a}$. These facts are illustrated in Figures 15.3–15.5.

FIGURE 15.3. Construction of $a+b$ and $a-b$.

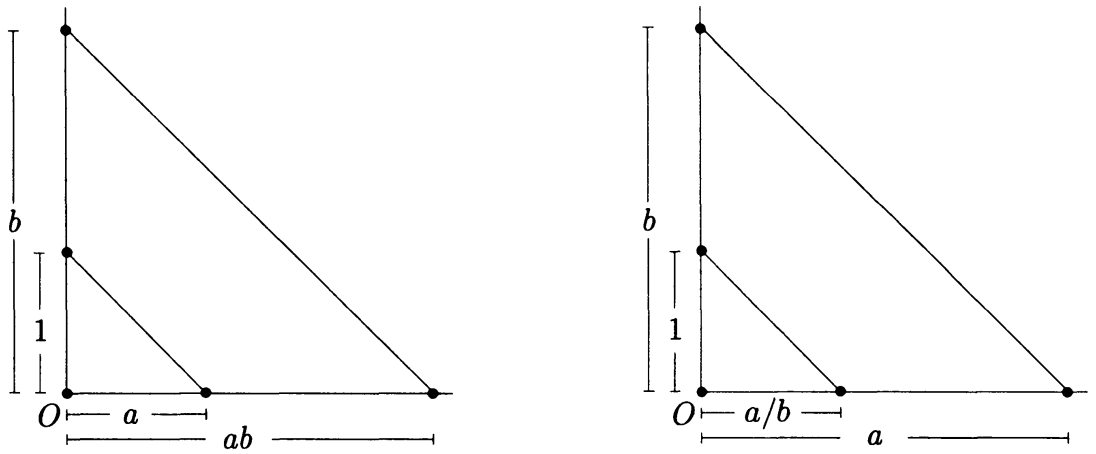

FIGURE 15.4. Construction of ab and a/b.

Suppose that P is a constructible point, and set $P = (a,b)$ in our coordinate system. We can construct the lines through P perpendicular to the x-axis and y-axis; hence, we can construct the points $(a,0)$ and $(0,b)$. Therefore, a and b are constructible numbers. Conversely, if a and b are constructible numbers, we can construct $(a,0)$ and $(0,b)$, so we can construct P as the intersection of the line through $(a,0)$ parallel to the y-axis with the line through $(0,b)$ parallel to the x-axis. Thus, $P = (a,b)$ is constructible if and only if a and b are constructible numbers.

In order to construct a number c, we must draw a finite number of lines and circles in such a way that $|c|$ is the distance between two points of intersection. Equivalently, we must draw lines and circles so that $(c,0)$ is a point of intersection. If we let K be the field generated over $\mathbb{Q}$ by all the numbers obtained in some such construction, we obtain a subfield of the field of constructible numbers. To give a criterion for when a number

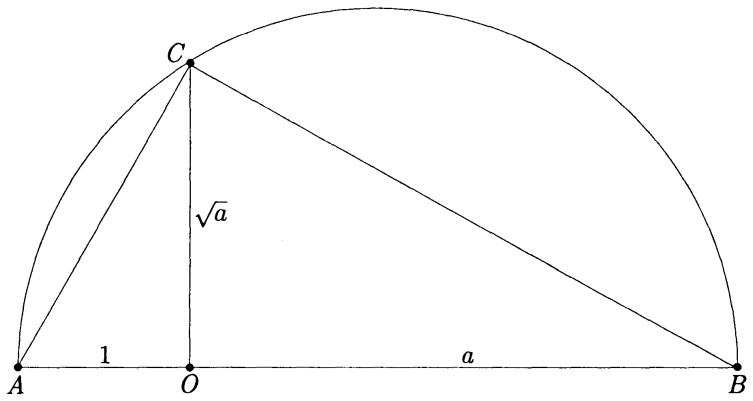

FIGURE 15.5. Construction of $\sqrt{a}$.

is constructible, we need to relate constructibility to properties of the field extension $K/\mathbb{Q}$. We do this with analytic geometry. Let K be a subfield of $\mathbb{R}$. Given any two points in the plane of K, we obtain a line through these points. This will be called a *line in* K. It is not hard to show that a line in K has an equation of the form $ax + by + c = 0$ with $a, b, c \in K$. If P and Q are points in the plane of K, the circle with center P passing through Q is called a *circle in* K. Again, it is not hard to show that the equation of a circle in K can be written in the form $x^2 + y^2 + ax + by + c = 0$ for some $a, b, c \in K$. The next lemma gives us a connection between constructibility and field extensions.

Lemma 15.1 *Let K be a subfield of $\mathbb{R}$.*

1. *The intersection of two lines in K is either empty or is a point in the plane of K.*

2. *The intersection of a line and a circle in K is either empty or consists of one or two points in the plane of $K(\sqrt{u})$ for some $u \in K$ with $u \geq 0$.*

3. *The intersection of two circles in K is either empty or consists of one or two points in the plane of $K(\sqrt{u})$ for some $u \in K$ with $u \geq 0$.*

Proof. The first statement is an easy calculation. For the remaining two statements, it suffices to prove statement 2, since if $x^2+y^2+ax+by+c = 0$ and $x^2 + y^2 + a'x + b'y + c' = 0$ are the equations of circles C and C', respectively, then their intersection is the intersection of C with the line $(a - a')x + (b - b')y + (c - c') = 0$. So, to prove statement 2, suppose that our line L in K has the equation $dx + ey + f = 0$. We assume that $d \neq 0$, since if $d = 0$, then $e \neq 0$. By dividing by d, we may then assume that $d = 1$. Plugging $-x = ey + f$ into the equation of C, we obtain

$$(e^2 + 1)y^2 + (2ef - ae + b)y + (f^2 - af + c) = 0.$$

Writing this equation in the form $\alpha y^2 + \beta y + \gamma = 0$, if $\alpha = 0$, then $y \in K$. If $\alpha \neq 0$, then completing the square shows that either $L \cap C = \varnothing$ or $y \in K(\sqrt{\beta^2 - 4\alpha\gamma})$ with $\beta^2 - 4\alpha\gamma \geq 0$. □

From this lemma, we can turn the definition of constructibility into a property of field extensions of $\mathbb{Q}$, and in doing so obtain a criterion for when a number is constructible.

Theorem 15.2 *A real number c is constructible if and only if there is a tower of fields $\mathbb{Q} = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_r$ such that $c \in K_r$ and $[K_{i+1} : K_i] \leq 2$ for each i. Therefore, if c is constructible, then c is algebraic over $\mathbb{Q}$, and $[\mathbb{Q}(c) : \mathbb{Q}]$ is a power of 2.*

Proof. If c is constructible, then the point $(c, 0)$ can be obtained from a finite sequence of constructions starting from the plane of $\mathbb{Q}$. We then obtain a finite sequence of points, each an intersection of constructible lines and circles, ending at $(c, 0)$. By Lemma 15.1, the first point either lies in $\mathbb{Q}$ or in $\mathbb{Q}(\sqrt{u})$ for some u. This extension has degree either 1 or 2. Each time we construct a new point, we obtain a field extension whose degree over the previous field is either 1 or 2 by the lemma. Thus, we obtain a sequence of fields

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_r$$

with $[K_{i+1} : K_i] \leq 2$ and $c \in K_r$. Therefore, $[K_r : \mathbb{Q}] = 2^n$ for some n. However, $[\mathbb{Q}(c) : \mathbb{Q}]$ divides $[K_r : \mathbb{Q}]$, so $[\mathbb{Q}(c) : \mathbb{Q}]$ is also a power of 2.

For the converse, suppose that we have a tower $\mathbb{Q} = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_r$ with $c \in K_r$ and $[K_{i+1} : K_i] \leq 2$ for each i. We show that c is constructible by induction on r. If $r = 0$, then $c \in \mathbb{Q}$, so c is constructible. Assume then that $r > 0$ and that elements of K_{r-1} are constructible. Since $[K_r : K_{r-1}] \leq 2$, the quadratic formula shows that we may write $K_r = K_{r-1}(\sqrt{a})$ for some $a \in K_{r-1}$. Since a is constructible by assumption, so is $\sqrt{a}$. Therefore, $K_r = K_{r-1}(\sqrt{a})$ lies in the field of constructible numbers; hence, c is constructible. □

With this theorem, we are now able to answer the four questions posed earlier. We first consider trisection of angles. An angle of measure θ is constructible if we can construct two intersecting lines such that the angle between them is θ. For example, a 60° angle can be constructed because the point $(\sqrt{3}/2, 1/2)$ is constructible, and the line through this point and $(0, 0)$ makes an angle of 60° with the x-axis. Suppose that P is the point of intersection on two constructible lines. By drawing a circle of radius 1 centered at P, Figure 15.6 shows that if θ is the angle between the two lines, then $\sin\theta$ and $\cos\theta$ are constructible numbers. Conversely, if $\sin\theta$ and $\cos\theta$ are constructible, then θ is a constructible angle (see Problem 2). In order to trisect an angle of measure θ, we would need to be able to construct an angle of $\theta/3$.

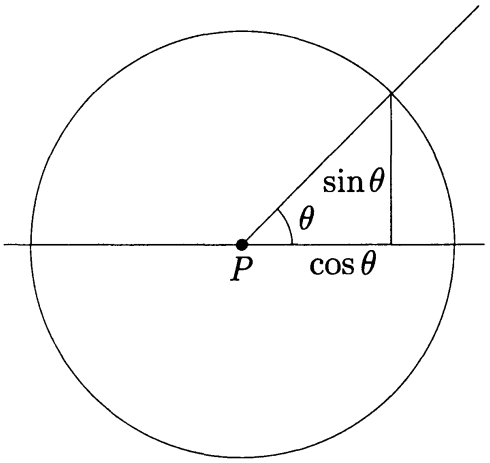

FIGURE 15.6. Construction of sines and cosines.

Theorem 15.3 *It is impossible to trisect a* 60° *angle by ruler and compass construction.*

Proof. As noted above, a 60° angle can be constructed. If a 60° angle can be trisected, then it is possible to construct the number $\alpha = \cos 20°$. However, the triple angle formula $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$ gives $4\alpha^3 - 3\alpha = \cos 60° = 1/2$. Thus, α is algebraic over $\mathbb{Q}$. The polynomial $8x^3 - 6x - 1$ is irreducible over $\mathbb{Q}$ because it has no rational roots. Therefore, $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$, so α is not constructible. A 20° angle cannot then be constructed, so a 60° degree angle cannot be trisected. □

This theorem does not say that no angle can be trisected. A 90° angle can be trisected, since a 30° angle can be constructed. This theorem only says that not all angles can be trisected, so there is no method that will trisect an arbitrary angle.

The second classical impossibility we consider is the doubling of a cube.

Theorem 15.4 *It is impossible to double a cube of length* 1 *by ruler and compass construction.*

Proof. The length of a side of a cube of volume 2 is $\sqrt[3]{2}$. The minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ is $x^3 - 2$. Thus, $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$ is not a power of 2, so $\sqrt[3]{2}$ is not constructible. □

The third of the classical impossibilities is the squaring of a circle. For this, we need to use the fact that π is transcendental over $\mathbb{Q}$.

Theorem 15.5 *It is impossible to square a circle of radius* 1.

Proof. We are asking whether we can construct a square of area π. To do so requires us to construct a line segment of length $\sqrt{\pi}$, which is impossible since $\sqrt{\pi}$ is transcendental over $\mathbb{Q}$ by the Lindemann–Weierstrauss theorem; hence, $\sqrt{\pi}$ is not algebraic of degree a power of 2. □

Our last question concerns construction of regular n-gons. To determine which regular n-gons can be constructed, we will need information about cyclotomic extensions. Recall from Section 7 that if ω is a primitive nth root of unity, then $[\mathbb{Q}(\omega) : \mathbb{Q}] = \phi(n)$, where ϕ is the Euler phi function.

Theorem 15.6 *A regular n-gon is constructible if and only if $\phi(n)$ is a power of 2.*

Proof. We point out that a regular n-gon is constructible if and only if the central angles $2\pi/n$ are constructible, and this occurs if and only if $\cos(2\pi/n)$ is a constructible number. Let $\omega = e^{2\pi i/n} = \cos(2\pi/n) + i\sin(2\pi/n)$, a primitive nth root of unity. Then $\cos(2\pi/n) = \frac{1}{2}(\omega + \omega^{-1})$, since $\omega^{-1} = \cos(2\pi/n) - i\sin(2\pi/n)$. Thus, $\cos(2\pi/n) \in \mathbb{Q}(\omega)$. However, $\cos(2\pi/n) \in \mathbb{R}$ and $\omega \notin \mathbb{R}$, so $\mathbb{Q}(\omega) \neq \mathbb{Q}(\cos(2\pi/n))$. But ω is a root of $x^2 - 2\cos(2\pi/n)x + 1$, as an easy calculation shows, so $[\mathbb{Q}(\omega) : \mathbb{Q}(\cos(2\pi/n))] = 2$. Therefore, if $\cos(2\pi/n)$ is constructible, then $[\mathbb{Q}(\cos(2\pi/n)) : \mathbb{Q}]$ is a power of 2. Hence, $\phi(n) = [\mathbb{Q}(\omega) : \mathbb{Q}]$ is also a power of 2.

Conversely, suppose that $\phi(n)$ is a power of 2. The field $\mathbb{Q}(\omega)$ is a Galois extension of $\mathbb{Q}$ with Abelian Galois group by Proposition 7.2. If $H = \mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q}(\cos(2\pi/n)))$, by the theory of finite Abelian groups there is a chain of subgroups

$$H_0 \subseteq H_1 \subseteq \cdots \subseteq H_r = H$$

with $|H_{i+1} : H_i| = 2$. If $L_i = \mathcal{F}(H_i)$, then $[L_i : L_{i+1}] = 2$; thus, $L_i = L_{i+1}(\sqrt{u_i})$ for some u_i. Since $L_i \subseteq \mathbb{Q}(\cos(2\pi/n)) \subseteq \mathbb{R}$, each of the $u_i \geq 0$. Since the square root of a constructible number is constructible, we see that everything in $\mathbb{Q}(\cos(2\pi/n))$ is constructible. Thus, $\cos(2\pi/n)$ is constructible, so a regular n-gon is constructible. □

This theorem shows, for example, that a regular 9-gon is not constructible and a regular 17-gon is constructible. An explicit algorithm for constructing a regular 17-gon was given by Gauss in 1801. If $n = p_1^{m_1} \cdots p_r^{m_r}$ is the prime factorization of n, then $\phi(n) = \prod_i p^{m_i - 1}(p_i - 1)$. Therefore, $\phi(n)$ is a power of 2 if and only if $n = 2^s q_1 \cdots q_r$, where $r, s \geq 0$, and the q_i are primes of the form $2^m + 1$. In order to determine which regular n-gons are constructible, it then reduces to determining the primes of the form $2^m + 1$.

Problems

1. Use the figures in this section to describe how to construct $a+b$, $a-b$, ab, a/b, and $\sqrt{a}$, provided that a and b are constructible.

2. If $\sin\theta$ and $\cos\theta$ are constructible numbers, show that θ is a constructible angle.

3. If an angle θ can be constructed, show that a line passing through the origin can be constructed such that the angle between this line and the x-axis is θ.

4. Use the figures of this section to answer the following questions.

   (a) Given two points a distance d apart and a constructible point P on a line ℓ, show that it is possible to construct a point Q on ℓ a distance d from P.

   (b) Given that some circle of radius r can be constructed, if P is a constructible point, show that the circle of radius r centered at P can be constructed.

   (c) Given a line ℓ and a point P on ℓ, show that it is possible to construct the line through P perpendicular to ℓ.

   (d) Given a line ℓ and a point P not on ℓ, show that it is possible to construct the line through P parallel to ℓ.

5. Let $c \in \mathbb{R}$ be a root of an irreducible quartic over $\mathbb{Q}$. Let N be the normal closure of $\mathbb{Q}(c)/\mathbb{Q}$.

   (a) If $\mathrm{Gal}(N/\mathbb{Q})$ is isomorphic to either D_4 or a group of order 4, show that c is constructible.

   (b) If $\mathrm{Gal}(N/\mathbb{Q})$ is isomorphic to either A_4 or S_4, show that c is not constructible.

6. Let $c \in \mathbb{R}$ be algebraic over $\mathbb{Q}$, and let N be the normal closure of $\mathbb{Q}(c)/\mathbb{Q}$. If $[N : \mathbb{Q}]$ is a power of 2, show that c is constructible.

7. This problem gives a partial converse to Theorem 15.2. If $c \in \mathbb{R}$ is algebraic over $\mathbb{Q}$ and if N is the normal closure of $\mathbb{Q}(c)/\mathbb{Q}$, then show that c is constructible if and only if $[N : \mathbb{Q}]$ is a power of 2.
   (The criterion for constructibility proven in this section is much like the definition of solvable by radicals given in Section 16. If you work this problem, some proofs of the next section will be easier to understand.)

8. A Fermat number is a number of the form $2^{2^r}+1$ for some r. Suppose that p is an odd prime such that a regular p-gon is constructible. Show that p is a Fermat number.

16 Solvability by Radicals

In this section, we address one of the driving forces of mathematics for hundreds of years, the solvability of polynomial equations. As we saw in Section