Appendix C
Ruler-Compass Constructions

We are going to show that the classical Greek problems: squaring the circle, duplicating the cube, and trisecting an angle, are impossible to solve. As we shall see, the discussion uses only elementary field theory; no Galois theory is required.

It is clear one that can trisect a 60° angle with a protractor (or any other device than can measure an angle); after all, one can divide any number by 3. Therefore, it is essential to state the problems carefully and to agree on certain ground rules. The Greek problems specify that only two tools are allowed, and each must be used in only one way. Let P and Q be points in the plane; we denote the line segment with endpoints P and Q by PQ, and we denote the length of this segment by $|PQ|$. A ***ruler*** (or ***straight-edge***) is a tool that can draw the line $L(P, Q)$ determined by P and Q; a ***compass*** is a tool that draws the circle with radius $|PQ|$ and center either P or Q; denote these circles by $C(P; Q)$ or $C(Q; P)$, respectively. Since every construction has only a finite number of steps, we shall be able to define "constructible" points inductively.

Given the plane, we establish a coordinate system by first choosing two distinct points, A and $\overline{A}$; call the line they determine the ***x-axis***. Use a compass to draw the two circles $C(A; \overline{A})$ and $C(\overline{A}; A)$ of radius $|A\overline{A}|$ with centers A and $\overline{A}$, respectively. These two circles intersect in two points; the line they determine is called the ***y- axis***; it is the perpendicular bisector of $A\overline{A}$, and it intersects the x-axis in a point O, called the ***origin***. We define the distance $|OA|$ to be 1. We have introduced coordinates in the plane; in particular, $A = (1, 0)$ and $\overline{A} = (-1, 0)$.

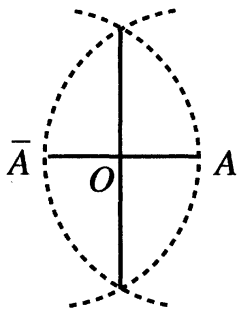

Figure 5

Informally, one constructs a new point T from (not necessarily distinct) old points P, Q, R, and S by using the first pair P, Q to draw a line or circle, the second pair R, S to draw a line or circle, and then obtaining T as one of the points of intersection of the two drawn lines, the drawn line and the drawn circle, or the two drawn circles. More generally, a point is called constructible if it is obtained from A and $\overline{A}$ by a finite number of such steps. Given a pair of constructible points, we do *not* assert that every point on the drawn line or the drawn circles they determine is constructible.

Here is the formal discussion.

Definition. Let E, F, G, and H be (not necessarily distinct) points in the plane. A point Z is ***constructible from*** E, F, G, and H if either

(i) $Z \in L(E, F) \cap L(G, H)$, where $L(E, F) \neq L(G, H)$;

(ii) $Z \in L(E, F) \cap C(G; H)$;

(iii) $Z \in C(E; F) \cap C(G; H)$, where $C(E; F) \neq C(G; H)$.

A point Z is ***constructible*** if $Z = A$ or $Z = \overline{A}$ or if there are points $P_1, \ldots, P_n$ with $Z = P_n$ so that, for all $j \geq 1$, the point P_{j+1} is constructible from points in $\{A, \overline{A}, P_1, \ldots, P_j\}$.

Example 38. Let us show that $Z = (0, 1)$ is constructible. We have seen above that the origin $P_1 = O$ is constructible. The points $P_2 = (0, \sqrt{3})$ and $P_3 = (0, -\sqrt{3})$ are constructible, for both lie in $C(A; \overline{A}) \cap C(\overline{A}; A)$, and so the y-axis $L(P_2, P_3)$ can be drawn. Finally,

$$Z = (0, 1) \in L(P_2, P_3) \cap C(O; A).$$

In our discussion, we shall freely use any standard result of euclidean geometry. For example, every angle can be bisected with ruler and compass; i.e., if $(\cos\theta, \sin\theta)$ is constructible, then so is $(\cos\theta/2, \sin\theta/2)$.

Definition. A complex number $z = x + iy$ is ***constructible*** if the point (x, y) is a constructible point.

Example 38 shows that the numbers $1, -1, 0, i\sqrt{3}, -i\sqrt{3}$, and i are constructible numbers.

Lemma R.1. *A complex number $z = x + iy$ is constructible if and only if its real part x and its imaginary part y are constructible.*

Proof. If z is constructible, then a standard euclidean construction draws the vertical line L through (x, y) which is parallel to the y-axis. It follows that x is constructible, for the point $(x, 0)$ is constructible, being the intersection of L and the x-axis. Similarly, the point $(0, y)$ is the intersection of the y-axis and a line through (x, y) which is parallel to the x-axis. It follows that $P = (y, 0)$ is constructible, for it is an intersection point of the x-axis and $C(O; P)$. Hence, y is a constructible number.

Conversely, assume that x and y are constructible numbers; that is, $Q = (x, 0)$ and $P = (y, 0)$ are constructible points. The point $(0, y)$ is constructible, being the intersection of the y-axis and $C(O; P)$. One can draw the vertical line through $(x, 0)$ as well as the horizontal line through $(0, y)$, and (x, y) is the intersection of these lines. Therefore, (x, y) is a constructible point, and so $z = x + iy$ is a constructible number. •

Definition. We denote by K the subset of $\mathbb{C}$ consisting of all the ***constructible numbers***.

Lemma R.2.

(i) *If $K \cap \mathbb{R}$ is a subfield of $\mathbb{R}$, then K is a subfield of $\mathbb{C}$.*

(ii) *If $K \cap \mathbb{R}$ is a subfield of $\mathbb{R}$ and if $\sqrt{a} \in K$ whenever $a \in K \cap \mathbb{R}$ is positive, then K is closed under square roots.*

Proof. (i) If $z = a + ib$ and $w = c + id$ are constructible, then $a, b, c, d \in K \cap \mathbb{R}$, by Lemma R.1. Hence, $a + c, b + d \in K \cap \mathbb{R}$, because $K \cap \mathbb{R}$ is a subfield, and so $(a + c) + i(b + d) \in K$, by Lemma R.1. Similarly, $zw = (ac - bd) + i(ad + bc) \in K$. If $z \neq 0$, then $z^{-1} = (a/z\bar{z}) - i(b/z\bar{z})$. Now $a, b \in K \cap \mathbb{R}$, by Lemma R.1, so that $z\bar{z} = a^2 + b^2 \in K \cap \mathbb{R}$, because $K \cap \mathbb{R}$ is a subfield of $\mathbb{C}$. Therefore, $z^{-1} \in K$.

(ii) If $z = a + ib \in K$, then $a, b \in K \cap \mathbb{R}$, by Lemma R.1, and so $r^2 = a^2 + b^2 \in K \cap \mathbb{R}$, as in part (i). Since r^2 is non-negative, the hypothesis gives $r \in K \cap \mathbb{R}$ and $\sqrt{r} \in K \cap \mathbb{R}$. Now $z = re^{i\theta}$, so that $e^{i\theta} = r^{-1}z \in K$, because K is a subfield of $\mathbb{C}$. That every angle can be bisected gives $e^{i\theta/2} \in K$, and so $\sqrt{z} = \sqrt{r}e^{i\theta/2} \in K$, as desired. •

Theorem R.3. *The set of all constructible numbers K is a subfield of $\mathbb{C}$ that is closed under square roots and complex conjugation.*

Proof. For the first two statements, it suffices to prove that the properties of $K \cap \mathbb{R}$ in Lemma R.2 do hold. Let a and b be constructible reals.

(i) *$-a$ is constructible.*

If $P = (a, 0)$ is a constructible point, then $(-a, 0)$ is the other intersection of the x-axis and $C(O; P)$.

(ii) *$a + b$ is constructible.*

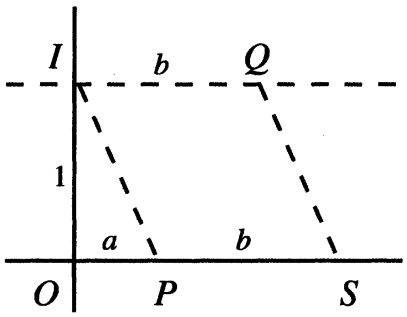

Figure 6

Let $I = (0, 1)$, $P = (a, 0)$ and $Q = (b, 1)$. Now Q is constructible: it is the intersection of the horizontal line through I and the vertical line through $(b, 0)$ [the latter point is constructible, by hypothesis]. The line through Q parallel to IP intersects the x-axis in $S = (a + b, 0)$, as desired. Although Figure 6 is drawn with a, b positive, it is clear that this construction works for any choice of signs of a, b.

(iii) *ab is constructible.*

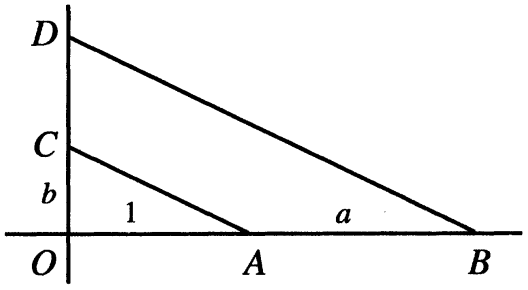

Figure 7

By (i), we may assume that both a and b are positive. In Figure 7, $A = (1, 0)$, $B = (1 + a, 0)$, and $C = (0, b)$. Define D to be the intersection of the y-axis and the line through B parallel to AC. Since the triangles OAC and OBD are similar,

$$|OB|/|OA| = |OD|/|OC|;$$

hence $(a + 1)/1 = (b + |CD|)/b$, and $|CD| = ab$. Therefore, $b + ab$ is constructible. Since $-b$ is constructible, by (i), we have $ab = (b+ab) - b$ constructible, by (ii).

(iv) *If $a \neq 0$, then a^{-1} is constructible.*

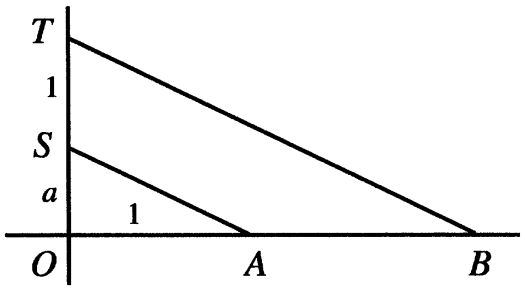

Figure 8

Let $A = (1, 0)$, $S = (0, a)$, and $T = (0, 1+a)$. Define B as the intersection of the x-axis and the line through T parallel to AS; thus, $B = (1 + u, 0)$ for some u. Similarity of the triangles OSA and OTB gives

$$|OT|/|OS| = |OB|/|OA|.$$

Hence, $(1 + a)/a = (1 + u)/1$, and so $u = a^{-1}$. Therefore, $1 + a^{-1}$ is constructible, and so $(1 + a^{-1}) - 1 = a^{-1}$ is constructible.

(v) *If $a \geq 0$, then $\sqrt{a}$ is constructible.*

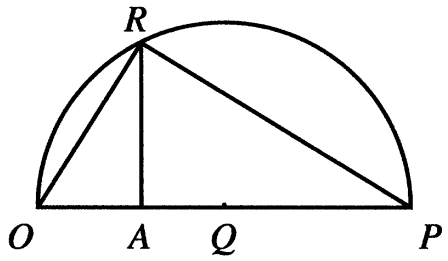

Figure 9

Let $A = (1, 0)$ and $P = (1 + a, 0)$; construct Q, the midpoint of OP. Define R as the intersection of the circle $C(Q; O)$ with the vertical line through A. The (right) triangles AOR and ARP are similar, so that

$$|OA|/|AR| = |AR|/|AP|,$$

and so $|AR| = \sqrt{a}$.

(vi) *If* $z = a + ib \in K$, *then* $\overline{z} = a - ib$ *is constructible.*

By Lemma R.2, K is a subfield of $\mathbb{C}$. Now $a, b \in K$, by Lemma R.1, and $i \in K$, by Example 38. Therefore, $-bi \in K$, and so $a - ib \in K$. •

Corollary R.4. *If* a, b, c *are constructible, then the roots of the quadratic* $ax^2 + bx + c$ *are also constructible.*

Proof. This follows from the theorem and the quadratic formula. •

We now consider subfields of $\mathbb{C}$ to enable us to prove an inductive step in the upcoming theorem.

Lemma R.5. *Let* F *be a subfield of* $\mathbb{C}$ *that contains* i *and that is closed under complex conjugation. Let* $z = a + ib, w = c + id \in F$, *and let* $P = (a, b)$ *and* $Q = (c, d)$.

(i) *If* $a + ib \in F$, *then* $a \in F$ *and* $b \in F$.

(ii) *If the equation of* $L(P, Q)$ *is* $y = mx + q$, *where* $m, q \in \mathbb{R}$, *then* $m, q \in F$.

(iii) *If the equation of* $C(P; Q)$ *is* $(x - a)^2 + (y - b)^2 = r^2$, *where* $a, b, r \in \mathbb{R}$, *then* $r^2 \in F$.

Proof. (i) If $z = a+ib \in F$, then $a = \frac{1}{2}(z+\overline{z}) \in F$ and $ib = \frac{1}{2}(z-\overline{z}) \in F$; since we are assuming $i \in F$, we have $b \in F$.

(ii) If $L(P, Q)$ is not vertical, its equation is $y - b = m(x - a)$. Now $m = (d-b)/(c-a) \in F$, since $a, b, c, d \in F$, and so $q = -ma+b \in F$.

(iii) The circle $C(P; Q)$ has equation $(x - a)^2 + (y - b)^2 = r^2$, and $r^2 = (c - a)^2 + (d - b)^2 \in F$. •

Lemma R.6. *Let* F *be a subfield of* $\mathbb{C}$ *that contains* i *and that is closed under complex conjugation. Let* P, Q, R, S *be points whose coordinates lie in* F, *and let* $\alpha = u + iv \in \mathbb{C}$. *If either*

$$\alpha \in L(P, Q) \cap L(R, S), \text{ where } L(P, Q) \neq L(R, S),$$
$$\alpha \in L(P, Q) \cap C(R; S),$$

or

$$\alpha \in C(P; Q) \cap C(R, S), \text{ where } C(P; Q) \neq C(R; S),$$

then $[F(\alpha) : F] \leq 2$.

Proof. If $L(P, Q)$ is not vertical, then Lemma R.5(ii) says that $L(P, Q)$ has equation $y = mx + b$, where $m, b \in F$. If $L(P, Q)$ is vertical, then its equation is $x = b$ because $P = (a, b) \in L(P, Q)$, and so $b \in F$, by

Lemma R.5(i). Similarly, $L(R, S)$ has equation $y = nx + c$ or $x = c$, where $m, b, n, c \in F$. Since these lines are not parallel, one can solve the pair of linear equations for (u, v), the coordinates of $\alpha \in L(P, Q) \cap L(R, S)$, and they also lie in F. In this case, therefore, $[F(\alpha) : F] = 1$.

Let $L(P, Q)$ have equation $y = mx + q$ or $x = q$, and let $C(R; S)$ have equation $(x - c)^2 + (y - d)^2 = r^2$; by Lemma R.5, we have $m, q, r^2 \in F$. Since $\alpha = u + iv \in L(P, Q) \cap C(R; S)$,

$$\begin{aligned} r^2 &= (u - c)^2 + (v - d)^2 \\ &= (u - c)^2 + (mu + q - d)^2, \end{aligned}$$

so that u is a root of a quadratic polynomial with coefficients in $F \cap \mathbb{R}$. Hence, $[F(u) : F] \leq 2$. Since $v = mu + q$, we have $v \in F(u)$, and, since $i \in F$, we have $\alpha \in F(u)$. Therefore, $\alpha = u + iv \in F(u)$, and so $[F(\alpha) : F] \leq 2$.

Let $C(P; Q)$ have equation $(x - a)^2 + (y - b)^2 = r^2$, and let $C(R; S)$ have equation $(x - c)^2 + (y - d)^2 = s^2$. By Lemma R.5, we have $r^2, s^2 \in F \cap \mathbb{R}$. Since $\alpha \in C(P; Q) \cap C(R; S)$, there are equations

$$(u - a)^2 + (v - b)^2 = r^2 \text{ and } (u - c)^2 + (v - d)^2 = s^2.$$

After expanding, both equations have the form $u^2 + v^2 +$ something $= 0$. Setting the something's equal gives an equation of the form $tu + t'v + t'' = 0$, where $t, t', t'' \in F$. Coupling this with the equation of one of the circles returns us to the situation of the second paragraph. •

Theorem R.7. *A complex number z is constructible if and only if there is a tower of fields*

$$\mathbb{Q} = K_0 \subset K_1 \subset \cdots \subset K_n,$$

where $z \in K_n$ and $[K_{j+1} : K_j] \leq 2$ for all j.

Proof. If z is constructible, there is a sequence of points $1, -1, z_1, \ldots, z_n = z$ with each z_j obtainable from $\{1, -1, z_1, \ldots, z_{j-1}\}$; since i is constructible, we may assume that $z_1 = i$. Define

$$K_j = \mathbb{Q}(z_1, \ldots, z_j).$$

Given $u = z_{j+1}$, there are points $E, F, G, H \in K_j$ with one of the following:

$$\begin{aligned} u &\in L(E, F) \cap L(G, H); \\ u &\in L(E, F) \cap C(G; H); \\ u &\in C(E; F) \cap C(G; H). \end{aligned}$$

We may assume, by induction on $j \geq 1$, that K_j is closed under complex conjugation, so that Lemma R.6 applies to show that $[K_{j+1} : K_j] \leq 2$. Finally, note that K_{j+1} is also closed under complex conjugation, for if z_{j+1} is a root of a quadratic $f(x) \in K_j[x]$, then $\overline{z}_{j+1}$ is the other root of $f(x)$.

To prove the converse, it suffices to prove that if $[B : F] = 2$, where $F \subset K$, then B/F is a pure extension of type 2, say, $B = F(\beta)$, where $\beta \in L(P, Q) \cap C(R; S)$ for $P, Q, R, S \in F$; it will then follow that $B \subset K$. Since $[B : F] = 2$, there is α with $B = F(\alpha)$, where α is a root of some irreducible quadratic $x^2 + bx + c \in F[x]$. If we define $\beta = \sqrt{b^2 - 4c}$, then $B = F(\beta)$ displays B/F as a pure extension of type 2. To see that β can be realized as a point in the intersection of a line and a circle, we use the construction in Theorem R.3(v). Let the line L be the vertical line through $A = (1, 0)$ and let the circle have center $Q = (\frac{1}{2}(1 + \beta^2), 0)$ and radius $\frac{1}{2}(1 + \beta^2)$. •

Corollary R.8. *If a complex number z is constructible, then $[\mathbb{Q}(z) : \mathbb{Q}]$ is a power of 2.*

Proof. This follows from the theorem and Lemma 49. •

Remark. The converse of this corollary is false. In Example 36, we saw that $p(x) = x^4 - 4x + 2$ is an irreducible polynomial over $\mathbb{Q}$ whose Galois group $\mathrm{Gal}(E/\mathbb{Q})$ is S_4, where $E/\mathbb{Q}$ is a splitting field of $p(x)$. Were every root of $p(x)$ constructible, then every element of E would be constructible, for all constructible numbers form a subfield of $\mathbb{C}$, by Theorem R.3. If H is a Sylow 2-subgroup of $G \cong S_4$, however, then $[G : H] = 3$; the intermediate field E^H thus has degree $[E^H : \mathbb{Q}] = [G : H] = 3$, and so none of its elements are constructible, by Corollary R.8. This contradiction shows that some root of $p(x)$ is not constructible, even though every root has degree 4 over $\mathbb{Q}$.

It is now a simple matter to dispose of some famous problems.

(1) *It is impossible to "square the circle."*

The problem is to construct, with ruler and compass, a square whose area is equal to the area of a circle of radius 1; in other words, one asks whether $\sqrt{\pi}$ is constructible. But it is a classical result, proved by F. Lindemann in 1882, that π, hence $\sqrt{\pi}$, is transcendental over $\mathbb{Q}$ (see [Hadlock, p. 47]), and so it does not lie in any finite extension of $\mathbb{Q}$, let alone one of degree a power of 2.

(2) *It is impossible to "duplicate the cube."*

The problem is to construct a cube whose volume is 2; in other words, is the real cube root of 2, call it α, constructible? Now $x^3 - 2$ is irreducible over $\mathbb{Q}$, by Eisenstein, and so $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$, which is not a power of 2. Corollary R.8 gives the result. This result was first proved by P. L. Wantzel in 1837.

(3) *It is impossible to trisect an arbitrary angle.*

An angle θ is given by two intersecting lines; it is no loss in generality to assume the lines intersect at the origin and that one line is the x-axis. If we could draw the angle trisector, then the point $(\cos\theta/3, \sin\theta/3)$, which is the intersection of the trisector and the unit circle, would be constructible; hence $\cos\theta/3$ would also be constructible, by Lemma R.1.

We will now show that 60° cannot be trisected. Computing the real part of $e^{3i\theta} = (\cos\theta + i\sin\theta)^3$ gives the trigonometric identity:

$$\cos 3\theta = 4\cos^3\theta - 3\cos\theta.$$

Defining $u = 2\cos\theta$ and $\theta = 20°$, we arrive at the equation

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

It is easy to see that this cubic is irreducible (it has no rational root, by Exercise 63), and so $[\mathbb{Q}(u) : \mathbb{Q}] = 3$. Corollary R.8 shows that u is not constructible. This result was also proved by P. L. Wantzel in 1837.

(4) *Regular p-gons.*

Galois theory will be used in discussing this problem.

Theorem R.9 (Gauss). *If p is an odd prime, then a regular p-gon is constructible if and only if $p = 2^{2^t} + 1$ for some $t \geq 0$.*

Proof. This is again a question of constructibility of a point on the unit circle, namely, $z = e^{2\pi i/p}$. Now the irreducible polynomial of z over $\mathbb{Q}$ is the cyclotomic polynomial $\Phi_p(x)$ of degree $p - 1$ (Corollary 41).

Assume z is constructible. By Corollary R.8, $p - 1 = 2^s$ for some s. We claim that s itself is a power of 2. Otherwise, there is an odd number $k > 1$ with $s = km$. But $x^k + 1$ factors over $\mathbb{Z}$ (because -1 is a root); setting $x = 2^m$ thus gives a forbidden factorization of p.

Conversely, assume $p = 2^{2^t} + 1$ is prime. Since z is a primitive pth root of unity, $\mathbb{Q}(z)$ is the splitting field of $\Phi_p(x)$ over $\mathbb{Q}$. Hence $\mathrm{Gal}(\mathbb{Q}(z)/\mathbb{Q})$ has order 2^{2^t}, and so the Galois group is a 2-group. But a 2-group has a normal series in which each factor group has order 2 (this follows easily from Theorem G.23); by the fundamental theorem of Galois theory, there is a tower of fields $\mathbb{Q} = K_0 \subset K_1 \subset \cdots \subset K_m = \mathbb{Q}(z)$ with $[K_{i+1} : K_i] = 2$ for all i, that is, z is constructible, by Theorem R.7. •

Remark. Primes of the form $2^{2^t} + 1$ are called ***Fermat primes***. The values $0 \leq t \leq 4$ do give primes (they are 3, 5, 17, 257, 65,537), the next few values of t do not give primes, and it is unknown whether any other Fermat primes exist.

Gauss actually gave a geometric construction of the regular 17-gon.

Corollary R.10. *It is impossible to construct a regular* 7*-gon, a regular* 11*-gon, or a regular* 13*-gon.*

Proof. 7, 11, and 13 are not Fermat primes. •

The following result is known (see [Hadlock, p. 106]):

Theorem R.11. *A regular n-gon is constructible if and only if n is a product of a power of* 2 *and distinct Fermat primes.*

It follows that regular 9-gons and regular 14-gons are not constructible; on the other hand, a regular 15-gon is constructible. It is possible that there are only finitely many constructible regular n-gons with n is odd, for there may be only finitely many Fermat primes.

Appendix D
Old-fashioned Galois Theory

Gimme that old-time Galois theory;

If it's good enough for Galois, then it's good enough for me!