ARNOLD'S ELEMENTARY PROOF OF THE INSOLVABILITY OF THE QUINTIC

LEO GOLDMAKHER

ABSTRACT. We give a proof (due to Arnold) that there is no quintic formula. Somewhat more precisely, we show that any finite combination of the four field operations $(+, -, \times, \div)$, radicals, the trigonometric functions, and the exponential function will never produce a formula for producing a root of a general quintic polynomial. The proof is elementary, requiring no knowledge of abstract group theory or Galois theory.

1. PREREQUISITE IDEAS AND NOTATIONS

To understand the arguments in this essay you don't need to know Galois theory. You also don't need to know abstract algebra or group theory. However, you *do* need to know about complex numbers and the complex plane $\mathbb{C}$. Let's begin with a basic

Question. *What is the definition of the imaginary number i?*

We all know the answer: it's the square-root of -1. There's only one problem. There are two square roots of -1! Thus, our definition defines two numbers $\pm i$ but doesn't give a way to distinguish between them. So how can we define i without simultaneously defining $-i$?

It turns out that this is impossible to do.[1] Of course, we can arbitrarily label one of the two square-roots as i, and then this forces the other to be $-i$. But this choice is arbitrary, and demonstrates that there's a fundamental symmetry between the two numbers $\pm i$; the two are yoked together and are algebraically indistinguishable. This observation lies at the heart of Galois theory, which studies other symmetries between numbers (not just pairs of numbers, but also triples and quadruples etc.) and, using group theory to extract properties of the symmetries, deduces properties that such numbers must have.[2]

These types of ideas will play a major role in our discussion below, as well. Unlike Galois theory, we won't require abstract group theory, but we will use a convenient notation – *cycle notation* – to describe permutations of objects. Rather than defining this formally, let's consider an example. Suppose you have six objects and you want to rearrange them so that

- the first object ends up in the third position,
- the second object ends up in the sixth position,
- the third object ends up in the fourth position,
- the fourth object ends up in the first position,
- the fifth object doesn't move, and
- the sixth object ends up in the second position.

In cycle notation, this permutation is written $(1\ 3\ 4)(2\ 6)$. Note that 5 doesn't appear here at all, since it ends up being fixed. The *trivial* permutation, denoted $()$, is the one which leaves all the objects where they are. One nice feature of cycle notation is that it's very easy to undo a given permutation: just write the numbers in reverse order inside each set of parentheses. For example, the permutation which undoes the one given above is $(4\ 3\ 1)(6\ 2)$. Note that $(4\ 3\ 1) = (1\ 4\ 3)$ and $(6\ 2) = (2\ 6)$, so we could have also written the overall permutation in the form $(1\ 4\ 3)(2\ 6)$. To learn more about cycle notation, check out the relevant subsection in Wikipedia's article on permutations.

Enough preliminaries. Let's do some math!

[1] Here I'm restricting the term *define* to mean solving polynomials whose coefficients are real numbers.

[2] More precisely, in Galois theory one derives information about how numbers can (and can't) be *described*.

2. Quadratic formulas aren't functions

The first step to understanding Arnold's idea is to build up some intuition. Navigate to the website

`duetosymmetry.com/tool/polynomial-roots-toy/`

Set the degree of the polynomial to 2. There are two copies of $\mathbb{C}$ drawn. On the left is *coefficient space* – there are two dots, labelled a_0 and a_1, corresponding to the coefficients of a quadratic polynomial (precisely, the polynomial $x^2 + a_1 x + a_0$). The $\mathbb{C}$ on the right is *root space*: the two dots displayed are the two roots of the polynomial defined by the current state coefficient space.

To get a feel for all this, drag the a_0 coefficient to -1 and the a_1 coefficient to $1/2$. You should have two real roots in root space (one at ≈ -1.28, the other at ≈ 0.78). Let's call r_1 the negative root, and r_2 the positive root. Now move the coefficient a_0 around in a small loop (i.e. move it around a little bit, and then return it to -1 where it started). Note that the roots move continuously, and then return to their original positions. Next, move a_0 in a big loop (big enough that it orbits around r_2). Something funny happens: the roots r_1 and r_2 switch places.

Pause and think about this for a second. This is really, really weird. Here's one immediate consequence of this observation:

Proposition 1. *There does not exist any continuous function from the space of quadratic polynomials to $\mathbb{C}$ which associates to any quadratic polynomial a root of that polynomial.*

Proof. Exercise. □

Wait, what? Don't we have a quadratic formula?! Well, here's the thing: the quadratic formula is

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

which isn't a function, since it outputs two values for a given input. You might think I'm cheating by writing the $\pm$ in there. What if we just write $+$? Well, it's still not a function, because *$\sqrt{z}$ isn't a function.* We saw an example of this at the very beginning of this essay, when discussing the definition of i. But even in cases when we're used to being able to canonically choose the value of the squareroot – for example, $\sqrt{1}$ – there's more than meets the eye. To convince yourself of this, consider $\sqrt{e^{2\pi i t}}$. No matter how we choose to define this, we end up with a problem if we allow t to vary broadly enough. For example, suppose we set $\sqrt{e^{2\pi i t}} := e^{\pi i t}$. Plugging in $t = 0$ gives $\sqrt{1} = 1$, unsurprisingly. But plugging in $t = 1$ gives $\sqrt{1} = -1$.

Thus, we see that $\sqrt{z}$ isn't a function over $\mathbb{C}$. By contrast, it turns out that $\sin$, $\cos$, and $\exp$ are all continuous functions over $\mathbb{C}$. In particular, Proposition 1 implies:

Corollary 2. *There's no quadratic formula built out of a finite combination of $+, -, \times, \div$, the functions $\sin, \cos, \exp$, and the coefficients of the polynomial.*

3. Connecting loops and permutations

Let's go back to the website and play more with coefficients and roots. Set the degree of the polynomial to be three, and maneuver the coefficients so that

$$a_0 = -\frac{1}{2} + \frac{1}{2}i \qquad a_1 = 0 \qquad \text{and} \qquad a_2 = -\frac{1}{2} - \frac{1}{2}i.$$

This produces three roots, which we label as

$$r_1 = -\frac{1}{2} - \frac{1}{2}i \qquad r_2 = i \qquad \text{and} \qquad r_3 = 1.$$

Moving a_0 around in different loops (making sure it ends up where it started), we see that we can induce at least two different permutations on the roots: the trivial permutation $()$ which fixes all three, and the nontrivial permutation $(1\ 2\ 3)$ which moves r_1 to second position, r_2 to third position, and r_3 to first position. Can we induce other permutations by moving around the other coefficients? For example, can we induce the permutation $(1\ 2)$?

Sure. The trick is to move the roots around, and track what happens to the coefficients. Let's manually induce the permutation $(1\ 2)$ by creating two (non-intersecting) paths connecting r_1 and r_2 in root space. Then we simultaneously move r_1 along one of the paths, and r_2 along the other, until they switch places. As we do this, we keep track of how the coefficients move. The key point is that the coefficients move in a continuous way, and they all return to their initial positions. They have to, since at the end of the day we have the same three roots as we started with (just permutated). Thus, there exist three loops in coefficient space – each based at one of the a_i's – which, when we move the coefficients around them simultaneously, induce the permutation $(1\ 2)$ on the roots. Similarly, we can induce *any* permutation of the roots we wish.

We can simplify the discussion by viewing the three coefficients as the coordinates of a single point. In other words, rather than coefficient space, we form *function space* $\mathbb{C}^3$, where each point (b, c, d) corresponds to a unique cubic $x^3 + bx^2 + cx + d$. This allows us to rephrase the conclusion of the previous paragraph as follows.

Proposition 3. *Pick any three distinct points $r_1, r_2, r_3 \in \mathbb{C}$, and let p denote the point of $\mathbb{C}^3$ corresponding to the polynomial $(x - r_1)(x - r_2)(x - r_3)$. For any permutation of the r_i there exists a loop in $\mathbb{C}^3$ based at p which induces this permutation.*

Note that even though $\mathbb{C}^3$ contains all cubic polynomials (up to a constant factor), the polynomial p in this proposition is of a special kind: all its roots are distinct. We'll be dealing with this type of polynomial a lot, so we give it a name:

Definition. A polynomial of degree n is called *separable* if and only if it has n distinct roots.

This notion allows us to restate Proposition 3 so that we are given the polynomial (as opposed to its roots). Let $\mathcal{F}_3$ denote the set of all points in function space $\mathbb{C}^3$ which correspond to separable cubic polynomials.

Proposition 4. *Pick any $p \in \mathcal{F}_3$ and denote the roots of the corresponding polynomial by r_1, r_2, r_3. For any permutation of the r_i there exists a loop in $\mathcal{F}_3$ based at p which induces this permutation.*

More generally, set

$$\mathcal{F}_n := \{(a_1, a_2, \ldots, a_n) \in \mathbb{C}^n : x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_0 \text{ is separable}\};$$

call this *n-dimensional function space*. Then we have

Proposition 5. *Pick any $p \in \mathcal{F}_n$ and denote the roots of the corresponding polynomial by $r_1, r_2, \ldots, r_n$. For any permutation of the r_i there exists a loop in $\mathcal{F}_n$ based at p which induces this permutation.*

This is one of the three building blocks of our proof of the insolvability of the quintic. Follow me, dear reader!

4. Cubic formulas require nested radicals

Let's return to $\mathcal{F}_3$ (the space of separable cubic polynomials). Pick any $p \in \mathcal{F}_3$, and denote its roots r_1, r_2, r_3. By Proposition 5, we can find a loop γ in function space which induces the permutation $(1\ 2)$ in root space. Since this is a nontrivial permutation, we arrive at the same conclusion as Proposition 1: there does not exist any continuous function which produces a root of every cubic polynomial. Given that this held even for quadratic polynomials, this is hardly surprising. The goal of this section is to go further: we'll show that even if we allow radicals to be used (in combination with continuous functions), any cubic formula requires *nested* radicals.

We've already seen an example of a loop in $\mathbb{C}$ whose image under $\sqrt{\cdot}$ is not a loop: the loop $\gamma : [0, 1] \to \mathbb{C}$ defined by $\gamma(t) := e^{2\pi it}$. No matter how we choose to define $\sqrt{\gamma(0)}$, we find that $\sqrt{\gamma(0)} \neq \sqrt{\gamma(1)}$, even though $\gamma(0) = 1 = \gamma(1)$. However, there are some types of loops whose image under $\sqrt{\cdot}$ *is* a loop. Here's a trivial way to construct such a loop: given any loop γ based at p, denote by γ^{-1} the loop based at p which follows the same path as γ but in reverse. It's not too hard to prove that the image of the composite loop $\gamma\gamma^{-1}$ (by which I mean: go around γ, then go around γ^{-1}) under $\sqrt{\cdot}$ is a loop. A more interesting version of such back and forth travel is called the *commutator*:

Definition. The *commutator* of two loops γ_1, γ_2 (both based at the same point) is defined to be

$$[\gamma_1, \gamma_2] := \gamma_1\gamma_2\gamma_1^{-1}\gamma_2^{-1}.$$

In words, the commutator means: go around γ_1, then around γ_2, then backwards around γ_1, then backwards along γ_2.

The reason we care about commutator loops is that they're trivial enough that their image under radicals is nice, but complicated enough that they can induce nontrivial permutations of roots. We expand on these two points. First:

Exercise 1. Suppose γ_1 and γ_2 are two loops based at the same point in $\mathcal{F}_n$, and pick any continuous function $f : \mathcal{F}_n \to \mathbb{C}$. Then for any $\alpha \in \mathbb{Q}$ the image of $f(p)^\alpha$ as p traverses the loop $[\gamma_1, \gamma_2]$ is a loop in $\mathbb{C}$.

Next, we show that commutator loops can induce nontrivial permutations of roots. More precisely, pick some separable cubic polynomial p. By Proposition 5 there exists some loop γ_1 in function space $\mathcal{F}_3$ which induces the permutation $(1\ 2\ 3)$ on the roots of p, and some other loop γ_2 in $\mathcal{F}_3$ which induces the permutation $(1\ 2)$ on the roots of p.

Exercise 2. What permutation does the commutator loop $[\gamma_1, \gamma_2]$ induce on the roots of p? In particular, show that no root is left fixed.

This immediately implies that the image of $[\gamma_1, \gamma_2]$ under *any* cubic formula cannot be a loop in $\mathbb{C}$! By contrast, Exercise 1 implies that any combination of the four field operations, continuous functions, and a single nesting of radicals *would* produce a loop. We have therefore proved

Proposition 6. *Any cubic formula built solely out of field operations, continuous functions, and radicals must contain nested radicals.*

5. Insolvability of the quintic

In this section we will adapt our arguments from the previous section to prove

Corollary 7. *There does not exist any quintic formula built out of a finite combination of field operations, continuous functions, and radicals.*

The inclusion of the word *finite* above is very important. For example:

Exercise 3. Express a solution to $x^5 - x - 1 = 0$ using just $+$, $\times$, and infinitely many nested radicals.

Our proof of Corollary 7 will look quite similar to that of Proposition 6. We begin with

Proposition 8. *There exists a commutator loop which induces the permutation* $(1\ 2\ 3\ 4\ 5)$.

There are many ways to prove this, the most direct of which is trial-and-error. Although not the most elegant solution, this makes explicit that we don't need any knowledge of abstract group theory for the proof. Here's the approach.

Given any two permutations of five elements, and any separable quintic polynomial p, Proposition 5 guarantees the existence of two loops γ_1 and γ_2 in function space $\mathcal{F}_5$ (based at the polynomial p) which induce these two permutations on the roots of p. Now, the commutator loop $[\gamma_1, \gamma_2]$ induces the commutator of the two corresponding permutations. Thus, given any permutation which is a commutator of two permutations, there exists a commutator loop inducing this permutation. We are thus led to

Question. *Which permutations on five elements can be expressed as the commutator of two permutations on five elements?*

There are 120 different permutations of five elements. If we compute all $14400 = (120)^2$ commutators of pairs of permutations (not so hard to program a computer to do), we see that

- there are precisely 60 different permutations which can be formed as commutators of permutations, and
- $(1\ 2\ 3\ 4\ 5)$ is one of these.

As was the case with cubics, right away this tells us that *any quintic formula built out of the field operations, continuous functions, and radicals must have nested radicals.*

But now we go a step further. Consider the 60 different commutator permutations, and look at the 3600 possible commutators of any pair of these. It turns out we get the same 60 permutations we started with! In particular, $(1\ 2\ 3\ 4\ 5)$ is a commutator of commutators, which means that any quintic formula built out of the field operations, continuous functions, and radicals must have *nested, nested* radicals. Clearly, we need some better way to discuss this. Let's say that an expression built out of the field operations, continuous functions, and radicals has *nesting of level* k iff the expression is of the form

$$\sqrt[a_1]{f_1 + \sqrt[a_2]{f_2 + \sqrt[a_3]{\cdots + \sqrt[a_k]{f_k}}}}.$$

In this language, we see that any quintic formula built out of the field operations, continuous functions, and radicals must have nesting of level ≥ 3.

Since we can keep taking commutators of commutators of commutators of... and still have the same 60 left, including $(1\ 2\ 3\ 4\ 5)$, we can iterate the argument about to conclude that

Theorem 9 (Arnold, 1963?)**.** *Fix any positive integer N. Any quintic formula built out of the field operations, continuous functions, and radicals must have nesting of level $\geq N$.*

Corollary 7 immediately follows.

One important point is that we are not disproving the possibility of solving any particular quintic using finitely many symbols. For example, we can easily express a solution to the quintic equation

$$x^5 - 1 = 0.$$

Instead, we proved the impossibility of a formula which produces a root for *arbitrary* separable quintics.

Exercise 4. Where in the proof did we require the quintic formula to be general? In other words, why doesn't the proof apply to a particular quintic?

6. History and relation to Galois theory

The ultimate goal of a typical introductory course on Galois theory is to develop an algorithm which, given a polynomial, outputs a certain group (the 'Galois group of the polynomial'). The shape of this group then tells you about the shape of the roots of the polynomial, in particular indicating the level of nesting of radicals. For example, the Galois algorithm proves that any root of $x^5 - x - 1$ requires infinite nesting of radicals to write down. In this sense, Galois theory says more than Arnold's theorem (our Theorem 9), which only asserts that any *general* quintic formula must have infinite nesting. However, in another sense, Arnold's approach gives a stronger result than Galois theory: Theorem 9 allows not only the use of radicals but also of $\sin()$, $\exp()$, and any other continuous function.

The other advantage of Arnold's approach is immediately evident to anyone who's taken a course on Galois theory: it's much more straightforward. In Galois theory, the algorithm comes out of a semester's worth of difficult proofs... not to mention that figuring out the Galois group of a given polynomial can be highly nontrivial!

Although Arnold's proof is surprisingly elementary, it should not come as a shock that one can prove such a theorem without Galois theory, since the insolvability of the quintic was originally proved *before* Galois. In 1799 – about 250 years after the discovery of the quartic formula – Paolo Ruffini announced a proof that no general quintic formula exists. However, his proof was 500 pages long (and, as it turned out later, had logical gaps). In 1824, the 22-year-old Norwegian genius Niels Henrik Abel published a six-page proof of Ruffini's assertion; an expanded version appeared two years later in Crelle's journal.[3] Neither Abel's nor Ruffini's proofs relied on Galois groups, since Galois' paper on the subject was only submitted in 1830, and not published until 1846. A wonderful account of Abel's proof and the history can be found in Rosen's essay [Ros95].

I wrote this essay because I was unable to find a suitable written version of Arnold's proof. The most famous written version is the problem book [Ale04], and a more standard description is given in the paper [Zol00].

[3]Apparently Gauss ignored an early manuscript Abel sent him, dismissing it as the work of a crank.

However, the former is hundreds of pages long, while the latter is short but written for a mathematically sophisticated reader. The goal of the present manuscript is meant to preserve the spirit of Arnold's original goal in devising the proof: to make this beautiful theorem accessible to a non-professional mathematician.[4]

Acknowledgements. I am indebted to Boaz Katz, whose beautiful video explanation

`https://www.youtube.com/watch?v=RhpVSV6iCko`

exposed me to Arnold's proof and inspired me to write this document. I'm also grateful to Leo C. Stein for creating and sharing his tool for exploring the interaction between coefficients and roots:

`duetosymmetry.com/tool/polynomial-roots-toy/`

REFERENCES

[Ale04] V. B. Alekseev, *Abel's theorem in problems and solutions*, Kluwer Academic Publishers, Dordrecht, 2004. Based on the lectures of Professor V. I. Arnold; With a preface and an appendix by Arnold and an appendix by A. Khovanskii.

[Ros95] Michael I. Rosen, *Niels Hendrik Abel and equations of the fifth degree*, Amer. Math. Monthly **102** (1995), no. 6, 495–505.

[Zol00] Henryk Zoladek, *The topological proof of Abel-Ruffini theorem*, Topol. Methods Nonlinear Anal. **16** (2000), no. 2, 253–265.

DEPT OF MATHEMATICS & STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA, USA
E-mail address: `leo.goldmakher@williams.edu`

[4] Arnold's goal was to make it accessible to high school students, while this document is probably better-suited to undergraduate math majors.