SYMMETRY AND THE CUBIC FORMULA

DAVID CORWIN

The purpose of this talk is to derive the cubic formula. But rather than finding the exact formula, I'm going to prove that there is a cubic formula. The way I'm going to do this uses symmetry in a very elegant way, and it foreshadows Galois theory. Note that this material comes almost entirely from the first chapter of `http://homepages.warwick.ac.uk/~masda/MA3D5/Galois.pdf`.

1. Deriving the Quadratic Formula

Before discussing the cubic formula, I would like to consider the quadratic formula. While I'm expecting you know the quadratic formula already, I'd like to treat this case first in order to motivate what we will do with the cubic formula. Suppose we're solving

$$f(x) = x^2 + bx + c = 0.$$

We know this factors as

$$f(x) = (x - \alpha)(x - \beta)$$

where α, β are some complex numbers, and the problem is to find α, β in terms of b, c. To go the other way around, we can multiply out the expression above to get

$$(x - \alpha)(x - \beta) = x^2 - (\alpha + \beta)x + \alpha\beta,$$

which means that $b = -(\alpha + \beta)$ and $c = \alpha\beta$. Notice that each of these expressions doesn't change when we interchange α and β. This should be the case, since after all, we labeled the two roots α and β arbitrarily. This means that any expression we get for α should equally be an expression for β. That is, *one* formula should produce *two* values. We say there is an *ambiguity* here; it's ambiguous whether the formula gives us α or β. This was the key insight of Evariste Galois, whose famous theory (later named "Galois theory") was originally called "the theory of ambiguity."

On the other hand, it turns out that this problem is exactly the solution. While we can't go directly from knowing $\alpha + \beta$ and $\alpha\beta$ to knowing α and β, notice that if we also knew $\alpha - \beta$, then we would be able to solve a linear system to get $\alpha + \beta$. We would therefore like to find $\alpha - \beta$ in terms of b and c, in order to achieve our goal of writing α and β in terms of b and c.

It follows that any formula in terms of $b = -(\alpha + \beta)$ and $c = \alpha\beta$ must be unchanged when we switch α and β. But $\alpha - \beta$ is manifestly *not* invariant under switching α and β. In particular, when we switch α and β, the expression $\alpha - \beta$ becomes $\beta - \alpha$, which is the *negative* of $\alpha - \beta$.

The key observation is that if $\alpha - \beta$ is multiplied by -1 upon switching α and β, then

$$(\alpha - \beta)^2$$

Date: January 5, 2021.

will be multiplied by $(-1)^2 = 1$, meaning it is *invariant*. Just as any expression in $b = -(\alpha + \beta)$ and $c = \alpha\beta$ is invariant upon switching α and β, we might guess that conversely, any expression that is invariant can in fact be written in terms of b and c. After a little ingenuity, one notices that

$$(\alpha - \beta)^2 = (\alpha + \beta)^2 - 4\alpha\beta = b^2 - 4c.$$

Our immediate impulse is to write $\alpha - \beta = \sqrt{b^2 - 4c}$, but we must remember that there are *two* square roots of any given nonzero number. It follows that

$$\alpha - \beta = \pm\sqrt{b^2 - 4c}.$$

This gives us the ambiguity we sought after - this method produces two different values for $\alpha - \beta$. And when $b^2 - 4c = 0$, we find precisely that $\alpha = \beta$, which makes sense from both ends. Going back, we see that

$$\begin{aligned} \alpha + \beta &= -b \\ \alpha - \beta &= \pm\sqrt{b^2 - 4c} \end{aligned}$$

And therefore

$$\alpha, \beta = \frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

2. Cubic Equations

We now try to do the same with the cubic equation. We suppose we are trying to solve

$$f(x) = x^3 + bx^2 + cx + d = 0$$

Then, as before, we can write it as

$$f(x) = (x - \alpha)(x - \beta)(x - \gamma) = x^3 - (\alpha + \beta + \gamma)x^2 + (\alpha\beta + \alpha\gamma + \beta\gamma)x - \alpha\beta\gamma,$$

where α, β, γ are the roots of f. Once again, we are trying to find α, β, γ in terms of b, c, d. As before, b, c, d are expressions in α, β, γ that are invariant under all *permutations* of them. There are $3! = 6$ permutations, and for an expression to be expressible in terms of b, c, d, it must remain the same when we apply any of these six permutations.

Before proceeding, we list the permutations:

$$\begin{gathered} \alpha, \beta, \gamma \\ \gamma, \alpha, \beta \\ \beta, \gamma, \alpha \\ \beta, \alpha, \gamma \\ \gamma, \beta, \alpha \\ \alpha, \gamma, \beta \end{gathered}$$

The first three are known as the *cyclic* permutations. The last three are *transpositions* and each fix one of the roots. Note that we can apply two permutations in succession. For example, if we apply the second and then the fifth, the resulting order is β, α, γ, which is the same as applying the fourth. We say that the composition of the second and the fifth is the fourth. We also notice that the composition of any two cyclic permutations is once again cyclic, and the composition of

anything with the first permutation is itself. (If you know what this means, the permutations form a group, S_3, and the cyclic permutations form a subgroup.)

We now proceed to finding expressions in α, β, γ invariant under permutations. As before, we already know $\alpha + \beta + \gamma$, and we want to find *other* linear combinations of the roots which we can solve for in terms of b, c, d.

Notice that in the quadratic case, while $\alpha - \beta$ is not invariant under permutations of the roots, it satisfies a weaker property: it is an eigenvector of the operation of switching the roots. For those of you who don't know linear algebra or don't immediately see a vector space lurking in the picture, that means that while interchanging α and β does not send $\alpha - \beta$ to *itself*, it sends $\alpha - \beta$ to a *multiple* $C(\alpha - \beta)$ of itself, where C is some constant. In this case $C = -1$. Note that $(\alpha - \beta)^2$ was invariant because C was a (second) root of unity.

As before, let's look for a linear combination of α, β, γ that is an eigenvector of a cyclic permutation. I.e., suppose we have some linear expression

$$L(\alpha, \beta, \gamma) = a_1\alpha + a_2\beta + a_3\gamma$$

such that

$$L(\gamma, \alpha, \beta) = CL(\alpha, \beta, \gamma)$$

for some constant C.

What about $L(\beta, \gamma, \alpha)$? Remembering that our variable names are arbitrary, we realize that the equality $L(\gamma, \alpha, \beta) = CL(\alpha, \beta, \gamma)$ just means that shifting everything right and putting the last entry at the front multiplies by C. But this means that $L(\beta, \gamma, \alpha) = CL(\gamma, \alpha, \beta)$, which is of course $C(CL(\alpha, \beta, \gamma)) = C^2L(\alpha, \beta, \gamma)$. Doing this again, we see that $L(\alpha, \beta, \gamma) = CL(\beta, \gamma, \alpha) = C^3L(\alpha, \beta, \gamma)$. So as long as L is not identically zero, this means that $C^3 = 1$.

In particular, this means that $F(\alpha, \beta, \gamma) := (L(\alpha, \beta, \gamma))^3$ is *invariant* under all cyclic permutations of α, β, γ, i.e.

$$F(\beta, \gamma, \alpha) = F(\alpha, \beta, \gamma) = F(\gamma, \alpha, \beta).$$

In other words, if ω is a third root of unity, we want $L(\gamma, \alpha, \beta) = \omega L(\alpha, \beta, \gamma)$. If $L(\alpha, \beta, \gamma) = a_1\alpha + a_2\beta + a_3\gamma$, then

$$a_1\gamma + a_2\alpha + a_3\beta = \omega(a_1\alpha + a_2\beta + a_3\gamma).$$

Equating coefficients, this means that $a_2 = \omega a_1$, $a_3 = \omega a_2$, and $a_1 = \omega a_3$, so assuming WLOG that $a_1 = 1$, we have

$$L(\alpha, \beta, \gamma) = \alpha + \omega\beta + \omega^2\gamma.$$

We set $F(\alpha, \beta, \gamma) = (\alpha + \omega\beta + \omega^2\gamma)^3$ and note that this is invariant under cyclic permutations of α, β, γ.

One can similarly define $M(\alpha, \beta, \gamma) = \alpha + \omega^2\beta + \omega\gamma$. We would get this if we used ω^2, the other primitive third root of unity, in place of ω. Then

$$G(\alpha, \beta, \gamma) := (\alpha + \omega^2\beta + \omega\gamma)^3$$

is also invariant under cyclic permutations of α, β, γ.

What do we do about the transpositions? Notice that the transposition switching β and γ *interchanges* F and G rather than leaving them invariant. What about the other transpositions? Let us switch α and β. Then

$$L(\beta,\alpha,\gamma)=\omega\alpha+\beta+\omega^2\gamma=\omega(\alpha+\omega^2\beta+\omega\gamma)=\omega M(\alpha,\beta,\gamma).$$

In particular,

$$F(\beta,\alpha,\gamma)=G(\alpha,\beta,\gamma).$$

One can similarly see that

$$F(\gamma,\beta,\alpha)=G(\alpha,\beta,\gamma).$$

In particular, while the transpositions do not fix F and G, they *interchange* F and G.

Where have we encountered permutations that swap two things?

Answer: In our derivation of the quadratic formula!

The key is to pretend that F and G are the roots of some quadratic polynomial $(y-F)(y-G)=y^2-(F+G)y+FG$. Then $F+G$ and FG are invariant under cyclic permutations of α,β,γ as F,G each are. As well, since transpositions swap F and G, these quantities are invariant under transpositions too! In particular, FG and $F+G$ are polynomials in α,β,γ that are invariant under *all* permutations of α,β,γ.

We therefore hope that we can express them as polynomials in $b=-(\alpha+\beta+\gamma)$, $c=\alpha\beta+\alpha\gamma+\beta\gamma$, and $d=-\alpha\beta\gamma$, just as we could express $(\alpha-\beta)^2$ in terms of $\alpha+\beta$ and $\alpha\beta$. It turns out that there is a theorem saying we can do this in general. If you don't like proving general theorems, go multiply out $F+G$ and FG and explicitly write it in terms of b,c,d (or if you actually want to explicitly find the cubic formula).

2.1. **The Theorem on Symmetric Functions.** We state the following theorem in the case of three variables α,β,γ. We note that it generalizes to any number of variables, but we will not need this generality.

Theorem 2.1. *Let $P(\alpha,\beta,\gamma)$ be a polynomial in α,β,γ invariant under all permutations of α,β,γ. Then P can be written as a polynomial in $\alpha+\beta+\gamma$, $\alpha\beta+\alpha\gamma+\beta\gamma$, and $\alpha\beta\gamma$.*

Example 2.2. Consider $\alpha^3+\beta^3+\gamma^3$. It is clearly invariant. It turns out we have

$$\alpha^3+\beta^3+\gamma^3=(\alpha+\beta+\gamma)-3(\alpha+\beta+\gamma)(\alpha\beta+\alpha\gamma+\beta\gamma)+3\alpha\beta\gamma.$$

For a proof, see the notes mentioned at the beginning.

2.2. **The Cubic Formula.** From the theorem on symmetric functions, we know that $F+G$ and FG can be expressed as polynomials in b, c, and d. We can then apply the quadratic formula to solve for F and G in terms of b,c,d. One we know F and G, we can take cube roots to get $\alpha+\omega\beta+\omega^2\gamma$ and $\alpha+\omega^2\beta+\omega\gamma$. Finally, using $\alpha+\beta+\gamma=-b$, we can solve for α, β, and γ, since the determinant

$$\begin{vmatrix}1&1&1\\1&\omega&\omega^2\\1&\omega^2&\omega\end{vmatrix}$$

is nonzero. In this way, we can solve for α, β, γ from b, c, d using the elementary operations of arithmetic and square roots and cube roots!

Therefore, there *is* a way to get the roots of a cubic polynomial from its coefficients using cube and square roots, i.e. there is a cubic formula!

Remark 2.3. Notice that there are three possibilities for each square root, making a total of nine possible expressions for the roots, when there are only three of them. It turns out there is a consistent way to choose the cube roots so that it all works out, but we will not detail this.

Remark 2.4. We found that we had to solve a quadratic equation in order to solve a cubic equation, which might seem surprising. It happened because of what happened with permutations. However, we should have expected it for the following reason. For any quadratic polynomial $x^2 + bx + c$, we can consider the cubic polynomial $x^3 + bx^2 + cx$, whose roots include the roots of the quadratic. Therefore, the problem of solving a cubic equation is at least as difficult as solving a quadratic equation.

3. Glimpses Ahead: The Quartic and Quintic

We recall that we found the cubic formula by finding expressions F and G which, while not invariant under *all* of S_3, were invariant under a *subset* of S_3. If an expression is invariant under two permutations, then it is automatically invariant under their composition. We therefore introduce the following terminology.

For any positive integer n, there is a set S_n of permutations on n letters (of which there are $n!$ such permutations). For any two permutations, we can compose them, meaning we first apply one and then the other; there is an identity permutation, doing nothing, and its composition with anything is that thing; and any permutation has an inverse, undoing that operation. For this reason we say that S_n forms a *group* and not just a set. A subset of permutations $H \subseteq S_n$ is called a *subgroup* if it contains the identity, and the composition of any two elements of H is once again an element of H. In particular, the set of permutations fixing a given expression is a *subgroup* of S_3.

In the case of F and G from the cubic formula, it is the subgroup of cyclic permutations that is the subgroup fixing F and G.

We would now like to sketch the idea behind the proof that there is no quintic formula. Suppose there were a quintic formula. Then we might suppose that we want to find polynomials in the five roots invariant under the group S_5 of permutations on five objects.

We would then be able to undo the process by taking roots of polynomials in the coefficients of the polynomial. We want to obtain these through successive linear combinations and powers.

The collection of roots itself is not invariant under anything other than the identity in S_5 (each individual root is fixed by a nontrivial subgroup, but there is no subgroup fixing all of the roots). We then might take some power n_1 of some linear combinations of the roots to get expressions invariant under more of S_5, say a subgroup H_1 (we can, in hindsight, define $H_0 = \{e\}$). Next, after taking linear combinations, we can put our linear combinations to another power n_2, to get expressions invariant under a larger subgroup H_2. We want to repeat, until we get an expression

invariant under all of S_5, i.e. $H_k = S_5$ for some k. In other words, we get a chain of subgroups

$$\{e\} = H_0 \subseteq H_1 \subseteq H_2 \subseteq \cdots \subseteq H_k = S_5.$$

The next key observation is that H_{i-1} is normal in H_i and has quotient injecting into the group of n_ith roots of unity. To see why, suppose some expression P_{i-1} of the roots is invariant under H_{i-1}, and $(P_{i-1})^{n_i}$ is invariant under H_i (in general there will be multiple expressions invariant under each subgroup, since at each stage we will need to solve a system of linear equations to go backwards, but I'll focus on one for simplicity). Then applying any element of H_i to P_{i-1} must multiply P_{i-1} by some n_ith root of unity, and that root of unity is trivial if our element of H_i is actually in H_{i-1}. Since H_{i-1} is presumed to be the group of all permutations that fix P_{i-1}, this map is *injective*. In particular, the quotient H_i/H_{i-1} maps into a cyclic group of order n_i and hence is *abelian*.

So what we've shown is that if there is a quintic formula obtained by the kind of process shown above, then there exists a sequence of subgroups

$$\{e\} = H_0 \trianglelefteq H_1 \trianglelefteq H_2 \trianglelefteq \cdots \trianglelefteq H_k = S_5$$

such that each successive quotient H_i/H_{i-1} for $1 \leq i \leq k$ is abelian. If there were such a sequence, we would say that the group S_5 is *solvable*.[1] One can then use some arguments in group theory to show that S_5 is not solvable (this follows by showing that A_5 is non-abelian and simple). In particular, this shows that there is no quintic formula!

While our argument assumed that any quintic formula could be obtained by a certain slightly contrived process, one can in theory make an argument like this rigorous if one is very careful! The more common way to formulate this argument, though, is to use a little bit of abstract algebra to explain where these permutation groups come from. One can talk about the base field $\mathbb{Q}(b, c, d, \cdots)$, which is the field of rational functions in the coefficients. This injects into the field of rational functions in the roots. The group of permutations of the roots can then be seen as the automorphisms of the larger field (i.e. bijective ring homomorphisms into itself) that *fix* the smaller field. There are then *intermediate* fields, containing the smaller field and contained in the larger field. For example, in the cubic case, $\mathbb{Q}(F, G)$ is an intermediate field containing $\mathbb{Q}(b, c, d)$ and contained in $\mathbb{Q}(\alpha, \beta, \gamma)$. The automorphisms of $\mathbb{Q}(\alpha, \beta, \gamma)$ fixing $\mathbb{Q}(F, G)$ is a subgroup of those fixing $\mathbb{Q}(b, c, d)$; it is the subgroup of cyclic permutations. The proof of the quintic formula proceeds by showing that a quintic formula would give rise to a series of intermediate fields which would in turn give rise to a series of subgroups satisfying the conditions above. These groups of automorphisms of fields are called *Galois groups* and form the subject matter of *Galois theory*.

[1] In fact, the term "solvable" in group theory comes from the fact that it is related to whether an equation is solvable in radicals.