Notes on automorphisms and the Galois group

Math 361, UMass Boston

May 11, 2022

In the previous lecture we introduced the idea of the *splitting field of a polynomial*, which is a field extension $F \to E$ which is "just big enough" to allow factorization of $f \in F[x]$ as a product of linear factors. In practice these are often used for lossless machine representation of expressions involving the roots of f; our construction of splitting fields implies that whenever the arithmetic of F can be exactly implemented on computers, so can the arithmetic of the splitting field.

In the present lecture we zoom back out to the theory of a *general* field extension $F \to E$. We shall be particularly concerned with the *structure* and *symmetry* of a general extension. By the former we mean the usual idea of determining all of the substructures and how they fit together—in this case we shall be concerned with the problem of describing the subfields of E that contain the image of F. By the latter we mean the theory of *automorphisms* of the field extension, which we shall define momentarily. The automorphisms of a given field extension turn out to form a group under composition, the so-called *Galois group* of the extension, and it is reasonable to think of this as describing the "symmetries of the extension" in the same way that groups of rigid motions describe the symmetries of geometric figures.

The key insights of Évariste Galois, summarized here in modern language that was not available to Galois himself, were these: (i) the problem of solving polynomial equations is closely related to the structure theory of splitting fields, and (ii) the structure theory of field extensions is closely related to the structure theory of their Galois groups (which in the case of splitting fields turn out to be finite objects). In this way, Galois was able to "reduce" the problem of solving polynomial equations to a structure problem in finite group theory. This may not sound like an advance! But in fact it was hugely successful and essentially closed the book on classical algebra, by fully solving its central problem. Furthermore, it marked the modern re-emergence of the ancient idea, now pervasive in the physical sciences, that difficult problems are best approached by first analyzing their symmetries, including (since the symmetries described by the Galois group are far from obvious) any possible "hidden symmetries."

We now turn to the actual definitions.

Definition 1. Let (F, E, ι) be a field extension. An *automorphism* of (F, E, ι) is an isomorphism from (F, E, ι) to itself, i.e. a unital ring isomorphism $\phi : E \to E$ satisfying $\phi \circ \iota = \iota$.

Example 2. Let (F, E, ι) be any field extension, and let $\phi : E \to E$ be the identity map $\phi(x) = x$. (Evidently it is best not to denote this by the usual letter ι in this context.) Then ϕ is a unital ring isomorphism and we certainly have $\phi \circ \iota = \iota$. Thus, ϕ is an automorphism of (F, E, ι), usually called the *trivial automorphism*.

Example 3. Define $\iota : \mathbb{R} \to \mathbb{C}$ by the formula $\iota(a) = a + 0i$. Many teachers of elementary mathematics regard $\mathbb{R}$ as an actual subset of $\mathbb{C}$ and would describe ι as an "inclusion map." As we now know, this is not strictly correct, but it is true that ι is a unital ring monomorphism, so that $(\mathbb{R}, \mathbb{C}, \iota)$ is a legitimate field extension. Now define $\gamma : \mathbb{C} \to \mathbb{C}$ by the formula $\gamma(a + bi) = a - bi$ (this map is usually called *complex conjugation*). It is interesting and not especially difficult to verify that γ is a unital ring isomorphism from $\mathbb{C}$ to itself. (If you are lazy then you can find this verification carried out for you in many places, including page 2 of this excellent introduction to complex arithmetic.) Moreover, we have $\gamma(\iota(a)) = \gamma(a + 0i) = a - 0i = a + 0i = \iota(a)$, so indeed $\gamma \circ \iota = \iota$ and γ is a legitimate automorphism of the extension $(\mathbb{R}, \mathbb{C}, \iota)$.

Remark 4. It is helpful in some ways to know that the automorphism of the last example has a geometric visualization. If we identify $\mathbb{C}$ with the Euclidean plane in the usual manner (i.e. the complex number $a + bi$ is identified with the plane point (a, b)), then γ is simply *reflection across the real axis*, which partly explains why people speak of automorphisms as "symmetries" of field extensions (even though in case of a general field extension there is no straightforward geometric visualization).

Remark 5. It is also helpful to meditate on the fact that the automorphism of the previous example *leaves all points of the real axis fixed in place*. This is related to the fact that $\gamma \circ \iota = \iota$. Indeed, any point of the real axis has the form $a + 0i = \iota(a)$, so the condition that all such points are fixed by γ is equivalent to the condition $\gamma(\iota(a)) = \iota(a)$. This is the real significance of the condition $\phi \circ \iota = \iota$ occurring in the definition—it is equivalent to the condition that ϕ *leaves all elements of the image of F within E fixed in place*. (This is a general principle not confined to this particular example, though in this example it is particularly easy to visualize.)

Example 6. Put

$$F = \{a + d\sqrt{2} \,|\, a, d \in \mathbb{Q}\}$$

and

$$E = \{a + b\sqrt[6]{2} + c\sqrt[6]{4} + d\sqrt[6]{8} + e\sqrt[6]{16} + f\sqrt[6]{32} \,|\, a, b, c, d, e, f \in \mathbb{Q}\}.$$

It is not too hard to see that both F and E are subfields of $\mathbb{R}$. (Indeed, F is isomorphic to the field $\mathbb{Q}[x]/\langle x^2 - 2\rangle$ while E is isomorphic to quotient ring $\mathbb{Q}[x]/\langle x^6 - 2\rangle$. One uses Eisenstein's Criterion (Theorem 23.15 of our text) to see that $x^6 - 2$ is irreducible over $\mathbb{Q}$ and thus that E really is a field.) Moreover, since $\sqrt[6]{8} = \sqrt{2}$ we see that F is actually a subset of E. Letting $\iota : F \to E$ denote the inclusion map, we obtain a field extension (F, E, ι).

It is possible though somewhat tedious to verify that the map

$$\phi(a+b\sqrt[6]{2}+c\sqrt[6]{4}+d\sqrt[6]{8}+e\sqrt[6]{16}+f\sqrt[6]{32}) = a-b\sqrt[6]{2}+c\sqrt[6]{4}-d\sqrt[6]{8}+e\sqrt[6]{16}-f\sqrt[6]{32}$$

is a unital ring isomorphism from E to itself. (I do not recommend attempting this verification by brute force; the verification is greatly aided by passing to the abstract models of E and F described a moment ago and then making a certain clever use of the Fundamental Theorem of Homomorphisms.)

However, ϕ is *not* an automorphism of the extension (F, E, ι) since $\phi(\iota(a + d\sqrt{2})) = \phi(a + d\sqrt[6]{8}) = a - d\sqrt[6]{8}$ which is typically not equal to $\iota(a + d\sqrt{2})$, and hence $\phi \circ \iota \neq \iota$. This is a rather technical way of expressing the simple observation that many points of the image of F are being moved around by ϕ, which is specifically forbidden by the definition. Again, the requirement that $\phi \circ \iota = \iota$ is equivalent to the requirement that *extension automorphisms must leave all points of the image of F fixed in place.* Philosophically, the idea is that an extension automorphism not only preserves the intrinsic structure of E itself, but also preserves the specific manner in which the base field is embedded inside E. On a technical level, this condition turns out to be important to establishing the basic properties of the Galois Correspondence (see below).

Theorem 7. *The set of automorphisms of a given field extension forms a group under composition.*

Proof. We already know that the identity map is always an extension automorphism. It is straightforward though tedious to verify that the inverse of a unital ring isomorphism is a unital ring isomorphism, and that the composition of two unital ring isomorphisms is a unital ring isomorphism. For the rest, suppose that ϕ and ψ are automorphisms of the extension (F, E, ι). Then

$$\begin{aligned}(\phi \circ \psi) \circ \iota &= \phi \circ (\psi \circ \iota) \\ &= \phi \circ \iota \\ &= \iota.\end{aligned}$$

Also, applying the function ϕ^{-1} to both sides of the equation $\phi(\iota(a)) = \iota(a)$ gives immediately that $\phi^{-1} \circ \iota = \iota$.

Taken together, these facts imply that the set of automorphisms of (F, E, ι) is in fact a subgroup of the permutation group $\mathrm{Sym}(E)$. □

Definition 8. The group of automorphisms of the field extension (F, E, ι) is called its *Galois group*, and is denoted $\mathrm{Gal}(F, E, \iota)$ or (if it is clear from context and we do not wish to waste a symbol on the embedding ι) $\mathrm{Gal}(F \to E)$ or in some books, $\mathrm{Gal}_F(E)$.

Example 9. Let us compute $\mathrm{Gal}(\mathbb{R}, \mathbb{C}, \iota)$.

We have already seen that the identity map $e : \mathbb{C} \to \mathbb{C}$ and the complex conjugation map $\gamma : \mathbb{C} \to \mathbb{C}$ are automorphisms of this extension. In fact these are the only automorphisms of this extension. To see this, suppose ψ is any automorphism, and consider the equation $i^2 = -1$. Applying ψ to both sides

gives $(\psi(i))^2 = -1$ (remember that ψ must leave real numbers fixed!) and thus $\psi(i) = \pm i$.

If $\psi(i) = i$ then we must have $\psi(a+bi) = \psi(a) + \psi(b)\psi(i) = a + bi$ (because ψ fixes real numbers) and thus $\psi = e$. On the other hand, if $\psi(i) = -i$ then a similar argument shows that $\psi(a+bi) = a - bi$ and thus $\psi = \gamma$.

This shows that $\mathrm{Gal}(\mathbb{R}, \mathbb{C}, \iota) = \{e, \gamma\}$ is a two-element group and is thus isomorphic to $\mathbb{Z}_2$. (It is easy and fun to play the Sudoku game for this group to find its table, and then to verify explicitly that $\gamma \circ \gamma = e$.)

Remark 10. It is a striking fact that even though both $\mathbb{R}$ and $\mathbb{C}$ are infinite fields, the Galois group of $\mathbb{R} \to \mathbb{C}$ turns out to be a *finite* group. Although there do exist field extensions with infinite Galois groups, it is not too hard to prove that *splitting fields* always have finite Galois groups.

Here is a sketch of the argument: suppose $F \to E$ is a splitting field for $f \in F[x]$, and let S denote the set of roots of f in E. Note that S is always a finite set; indeed by the Factor Theorem, its cardinality is bounded above by $\deg(f)$. By an argument similar to the one given in the previous example, any extension automorphism must permute the elements of S. Better still, specifically because E is generated by F and S, any extension automorphism is *determined everywhere* by the manner in which it permutes the elements of S. (This is analogous to the principle that a symmetry of a polygon is determined everywhere by the manner in which it permutes the vertices.) This not only shows that the Galois group is finite, it also exhibits an explicit finite permutation model for it, enabling machine calculations in the Galois group if these are necessary.

The previous example is a special case of this idea. Indeed, $\mathbb{R} \to \mathbb{C}$ is the splitting field of x^2+1, and we have $S = \{i, -i\}$, a two-element set. The identity automorphism e is modeled by the identity permutation, while the conjugation automorphism γ is modeled by the swap.

Definition 11. (The Galois Correspondence) Put $G = \mathrm{Gal}(F \to E)$, and suppose H is a subgroup of G. The *fixed field* of H is the set of all points of E that are left fixed by every element of H. In symbols,

$$\phi(H) = \{e \in E \,|\, \forall h \in H, h(e) = e\}.$$

(N.B.: here we are recycling the symbol ϕ—the function ϕ defined here is not itself an extension automorphism. In this context the symbol ϕ is merely a Greek mnemonic for *fixed*.)

The set $\phi(H)$ is evidently a subset of E, and it is straightforward to show that it is in fact a subfield of E that contains the image of F, i.e. it is a so-called *subextension* of (F, E, ι). Thus, ϕ itself is a function from the set of subgroups of $\mathrm{Gal}(F, E, \iota)$ to the set of subextensions of (F, E, ι). This function is called the *Galois Correspondence*.

Example 12. Let us compute the Galois Correspondence for $\mathbb{R} \to \mathbb{C}$. The Galois group is the two-element group $G = \{e, \gamma\}$. This has only two subgroups, namely the trivial subgroup $\{e\}$ and the improper subgroup $\{e, \gamma\}$.

First let us compute $\phi(\{e\})$. By definition, this consists of all complex numbers that are fixed by e. But *every* complex number is fixed by e, yielding $\phi(\{e\}) = \mathbb{C}$.

Next we tackle $\phi(\{e, \gamma\})$. This consists of all complex numbers which are fixed by *both* e and γ. As we just saw, being fixed by e is a vacuous condition, so we need only determine which complex numbers are fixed by γ. But the complex number $a + bi$ is fixed by γ if and only if $a - bi = a + bi$, i.e. if and only if $b = 0$. This shows that $\phi(\{e, \gamma\}) = \mathbb{R}$. (Technically speaking, it is not truly $\mathbb{R}$ but the image of $\mathbb{R}$ inside $\mathbb{C}$—but life is too short to keep making this distinction explicitly.)

Remark 13. This example already suggests the general fact that the Galois Correspondence is *inclusion-reversing,* i.e. large subgroups correspond to small subextensions and vice-versa.

At last we are ready to state, or at least to summarize, the Fundamental Theorem of Galois Theory: **under certain mild technical hypotheses which we shall not state here, the Galois Correspondence is bijective.**

This result means that to find all subextensions of a field extension, it often suffices to find all subgroups of its Galois group. At at least for splitting fields, the latter is a *finite* problem, and the search could in principle be carried out by machines. (In practice this brute force approach tends to be impractical, which is one of the reasons why mathematicians developed a much more sophisticated theory of finite groups than this course has hinted at.)

No doubt all of this seems quite far removed from the problem of solving polynomial equations. But along with other mathematicians of the early nineteenth century, Galois also realized that solving a polynomial equation "by radicals" (i.e. in roughly the same sense in which the quadratic formula solves quadratic equations) is equivalent to finding a "tower" of field extensions

$$F \to F_1 \to F_2 \to F_3 \to \cdots \to F_n$$

in which F_n is a splitting field for the equation, and each extension in the tower has a particularly simple form. (Believe it or not, the condition is that each individual extension should have an abelian Galois group.) In this way, the problem of solving polynomial equations by radicals was reduced to searching for certain towers of subgroups of the Galois group of the splitting field—a finite problem.

The details of the resulting "Theory of Equations" (a.k.a. "Galois Theory") are beyond the scope of Math 361. However, having completed Math 361, you are well-prepared to begin the study of Galois Theory on your own if you so desire.

Our textbook does contain a presentation of the elements of this theory, but I cannot recommend it. The author, who in every other chapter of his book displays good pedagogical skill, makes a number of unfortunate choices in his presentation of Galois Theory which in my opinion mar the exposition, concealing the essential simplicity and beauty of the theory and making it unnecessarily difficult to learn.

For a more skillfully presented introduction, I recommend one of two options. If you would prefer a leisurely yet thorough presentation of the theory, with many examples and with some discussion of the astonishing history of the subject, consider reading David A. Cox, *Galois Theory.* Or, if you are looking for something much more concise, with only the bare essentials of the theory presented in as few pages as possible, consider Chapter 4 of Nathan Jacobson's *Basic Algebra I* (surely one of the most misleading titles in the whole history of mathematics textbooks).

Galois Theory is the time-honored "next step" after studying the basics of Abstract Algebra, and preparing students to study it was one of the original design goals of the Math 360-361 sequence. However, the emergence of fast computers, and most especially the emergence of long-distance computer networks roughly in the period 1960–1995, with all of the communication and information-security problems that this generated, have made Abstract Algebra relevant to non-mathematicians in new ways. If your tastes run in the direction of computer science or communications engineering, you may wish to forego the study of Galois Theory in favor of Coding Theory, or of Cryptography, or even of the simple and beautiful theory of finite fields. Good introductory books in these three areas are, respectively, Norman L. Biggs, *Codes: An Introduction to Information Communication and Cryptography;* Jeffrey Hoffstein, Jill Pipher, and Joseph Silverman, *An Introduction to Mathematical Cryptography;* and Rudolf Lidl and Harald Niederreiter, *Introduction to Finite Fields and their Applications.*

This concludes Math 361. Each of you has my sincere wish for success in your endeavors, and for lasting joy in the study of beautiful mathematics.