GALOIS THEORY AFTER GALOIS

TAMÁS SZAMUELY

The editor of *Lettera Matematica Pristem* has kindly asked me to write an informal survey of some modern aspects of Galois theory, including infinite Galois theory and Grothendieck's approach. Here is my attempt at fulfilling his request. Of course, since I shall be dealing with advanced subjects, I'll have to assume that readers are familiar with some concepts in higher mathematics, including basic Galois theory as presented in standard algebra textbooks. Later sections will also require some familiarity with the fundamentals of point set topology and complex analysis.

But let us begin at the beginning.

1. What is a Galois extension?

In the latter half of the 19th century eminent mathematicians such as Liouville, Jordan, Dedekind and Weber devoted considerable effort to understanding and developing the seminal work of Galois. During this work of clarification the focus of Galois theory gradually shifted from equations to field extensions until it arrived at the form we use nowadays. In the course of this process the concept of a Galois extension of fields was redefined several times. Let us list four equivalent formulations, in increasing order of abstraction.

The definition which is closest to the original approach of Galois is:

(1) *A Galois extension is a field extension $L|K$, where L is obtained by adjoining to K all roots of an irreducible separable polynomial with coefficients in K.*

Recall that an irreducible polynomial is separable if it has no multiple roots. Correspondingly, a field extension is called separable if the minimal polynomials of all elements are separable.

As a next step, it was noticed that the above defining property can be made independent of the choice of a polynomial.

(2) *A Galois extension is a finite separable field extension $L|K$ having the following property: if an irreducible polynomial with coefficients in K has a root in L, then all of its roots lie in L.*

Here the emphasis is already on fields but polynomials still lurk behind. A further step towards abstraction became possible when Dedekind defined the Galois group as the automorphism group of a

Date: February 6, 2012.

Galois field extension. In previous approaches it was always defined as a group permuting the roots of a defining polynomial.

(3) *A Galois extension is a finite separable field extension $L|K$ where every element of L not lying in K is moved by a field automorphism of L fixing K.*

The link with the previous definition is as follows: an automorphism of L fixing K must take an element $\alpha \in L$ to another root of its minimal polynomial over K. If α is not in K, there is at least one other such root.

The last step was taken by Emil Artin in the 1940's. It is the most elegant definition of a finite Galois extension.

(4) *A Galois extension is a field extension $L|K$, where K arises as the fixed field of a finite group G acting on L via field automorphisms.*

To relate this to the previous definition, one has to prove that a field extension as above is automatically separable. Afterwards the defining property follows. The group G is called the Galois group of the extension $L|K$.

It was also Artin who formulated the Galois correspondence as we know it today. Let us recall the statement:

Let $L|K$ be a finite Galois extension with Galois group G. There is a one-to-one correspondence between extensions $M|K$ contained in L and subgroups $H \subset G$. Galois extensions correspond to normal subgroups. In this case the Galois group of M over K is isomorphic to G/H.

The correspondence is given by mapping M to the subgroup of G whose elements fix M elementwise, and conversely, by mapping a subgroup $H \subset G$ to the subfield of L it fixes.

This statement is purely in the spirit of the great German algebra school of the early 20th century. It is still the one we teach and apply today. There was, however, one further step to be taken: to relax the condition of finiteness.

2. The absolute Galois group

There have been several motivations for studying infinite algebraic extensions. One came from the investigation of special classes of extensions in algebraic number theory. For instance, cyclotomic fields, which are obtained by adjoining to the field $\mathbf{Q}$ of rational numbers some root of unity ω, have always played a prominent role since the discovery of their relevance to Fermat's last theorem. Any n-th root of ω is again a root of unity, so cyclotomic fields give naturally rise to towers of field extensions. For instance, one may fix a prime number p, and consider the tower $\mathbf{Q}(\mu_p) \subset \mathbf{Q}(\mu_{p^2}) \subset \mathbf{Q}(\mu_{p^3}) \subset \cdots$ of fields obtained by adjoining p-th, p^2-th, p^3-th roots of unity, and so on. Their union is a very interesting infinite algebraic extension whose study gave

rise to one of the most important branches of present-day arithmetic, *Iwasawa theory.*

Another motivation comes from a purely algebraic question: how to find fields over which every polynomial equation has a solution? That this question is not purely academic goes back to a famous theorem of Gauss, nowadays known as the *fundamental theorem of algebra:*

Over the field $\mathbf{C}$ *of complex numbers every polynomial equation* $f = 0$ *has a root* $\alpha \in \mathbf{C}$.

Moreover, all roots of the equation must lie in $\mathbf{C}$, as one sees by dividing f by factors of the form $(x - \alpha)$.

But how to find other examples? There is a natural idea: start with a field K and try to find a means of adjoining to K *all* solutions of polynomial equations with coefficients in K. The resulting field $\overline{K}$ will then have the required property. Indeed, given a polynomial f with coefficients in $\overline{K}$, its finitely many coefficients lie in some finite extension L of K by construction of $\overline{K}$. But then a root α of f lies in a finite extension of L which is again a finite extension of K. As such, it is generated by roots of polynomials with coefficients in K, so α must lie in $\overline{K}$.

For $K = \mathbf{Q}$ the construction of $\overline{K}$ is easy. It suffices to view $\mathbf{Q}$ as a subfield of $\mathbf{C}$ and take all complex numbers which are roots of some polynomial with coefficients in $\mathbf{Q}$. One shows that these numbers form a field, the *field of algebraic numbers.* Since we know by the fundamental theorem of algebra that the roots of every polynomial with coefficients in $\mathbf{Q}$ lie in $\mathbf{C}$, we have solved our problem.

For those fields that cannot be embedded in $\mathbf{C}$ (for instance those of positive characteristic or transcendental extensions of $\mathbf{C}$) the above method breaks down. The general solution was found in a fundamental paper of Steinitz [6] which exploited the then fairly new techniques of set theory, in particular the axiom of choice. He showed:

Every field K *has an* algebraic closure $\overline{K}$ *which is an algebraic extension of* K *such that every polynomial equation with coefficients in* $\overline{K}$ *has a root in* $\overline{K}$. *Moreover,* $\overline{K}$ *is unique up to a* non-unique *isomorphism.*

It is this non-unicity which is crucial for developing infinite Galois theory. The fact that $\overline{K}$ is algebraic over K means that it is generated by roots of polynomials with coefficients in K. We thus see that $\overline{K}$ indeed arises by adding all roots of such polynomials in a systematic way.

Once we have an algebraic closure $\overline{K}$ at hand, we may define a separable closure K_s of K as being those elements in $\overline{K}$ whose minimal polynomial over K is separable. It can be checked that these elements form a field over which every separable polynomial has a root. Consider now the group of field automorphisms of K_s fixing K. It is the

absolute Galois group of K; we denote it by Γ. It depends on the choice of K_s but its isomorphism class does not.

The absolute Galois group Γ has several important features. Firstly, it has the property of definition (3) above: any element $\alpha \in K_s$ not in K is moved by some element of Γ. The proof of this fact uses another nontrivial theorem of Steinitz: if we map α to another root α' of its minimal polynomial, the resulting isomorphism $K(\alpha) \xrightarrow{\sim} K(\alpha')$ can be extended to an automorphism of K_s. Another important point is that K_s is the union of all finite Galois extensions of K contained in $\overline{K}$; this is because one may embed any finite separable extension in a finite Galois extension. Furthermore, the consideration of Γ gives us a means for giving yet another definition of a finite Galois extension:

(5) *A finite separable extension $L|K$ contained in K_s is Galois if $\sigma(L) \subset L$ for all $\sigma \in \Gamma$.*

That L should be contained in K_s is not a serious restriction, as any L separable over K can be embedded in K_s. We thus get another insight into the basics of Galois theory: one can decide whether an extension is Galois by using the absolute Galois group which is 'always there'. The next step is to use it to describe *all* subfields of K_s.

3. Infinite Galois Theory

The discussion of the previous two sections points to several equivalent ways of defining a possibly infinite Galois extension. We mention two of them. The first is a variant of definition (3): a Galois extension is a separable algebraic extension $L|K$ such that every element of L not lying in K is moved by a field automorphism of L fixing K. The second one is motivated by definition (5): an extension $L|K$ contained in K_s is Galois if $\sigma(L) \subset L$ for all σ in the absolute Galois group Γ. The most interesting infinite Galois extension of K is of course the separable closure K_s.

Whichever definition we adopt, there is only one way to define the Galois group $\mathrm{Gal}(L|K)$: it is the group of field automorphisms of L fixing K. However, it is the second definition that gives us an easy key to a fundamental property of the Galois group. Namely, given a Galois extension $M|K$ contained in L, one has a natural group homomorphism $\mathrm{Gal}(L|K) \to \mathrm{Gal}(M|K)$ given by restricting automorphisms of L to M. This homomorphism is moreover surjective by the theorem of Steinitz already mentioned in the previous section: one may extend any automorphism of M over K to an automorphism of K_s which must then preserve L by definition. Thus $\mathrm{Gal}(M|K)$ arises as a quotient of $\mathrm{Gal}(L|K)$; in particular this applies to all finite Galois extensions contained in L.

The basic fact is now that $\mathrm{Gal}(L|K)$ is completely determined by its finite quotients. This is not so surprising if we recall from the

previous section that L is the union of all the finite Galois extensions it contains. However, giving a precise formulation requires a sophisticated algebraic tool called the *inverse limit:* one says that $\mathrm{Gal}(L|K)$ is the inverse limit of its finite quotients. A group that arises as an inverse limit of finite groups is called a *profinite group.* We shall not give the detailed definition of inverse limits and profinite groups here. What is important to bear in mind is that profinite groups are determined by their finite quotients.

Profinite groups carry an important additional structure: they are *topological groups.* It is possible to introduce this topology by a general method that starts by putting the discrete topology on finite quotients. However, in the case of Galois groups there is a more direct approach introduced by Krull in his groundbreaking paper [4] on infinite Galois theory. Consider the kernels of the homomorphisms $\mathrm{Gal}(L|K) \to \mathrm{Gal}(M|K)$ considered above for all finite Galois extensions $M|K$ contained in L, and declare these to be a system of open neighbourhoods of the identity in $\mathrm{Gal}(L|K)$. For a general element $\sigma \in \mathrm{Gal}(L|K)$ a system of open neighborhoods is given by the cosets σU, where U is an open neighbourhood of the identity. One checks that these open sets form the basis of a topology on $\mathrm{Gal}(L|K)$. In honour of its father it is called the *Krull topology.* It can be shown that with the Krull topology $\mathrm{Gal}(L|K)$ becomes a compact Hausdorff topological group.

Any open subgroup in $\mathrm{Gal}(L|K)$ is also closed because its complement is a union of cosets which must be open as well. Hence any, possibly infinite, intersection of open subgroups is a closed subgroup. On the other hand, it is not hard to show that every closed subgroup is the intersection of the open subgroups containing it. This gives a hint at how one should develop the Galois correspondence for infinite extensions: since open normal subgroups correspond to finite Galois extensions, closed subgroups should correspond to arbitrary Galois extensions, again because the latter are unions of finite Galois extensions. Indeed, Krull proved the following generalized Galois correspondence:

Let $L|K$ be a Galois extension with Galois group G. There is a one-to-one correspondence between extensions $M|K$ contained in L and closed subgroups $H \subset G$. Finite extensions correspond to open subgroups and Galois extensions to closed normal subgroups. In the latter case the Galois group of M over K is isomorphic to G/H.

At this point the natural question arises whether there exist non-closed subgroups in the Galois group or, in other words, whether there exist subgroups that do not arise as the subgroup of elements of G fixing some field extension. In fact, this question had been solved by Dedekind [2] well before Krull set up his theory. He argued as follows:

If $L_1 \subset L_2 \subset L_2 \subset \dots$ is a strictly increasing chain of finite Galois extensions of K, then each automorphism in $\mathrm{Gal}(L_i|K)$ can be extended in at least two different ways to L_{i+1} by classical Galois theory. An infinite chain thus gives rise to an uncountable Galois group which has uncountably many subgroups. But if the base field K is countable, e.g. $K = \mathbf{Q}$, then there are only countably many finite field extensions of K because there are only countably many polynomials with coefficients in K. It should be remarked in passing that the aforementioned work of Dedekind was the main inspiration for Krull's theory. In fact, Dedekind already had the insight that infinite Galois groups should enjoy some continuity property.

In concrete situations it is easy to exhibit non-closed subgroups. For instance, if $\mathbf{F}_q$ denotes the finite field of order q, then the Frobenius automorphism $F : x \mapsto x^q$ of $\overline{\mathbf{F}}_q$ does not generate a closed subgroup. In fact, it is easy to describe the Galois theory of the infinite extension $\overline{\mathbf{F}}_q|\mathbf{F}_q$ without mentioning profinite groups. As we know from the theory of finite fields, for each integer $r > 0$ there is a unique subextension $\mathbf{F}_{q^r}|\mathbf{F}_q$ of $\overline{\mathbf{F}}_q|\mathbf{F}_q$ which has degree r over $\mathbf{F}_q$, and moreover it is a Galois extension of $\mathbf{F}_q$ with group $\mathbf{Z}/r\mathbf{Z}$. It follows that the open subgroups of $\Gamma = \mathrm{Gal}(\overline{\mathbf{F}}_q|\mathbf{F}_q)$ are totally ordered by inclusion, and therefore every system of open subgroups either has trivial intersection or has a smallest element. Hence every nontrivial closed subgroup $H \subset \Gamma$ is in fact open; moreover, it is normal with Γ/H cyclic. The infinite cyclic subgroup of Γ generated by F is not open: its fixed field is $\mathbf{F}_q$ but it does not equal Γ.

4. Grothendieck's Reformulation

Alexander Grothendieck, whose influence on mathematics in the latter half of the 20th century was comparable to that of Galois in the 19th, found a very useful reformulation of the main theorem of Galois theory which can be generalized to many other settings as well. In his seminar [3] he gave a general categorical formulation encompassing several situations. We stick here to the already explored case of field extensions. From Grothendieck's viewpoint the aim of Galois theory is to classify finite separable extensions of a given field by means of permutation representations.

To explain his idea, let K be a base field and $L|K$ a finite separable extension. Fix a separable closure K_s of K. As we know from the work of Galois himself, L is generated over K by a single element α. Let f be the minimal polynomial of α over K, and $\alpha_1, \dots, \alpha_n$ the roots of f in K_s. The absolute Galois group $\Gamma := \mathrm{Gal}(K_s|K)$ acts on the finite set $\alpha_1, \dots, \alpha_n$ via permutations: for each $\sigma \in \Gamma$ the n-tuple $\sigma(\alpha_1), \dots, \sigma(\alpha_n)$ is just the system of the α_i listed in a possibly different order. There are two important properties of this action. Firstly, it is *transitive*, which means that for given α_i, α_j we may find a $\sigma \in \Gamma$ with

$\sigma(\alpha_i) = \alpha_j$. Secondly, it is *continuous* for the topology of Γ. This means that for each α_i the set of those $\sigma \in \Gamma$ for which $\sigma(\alpha_i) = \alpha_i$ is an *open* subgroup of Γ.

Grothendieck's reformulation now can be expressed as follows:

There is a one-to-one correspondence between isomorphism classes of finite separable extensions of K and finite sets S equipped with a continuous transitive action of Γ.

We have already seen how to associate a finite continuous Γ-set to a field extension. To get a map in the reverse direction, one picks an element $\alpha_i \in S$ and considers its stabilizer in Γ, i.e. those $\sigma \in \Gamma$ for which $\sigma(\alpha_i) = \alpha_i$. It is an open subgroup in Γ, so by infinite Galois theory it fixes a finite separable extension $L|K$. If we choose the stabilizer of another element, we arrive at the fixed field of a conjugate open subgroup, which fixes a field extension isomorphic to L.

In a sense the above formulation of Galois theory is closer to Galois' original approach than Artin's Galois correspondence because Galois also considered the permutation representation of the Galois group on the roots of the equation – this aspect is somewhat obscured if one only looks at the subgroups/subfields correspondence.

One can relax the condition of transitivity of the Γ-action by considering not just finite separable extensions of K but also finite direct products of these. Following Grothendieck they are usually called *finite étale k-algebras.*

5. Monodromy Representations

Another situation where equations can be classified by means of group representations that was already abundantly studied in the 19th century is the monodromy theory of differential equations. Consider a linear differential equation

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y = 0 \tag{1}$$

in the complex plane $\mathbf{C}$, where the coefficients a_i are complex functions that are holomorphic except in finitely many points $x_1, \ldots, x_r$ where they extend meromorphically. By a basic existence theorem of Cauchy each point $x \neq x_i$ has an open neighbourhood U not containing any of the x_i where the equation has n local holomorphic solutions $y_1, \ldots, y_n$ that are linearly independent over $\mathbf{C}$ and moreover every local solution over U is a linear combination of these. In other words, locally around x the solutions of the equation form an n-dimensional $\mathbf{C}$-vector space.

The trouble is that when we move the point x the solution space may not remain the same. To see this, consider the simplest example where $r = 1$ and the equation is of the form

$$y' = fy$$

with a meromorphic function f that is holomorphic outside x_1. As is well known, all local solutions of the equation are constant multiples of the function $\exp \circ F$, where F is a primitive of f. However, as we learn in basic complex analysis, a primitive of F does not exist on the whole of $X \setminus \{x_1\}$, only in $X \setminus D$, where D is a half-line starting from x_1. For instance, if $x_1 = 0$, then a well-defined primitive F_- of f exists over $U_- = \mathbf{C} \setminus [0, -i\infty)$ and another primitive F_+ over $U_+ = \mathbf{C} \setminus [0, i\infty)$. The intersection $U_- \cap U_+$ splits in two connected components $C_- = \{z : \mathrm{Re}(z) < 0\}$ and $C_+ = \{z : \mathrm{Re}(z) > 0\}$. As F_+ and F_- may differ only by a constant on each component, we are allowed to choose them in such a way that $F_- = F_+$ on C_-. However, they will then differ by a constant c on C_+! So if we choose a closed loop around 0, say the circle of radius 1, then in a neighbourhood of 1 we may take $\exp \circ F_+$ as a local solution. Moving along the circle away from 1 we may still use the same local solution for some time around each point of the circle, but at some point in C_- we must switch from F_+ to F_- because F_+ is not defined along the whole circle. So when we arrive back to 1, we get $\exp \circ F_-$ as the 'continuation' of the local solution, which is $e^c \cdot (\exp \circ F_+)$. This constant e^c is the *monodromy* of the equation around 0.

The above procedure can be generalized to the general equation (1). Choose closed loops γ_i passing through x such that each γ_i has only the point x_i in its interior but not the other x_j. For fixed i we can do the same as in the special case discussed above: we start with a local solution y_x of (1) around x and we move along γ_i continuing it to local solutions around each point. When we return to x we obtain another local solution z_x around x. It is usually not a constant multiple any more because the equation can be more complicated but it lies in the same finite-dimensional vector space of local solutions around x.

We can proceed likewise for all i. An apparent drawback is that the results *a priori* depend on the choice of the loops γ_i. However, the classical *monodromy theorem* of complex analysis says that this is not the case: if we replace γ_i by another loop γ_i' having the same properties, then the resulting z_x will be the same. From the modern point of view this is because γ_i and γ_i' are *homotopic*: they can be continuously deformed into each other without touching any of the x_i along the way. This observation is the germ of the notion of the *fundamental group*: one may consider *all* closed loops passing through x and not touching the x_i up to continuous deformation. It is possible to introduce a group operation on this set by taking as the product of two loops γ and δ the loop obtained by going through δ and then γ. (It goes through x twice but one may deform it so that it is not the case any more. For instance, one may represent the product $\gamma_i \cdot \gamma_j$ by a loop around x not meeting itself and having x_i and x_j in its interior.) The resulting group Π is generated by the classes of the γ_i. Observe now that the operation of

continuing local solutions of (1) along loops representing elements of Π is $\mathbf{C}$-linear. In other words, we obtain an action of Π on the $\mathbf{C}$-vector space of solutions of (1) around x. Fixing a basis $y_1, \ldots, y_n$ of this space we thus obtain a homomorphism $\rho : \Pi \to \mathrm{GL}(n, \mathbf{C})$. It is called the *monodromy representation* of (1) around x.

The question now arises: can one classify differential equations by their monodromy representations in the same way as permutation representations of the Galois group classify finite field extensions? The perceptive reader will immediately notice that the question is not correctly formulated. Giving a field extension is not the same as giving a polynomial equation, although one knows that every finite field extension does indeed come from the polynomial. The question for differential equations must be similarly split in two parts.

The first part is usually called the *Riemann–Hilbert correspondence.* An elementary form of it says:

There is a one-to one correspondence between isomorphism classes of monodromy representations $\rho : \Pi \to \mathrm{GL}(n, \mathbf{C})$ *and* local systems*, i.e. systems of holomorphic functions on open subsets of* $\mathbf{C}$ *avoiding the* x_i *which locally around each point form an* n*-dimensional* $\mathbf{C}$*-vector space.*

The second part is the *Riemann–Hilbert problem* which asks:

Does every local system (and hence every monodromy representation) come from a linear differential equation (1)*?*

The answer is not unambiguous, especially if one requires additional properties of the equation. Let us cite the most famous classical result, that of Plemelj [5]:

Every local system as above comes from a linear differential equation (1) *whose coefficients are meromorphic functions in* $\mathbf{C}$*. Moreover, the equation can be chosen to be* Fuchsian*, which means that at each singular point the the coefficient* a_i *has a pole of order at most* i*.*

It is important to observe that though in the above statement the coefficients of the equation have at most poles at the x_i, they may have poles at finitely many other points as well. One may consider a monodromy representation taking these extra points into account as well, but with the additional condition that the action of group elements coming from loops around the extra points is trivial. In other words, continuing a local solution around the extra points does not change the solution. Such points are classically called *apparent singularities.* If one does not allow apparent singularities, the statement does not hold. For a beautiful introduction to the Riemann–Hilbert problem and its generalizations, see Beauville's report [1].

We have thus indeed arrived at statements that resemble the modern formulation of Galois theory. Thanks largely to insights of Grothendieck it is now possible to develop several general theories encompassing both;

see e.g. my book [7] where the concepts surveyed in this article are also explained in more detail. But the ties between these branches of mathematics are stronger. In recent decades differential equations, and even more the topological considerations arising from their theory, have been successfully applied to the construction of interesting Galois extensions of fields like $\mathbf{C}(t)$ and even $\mathbf{Q}$. On the other hand, methods imported from Galois theory have proven to be fundamental for analyzing differential equations. We can thus happily observe that two centuries after Galois his ideas are not only more alive than ever but have also invaded a large part of present-day mathematical research.

References

[1] A. Beauville, Monodromie des systèmes différentielles linéaires à pôles simples sur la sphère de Riemann [d'après A. Bolibruch], Séminaire Bourbaki, exposé 765, *Astérisque* 216 (1993), 103–119.

[2] R. Dedekind, *Über die Permutationen des Körpers aller algebraischen Zahlen*, Abhandlungen der Gesellschaft der Wissenschaften zu Göttingen, 1901.

[3] A. Grothendieck, *Revêtements étales et groupe fondamental* (SGA 1), Lecture Notes in Mathematics, vol. 224, Springer-Verlag, Berlin-New York, 1971. New annotated edition: Société Mathématique de France, Paris, 2003.

[4] W. Krull, Galoissche Theorie der unendlichen algebraischen Erweiterungen, *Math. Ann.* **100** (1928), 687–698.

[5] J. Plemelj, Riemannsche Funktionenscharen mit gegebener Monodromiegruppe, *Monatshefte Math. Phys.* 19 (1908), 211–246.

[6] E. Steinitz, Algebraische Theorie der Körper, *J. reine angew. Math.* 137 (1908), 167–309.

[7] T. Szamuely, *Galois Groups and Fundamental Groups*, Cambridge Studies in Advanced Mathematics, vol. 117, Cambridge University Press, 2009.