DIFFERENTIAL GALOIS THEORY: PROVING ANTIDERIVATIVES AREN'T ELEMENTARY

ARUN DEBRAY AND ROK GREGORIC
AUGUST 5, 2019

TODO: standard blurb

0. Overview

Every year we tell our calculus students that the Gaussian e^{-x^2} has no elementary antiderivative. It's striking and accessible. But the proof is not well known, even though it's absolutely within reach of graduate students. The two of us were interested in learning the proof (and a few other things related to differential algebra); the lecture notes are currently Arun's notes (in progress!) for his talks, leading to a proof of Liouville's theorem, following Hubbard-Lundell (http://pi.math.cornell.edu/~hubbard/diffalg1.pdf). Please let us know if you find any mistakes or typos.

1. Lightning review of Galois theory

Our first goal is to prove that functions such as e^{-x^2} have no elementary antiderivatives; we may more generally consider elementary solutions to differential equations. The proof follows a similar line of reasoning as in Galois theory: study the group of symmetries of a minimal field containing solutions to the equations, and prove that only certain symmetry groups can arise if we want elementary functions. If it's been a while since you've seen Galois theory, you are in good company, so let's begin with a quick review.

Galois theory studies the symmetries of polynomials over fields. It works in great generality, but to simplify the exposition we will assume the base field k has characteristic zero.

Definition 1.1. A *(field) extension* is a map of fields $j\colon k \hookrightarrow L$ (i.e. a ring homomorphism, where L is also a field). Such a map is necessarily injective.

For now, assume for simplicity that this is a *finite* field extension, meaning j makes L into a finite-dimensional k-vector space.

Definition 1.2. A *splitting field* for a collection of polynomials $S \subset k[x]$ is a field extension $k \hookrightarrow L$ such that all $f \in S$ factor completely (i.e. into linear functions), and that L is minimal with respect to this property. A *normal extension* is one isomorphic to the splitting field of some collection of polynomials.

The idea is that a splitting field of f is the minimal field containing all of the roots of f. Abstractly, splitting fields exist and are unique up to unique isomorphism, but you could also just always work inside $\mathbb{C}$.

Example 1.3. If $f(x) = x^3 - 2 \in \mathbb{Q}[x]$, then $\mathbb{Q}(\sqrt[3]{2})$ is not the splitting field of f: the other two roots of unity are $e^{\pm 2\pi i/3}\sqrt[3]{2}$. Therefore the splitting field of f is $\mathbb{Q}(\sqrt[3]{2}, e^{2\pi i/3})$. ◀

(Finite) normal extensions are examples of *Galois extensions*; in our setting these are synonymous, but not in characteristic p. In this case, the group of symmetries is nice.

Definition 1.4. If $j\colon k \hookrightarrow L$ is a Galois extension, its *Galois group* $\mathrm{Gal}(L/k)$ is the group of automorphisms of L (as a field) which fix k.

The Galois group of the splitting field of $f \in k[x]$ permutes the roots of f, and in fact is a subgroup of $S_{\deg f}$

For example, for $\mathbb{Q} \hookrightarrow \mathbb{Q}(\sqrt[3]{2}, e^{2\pi i/3})$, the Galois group is S_3: complex conjugation swaps the two complex roots, giving us a transposition, and we get the 3-cycle from the automorphism

$$a + b\sqrt[3]{2} + ce^{2\pi i/3}\sqrt[3]{2} \longmapsto a + e^{2\pi i/3}\Big(b\sqrt[3]{2} + ce^{2\pi i/3}\sqrt[3]{2}\Big). \tag{1.5}$$

The idea of a Galois group leads quickly to two important theorems.

Given a group G, let $\mathcal{L}(G)$ denote its poset of subgroups, ordered by inclusion Given a field extension $k \hookrightarrow L$, let $\mathrm{Ext}(L/k)$ denote the poset of subextensions of $k \hookrightarrow L$; this is a poset, ordered by inclusion. If $k \hookrightarrow L$ is Galois, then given a subgroup $H \leq \mathrm{Gal}(L/k)$, let L^H denote the subfield fixed by the action of H.

Theorem 1.6 (Fundamental theorem of Galois theory)**.** *Let $k \hookrightarrow L$ be a Galois extension. The assignments*

$$\begin{aligned} \mathcal{L}(\mathrm{Gal}(L/k))^{\mathrm{op}} &\longrightarrow \mathrm{Ext}(L/k) \\ H &\longmapsto L^H \end{aligned} \tag{1.7a}$$

and

$$\begin{aligned} \mathrm{Ext}(L/k)^{\mathrm{op}} &\longrightarrow \mathcal{L}(\mathrm{Gal}(L/k)) \\ L' &\longmapsto \mathrm{Aut}(L'/k) \end{aligned} \tag{1.7b}$$

define an order-reversing isomorphism of posets. Moreover, the degrees match: $\dim_{L^H} L = |H|$ *and* $\dim_k L^H = |\mathrm{Gal}(L/k)|/|H|$. $k \hookrightarrow L^H$ *is Galois iff* $H \trianglelefteq \mathrm{Gal}(L/k)$*; in this case,* $\mathrm{Gal}(L^H/k) \cong \mathrm{Gal}(L/k)/H$.

But our immediate focus is a different theorem.

Definition 1.8. Let $\mathbb{Q} \hookrightarrow L$ be a field extension. An $x \in L$ is *solvable by radicals* if:

- $x \in \mathbb{Q}$,
- x is the sum, product, difference, or quotient of two numbers solvable by radicals, or
- x is the n^{th} power or n^{th} root of a number solvable by radicals.

A polynomial $f \in \mathbb{Q}[x]$ is *solvable by radicals* if its roots are, where L is its splitting field.

So the quadratic formula, cubic formula, and quartic formula show all polynomials of degree at most four are solvable by radicals.

Theorem 1.9 (Abel-Ruffini)**.** $f \in \mathbb{Q}[x]$ *is solvable by radicals if and only if the Galois group of its splitting field is solvable. In particular, for $d \geq 1$, there are degree-d polynomials with Galois group S_d; hence, for $d \geq 5$, there exist degree-d polynomials not solvable by radicals.*

Recall that a finite group G has a Jordan-Hölder composition series $1 \trianglelefteq G_1 \trianglelefteq G_2 \trianglelefteq \cdots \trianglelefteq G_n = G$, where the quotients G_i/G_{i-1} are simple groups (i.e. they have no nontrivial normal subgroups). We say G is *solvable* if said quotients are all abelian.

How does the proof of the Abel-Ruffini theorem go? The vague basic idea is: both solvability by radicals and solvability of the Galois group are describing your splitting field as an iterated sequence of particularly nice field extensions. Specifically, adjoining an n^{th} root of some element of your field is an *abelian extension*, i.e. $\mathrm{Aut}(k(\sqrt[n]{a})/k)$ is abelian, and, using Theorem 1.6, if the Galois group of $k \hookrightarrow L$ is solvable, the Jordan-Hölder decomposition describes it as a composition of abelian extensions.

2. Basics of differential algebra

To discuss differential equations we need derivatives.

Definition 2.1. A *differential field* is a field k together with a *derivation* $\delta\colon k \to k$, i.e. a k-linear map satisfying the *Leibniz rule* $\delta(fg) = f\delta(g) = \delta(f)g$. The *constants* in k are those elements with $\delta(k) = 0$; these form a subfield.

So any field of functions with the usual derivative works. We will think of $\mathbb{C}(t)$ as our "base field," analogous to $\mathbb{Q}$ in Galois theory. Another good example is the field $\mathcal{M}(U)$ of meromorphic functions on some connected open set $U \subset \mathbb{C}$: the existence and uniqueness theorem for ODEs tells us that any system of differential equations has a solution in $\mathcal{M}(U)$ for some U. This will play the role that $\mathbb{C}$ did in Galois theory, sidestepping a lot of existence and uniqueness questions at once. In fact, given a differential operator L acting on $\mathbb{C}(t)$, let $U_L \subset \mathbb{C}$ denote the maximal open subset on which $Lu = 0$ has solutions.

In the rest of this section and the next, k is a differential field containing $\mathbb{C}(t)$ and contained in $\mathcal{M}(U)$ for some U.

Definition 2.2. Let L be a differential operator on k. The *(differential) splitting field* for L, denoted E_L, is the smallest subfield of $\mathcal{M}(U_L)$ containing k and the solutions of L.

Example 2.3. Consider the differential operator $L(u) \coloneqq u - u'$. Of course, the solutions to $Lu = 0$ are the functions $u(t) = Ce^t$. A general element of E_L is of the form

$$\frac{p_1(t)e^t + \cdots + p_m e^{mt}}{q_1(t)e^t + \cdots + q_n(t)e^{nt}}. \tag{2.4}$$

That is, it's "rational functions" in the solutions of L and their derivatives. For intuition, think of this as $\mathbb{C}(t)$ adjoin e^t in the sense of a differential field. ◂

Recall that the *transcendence degree* of a field extension $k \hookrightarrow L$ is the maximal cardinality of an algebraically independent subset of L.

Lemma 2.5. *The extension $k \hookrightarrow E_L$ has finite transcendence degree.*

Proof. If L is an n^{th}-order differential operator, then $\{u, u', \ldots, u^{(n-1)}\}$ contains a transcendence basis for E_L over k. ⊠

We will now construct a canonical subfield of E_L using the Wronskian.

Definition 2.6. Fix a differential operator L, which is *a priori* a higher-order operator, and rewrite it if necessary to a system of first-order operators $W' = A(t)W$, where A and W are matrices which may depend on time. Let $W(t)$ be the particular solution with $W(0) = I$. Then the *Wronskian* of L is $\mathrm{Wr}_L(t) \coloneqq \det(W(t)) \in E_L$.

Proposition 2.7. $\mathrm{Wr}'_L(t) = \mathrm{tr}(A(t))\mathrm{Wr}_L(t)$.

Proof. First assume constant coefficients, i.e. that $A(t) = A \coloneqq A(0)$. The solution to $W' = AW$ is of the form $W(t) = e^{At}$, and $\det(e^{At}) = e^{\mathrm{tr}(At)} = e^{t\,\mathrm{tr}(A)}$.

If $A(t)$ does depend on time, you can "freeze" $A(t)$ at a given time t_0, i.e. run the above argument with constant coefficients $A = A(t_0)$. Thus the theorem is true at time t_0, and of course t_0 is arbitrary. ⊠

One upshot is that the Wronskian can always be expressed in terms of elementary functions (as antiderivatives of rational functions are elementary, and then we exponentiate).

Therefore we may consider the minimal differential subfield of E_L containing $\mathbb{C}(t)$ and Wr_L; call this $K(\mathrm{Wr}_L)$. This will be useful when we think about differential Galois groups (next).

The Wronskian plays the role in differential Galois theory that the discriminant plays in ordinary Galois theory.

3. Differential Galois groups

Now let's define differential Galois groups. The major conclusion of this section are that this is a linear (i.e. affine) algebraic group. As before, k is a differential field containing $\mathbb{C}(t)$ and contained in $\mathcal{M}(U)$ for some $U \subset \mathbb{C}$.

By an automorphism of a differential field we mean a field automorphism which commutes with the derivation.

Definition 3.1. The *(differential) Galois group* of an extension of differential fields $k \hookrightarrow F$ is the group $\mathrm{Gal}(F/k)$ of differential field automorphisms of F which fix k.

Typically F is the splitting field of a differential operator on k. In this case, the elements of the Galois group permute the solutions to $Lu = 0$, so if V_L denotes the vector space of solutions to $Lu = 0$, then $\mathrm{Gal}(E_L/k) \le \mathrm{GL}(V_L)$.

Example 3.2. Consider $L(u) = u' - u$. An element of $\mathrm{Gal}(E_L/\mathbb{C}(t))$ must send $e^t \mapsto e^{Ct}$ for some $C \in \mathbb{C}^\times$ – and the choice of C determines the automorphism (recall that a general element of E_L has the form (2.4)). Thus $\mathrm{Gal}(E_L/\mathbb{C}(t)) \cong \mathbb{C}^\times$. ◂

This is a lot bigger than the groups we encountered in Galois theory!

Theorem 3.3. $\mathrm{Gal}(E_L/k)$ *is in fact an algebraic subgroup of* $\mathrm{GL}(V_L)$*; in particular it has finitely many connected components.*

Here by "an algebraic subgroup" we mean that it's cut out by finitely many algebraic equations. The rest of the theorem follows simply because it's an affine variety.

Proof. Let ℓ be the order of L, and choose $\{f_1, \dots, f_\ell\}$ a basis of V_L. This sits inside E_L, and the set $\{f_i^{(j)} \mid 1 \le i, j \le \ell\}$ contains a transcendence basis for E_L over k.

Introduce ℓ^2 formal variables x_{ij}, $1 \le i, j \le \ell$, and consider the ring homomorphism

$$\begin{aligned} K[X] \coloneqq K[x_{ij} \mid 1 \le i, j \le \ell] &\xrightarrow{\Phi} \mathcal{M}(U_L) \\ x_{ij} &\longmapsto f_i^{(j)}. \end{aligned} \tag{3.4}$$

Hilbert's basis theorem says $K[X]$ is Noetherian, so $\ker(\Phi)$ is finitely generated. Let $P_1, \dots, P_m$ be a generating set. Intuitively, we've started with a bunch of abstract functions and imposed on them the relations that they satisfy as solutions to $Lu = 0$; $\ker(\Phi)$ contains those relations.

Our choice of a basis of V_L identifies $\mathrm{GL}(V_L) \cong \mathrm{GL}_\ell(\mathbb{C})$; explicitly, the $\ell \times \ell$ matrix $A = (a_{ij})$ acts by

$$f_i \longmapsto \sum_j a_{ij} f_j. \tag{3.5}$$

Inside $\mathrm{GL}(V_L)$, $\mathrm{Gal}(E_L/k)$ is precisely the subgroup of elements that send solutions to solutions. Formally, this is the same as specifying

$$P_a\left(\sum_j a_{j1} X_j^0, \dots, \sum_j a_{jk} X_j^{k-1}\right) = 0 \tag{3.6}$$

for $1 \le a \le m$, which is a finite set of polynomials in the variables a_{jk}. ⊠

Remark 3.7. Sometimes the differential Galois group is a finite group. This happens precisely when the solutions to $Lu = 0$ are algebraic functions. ◂

Finally, we'll need the following lemma later.

Lemma 3.8. $\mathrm{Gal}(E_L/L(\mathrm{Wr}_L)) = \mathrm{Gal}(E_L/K) \cap \mathrm{SL}(V_L)$. *In particular, if* $\mathrm{Wr}_L \in \mathbb{C}(t)$, *then* $\mathrm{Gal}(E_L/\mathbb{C}(t)) \subset \mathrm{SL}(V_L)$.

Proof. If $\tau \in \mathrm{GL}(V_L)$, then τ acts on Wr_L by multiplication by $\det \tau$. ⊠

This is analogous to the following fact from Galois theory: the Galois group of a degree-n irreducible polynomial f is manifestly a subgroup of S_n in that it permutes the roots of f. It lies within $A_n \le S_n$ iff the discriminant of f is zero. This fact is often useful in practice for computing Galois groups, and the analogous fact about the Wronskian in differential Galois theory is also true.

4. Liouville's theorem

Definition 4.1. An extension $k \hookrightarrow L$ of differential fields is *Liouvillian* if it factors as a sequence $k = L_0 \hookrightarrow L_1 \hookrightarrow \cdots \hookrightarrow L_m = L$ such that each iterate $L_i \hookrightarrow L_{i+1}$ is one of

(1) a finite extension;[1]
(2) adjoining an antiderivative of some $f \in L_i$, i.e. a splitting field for the operator $D(u) = u' - f$; or
(3) adjoining the exponential of an antiderivative of f, i.e. a splitting field for the operator $D(u) = u' - fu$.

We're thinking of differential extensions of $\mathbb{C}(t)$ as contained inside $\mathcal{M}(U)$ for some U. Sometimes one has to restrict to functions on some $U' \subset U$ because of branch cuts: if f has a pole at some $z \in U$ with nonzero residue, it can only have an antiderivative on simply connected subsets of $U \setminus z$.

Proposition 4.2. *Every elementary function f is contained in some Liouvillian extension of* $\mathbb{C}(t)$ *(which may depend on f).*

Proof. What makes this interesting is composition. Let's induct on the "length" of an elementary function f, i.e. the number of symbols needed to define it. The base cases are rational functions and exponentials, which we have by definition. A general f is of one of the following forms.

(1) A sum, product, difference, or quotient of functions which have smaller length, hence contained in a Liouvillian extension by the inductive assumption.

[1] By this I mean a finite extension of ordinary fields that's also a differential field extension. I don't know if this is standard notation in differential algebra.

(2) $f = e^g$ where g is elementary of smaller length, which we get from the definition of a Liouvillian extension.
(3) $f = \ln(g)$ where g is elementary of smaller length, which we get by adjoining the antiderivative of g'/g.
(4) $f = \sin(g)$ where g is elementary of smaller length, which we get with Euler's identity $2i\sin(t) = e^{it} - e^{-it}$. ☒

Here's the analogue of (one part of) the Abel-Ruffini theorem.

Theorem 4.3. *Let k be a differential field containing $\mathbb{C}(t)$ and contained in $\mathcal{M}(U)$ for some $U \subset \mathbb{C}$, and let L be a differential operator on k. If $k \hookrightarrow E_L$ is contained in a Liouvillian extension, then $G := \mathrm{Gal}(E_L/k)$ has a sequence of subgroups*

$$\{1\} = G_n \trianglelefteq G_{n-1} \trianglelefteq \cdots \trianglelefteq G_0 \trianglelefteq G_{-1} = G, \tag{4.4}$$

such that each G_j/G_{j+1} is either finite, isomorphic to $\mathbb{C}$, or isomorphic to $\mathbb{C}^\times$.

Our proof leans on the fundamental theorem of differential Galois theory, which you will hear about later this week.

Proof. Using the Liouvillian hypothesis, we have a sequence $k = L_0 \hookrightarrow L_1 \hookrightarrow \cdots \hookrightarrow L_n \supset E_L$ of differential field extensions, where each successive extension is either finite, adjoining an antiderivative, or adjoining the exponential of an antiderivative. These correspond to the three cases finite quotient, $\mathbb{C}$ quotient, $\mathbb{C}^\times$ quotient, respectively.

Without loss of generality we can assume $L_n = E_L$; $\mathrm{Gal}(E_L/k)$ is an algebraic subgroup of $\mathrm{Gal}(L_n/k)$,[2] so when we compute the intersections of the groups in (4.4) with $\mathrm{Gal}(E_L/k)$, we'll end up asking about algebraic subgroups of finite groups, of $\mathbb{C}$, or of $\mathbb{C}^\times$, and these are again finite, $\mathbb{C}$, or $\mathbb{C}^\times$.

Now you can probably see where this is going: $G_j := \mathrm{Gal}(L_j/k)$, and $G_j/G_{j+1} \cong \mathrm{Gal}(L_{j+1}/L_j)$. TODO: make sure the details are right. TODO: prove normality separately? Or just punt to fundamental theorem?

Now the three cases.

(1) The differential Galois group of a finite extension is finite; this follows from ordinary Galois theory.
(2) Suppose we're adjoining an antiderivative F of f. All other antiderivatives of F are of the form $F = f + C$, and $\mathbb{C}$ acts by addition on the constant.
(3) Suppose we're adjoining the exponential of an antiderivative F of f, generalizing Example 2.3. Then the other solutions to the same differential equation are e^{CF} for all $C \in \mathbb{C}^\times$, so we add these and nothing more (and then take the smallest differential field containing those and L_j). Then the Galois group is $\mathbb{C}^\times$, acting by multiplication on the constant. ☒

Now let's use this.

Definition 4.5. The *Airy equation* is the second-order differential equation

$$u''(t) - tu(t) = 0. \tag{4.6}$$

Its space of solutions is two-dimensional; the standard basis is $\{\mathrm{Ai}(t), \mathrm{Bi}(t)\}$, where $\mathrm{Ai}(t)$ is the solution with $\mathrm{Ai}(0) = 1/(3^{2/3}\Gamma(2/3))$ and $\mathrm{Ai}'(0) = -1/(3^{1/3}\Gamma(1/3))$ and $\mathrm{Bi}(t)$ is the solution with $\mathrm{Bi}(0) = 1/(3^{1/6}\Gamma(2/3))$ and $\mathrm{Bi}'(0) = 3^{1/6}/\Gamma(2/3)$. $\mathrm{Ai}(x)$ is called the *Airy function (of the first kind)* and $\mathrm{Bi}(x)$ the *Airy function of the second kind*.

Remark 4.7. Here are some non-closed-form expressions for the Airy functions:

$$\begin{aligned} \mathrm{Ai}(x) &= \frac{1}{\pi}\int_0^\infty \cos\left(\frac{t^3}{3} + xt\right) \mathrm{d}t \\ \mathrm{Bi}(x) &= \frac{1}{\pi}\int_0^\infty \left(\exp\left(-\frac{t^3}{3} + xt\right) + \sin\left(\frac{t^3}{3} + xt\right)\right) \mathrm{d}t. \end{aligned} \tag{4.8}$$ ◀

We will prove that these functions (in fact, all nonzero solutions to the Airy equation) are not elementary, first by calculating the differential Galois group of the Airy equation, then showing it doesn't factor as in Theorem 4.3.

[2] We didn't prove this; it's part of the fundamental theorem for differential Galois theory, and proceeds in a similar way to Theorem 3.3.

Proposition 4.9. *If E_L denotes the splitting field of the Airy equation,* $\mathrm{Gal}(E_L/\mathbb{C}(t)) \cong \mathrm{SL}_2(\mathbb{C})$.

Proof. TODO ⊠

Proposition 4.10. *There is no chain of subgroups of* $\mathrm{SL}_2(\mathbb{C})$ *satisfying Theorem 4.3.*

Proof. TODO ⊠

Remark 4.11. This approach does not work for showing that antiderivatives of elementary functions aren't elementary, as adjoining an antiderivative is an example of a Liouvillian extension. We will (probably) be able to discuss the case $f(x)e^{g(x)}$, f and g rational, using a related but different method (TODO: fill this bit in), by reducing it to an algebraic question: $f(x)e^{g(x)}$ has an elementary antiderivative iff there is a rational function h with $f = h' + hg$. This kills, for example, $g(x) = \pm x^2$.

For other functions, such as $(\sin x)/x$ or x^x, the same broad ideas apply but the details are different. ◂