ROBERT LANGLANDS: ABEL LAUREATE 2018

ERIC PETERSON

ABSTRACT. These are companion notes for a slide lecture I gave about the works of Robert Langlands to an audience of experimental physicists. They are, more or less, the words I said aloud.

1. INTRODUCTION

My goal in this talk is to describe to you the life and times of Robert Langlands, Abel Prize laureate for this calendar year. To set the stage, the Abel Prize is the most prestigious award given out in mathematics. There another prize, the Fields Medal, which is more forward in the scientific-public's consciousness, and which is often compared to the Nobel Prize in the sciences, but which has some peculiarties that separate it from the Abel Prize. The Fields Medal is awarded based on the strength of a single, crystallized accomplishment: if someone resolves a conjecture that has stood open for decades and is known to have important consequences for other areas of mathematics, then that's the kind of thing that puts them in the running for the Fields Medal. It also carries a quirk: the Fields Medal is limited to those of age 40 and below.

By constrast, the Abel Prize is awarded based on the merit of a lifetime of work. The sort of person who wins the Abel Prize is the sort of person whose name is still appearing in math textbooks a century later, which is something that may or may not be true of a Fields Medalist. By consequence, if you want to judge someone's lifetime of work, you have to wait until they're old enough to tell, and so there's correspondingly no age limit imposed on the Prize—and, indeed, the average age of the 19 recipients thus far is 72. Additionally, it's difficult to accomplish a lifetime of such exceptionally important work without having also proven some amazing theorems, and so it's quite often been the case that by the time someone wins an Abel Prize, they've already won a Fields Medal.

In Langlands's case, he actually had a sense of what he was going to do with his life in his 30s, and he wrote down his vision in a letter to another famous mid-20th century geometer, Andre Weil. His 17 page research program was prefaced by the following:

> "In response to your invitation to come and talk I wrote the enclosed letter. After I wrote it I realized there was hardly a statement in it of which I was certain. If you are willing to read it as pure speculation I would appreciate that; if not—I am sure you have a waste basket handy."

———Robert Langlands

You should read this in its embarrassed context: mathematicians are trained to speak only hyper-rigorously, only in terms of things they can prove, and then only in terms of their proofs. One isn't supposed to write 17 pages of unfounded, fantastical conjecture and address it to a famous geometer. Nonetheless, this is the material that he (and much of

Date: May 2nd, 2018.

the rest of the mathematical world!) pursued for the rest of his life, and it's the material for which he was ultimately awarded the most prestigious prize in mathematics.

Additionally, before we really get started, I should say a few things about what this talk is and what it isn't. I'd like to start with this quote:

> "Mathematics is the art of giving the same name to different things."
>
> ———Henri Poincaré

Poincaré meant something very precise when he said this. As a mathematician, when you begin your investigations into something, you typically have a very clear idea of what you're setting out to probe: you're going to understand everything there is to know about the Euclidean plane; or you're going to figure out what the deal is with algebraic equations; or whatever. As you start proving theorems, you look over at what your friend is doing, and you notice that a bunch of your theorems look awfully similar: you can perform a kind of word-substitution on them and trade all of your theorems for theirs, and vice versa. Once you've noticed this, then the game is really on: the real object of pursuit is the explanation as to why these seemingly disparate fields are, in fact, the "same".

This is why the following quote is so striking:

> "...The Langlands Program is a Grand Unified Theory of mathematics."
>
> ———Edward Frenkel

Robert Langlands's work has this property in spades. As inspiring as this is, it should also have a note of foreboding: mathematics is a very broad subject, and so a "Grand Unified Theory" will necessarily have to cover a lot of ground. Accordingly, this talk may feel like a bit of a whirlwind. Additionally, this talk covers an area of modern mathematical research, and so the material is quite dense. If you were to stop the talk on a particular slide and pour five years down whatever rabbit hole we're currently circling, you're liable to end up with a PhD. You're thus likely to feel lost at times, but I hope you hang on for the ride, since the views are beautiful all the same. On top of *that*, I have a math PhD, but it's not in any of these fields—and so my understanding of some of the points I'll present is rather superficial. I've done my best to arrange the material in a way that makes sense to me, and I hope that some of this transfers to the audience.

Finally, as a warning to any true mathematicians reading this document, I have certainly fudged the truth where I thought I could get away with it. In particular, I didn't want to mention such conceptually difficult objects as the adèles. Please forgive me, and tread lightly.

2. Geometry and number theory

Let's begin with a discussion of complex geometry, and in particular the theory of Riemann surfaces. This subject has the pleasant properties that I'm able to actually draw some pictures[1] as well as that we have a really firm grip of what's going on in this area of math.

A *Riemann surface* is an object that one can build by gluing together patches of the complex plane (without getting any individual patch stuck on itself, and without edges). The stereotypical example of this is a sphere: if I cut out a disk from the complex plane,

[1] See slides.

I can wrap it up along the surface of the sphere, covering more and more of it. I can't cover everything, though, without overlapping myself—in fact, the best that I can do is to cover everything except the north pole. On the other hand, I can do the same thing by draping a disk over the top of the sphere and tugging it down to cover everything but the south pole. Doing both of these together is what I mean by a Riemann surface: I have two patches of the complex plane, and I've glued them along their overlap at the equator in order to build a sphere out of them.

There are some restrictions to this process—I'm not allowed to just go gluing patches willy-nilly. The main restriction here is that my complex plane pieces have an inherent notion of "multiplication by i", or counterclockwise rotation by 90 degrees. If I pick a point in an overlap (e.g., on the equator of the sphere) and a tangent vector at that overlap, my two complex patches give me two competing notions of what it means to "multiply the tangent vector by i", and I demand that I've glued these things together in a way that the two notions agree.

It turns out that (topologically) we can give a total description of what Riemann surfaces look like: they're all donuts with varying numbers of holes. The sphere is a donut with no holes; a normal donut has one hole; then there are donuts with two, three,—however many holes you like. What *isn't* a Riemann surface is the Möbius band: the whole business of the Möbius band is that it doesn't have a consistent notion of the word "counterclockwise", and accordingly it can't have a consistent notion of "multiplication by i". (Also, it has an edge, which isn't allowed.)

Once you've settled on an object to study in mathematics, you always make the same next move: supposing someone gave you two such objects, how would you tell if they're the same? This question presupposes an important notion: you have to have a definition of what "the same" means. The typical way to approach this is to introduce a notion of "function", and then "the same" means that there exists an invertible function that compares the two objects. In the context of complex geometry, the relevant brand of functions are those which are determined by their Taylor expansions.

Fixing your favorite Riemann surface C, the set of functions from C to the sphere $\mathbb{P}^1$ turns out to be especially important, not least because it has the following interesting alternative description. First, a theorem from complex analysis says that a nonconstant function $f\colon C \to \mathbb{P}^1$ can only strike any particular point on the sphere at most finitely many times—this is true, in particular, of the north pole. So, f mostly takes values in the "bottom region" of the sphere. On the other hand, from our discussion of the sphere as a Riemann surface above, we have a way to re-envision a point on the bottom region of the sphere as a point in the complex plane—and so f can mostly be considered to be a function valued in $\mathbb{C}$. What do we do with those points that map to the north pole? The north pole is where you end up on the sphere if you try to walk "to the edge" of the bottom part of the sphere—which, in terms of the complex plane, means picking a direction and walking towards infinity. This tells us how to interpret the remaining points: they're singularities of f. Mathematicians have vocabulary for everything, and such complex functions f with singularities are referred to as *meromorphic*.

One of the pleasant features of meromorphic functions is that they can be drawn. In the slides, you'll find a picture of the "cubing function", $w = z^3$—but it's not drawn as one normally draws graphs. What we've done is take a "source" complex plane and a "target" complex plane, and we've twisted up the source plane and moved it atop the target plane in such a way that all the points that end up in the same place are all stacked on top of each other. The center of the picture corresponds to the value 0, so if you move a bit away from

0 you'll find 1, so that the three points stacked above 1 correspond to the three complex cube-roots of unity: $e^{2\pi i/3}$, $e^{2\cdot 2\pi i/3}$, and 1. At the center of the picture, you see a sort of pinch point, and this corresponds to the fact that 0 has a unique cube root.

This behavior is actually quite generic for meromorphic functions. Over almost all points, the graph looks like a bunch of sheets collapsing cleanly down. At some stray points, you'll find some premature collapsing or "pinching" behavior; we call these points *ramified*, and we call the general phenomenon *ramification*. Finally, there's a more subtle issue: even if you stay away from the ramified points and in the "good" region of the graph, it's impossible to pick consistent preimages. If you pick your favorite cube root of 1, you can use that choice to wander out and make a continuous choice of a cube root of i, then continue on to infer a choice of cube root of -1, then one for $-i$, and when you return to 1 you discover that you're obligated to pick a different choice of cube root of 1. This corresponds to the kind of "staircase" pattern of the graph, and this phenomenon is referred to as *monodromy*.

The reason that we've belabored this example so long is the following result:

Theorem 1. *The set of meromorphic functions totally determines the Riemann surface.*

What I mean by this is that if you have a secret Riemann surface, and you reveal to me only what meromorphic functions you can construct on it, I can determine what Riemann surface you're thinking of. In order to really make this precise, I have to name some important properties of what collections of meromorphic functions generally look like (lest you try to trick me by giving me an illegal set):

- I can scale them by complex numbers.
- I can do arithmetic with them: I can add, subtract, multiply, or divide them. Such arithmetical contexts are called fields (and, unfortunately, this is not the physicist's use of the word). The last point is especially interesting; normally you're not allowed to do things like divide two functions because you're worried about dividing by zero, but we've already admitted singularities into our lives, and so we're not bothered by potentially picking up more poles.
- There's a more technical condition: they're of "transcendence degree 1". Very imprecisely, what this means is that there is (more or less) one choice of function that is wholly unrelated to all other available functions.

The situation is actually more elaborate than what I've said so far: not only is there a correspondence between Riemann surfaces and these complex fields through meromorphic functions, but if you have a map of Riemann surfaces then this produces a corresponding map of complex fields. For example, the cubing function gets converted into the inclusion into all rational functions of just those rational functions whose exponents are divisible by 3.

Where else in math do fields show up? Well, if they're all about adding, subtracting, multiplying, and dividing, there's a whole business of people interested in the same operations: the number theorists. A *number field* is what you get if you start with the rationals, select a bunch of your favorite polynomials, and throw in some of the roots of those polynomials into the mix as well. For instance, the polynomial (x^2-2) has a root $\sqrt{2}$, and so I can form the number field $K = \mathbb{Q}(\sqrt{2})$ and start thinking about doing arithmetic in it.

An interesting feature of this field is that if I grant myself access to $\sqrt{2}$, I automatically also get access to the other root, $-\sqrt{2}$, since I'm allowed to multiply things in my field: $-\sqrt{2} = -1 \cdot \sqrt{2}$. These two roots have an interesting symmetry property: if I write down a true sentence involving rational numbers and these two roots of 2, then I can form another sentence by replacing all the instances of $\sqrt{2}$ with $-\sqrt{2}$ and vice versa, and this new sentence also turns out to always be true. The total of these kinds of symmetry properties, where I'm allowed to swap roots around, is tracked in the *small Galois group* of the number field, $\mathrm{gal}(K)$. You can compute this in some other small examples. For example, once you get tired of $\sqrt{2}$, you might move on to $\sqrt[3]{2}$ and $x^3 - 2$. Here it's possible to add one root without adding the others, and that number field has no small Galois group: there are no interesting symmetries of a single thing. If you go ahead and add all of the roots in, you end up with a small Galois group of Σ_3, the full permutation group.

In the previous section of complex geometry, not only did fields show up, but so did inclusions of fields. Taking inspiration from this, we also note that there's a relative version of this construction: given a number field (i.e., the rationals with some extra roots), we can construct a yet bigger number field that contains it by throwing in more roots of more polynomials, and we can study those symmetries of the big root system that leave invariant the small root system that we started with. These symmetries consistute the *relative Galois group*. You can also take the auxiliary larger field out of the picture: by throwing in more and more roots, you get bigger and bigger relative Galois groups, and you can sort of "send the big field to infinity" by taking a union, and the corresponding union of all the relative Galois groups is called the *big Galois group*, $\mathrm{Gal}(K)$. As its name implies (and as indicated by the construction), the big Galois group is often very big.

One of the first theorems proven in an algebraic number theory course is:

Theorem 2. *Given a surjection* $\mathrm{Gal}(K) \to G$*, there exists a number field* $L \supset K$ *whose relative Galois group is* G.

This is rather opaquely stated, but I want you to take it as evidence for the following maxim: if the big Galois group of K is very large and complicated, this means that it has a high probability of admitting lots of interesting surjections. Using the Theorem, this means that there exists a lot of number fields L that lie atop K. By consequence, if L is constructed by adjoining roots to K, that must mean that K itself is missing a lot of roots. If K is missing a lot of roots, that means it will be difficult to solve polynomial equations over K. Altogether, then, the big Galois group can be taken as a measure of the relative difficulty of getting things done internally to K.

It turns out that you can define a relative Galois group of a complex field as well, and there the relative Galois group corresponds to the monodromy of the equivalent map of Riemann surfaces. You can continue to port over other ideas from complex geometry, including that of ramification. The main point of ramification is that the polynomial we were studying, z^3, had a repeated root at $w = 0$, whereas other values of w didn't have repeated roots. A critical feature of ramification is that it is a phenomenon that happens "at a point", which we encoded into the equations by setting w to a particular value. The analogue of these observations on the number-theoretic side is not so obvious without a more careful algebraic statement, but it turns out that the thing to do is to pass from polynomials defined over the integers to polynomials modulo a particular prime.[2] When you do this, your polynomial might accidentally pick up a repeated root where it

[2] Rather than setting $w = 0$, we are setting $p = 0$.

previously had none. For example, working with the polynomial x^3-2 and modulo 2, this polynomial becomes merely x^3, which has a repeated root at zero. This has significant consequences for how equations involving cube roots of 2 behave modulo 2.

We've started to see how one can give the same name to two very different sets of ideas, in a way that bears conceptual fruit: we can move concepts from one arena to the other as across a Rosetta stone connecting complex geometry and algebraic number theory, often to great effect. What's less obvious except through experiment is that not only do the *concepts* transfer, but actual *computations* transfer. Sometimes the monodromy group of a particular Riemann surface behaves very mysteriously like the Galois group of a particular number field. This is actually not so mysterious: these two families of objects have a common origin. Given a polynomial with integer coefficients (something from the realm of *Diophantine equations*), one can extract two things from it: by setting it equal to w and calculating the set of complex solutions, one produces a Riemann surface; and by adjoining roots of this polynomial to $\mathbb{Q}$, one produces a number field. Because the *individual objects* in these two realms can be connected by a single *background object* pulling the strings, it is not just the two ideas of (e.g.) ramification that line up, but its uses in the two fields exert a kind of gravitational pull on each other. This is an exceedingly fruitful technique: if you are confused about the behavior of some number field, you can trade it for the corresponding Riemann surface, which you can study by drawing pictures. Once you're satisfied with your study of that particular surface, you can return to the number field, where you find that "things work much the same way."

3. Analysis and number theory

Now, for something completely different: there are other avenues of attack on the questions of number theory, and a particularly fruitful one comes in the form of *analytic number theory*. One of the basic objects in number theory enjoys popular fame: the ζ–function,

$$\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}=\frac{1}{1^s}+\frac{1}{2^s}+\frac{1}{3^2}+\frac{1}{4^s}+\frac{1}{5^s}+\frac{1}{6^s}+\cdots.$$

Calculus students actually learn some basic facts about this function:

- Setting $s=1$ recovers the harmonic series, which is known to diverge.
- Setting $s>1$ is approachable by the "p–test", which shows the series to converge.

A property that calculus students don't learn about is the product formula

$$\zeta(s)=\prod_{p\text{ prime}}\frac{1}{1-p^{-s}}=\prod_{p}\left(1+\frac{1}{p^s}+\frac{1}{(p^2)^s}+\frac{1}{(p^3)^s}+\cdots\right).$$

Moving from the sum formula to the product formula is a little confusing; it is easier to see what's happening the other way around, if you take the product formula as given and start expanding its terms out. How would you get the terms in the sum formula? Well, for instance, the sixth term appears as

$$\frac{1}{6^s}=\frac{1}{2^s}\cdot\frac{1}{3^s}\cdot 1\cdot 1\cdot 1\cdots\cdot.$$

In general, the appearance of the nth term is governed by its prime factorization—indeed, the product formula is a kind of analytic witness to the arithmetic *idea* of prime factorization.

One can use these formulas to prove goofy results, like the following:

Theorem 3. *There are infinitely many prime numbers.*

This is something you all knew before you walked in the room, and it's something that Euclid knew back in the B.C.E. era. Still, you can use the above methods to produce a new proof. The main idea is to apply a logarithm to the product formula; logs are supposed to play well with products, so this is not a totally insane thing to do.

$$\log \zeta(s) = \log\left(\prod_{p \text{ prime}} \frac{1}{1-p^{-s}}\right) = \sum_p \log\left(\frac{1}{1-p^{-s}}\right) = \sum_p \left(\frac{1}{p^s}\right) + O(1)$$

After moving the product out past the log as a sum, we have also made use of the the Taylor expansion of the logarithm: the leading term of the Taylor expansion is p^{-s}, and the remaining terms (owing, essentially, to the square in their exponent) all gather together to form a function bounded in s. After this, we take the limit as s approaches 1:

$$\infty \xleftarrow{s\to 1^+} \log \zeta(s) = \sum_p \left(\frac{1}{p^s}\right) + O(1) \xrightarrow{s\to 1^+} \sum_p \frac{1}{p} + \text{const.}$$

On the left, we know that $\zeta(1)$ diverges, and the logarithm of infinity is infinity. On the right, replacing s with 1 gives us the exact sum of the prime reciprocals, plus whatever junk constant value comes out of the bounded function. The main point is that the sum must have infinitely many terms in order to get the infinity on the left: if the sum had only finitely many terms, it would give a finite value. If nothing else, this is quaint.

Let's try for something more serious. If we're going to approach number theory via analysis, then we should make use of real tools from analysis, and one of the most effective such tools is harmonic analysis—Fourier theory, that is, by any other name. The basic unit of the general theory of harmonic analysis is the *character*: a character on a group G (where a group is a context in which you can multiply) is a multiplicative complex function $\chi\colon G \to \mathbb{C}$, i.e., it satisfies

$$\chi(g \cdot g') = \chi(g) \cdot \chi(g').$$

The main theorem in this context is:

Theorem 4. *A nice enough complex function on a nice enough* commutative *group can be written as a sum of characters. (That is, the set of characters forms a basis for the space of functions.) Using the inner product on the space of functions, this gives the formulas*

$$f = \sum_\chi \langle \chi, f\rangle \cdot \chi, \qquad \langle \chi, f\rangle = \int_G \chi(g^{-1}) f(g)\, \mathrm{d}\mu.$$

Let's expand this in the most familiar example we can think of: let's set G to be the real line with addition. A character, then, is a complex function on the real line that carries addition to multiplication—and the complex exponential $\chi_a(x) = e^{\pi i \cdot a \cdot x}$ is an example of such a function. In fact, these are all such functions, and substituting this into the integral formula shows that the expression

$$\mathscr{F}\{f\}(a) = \langle \chi_a, f\rangle = \int_{\mathbb{R}} e^{-\pi i a x} f(x)\, \mathrm{d}x$$

considered as a function of a recovers the Fourier transform of f.

The point of setting this up in generality is not to restrict ourselves to this particular example. The *next* most familiar thing you might do is think about the (positive) real numbers with multiplication, which also form a commutative group. Characters in this setting look like $\chi_a(x) = x^a$, owing to the formal property

$$\chi_a(xy) = (xy)^a = x^a y^a = \chi_a(x)\chi_a(y).$$

Writing out the multiplicative analogue of the Fourier transform in this setting gives rise to what number theorists call the *Mellin transform*, $\mathscr{M}$.

Once made aware of the Mellin transform, you might set about trying to compute some examples for it. Here's one:

$$\mathscr{M}\{e^{-\pi n^2 z}\}(s) = \pi^{-s}\Gamma(s)n^{-2s}.$$

Something you can notice on the right is that there's just one thing that depends on n: the factor n^{-2s}, which looks rather like a single summand of the ζ–function. Whatever the mysterious function $\Gamma(s)$ is, it at least has nothing to do with n. With this in mind, you can try to get all the terms in the ζ–function just by summing together the right exponentials:[3]

$$\mathscr{M}\left\{\sum_{n=1}^{\infty} e^{-\pi n^2 z}\right\}(s/2) = \pi^{-s/2}\Gamma(s/2)\zeta(s) =: Z(s).$$

This is a pretty interesting equation: number theory is all about the mix of additive and multiplicative properties; just a moment ago, we were already talking about exponentials in the context of *additive* characters; then we applied the *mulitplicative* transform; and, lo' and behold, some important number-theoretic object popped out.

This Fourier-theoretic perspective leads you to some other important identities. For instance, this exponential has the property that it is self-dual under the additive Fourier transform, and this combines with the Poisson summation formula

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \mathscr{F}\{f\}(n)$$

to give the reflection equation

$$Z(s) = Z(1-s).$$

One sense in which this is interesting is that it extends the range of definition of the ζ–function; we previously claimed only that it made sense for s with large positive real part, but this equation also tells us that if s has large negative real part, we're fine. Another thing to come out of this set-up is a variant of the Riemann hypothesis, which states[4]

$$\mathscr{M}\{\text{Dirac comb at zeroes of Z}\} \approx \text{Dirac comb at } \log p.$$

Finally, a result known popularly as "Tate's thesis" states that you can commingle these ideas with those from algebraic number theory. He constructs the product formula

$$\zeta(s) = \prod_p \frac{1}{1-p^{-s}}$$

one piece at a time: he develops a kind of Fourier transform which operates "at a prime", takes the multiplicative transform of an additive character, and what comes out is the

[3] There seems to be a perspective in the literature that the sum should be considered as a θ–function, but trying to understand this gives me vertigo.

[4] The "$\approx$" should be taken seriously; there is a significant and obscuring error term which I have hidden.

function $(1-p^{-s})^{-1}$. Additionally, he explains the extra factors in the function Z as naturally occurring, and gives a compelling argument that Z is the truly natural function over ζ. Finally, he does this in a way that makes it possible to do this at all primes simultaneously, and the product—i.e., function Z itself—pops out.

This isn't the only ζ–function around; rather, number theorists have discovered that this is but one member of a large class of them, though exactly what counts as a ζ–function and what doesn't has been uncovered through trial and error rather than through a solid general definition.

The first generalization made is through *Dirichlet characters*, which are "characters"

$$\chi : (\mathbb{N}, \times) \to \{\text{complex roots of unity}\}.$$

These aren't quite characters in the sense meant previously, since $\mathbb{N}$ isn't a group, but we still intend the same thing: χ is required to be a multiplicative function. Given one of these objects, the associated Dirichlet ζ–function is given by the formula

$$\zeta_\chi(s) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s},$$

where we've just thrown a rescaling by χ into the numerator. The first sign that this is a sane thing to do comes from a corresponding product formula

$$\zeta_\chi(s) = \prod_p \frac{1}{1-\chi(p)p^{-s}},$$

which is proved in basically the same way: when multiplying the terms out, you can lean on the multiplicativity of χ to decompose its argument in the same way you're decomposing the denominator of the relevant fraction. The first nontrivial theorem proven about these functions is:

Theorem 5. *For χ nontrivial, $\zeta_\chi(1)$ converges.*

In other words, the χ–analogue of the harmonic series, into which χ has stuck a bunch of signs, is no longer divergent. This seemingly innocuous fact has a number of striking consequences, including the following:

Theorem 6. *For any fixed coprime d and r, there are infinitely many primes of the form $p = qd + r$.*

You should read this statement as a kind of even distribution of the primes through the "space" of the integers. The right-hand side of this equation looks like the slope-intercept form for a line, and so the claim is that any "line" you draw through the integers collides with infinitely many primes.

The next step in generalization is to think of the target of χ, originally the complex numbers, as 1×1 matrices and to replace those with $n \times n$ matrices instead. Such a multiplicative function is called a *representation* (of the domain group G), and we denote such a function by ρ. In the specific case that the representation has $G = \mathrm{Gal}(K)$ as its domain for some number field K, there is a recipe which associates to this data a ζ–function, called its *Artin ζ–function*.

This is less arbitrary than it sounds: it turns out that number theory is positively littered with sources of Galois representations,[5] and this ζ–function construction is super well-behaved and indeed appears to capture interesting arithmetical data about the object from which the representation was extracted—even though it's two steps removed! The construction also has an odd selectivity to it: if you write down a *random* Galois representation and extract its Artin ζ–function, the function you get out appears to have essentially random behavior and no discernable nice property. In some sense, the construction is conscious of when (and how) the representation "came from geometry". This is obviously intriguing but also simultaneously discouraging: mathematicians are discovering each day a new sense in which something ought to be considered "geometric", which means we have a very poor grasp of when the Artin ζ–function construction behaves well and when it doesn't. It is, correspondingly, very difficult to prove theorems about a class of functions with no well-defined boundaries. One of the major open questions is whether $\zeta_\rho(1)$ converges—indeed, this was one of the questions that prompted Langlands to begin his work.

Rather than describe the general recipe for Artin ζ–functions in detail, I would like to illustrate them in an example: that of the elliptic curve

$$C = \{y^2 + y = x^3 - x^2\}.$$

Elliptic curves have risen to promenance in the scientifically literate public's consciousness through their applications in cryptography, but they also arise quite naturally in our story: back when we were discussing Riemann surfaces, we completely covered the case of a sphere (or a "donut with no holes"), and if we felt bold we might move on to the next case: a donut with one hole. Elliptic curves are exactly this next case, and they're given in terms of equations that are quadratic in y and cubic in x.[6]

Given an arbitrary such elliptic curve C, it's possible to extract a 2–dimensional Galois representation from it, and from that extract an Artin ζ–function:

$$\zeta_C = \sum_{n=1}^{\infty} \frac{a_n}{n^{-s}}.$$

The coefficients of this function have the following intriguing property:

Theorem 7. *The number of solutions to the equation defining C over $\overline{\mathbb{F}}_p$ is given by $p - a_p$.*[7] *(That is, a_p tracks a kind of "defect" in the solution count.)*

This should give us pause: the whole business of elliptic curve cryptography is founded on the idea that, for large primes, it is difficult to write out all the points associated to an elliptic curve (and difficult to exhaustively write out their multiplication tables). This gives strong indication that it will be difficult to figure out the coefficients in the ζ–function; certainly we can't approach this by brute force.

[5] Essentially any time you take any variant of "cohomology" of any arithmetic object defined over K, you get such a representation.

[6] In the slides, you can see a graph of the *real* solutions to this equation. This corresponds to taking a vertical cross-section of the donut, where the cut is placed just off-center enough so that you don't see the hole in the donut, but you still see the divot it makes in the altitude. The "point at infinity" is located at the far right of the cross-section, which is why that half of the graph of the real points blooms open.

[7] Again, for mathematicians in the audience: this is an application of the arithmetician's version of the Lefschetz trace formula.

For elliptic curves arising through a particular construction (for instance, our choice of example curve C), there is a related series $f_C(q)$, called a *modular form*, with a number of very remarkable properties:

- The coefficients in the series expansion of $f_C(q)$ are precisely those of $\zeta_C(s)$:
$$f_C(q) = \sum_{n=1}^{\infty} a_n q^n, \qquad \zeta_C(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^{-s}}.$$
Thinking of q as the exponential $e^{2\pi i z}$, one can interpret this as saying roughly $\mathscr{M}\{f_C\}(s) = \zeta_C(s)$, or that f_C is a kind of "inverse Mellin-transform" to ζ_C.
- The series $f_C(q)$ converges on the complex unit disk to determine a complex function there. This function has incredible symmetry properties: given an Möbius transform μ fixing the unit disk (i.e., when studying the action of $SL_2(\mathbb{Z})$), the relation between $f_C(q)$ and $f_C(\mu \cdot q)$ is given by a completely predictable rescaling. This property is called *modularity*. There is no reason to expect this behavior to come out of some random inverse Mellin-transformed function.
- $f_C(q)$ comes with a polite product expansion. In the specific case of our example C above, this product expansion is given by
$$f_C(q) = q \prod_{j=1}^{\infty} (1-q^j)^2 (1-q^{11j})^2.$$
In particular, this means that the coefficients a_n—moments ago thought to be quite difficult to compute—can be computed to arbitrary order, even by hand.

At this point, you might raise an objection to getting too excited about this auxiliary function f_C: I introduced it by claiming that we only have access to such a miraculous function for certain nice C, and so perhaps I'm pulling the wool over your eyes by working through a particularly nice example. The following theorem (which won the Abel Prize in 2016) rewards the optimistic viewpoint instead:

Theorem 8. *For every elliptic curve C, there is an associated modular form f_C with the above properties. Additionally, if C and C' have the same modular form, they're equivalent as elliptic curves.*

In other words, the above situation is generic over elliptic curves.

4. The Langlands program

We now move into Langlands's domain. His essential claim is that the above situation is not just generic for elliptic curves, but for all geometric Galois representations. His main observation is that, in the general case, we are missing an analogue of the Mellin transform: $n \times n$ matrices for $n \geq 2$ no longer form a commutative group, and in general a big Galois group $\mathrm{Gal}(K)$ isn't commutative either,[8] which stifles the discussion of harmonic analysis from earlier. Instead, Langlands proposes an analogue of the Mellin transform that happens directly at the level of representations:

[8] For the mathematicians in the audience: the noncommutative big Galois group has been with us all along, but when we were studying Dirichlet characters valued in mere complex numbers, our representation factored through the abelianization of $\mathrm{Gal}(K)$, and hence we could employ classical Fourier theory.

$$\begin{tikzcd}
\left\{ d\text{-dim'l Galois rep's} \right\} \arrow[r, leftrightarrow, "\text{noncomm. } \mathscr{M}"] \arrow[d, "\text{Artin}"] & \left\{ \text{automorphic rep's of } d \times d \text{ matrices} \right\} \arrow[d] \\
\{\zeta\text{-functions}\} \arrow[r, equal] & \{\zeta\text{-functions}\}
\end{tikzcd}$$

The idea is to replace the function f_C with a kind of Mellin transformed representation and to try to find the analogues of the "modularity" properties directly on it. The exact definition of one of these resulting "automorphic representations" is extremely technical, and so we will skip it—suffice it to say that that's the idea.

He gave a recipe for how this correspondence should go, including a recipe for how to extract a ζ-function from an automorphic representation, but couldn't immediately prove all of the nice properties he expected it to have—in particular, he couldn't show that it has the same "invertibility" property that the classical Mellin transform enjoys. However, he could explore it in a number of special cases (including that of 2-dimensional representations coming from elliptic curves) and see that it unified a great deal of classical discussion about Artin ζ-functions.

While in the neighborhood, he he did notice a couple of other very important properties of his proposed correspondence that weren't classically visible. It is often the case that a Galois representation doesn't take values in arbitrary $d \times d$ complex matrices, but rather that the matrices enjoy some extra structure—for instance, they might all be unitary matrices. Langlands noticed that, when this is the case, the automorphic representation *also* carries some extra structure—but this extra structure is *different*.

$$\begin{tikzcd}
\left\{ \begin{array}{c} d\text{-dim'l Galois rep's} \\ \text{with extra structure} \end{array} \right\} \arrow[r, leftrightarrow, "\text{noncomm. } \mathscr{M}"] \arrow[d, "\text{Artin}"] & \left\{ \begin{array}{c} \text{automorphic rep's of } d \times d \text{ matrices} \\ \text{with \textit{different} extra structure} \end{array} \right\} \arrow[d] \\
\{\zeta\text{-functions}\} \arrow[r, equal] & \{\zeta\text{-functions}\}
\end{tikzcd}$$

Here's a table of some of the exchanges in structure:

left structure, G	SL_n	$SO(2n+1)$	$Spin(2n)$	$SO(2n)$	$SU(n)$	E_8	$\cdots$
right structure, $G^\vee$	PGL_n	$Sp(2n)$	$SO(2n)/Z$	$SO(2n)$	$SU(n)/Z$	E_8	$\cdots$

For example, if your Galois representation consists of rotation matrices, then the automorphic representation will involve symplectic matrices.[9]

This is quite inspiring, and teasing out all the details of this conjectured correspondence illuminates a lot of yet-unexplored number theory. It's also breath-taking in how broad-reaching it is: it is a simultaneous application of harmonic analysis, analytic number theory, algebraic number theory, Lie theory, and representation theory. There is a neglected aspect of this picture, though, which harkens back to the beginning of this talk: we began by discussing a "Rosetta stone" that translated between algebraic number theory and complex geometry. What happens if we try to push the whole of Langlands's ideas across that bridge and into complex geometry?

[9] In general, this correspondence is informed by behavior at the level of Lie algebras: a Lie algebra $\mathfrak{g}$ and its root system is graded for its linear dual $\mathfrak{g}^\vee$ and its coroot system, which again determines a Lie algebra and hence a Lie group.

In fact, this is possible, and the resulting body of conjectures is known as the "geometric Langlands program". The translation is not especially easy to follow, but we can gesture at parts of it:

- The d–dimensional Galois representations from number theory are replaced by d–dimensional vector bundles over the Riemann surface with a connection ∇.
- The automorphic representation is replaced by "a $\mathscr{D}$–module over the moduli of d–dimensional vector bundles over the Riemann surface". Just as automorphic representations were impossible for us to pull apart into non-technical terms, this is also inaccessible to us.

Just as in the number theoretic setting, you can introduce extra structure: the vector bundle might carry an action of a Lie group G, in which case the data on the other side of the correspondence also acquires some interaction with $G^\vee$, the same Langlands dual group as before.

To an audience of physicists, this is pretty interesting: a vector bundle with a connection and an action of a Lie group are precisely the ingredients of a gauge theory. To set the stage for exploring this connection, there is a classical duality in physics between electric and magnetic fields: if (E,B) is a pair of electric and magnetic fields satisfying Maxwell's equations, then their transposition $(B,-E)$ also satisfies Maxwell's equations. One interesting feature of this exchange is that an electric monopole in E with electric charge e becomes a magnetic monopole with magnetic charge $1/e$.

There is also a quantum version of this statement, first observed by Montonen and Olive and connected with the geometric Langlands program by Kapustin and Witten: there is a sort of "toy model" of quantum electrodynamics: the $\mathscr{N} = 4$ supersymmetric Yang–Mills field theory, with a noncommutative gauge group G and complexified coupling constant τ. They noticed that solutions to this field theory have a correspondence with solutions for $\mathscr{N} = 4$ Yang–Mills with gauge group $G^\vee$ and coupling constant $-1/(n_G \tau)$. This last statement is quite striking: a solution with strong coupling on one side is governed by a solution with weak coupling on the other side. This is a computational physicist's dream: solutions under strong coupling are very difficult to come by, whereas solutions under weak coupling we have loads of methods to handle, all under the broad heading of "perturbative methods". That strong solutions for one system can be extracted from weak solutions to another system is a really big deal. This is supposed to be one instance of many of such exchanges, under the banner of "S–duality"—and so the geometric Langlands program portends to have significant impact on our understanding of physics.

Finally, there is one last intriguing exchange of ideas: rather than importing mathematical ideas into physics, we can also try to export them. In number theory, we became very concerned with Galois representations, and in the context of gauge theories we are also very concerned with G–representations: irreducible such representations classify the fundamental particles of the system. In some sense, then, what we are doing in mathematics is trying to understand the "fundamental particles" of number theory. This, too, is an intriguing analogy to pursue.