INTRO TO THE RIEMANN-HILBERT CORRESPONDENCE

CALEB JI

1. Euler and Gauss: Origins

1.1. **The hypergeometric series.** In pre-20th century mathematics, it was of great interest to compute functions that satisfy various interesting algebraic and differential equations. In this context, the hypergeometric series naturally arose as a class of functions that specializes to several other common ones. While related series had been studied by others, Euler was the first to introduce the hypergeometric series as it is defined today. To define it we will use the following notation, known as the Pochhammer symbol.

$$(q)_n := \begin{cases} 1 & n = 0 \\ q(q+1)\cdots(q+n-1) & n > 0. \end{cases}$$

Definition 1.1. *Given real numbers a, b, c with c not a non-positive integer, the hypergeometric series is given by the power series*

$${}_2F_1(a,b;c;z) := \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!} = 1 + \frac{ab}{c}\frac{z}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{z^2}{2!} + \cdots .$$

Exercise 1.2. *Show that the hypergeometric series converges for $|z| < 1$.*

Taking $b = c$, we obtain

$${}_2F_1(a,b;b;z) = \sum_{n=0}^{\infty} (a)_n \frac{z^n}{n!},$$

which is equivalent to Newton's generalized binomial expansion for $(1-z)^{-a}$. Note that if $a = 1$, this gives the geometric series

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + \cdots ,$$

which is one reason for the name 'hypergeometric.' On the other hand, setting $a = b = 1, c = 2$ yields ${}_2F_1(1,1;2;z) = \ln(1+z)z^{-1}$ and setting ${}_2F_1(1/2,1/2;3/2;z) = \arcsin(z)z^{-1}$, showing that many other types of functions are obtainable. In fact, when Euler introduced these series, he used them to help evaluate certain integrals arising from the computation of certain planetary orbits. For more on the early history of the hypergeometric series, see [1].

1.2. **The hypergeometric differential equation.** Euler showed that the hypergeometric series ${}_2F_1(a,b;c;z)$ give a solution to what it now known as the hypergeometric differential equation:

$$z(1-z)w'' + (c - (a+b+1)z)w' - abw = 0. \tag{1}$$

This is a second order linear differential equation, and we can compute its singularities. It has regular singularities at $0, 1$, and ∞.

Date: June 4, 2024.

Remark. We recall what we mean by the singularity at ∞. In mathematics it is almost always better to work over $\mathbb{C}$ instead of over $\mathbb{R}$. (This is why we use the variable z rather than x.) But sometiems even $\mathbb{C}$ doesn't have all we want; it is not compact as we can go off to infinity and never reach it. If we add in the point at infinity, with a simple rigorous procedure known as the one-point compactification, we obtain the Riemann sphere $\mathbb{P}^1_{\mathbb{C}}$. A concrete way of thinking about this is to take your coordinate z and invert it to get $v = z^{-1}$, now ∞ is the new 0-point in the new coordinate system. We also have to replace w' with $-v^2\frac{dw}{dv}$ and w'' with $v^4\frac{d^2w}{dv^2} + 2v^3\frac{dw}{dv}$. Doing this to the hypergeometric differential equation shows that $v = 0$, which corresponds to $z = \infty$, is a regular singular point.

Let us now check that the hypergeometric series does indeed satisfy Equation 1. We will look for a power series solution around 0, so using the Frobenius method we set

$$w = \sum_{n=0}^{\infty} a_n z^{n+r}.$$

This gives

$$w' = \sum_{n=0}^{\infty} a_n(n+r)z^{n+r-1}, \quad w'' = \sum_{n=0}^{\infty} a_n(n+r)(n+r-1)z^{n+r-2}.$$

Plugging these in gives

$$z(1-z)\sum_{n=0}^{\infty} a_n(n+r)(n+r-1)z^{n+r-2} + (c-(a+b+1)z)\sum_{n=0}^{\infty} a_n(n+r)z^{n+r-1} - ab\sum_{n=0}^{\infty} a_n z^{n+r} = 0.$$

The lowest degree term, z^{r-1}, gives the indicial equation

$$a_0 r(r-1) + ca_0 r = 0 \Rightarrow r = 0 \text{ or } 1-c.$$

Looking at the coefficient of z^{n+r}, we obtain

$$a_{n+1}(n+r+1)(n+r) - a_n(n+r)(n+r-1) + ca_{n+1}(n+r+1) - (a+b+1)a_n(n+r) - aba_n = 0.$$

This yields

$$a_{n+1} = \frac{(n+r)(n+r-1) + (a+b+1)(n+r) + ab}{(n+r+1)(n+r) + c(n+r+1)}a_n = \frac{(n+r+a)(n+r+b)}{(n+r+1)(n+r+c)}a_n.$$

This implies that $a_n = \frac{(r+a)_n(r+b)_n}{(r+1)_n(r+c)_n}a_0$. In particular, for the solution to the indicial equation given by $r = 0$, this gives $a_n = \frac{(a)_n(b)_n}{n!(c)_n}a_0$. Scaling by setting $a_0 = 1$ shows that the hypergeometric series

$${}_2F_1(a,b;c;z) := \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n}\frac{z^n}{n!}$$

is indeed a solution to the hypergeometric differential equation.

What about the other solutions? If c is not an integer, then one can do a similar procedure with $r = 1-c$ and obtain a linearly independent solution; otherwise there is a log term. There are two other singularities, at 1 and ∞, and they can be solved for in a similar fashion. We can summarize the solutions as follows.

Proposition 1.3. *Assume $c, a-b, c-a-b$ are not integers. Around 0, we have local solutions*

$$ {}_2F_1(a,b;c;z), \quad z^{1-c}{}_2F_1(1+a-c,1+b-c,2-c;z), $$

around 1, we have local solutions

$$ {}_2F_1(a,b;1+a+b-c;1-z), \quad (1-z)^{c-a-b}{}_2F_1(c-a,c-b,1+c-a-b;1-z), $$

and around ∞, we have local solutions

$$ z^{-a}{}_2F_1(a,1+a-c;1+a-b;1/z), \quad z^{-b}{}_2F_1(b,1+b-c;1+b-a;1/z). $$

One can ask how these are related. This was somewhat addressed by Gauss, who performed a detailed analytic study of hypergeometric functions and obtained many interesting formulas regarding them, showing their ubiquity in the mathematics of his time. It turns out there is a natural symmetry group acting on the solutions of order 24, each of the 6 solutions can be represented in 4 ways. The keywords for this are *Euler's hypergeometric transforms* and *Kummer's 24 solutions*. Indeed, building off of work of Gauss, Kummer showed that these solutions satisfy remarkable transformation properties, which essentially amounted to a complete description of all solutions to the hypergeometric equation in 24 forms and relations between them. As a simple example of these transformations, known to Euler, we have

$$ {}_2F_1(a,b;c;z) = (1-z)^{-a}{}_2F_1(a,c-b;c;z/(z-1)). $$

It might appear that this marks the end of the story of what the hypergeometric equation has to teach us. But this is far from true: the genius of Riemann opened up an entirely new perspective only a few years later, which has led to incredible developments in mathematics.

2. Riemann and Hilbert: The advent of monodromy

2.1. **Local systems, monodromy, and the fundamental group.** The basic idea behind monodromy is quite simple: if you have a potentially multi-valued function and evaluate it along a loop, once you come back to your starting point you might end up with a different value. One can see this by trying to follow the square root function around the unit circle. By considering various functions such as this, Riemann was led to develop the theory of Riemann surfaces. But these functions can also be considered as solutions to differential equations, which is what we are interested in here. If we take a solution to a differential equation and follow it around a loop, it might become a different solution when we get back to the starting point. But when, why, and how do they change? These are interesting questions which we can give fairly satisfactory answers to.

Remark. It is important that the base we work over is defined over the complex numbers at this point (rather than just the reals) if we want to do anything interesting. For example, there is not much point to going around a loop in a subset of the reals.

To understand the big picture, it will be helpful to generalize what we consider to *local systems* of vector spaces. The set of solutions to a homogeneous linear differential equation form a vector space because we can add them and multiply them by constants. But if we fix any single point, we can say more about the solutions which are valid in the neighborhood of a point: the existence and uniqueness theorem essentially tells us that the dimension of this vector space is given by the order of the differential equation. However, such a solution is not guaranteed to exist on the whole space. Thus we have a collection of vector spaces at each point, where the vector spaces at points close enough to each other 'fit together.' This is precisely the notion of a local system of vector spaces.

To be more precise, a local system is a locally constant sheaf. Rather than give the precise definition of these terms here, for those who have not encountered them yet it is helpful to think of two examples: the trivial local system and the Möbius strip with infinite length. There are two equivalent ways to think of a local system. First, a family of varying vector spaces over a base space X. In this case each point x gives us a vector space $\mathcal{F}_x$. Second, we can think of the total space $E \to X$ of all points in those vector spaces with a projection map down to X, along with some way of connecting the fibers (we will discuss this more later). The total space of the trivial local system (of finite-dimensional real vector spaces) over any space X is simply $X \times \mathbb{R}^n \to X$. The Möbius strip can be represented by an infinitely long sheet of paper that you fold into a cylinder but twist it. The base space is a circle, but if you start at any point on the circle at the point with height h on the fiber, as you trace that same height h point around the circle you end up at height $-h$ when you come around.

This illustrates a notion which we might term *analytic continuation* of local systems along paths. Given a local system $\mathcal{F}$ on X and a path $[0,1] \to X$, we want to explain how to give a map from $\mathcal{F}_0$ to $\mathcal{F}_1$. If $\mathcal{F}$ is trivial, i.e. of the form $X \times \mathbb{R}^n$, then this is easy: just let $(f(0), v)$ go to $(f(1), v)$ for $v \in \mathbb{R}^n$. As it is, though, $\mathcal{F}$ is only locally trivial. But since $[0,1]$ is compact, that means we can pick finitely many open subsets connecting $f(0)$ to $f(1)$ over which $\mathcal{F}$ is trivial, so we can just make this sort of identification finitely many times to get from $f(0)$ to $f(1)$.

Anyways, this gives an answer to our original question of *why* monodromy exists: local systems do not need to be globally trivial, so there is no reason to suppose that you end up in the same state after you go around a loop on your base space. Now we want to know: *when* does this happen? Are there some loops which do not change where you end up? To answer this, one is naturally led to the notion of the fundamental group.

Definition 2.1. *The fundamental group $\pi_1(X, x)$ of a topological space X based at a point $x \in X$ is the group whose elements consist of the equivalence classes of loops (i.e. continuous functions $f : [0,1] \to X, f(0) = f(1) = x$) up to homotopy fixing endpoints, with composition as the binary operation, the trivial loop as the identity, and reversal as the inverse.*

The idea of the fundamental group is simple: you want to describe all the loops based at a point in the space that are not topologically trivial. For example, if you take a contractible space like $[0,1]$ or $\mathbb{R}^n$, the fundamental group (at any point) is trivial. If you take S^1, it is a good exercise to prove that the fundamental group is $\mathbb{Z}$, generated by going around the circle.

You might guess why this is relevant to monodromy: if you go around a loop that is homotopically trivial, i.e. trivial in $\pi_1(X, x)$, then the action on the fiber of a local system at x is trivial. The reason is essentially because of the compactness of the square $[0,1]\times[0,1]$. Local systems are locally trivial, so given two homotopic paths between the same two points you can find a finite open covering on which the local system is trivial, and identify the two linear transformations given by the two parths through them. This allows us to make the following definition.

Definition 2.2. *Given a local system $\mathcal{F}$ on X, the monodromy representation at $x\in X$ is the linear representation $\pi_1(X,x)\to \mathrm{GL}(\mathcal{F}_x)$ given by analytic continuation.*

In fact, giving the data of a local system is equivalent to giving a representation of its fundamental group. This is a very classical fact that is relatively straightforward to show, and is the first incarnation of the Riemann-Hilbert correspondence.

Proposition 2.3. *On any reasonable connected topological space (e.g. manifolds), there is an equivalence of categories between local systems of vector spaces and representations of the fundamental group.*

2.2. **The Riemann-Hilbert problem.** Riemann introduced the notion of a local system to study the solutions to linear ordinary differential equations, and in particular the hypergeometric equation. Because the hypergeometric equation has singularities at $0,1,\infty$, the base space is given by $U=\mathbb{P}^1_{\mathbb{C}}-\{0,1,\infty\}$. To understand the monodromy representation, we first compute the fundamental group of the base. Fix any base point $u\in U$. One can perform a deformation retract from U to a space consisting of two circles joined at a point, which has fundamental group F_2: the free group on two elements. However, a more suggestive way to describe the fundamental group is

$$\pi_1(U,u)\cong\{\rho_0,\rho_1,\rho_\infty|\rho_\infty\rho_1\rho_0=1\}.$$

Here we think of ρ_z as a loop going around z, and the relation expresses the fact that if you go around them all, you get a contractible loop.

Now we can consider the effect of each ρ_i, with the caveat that we are doing it with respect to different bases. A power series itself doesn't have monodromy as you go around; this is essentially because each exponent is a nonnegative integer. Instead, it's the fractional powers that give monodromy. In particular, z^a will have monodromy $e^{2\pi ia}$. Looking at the solutions in Proposition 1.3, we see that the 'local mondoromy' is given by the matrices

$$\begin{pmatrix}1 & 0\\ 0 & e^{2\pi i(1-c)}\end{pmatrix},\quad \begin{pmatrix}1 & 0\\ 0 & e^{2\pi i(c-a-b)}\end{pmatrix},\quad \begin{pmatrix}e^{-2\pi i-a} & 0\\ 0 & e^{-2\pi i-b}\end{pmatrix},$$

again with the caveat that these are not in the same basis. To get them into the same basis, one has to use Kummer's transformation formulae, which can be quite complicated.

So what did Riemann do? He turned the problem on its head, beginning with the abstract notion of mondoromy, and used them to deduce Kummer's transformation formulas of the hypergeometric without actually needing to compute the hypergeometric solutions themselves! To be more precise, he began with a collection of three local mondoromy matrices, subject to the condition that the product is the identity. He showed that this information was enough to determine the global representation, which describes the entirety of the associated local system. This allowed him to reproduce Kummer's results with little computation needed.

Remark. As Nick Katz remarks in his famous monograph *Rigid Local Systems*, Riemann was "lucky", in that his method wouldn't immediately work in more general scenarios (e.g. with n singular points). The reason it worked in the case of $U = \mathbb{P}^1 - \{0, 1, \infty\}$, in modern terms, is that rank 2 local systems on U are *rigid*. Far from being a defect, this was rather the beginning of a whole new beautiful theory largely developed by Katz in its modern form.

In Hilbert's list of problems for the 20th century, he included one inspired by this work. It is Hilber'ts 21st problem, but it has gone on to also be known as the Riemann-Hilbert problem. We quote from a translation.

Problem: [Riemann-Hilbert problem] "*Proof of the existence of linear differential equations having a prescribed monodromic group*
In the theory of linear differential equations with one independent variable z, I wish to indicate an important problem one which very likely Riemann himself may have had in mind. This problem is as follows: To show that there always exists a linear differential equation of the Fuchsian class, with given singular points and monodromic group. The problem requires the production of n functions of the variable z, regular throughout the complex z-plane except at the given singular points; at these points the functions may become infinite of only finite order, and when z describes circuits about these points the functions shall undergo the prescribed linear substitutions. The existence of such differential equations has been shown to be probable by counting the constants, but the rigorous proof has been obtained up to this time only in the particular case where the fundamental equations of the given substitutions have roots all of absolute magnitude unity. L. Schlesinger (1895) has given this proof, based upon Poincaré's theory of the Fuchsian zeta-functions. The theory of linear differential equations would evidently have a more finished appearance if the problem here sketched could be disposed of by some perfectly general method."

Another way to ask the question is: for any punctured Riemann surface U, can every finite-dimensional representation of the fundamental group of $\pi_1(U)$ be obtained as the monodromy representation of a differential equation on U with regualr singular points? The case of $U = \mathbb{P}^1 - \{0, 1, \infty\}$ and 2-dimensional representations follows from Riemann's work on the hypergeometric.

As Hilbert stated it, the problem was essentially one of function theory, which was solved by various mathematicians in the years to come. But the first statement of a problem isn't always the deepest or most fruitful. Just as Riemann's originality had transformed the study of the hypergeometric equation, the work of mathematicians such as Grothendieck paved the way for a deeper understanding of what the Riemann-Hilbert problem ought to truly be about.

References

[1] Jacques Dutka. “The Early History of the Hypergeometric Function”. In: *Archive for History of Exact Sciences* 31.1 (1984), pp. 15–34. issn: 00039519, 14320657. url: http://www.jstor.org/stable/41133728 (visited on 05/31/2024).