The arithmetic and topology of differential equations

Don Zagier

Abstract. This survey paper attempts to present in as elementary a way as possible a wide panorama of results concerning the relations between differential equations on the one hand and algebraic geometry, number theory and topology on the other. We use the famous Apéry numbers as a running example to illustrate the connections with, among other things, the theory of periods (Picard–Fuchs differential equations), the theory of modular forms (and the special values of their L-series), the theory of motives (starting with counting points on varieties over finite fields), and mirror symmetry (in particular, the Gamma Conjecture relating the asymptotics of solutions of differential equations to the multiplicative "Gamma class" of a variety). A number of relations to works by Friedrich Hirzebruch, in whose honor the lecture was given, are also described.

My principal aim in this paper is to present, in as simple a way and to as wide a readership as possible, some of the beautiful ways in which differential equations are related to number theory, algebraic geometry, and topology. But – since this is the first of the "Hirzebruch Lectures" that are now planned to be given at each future European Mathematical Congress – a secondary goal was to present topics that would have appealed to Friedrich Hirzebruch and that are related to some of his major discoveries. Three such topics that will appear are: his resolution of the cusp singularities of Hilbert modular surfaces, which serve to explain a puzzling integrality property of a differential equation connected with a certain non-arithmetic modular curve (in Section 4); his theory of multiplicative characteristic classes and genera, which are needed to define the Gamma Class and formulate the Gamma Conjecture in the theory of mirror symmetry (in Section 9); and, very briefly in Section 10, the study of ramified coverings of the projective plane and surfaces with $c_1^2 = 3c_2$, which are related via his proportionality principle to quotients of the complex 2-ball but also to the monodromy groups of certain higher-dimensional hypergeometric differential equations. The last topic was of particular interest to Hirzebruch in his later years, and together with Paula Beazley Cohen (later Tretkoff) he wrote a long review of the monograph of Deligne and Mostow on the subject [2] and also originally planned a joint book with her based on a course that he gave in Zürich in 1996, though he eventually abandoned the project and asked her to complete it alone [68].

In accordance with this double aim, I have tried to explain at least the basic ideas of each topic occurring in as elementary a way as possible, and to include even definitions that will be familiar to many readers. In particular, in the second section, which is concerned with the relations of differential equations to algebraic geometry, I include a discussion of differential forms and of the fact that the periods associated to a family of algebraic varieties always satisfy a linear differential equation with integral monodromy (Picard–Fuchs equation), and in the following section, which treats the relation of differential equations to modular forms, I include a brief review of the definitions and main features of modular forms. Also in the case of some of the more advanced topics that are discussed later in the paper, such as the theory of motives or mirror symmetry and quantum cohomology, I have tried to explain the main concepts from scratch.

To whet the reader's appetite, the paper begins with Apéry's famous proof of the irrationality of $\zeta(3)$ and a discussion, from six different points of view, of the reasons for the "miraculous" integrality that makes it work. This example belongs to our subject because the generating function of the Apéry numbers satisfies a differential equation of Picard–Fuchs type, and in the main body of the paper, where the links between differential equations and other fields of mathematics – algebraic geometry, modular forms, number theory, mirror symmetry, and topology – are treated, each of these viewpoints reappears in a more general setting. Specifically, the further sections of the paper are as follows. Section 2 discusses the notions of periods and period functions and the proof that the latter always satisfy a linear differential equation, with many examples. Section 4 contains a brief overview of the theory of modular forms and a sketch of the proof that every modular form, when written as a function of a modular function, also satisfies a linear differential equation (in fact, of Picard–Fuchs type), again with many examples, including the non-arithmetic one related to cusps on Hilbert modular surfaces that was mentioned in the opening paragraph. In the next two sections we turn to number theory, with a discussion of the zeta function of a variety defined over a number field and its conjectural relationship to periods and with a brief introduction to the concept of motives. Here we illustrate the *usefulness* of the motivic point of view by discussing in some detail three concrete predictions that it yields, each of which could be verified numerically or theoretically or both. In the last two sections of the paper proper, the central role is played by topology rather than number theory, although both appear. Here we discuss mirror symmetry and the way that the quantum cohomology of a variety (defined in terms of its Gromov–Witten invariants, i.e., the counting functions of holomorphic embeddings of curves into the variety) leads to a differential equation that is conjecturally equivalent to the Picard–Fuchs equation of the "mirror" family of varieties. We also explain the Gamma Conjecture, which makes a specific link between differential equations (specifically, the asymptotics at infinity of the solutions of the quantum differential equation of a variety) and topology (specifically, the Hirzebruch characteristic class of the variety defined using the power series expansion of $\Gamma(1+x)$). A final section treats a few miscellaneous topics, including a very brief discussion of the connection between higher-dimensional hypergeometric differential equations and the geometry of quotients of the complex 2-ball as mentioned above. With the exception of some results proved in Section 7, the paper is entirely expository.

In preparing this paper I was helped by many people, and I would like to thank Sasha Beilinson, Spencer Bloch, Philip Candelas, Alessio Corti, Boris Dubrovin, Javier

Fresán, Sergei Galkin, Stavros Garoufalidis, Günter Harder, Albrecht Klemm, Maxim Kontsevich, Anton Mellit, Martin Möller, Danylo Radchenko, Emanuel Scheidegger, Fernando Rodriguez Villegas, Masha Vlasenko, Di Yang, Shing-Tung Yau, Noriko Yui, Federico Zerbini, Wadim Zudilin and my wife Silke, as well as two very careful anonymous referees, for their many suggestions and comments. I am especially grateful to Duco van Straten and Vasily Golyshev, both of whom were unstinting in sharing their deep insight and knowledge of the field with me. Golyshev's influence is visible everywhere in the paper: several of the results (in particular those about the Gamma Conjecture in Section 9) were obtained jointly with him, and all three of the "motivic predictions" discussed in Section 7 came from him. But above all I would like to thank the late Friedrich Hirzebruch, my teacher and friend, who taught me and a whole generation of mathematicians better ways to think about mathematics and better ways to think about the world.

1 Prelude: The Apéry integrality miracle

In 1978, Roger Apéry created a sensation in the mathematical world by proving that the number

$$\zeta(3) = \sum_{n=1}^{\infty} \frac{1}{n^3} = 1.2020569031595\cdots$$

is irrational. He obtained this as a consequence of the following three remarkable facts. Let

$$\{A_n\}_{n\ge 0} = \{\,1,\ 5,\ 73,\ 1445,\ 33001,\ \ldots\}$$

and

$$\{B_n\}_{n\ge 0} = \Big\{0,\ 6,\ \frac{351}{4},\ \frac{62531}{36},\ \frac{11424695}{288},\ \ldots\Big\}$$

be the solutions of the recursion

$$(n+1)^3 u_{n+1} - (34n^3 + 51n^2 + 27n + 5)\, u_n + n^3\, u_{n-1} = 0 \qquad (1.1)$$

with initial conditions $(A_0, A_1) = (1, 5)$ and $(B_0, B_1) = (0, 6)$. Then

(a) $A_n \in \mathbb{Z}$ for all $n \ge 0$;
(b) $d_n^3 B_n \in \mathbb{Z}$ for all $n \ge 0$, where $d_n = \mathrm{l.c.m.}\{1, 2, \ldots, n\}$;
(c) $B_n/A_n \to \zeta(3)$ as $n \to \infty$.

Together, these three facts quickly imply the irrationality of $\zeta(3)$: any solution of (1.1) must behave asymptotically like a constant times $n^{-3/2}C^{\pm n}$ as $n \to \infty$, where $C = (1+\sqrt{2})^4 = 33.97\cdots$, so the difference between B_n/A_n and its limiting value $\zeta(3)$ is $\mathrm{O}(C^{-2n})$ as $n \to \infty$, and since the denominator of B_n/A_n is $\mathrm{O}(e^{3n}C^n) = \mathrm{o}(C^{2n})$ (because $d_n = e^{n+\mathrm{o}(n)}$ and $e^3 < C$) this degree of approximability is not compatible with the rationality of $\zeta(3)$. As a numerical illustration of the rapidity of Apéry's approximations, we have

$$\frac{B_4}{A_4} = \frac{11424695}{288 \times 33001} = 1.2020569031578\cdots .$$

Thus (a)–(c) imply Apéry's sensational discovery that $\zeta(3)$ is irrational. But why are they true? In particular, where does the integrality statement (a) come from? In computing A_n recursively from (1.1) we must divide by n^3 at each stage and hence should expect a priori that A_n has denominator $n!^3$. So the integrality assertion (a) (and to a somewhat lesser degree the denominator bound (b)) is very surprising, and we will describe in a moment a numerical experiment showing that this phenomenon is indeed exceedingly rare. In the rest of this section, which serves as motivation for the rest of the paper, we will list some of the explanations for the integrality that have been found, giving more details of each in later sections.

First, to justify the word "miracle" in the title of this section we should say something about the numerical evidence showing how special the integrality is. As well as his proof of the irrationality of $\zeta(3)$, Apéry had found a completely similar proof of that of $\zeta(2)$ (which was of course already known as a consequence of Euler's formula $\zeta(2) = \pi^2/6$) based on the 3-term recursion

$$(n+1)^2 u_{n+1} - (11n^2 + 11n + 3) u_n - n^2 u_{n-1} = 0 \tag{1.2}$$

instead of (1.1), where the two solutions defined by the initial values $(A_0, A_1) = (1, 3)$ and $(B_0, B_1) = (0, 5)$ satisfy the same properties as before, but now with d_n^3 replaced by d_n^2 and $\zeta(3)$ by $\zeta(2)$. As part of his study of the congruence properties of such recursions, about which we will say more later, Beukers [8] generalized (1.2) to the 3-parameter family of "Apéry-like" recursions

$$(n+1)^2 u_{n+1} - (An^2 + An + B) u_n + Cn^2 u_{n-1} = 0 \tag{1.3}$$

with A, B, $C \in \mathbb{Z}$. In [74], I looked at the first 100 million triples (A, B, C) and found (up to scaling and assuming $C(A^2 - 4C) \neq 0$ to avoid degenerate cases) that only seven of them gave recursions with integral solutions: the initial one $(11, 3, -1)$ found by Apéry and the six further cases

$$(0,0,-16),\ (7,2,-8),\ (9,3,27),\ (10,3,9),\ (12,4,32),\ (17,6,72)\,! \tag{1.4}$$

So the phenomenon we are talking about is indeed an extremely rare one.

We now return to equation (1.1) and describe five different methods that can be used to prove the integrality of the Apéry numbers A_n (and also in each case the further properties (b) and (c), though we will not describe this), and one further interpretation of these numbers.

- Apéry's own proofs of (a)–(c) were based on his *explicit formula*

$$A_n = \sum_{k=0}^{n} \binom{n}{k}^2 \binom{n+k}{k}^2 \tag{1.5}$$

  for A_n and a similar but more complicated expression for B_n. These two formulas immediately implied all three of his assertions, but his explanation of where they came from and why they were true was rather obscure and did not convince everybody. (To quote from [4], "The proof was elementary but the complexity and the unexpected nature of Apéry's formulas divided the audience into believers and disbelievers.") A more standard proof was quickly found, and was presented by Henri Cohen at the International Mathematical Congress in 1978 only two months after Apéry's announcement (the full story, in which I was peripherally involved, is told very amusingly in the article [57] by Alf van der Poorten), and will be reproduced and slightly generalized in Section 7. But it did little to dispel the mystery and everybody felt that there had to be more enlightening explanations.

- The first such explanation was found by Frits Beukers, only three months after the Helsinki congress. In [4], he showed that the difference $B_n - A_n\zeta(3)$ has the *integral representation*

$$B_n - A_n\zeta(3) = \int_0^1\int_0^1\int_0^1 \left(\frac{x(1-x)y(1-y)z(1-z)}{1-z+xyz}\right)^n \frac{dx\,dy\,dz}{1-z+xyz} \tag{1.6}$$

and that all three properties (a)–(c) follow from this. Of course this formula also has a somewhat "rabbit-out-of-a-hat" appearance (although it does fit better into the framework of earlier known irrationality proofs than just the unmotivated recursion (1.1) for the numbers (1.5)), but as we will see in the next-but-one bullet, it does in fact have a clear algebraic-geometric meaning.

- A yet more beautiful explanation of what was "really" behind Apéry's discovery was found a few years later, again by Beukers, and relies on the theory of *modular forms*. We will recall the definition and main properties of modular forms in Section 4, and here merely reproduce the relevant formulas from [5]. Set

$$T(q) = q\prod_{n=1}^{\infty}\frac{(1-q^n)^{12}(1-q^{6n})^{12}}{(1-q^{2n})^{12}(1-q^{3n})^{12}} = q - 12q^2 + 66q^3 - 220q^4 + \cdots,$$

$$F(q) = \prod_{n=1}^{\infty}\frac{(1-q^{2n})^7(1-q^{3n})^7}{(1-q^n)^5(1-q^{6n})^5} = 1 + 5q + 13q^2 + 23q^3 + \cdots.$$

Then, as will be explained in Section 4, the theory of modular forms implies that

$$F(q) = 1 + 5\,T(q) + 73\,T(q)^2 + 1445\,T(q)^3 + \cdots = \sum_{n=0}^{\infty} A_n\,T(q)^n, \tag{1.7}$$

which gives the integrality of A_n since both F and T have integral coefficients.

- In 1984, Beukers and Chris Peters [11] returned to the integral representation (1.6) and understood its underlying geometry in terms of *periods* in a family of algebraic varieties. The recursion (1.1) is equivalent to the statement that the generating function

$$\mathcal{A}(t) = \sum_{n=0}^{\infty} A_n t^n = 1 + 5\,t + 73\,t^2 + 1445\,t^3 + \cdots$$

is a solution of the differential equation $\mathcal{L}(\mathcal{A}) = 0$, where $\mathcal{L}$ is the differential operator

$$\mathcal{L} = D^3 - t\,(34D^3 + 51D^2 + 27D + 5) + t^2\,(D+1)^3 \qquad \left(D = t\,\frac{d}{dt}\right). \tag{1.8}$$

(It is of course this connection to differential equations that explains why we are using the Apéry proof as entry point into the subject of this paper.) If we substitute the integral representation (1.6) into this generating function, then we obtain an integral expression for $\mathcal{B}(t) - \mathcal{A}(t)\zeta(3)$ (where $\mathcal{B}(t) = \sum_{n=0}^{\infty} B_n t^n$ is the companion generating series to $\mathcal{A}(t)$) in which the integrand has singularities along the surface

$$V_t: \quad 1 - t\,\frac{x(1-x)y(1-y)z(1-z)}{1-z+xyz} = 0\,. \tag{1.9}$$

Beukers and Peters showed that V_t for generic t is birationally equivalent to a K3 surface with Picard number 19 and that $\mathcal{L}\mathcal{A} = 0$ is the Picard–Fuchs differential equation corresponding to this family. (We will recall this notion in the next section.) This is the point of view that will be the most important for this paper.

- Next, and as a consequence of the above, there is a simple description of the numbers A_n in terms of *Laurent polynomials*. If we make the substitution $x \to (x+z-1)/yz$ in the defining equation of V_t, then the new equation has the easy form $1 - tL(x,y,z) = 0$, where $L(x,y,z)$ is the Laurent polynomial

$$L(x,y,z) = \frac{(y-1)(z-1)(x+z-1)(yz-x-z+1)}{xyz} \tag{1.10}$$

("Landau–Ginzburg model"). The formula for A_n then becomes simply

$$A_n = \text{c.t.}(L^n)\,, \tag{1.11}$$

where "c.t." denotes the constant term in a Laurent polynomial in several variables. This expression gives the integrality of A_n instantly since L has integral coefficients, and we also see by a short calculation that it reproduces Apéry's formula (1.5).

- Finally, the numbers A_n have an interpretation in terms of *Gromov–Witten theory*. This is considerably less elementary to explain than the other points of view, so we give only a brief description in words here, referring the reader to Section 8 and the literature cited there for more definitions and explanations. To any Fano variety F one can associate a sequence of rational numbers $v_n(F)$ that satisfy a linear recursion with coefficients that are determined by the Gromov–Witten invariants of F (= enumeration of holomorphic maps $\mathbb{P}^1(\mathbb{C}) \to F$; the number $v_n(F)$ is an appropriately defined volume of the moduli space of all such maps of degree n). On the other hand, mirror symmetry predicts that F has an associated "mirror manifold" (actually a family of algebraic manifolds) and that the numbers $n!\,v_n(F)$ agree with the Taylor coefficients of a period of this mirror. If we take for F a particular Fano 3-fold called V_{12}, then the mirror exists and is precisely the Beukers-Peters family of K3 surfaces discussed above, so the numbers $n!\,v_n(V_{12})$ coincide with the Apéry numbers A_n. This gives a completely different explanation of the meaning of the Apéry numbers, and also another reason to at least expect them to be integral, since for geometric reasons the invariants $v_n(F)$ of any Fano variety are believed (though not known) to have denominator at most $n!$.

2 Differential equations and algebraic geometry

The connections of differential equations with both arithmetic and topology arise through the *periods* of algebraic varieties. By definition, a period on an algebraic variety X defined over $\mathbb{Q}$ (or any number field, but in this paper we stick to $\mathbb{Q}$ for simplicity) is a number defined by integrating an algebraic differential form on X over a submanifold (either closed or with a boundary defined over $\mathbb{Q}$). The class of these numbers forms a countable subring $\mathcal{P} \subset \mathbb{C}$ that contains all algebraic numbers and many of the numbers of greatest interest in mathematics, such as π, $\log 2$, $\zeta(n)$ (or more generally multiple zeta values, which will occur in Section 9), periods of modular forms, Mahler measures of polynomials with rational coefficients, and many others. They are also related to special values of motivic zeta functions (see Section 6). A survey of periods was given in [50].

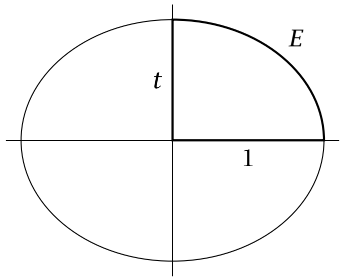
Just as the notion of algebraic numbers (numbers satisfying a polynomial equation over $\mathbb{Q}$) can be generalized to the notion of algebraic functions (functions that satisfy a polynomial equation over $\mathbb{Q}(t)$ and whose values at algebraic arguments are then automatically algebraic), the notion of periods can be extended to that of *period functions* (functions of t that are defined by the integral over a submanifold of a differential form depending algebraically on a parameter t, and which then automatically assume values in $\mathcal{P}$ at algebraic arguments). The key fact, whose proof will be recalled below, is that any period function satisfies a linear differential equation with algebraic coefficients (Picard–Fuchs differential equation). It is this fact that creates the link between the three subjects constituting the title of this paper.

In this section, after reviewing differential forms and the de Rham theorem, we will discuss the definition and differential equation of period functions in a little more detail and give a number of examples. These notions, and also these examples, will then recur in the rest of the paper in connection with other more specific topics like modular forms, zeta functions, and mirror symmetry. First, however, we present a simple and prototypical example.

Example: The circumference of an ellipse

The historically earliest example of the differential equation satisfied by a period function was given by Euler in 1733,[1] who showed that the quarter-length E of an ellipse of eccentricity $t < 1$,

1. See pp. 85 ff. of [39]. Both this book by Christian Houzel and the book [33] by Jeremy Gray contain a wealth of information about the early history of elliptic and hypergeometric functions, and of differential equations in general. See also the very nice expository paper [67] by Burt Totaro.



considered as a function of the parameter $k = \sqrt{1-t^2}$, satisfies the differential equation

$$k(k^2-1)\,E''(k) + (k^2-1)\,E'(k) - k\,E(k) = 0\,. \tag{2.1}$$

This is a consequence of the following calculation

$$\begin{aligned}
E(k) &= \int_0^{\pi/2} \sqrt{\cos^2\theta + t^2\sin^2\theta}\;d\theta = \int_0^{\pi/2}\sqrt{1-k^2\sin^2\theta}\;d\theta \\
&= \int_0^1 \sqrt{\frac{1-k^2x^2}{1-x^2}}\,dx \qquad\qquad (x=\sin\theta) \\
&= \frac{\pi}{2}\sum_{n=0}^{\infty}\binom{1/2}{n}\binom{-1/2}{n}k^{2n} = \frac{\pi}{2}\Bigl(1-\frac{k^2}{4}-\frac{3k^4}{64}-\frac{5k^6}{256}-\cdots\Bigr)\,,
\end{aligned}$$

in which the first equation comes directly from the definition of E, the second expresses E as a period, and the third expresses it as a power series in k^2, from which the differential equation (2.1) follows immediately by term-by-term differentiation. The power series for $\frac{2}{\pi}E(k)$ occurring here is the special case $F(-\frac12,\frac12;1;k^2)$ of the Euler–Gauss hypergeometric function

$$F(a,b;c;t) = \sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{(c)_n\,n!}\,t^n \qquad \bigl(\,(a)_n := a(a+1)\cdots(a+n-1)\,\bigr), \tag{2.2}$$

a very important and beautiful class of special functions, which the reader can learn more about from the wonderful book [72] by Masaaki Yoshida.

Review of differential forms and de Rham's theorem

The fact that period functions satisfy a differential equation is closely related to the de Rham theorem describing the cohomology of a manifold in terms of differential forms, so we begin by giving a brief review of this. This material is standard and can be skipped by any reader who is familiar with it.

Let X be a (compact, smooth, oriented) manifold of real dimension m. Then for every integer r between 0 and m we have the rth *homology group* $H_r(X;\mathbb{Z})$, which is an abelian group of finite rank $b_r(X)$ (called the rth *Betti number* of X)

whose elements are represented by closed r-dimensional "chains" modulo boundaries of $(r + 1)$-dimensional chains. (For an intuitive picture, one can think of r-dimensional oriented submanifolds of X modulo the relation that such a submanifold is equated to 0 if it is the boundary of an oriented $(r + 1)$-dimensional submanifold of X.) We also have the corresponding cohomology group $H^r(X;\mathbb{Z})$, which is an abelian group of the same rank that up to torsion can be identified with the dual of $H_r(X;\mathbb{Z})$, as well as the $b_r(X)$-dimensional vector space $H^r(X;\mathbb{C}) = H^r(X;\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{C} = \mathrm{Hom}(H_r(X;\mathbb{Z}),\mathbb{C})$ (rth cohomology group with complex coefficients).

Differential forms constitute the language that is needed to formulate the classical theorems of many-variable calculus for general manifolds. Let X and m be as above. By definition, a 0-*form* on X is a smooth function on X, represented in local coordinates $\mathbf{x} = (x_1,\dots,x_m)$ on X by a C^∞-function $f(\mathbf{x}) = f(x_1,\dots,x_m)$, a 1-*form* is a formal linear combination $\sum_{i=1}^n f_i(\mathbf{x})\,dx_i$, where each f_i is C^∞ and "dx_i" is a formal symbol meant to suggest a small change of the coordinate x_i with all of the other coordinates being kept constant, and an r-*form* for arbitrary $0 \le r \le m$ is a formal linear combination of r-fold products $dx_{i_1}\cdots dx_{i_r}$ with C^∞-functions of $\mathbf{x}$ as coefficients, where the multiplication of the symbols dx_i is required to satisfy the anticommutativity property $dx_i\,dx_j = -dx_j\,dx_i$. (In particular, $(dx_i)^2 = 0$, so that the i_ν have to be distinct and one can assume that $1 \le i_1 < \cdots < i_r \le m$.) The basic operation on differential forms is the "exterior derivative" d, which sends r-forms to $(r + 1)$-forms. It is defined on 0-forms by the formula $d(f) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}\,dx_i$ (corresponding to the gradient in multi-variable calculus), on the special 1-forms dx_i by $d(dx_i) = 0$, and then inductively on all forms by requiring d to be linear and to satisfy the derivation property $d(\omega_1\omega_2) = d(\omega_1)\omega_2 \pm \omega_1 d(\omega_2)$ for any two differential forms ω_1 and ω_2 on X. A simple calculation then shows that $d(d\omega) = 0$ for any differential form ω. In particular, if we define an r-form ω to be *closed* if $d\omega = 0$ and *exact* if $\omega = d\eta$ for some $(r - 1)$-form η, then all exact forms are closed.

The key fact about differential forms is Stokes's theorem $\int_A d\eta = \int_{\partial A} \eta$, which generalizes the fundamental theorem of calculus $\int_a^b f'(x)dx = f(x)\big|_{x=a}^{x=b}$ as well as many classical theorems of multivariate calculus like the divergence and curl theorems. This shows immediately that the integral of a closed r-form ω over a closed r-dimensional submanifold C of X depends only on the homology class $[C]$ of C in $H_r(X;\mathbb{Z})$ (because if $C = \partial B$ is homologous to 0 then $\int_C \omega = \int_{\partial B} \omega = \int_B d\omega = 0$), and that this integral vanishes if ω is exact (because $\int_C \omega = \int_C d\eta = \int_{\partial C} \eta = 0$). In other words, every closed form on X represents a cohomology class with complex coefficients, and this cohomology class is unchanged if the form is changed by the addition of an exact form. De Rham's theorem is the converse statement, that every class in $H^r(X;\mathbb{C})$ can be obtained by integrating a closed r-form on X which is uniquely determined up to an exact form. In more formal language, de Rham's theorem expresses $H^r(X;\mathbb{C})$ as the (finite-dimensional) quotient of the (infinite-dimensional) space of closed complex-valued r-forms on X by its (infinite-dimensional) subspace of exact r-forms.

The Picard–Fuchs differential equation

We first give a rough and not quite correct explanation of this to give the main idea, and then a more careful one. The situation we are now interested in is when the manifold denoted X above is replaced by a family $\{X_t\}_{t\in U}$ of manifolds depending smoothly on a parameter t in some "base space" U and we have a closed r-form ω_t on each X_t depending smoothly on t. We can then think of the X_t as the pre-images $\pi^{-1}(t)$ of a smooth map π from some larger "ambient space" $\mathcal{X}$ to U and take ω_t to be the restriction $\omega_t = \omega|_{X_t}$ of some (not necessarily closed!) r-form ω on $\mathcal{X}$ to X_t. The statement we want to make is a local one, so we can assume that t moves in a small open subset of U over which the mapping π is locally a product. Then the fibres X_t are smooth and diffeomorphic to each other, so the homology groups $H_r(X_t;\mathbb{Z})$ can be canonically identified with one another and we can speak (locally) of a "constant" cycle Y, meaning a smoothly varying family of submanifolds $Y_t \subset X_t$ with $[Y_t] \in H_r(X_t;\mathbb{Z})$ constant. We denote by $P(t)$ the integral of our chosen r-form ω_t over the submanifold Y_t. The argument producing a differential equation satisfied by $P(t)$ then runs as follows. Since our map is locally a product, we can take t and a coordinate system $\mathbf{x} = (x_1,\dots,x_m)$ on X_t as local coordinates on $\mathcal{X}$, and then $\partial/\partial t$ commutes with the partial derivatives $\partial/\partial x_i$ and hence with the exterior derivative d on X_t as introduced above. It follows that if the r-form $\omega_t(\mathbf{x})$ is closed, then so is the r-form $\partial^i\omega_t(\mathbf{x})/\partial t^i$ for $i \ge 0$, so by de Rham's theorem these forms for $i = 0,1,\dots,b$, where $b = b_r(X_t) = \dim_{\mathbb{C}} H^r(X_t;\mathbb{C})$, have a linear combination (with coefficients that are smooth functions of t) that is exact. The integral over Y of this linear combination is then 0, and this is our desired differential equation:

$$\sum_{i=0}^{b} a_i(t)\,\frac{\partial^i\omega_t(\mathbf{x})}{\partial t^i} = d(\eta_t) \quad\Rightarrow\quad \sum_{i=0}^{b} a_i(t)\,\frac{d^iP(t)}{dt^i} = 0\,.$$

If ω_t depends algebraically on t as well as on the coordinates in X_t, then the coefficients $a_i(t)$ are also algebraic functions and we get the desired Picard–Fuchs differential equation. In our examples, the base space U will always be a complex curve, but in practice it will usually be the complex line $\mathbb{C}$ or projective line $\mathbb{P}^1(\mathbb{C})$, in which case the coefficients of the differential equation will be rational functions and hence, after multiplying by a common denominator, polynomials in t, implying that the Taylor coefficients of any local (power series) solution of the differential equation will satisfy a recursion of finite order with polynomial coefficients.

In fact the argument just given is not correct as it stands, for two reasons. First of all, the classical de Rham theorem applies to real manifolds and C^∞ forms, while we want to work with algebraic manifolds over $\mathbb{C}$ and holomorphic (algebraic) forms. If X is a smooth complex variety of complex dimension n, so real dimension $m = 2n$, we can choose local complex coordinates $\mathbf{z} = (z_1,\dots,z_n)$ and take $\mathbf{x}$ to be $(x_1,y_1,\dots,x_n,y_n)$, where $z_j = x_j + iy_j$. Then we can take $dz_j = dx_j + i\,dy_j$ and $d\bar z_j = dx_j - i\,dy_j$ rather than dx_j and dy_j as our basis of 1-forms over the algebra of smooth functions on, and similarly $dz_{j_1}\dots dz_{j_p}d\bar z_{k_1}\cdots d\bar z_{k_q}$ (with $p,\,q \ge 0$, $p+q = r$) as our basis of r-forms. The cohomology classes represented by linear combinations of such forms with given p and q are said to be of *Hodge type* (p,q), the complex subspace of $H^r(X;\mathbb{C})$ consisting of such classes is denoted $H^{p,q}(X)$, and a fundamental theorem of Hodge says that for a smooth projective variety X the full cohomology group $H^r(X;\mathbb{C})$ is the direct sum of the spaces $H^{p,q}(X)$ with $p+q=r$. (If X is not smooth or not compact, then there is still a Hodge theory but it has a more complicated structure, due to Deligne, that we will not describe.) When we speak of an "algebraic" r-form on X, we mean a form of type $(r,0)$ with algebraic (and in particular holomorphic) coefficients, but then we are only getting part of the cohomology. For instance, if X is a compact Riemann surface of genus g, then the first cohomology group $H^1(X;\mathbb{C})$ has dimension $2g$, while the part $H^{1,0}(X)$ representable by algebraic forms has dimension only g.

The second problem is that if our variety X belongs to a family $\{X_t\}$ and we have an algebraically vanishing family of closed r-forms ω_t on X_t, then when we differentiate with respect to the parameter t we can create poles, as we will see explicitly in Example 1 (Legendre elliptic curve) in the next section. This is in fact connected with the first point, since if all derivatives $d^i\omega_t/dt^i$ were holomorphic then we would find a differential equation of order $\dim H^{r,0}(X)$, rather than the correct $\dim H^r(X)$, for the periods. In the case of curves (which is the context of the classical Picard–Fuchs differential equation) there is a simple solution. If X is a Riemann surface (complex curve), then instead of using the de Rham theorem to represent $H^1(X)$ in terms of holomorphic and antiholomorphic 1-forms we can represent it by meromorphic forms ω of the second kind modulo exact forms df, where "second kind" means that the residue of ω at each of its poles vanishes. Such a form represents a well-defined cohomology class on X because the value of its integral over a closed real curve $Y \subset X$ does not change as Y moves across a singularity of ω and hence depends only on the homology class of Y, and the class of meromorphic 1-forms of the second kind is also closed under differentiation with respect to a parameter in an algebraically defined family, so that everything is okay. For higher-dimensional varieties there is still a notion of differentials of the second kind (locally the sum of a smooth form and an exact one; see pp. 454–6 of [34]), but it is no longer sufficient to work with these, and one needs instead a general algebraic de Rham theory, in which the non-algebraically defined space $H^{p,q}(X)$ is replaced by qth cohomology group of X with coefficients in the sheaf of holomorphic r-forms, which is defined algebraically. (In general, one needs hypercohomology to define it correctly, but if X is affine – i.e., a subvariety of some $\mathbb{C}^N$ defined by polynomial equations – then ordinary cohomology suffices.) The requisite algebraic de Rham theorem was proved by Alexander Grothendieck [35], and a full algebraic treatment of Picard–Fuchs differential equations in arbitrary dimensions and over arbitrary fields was given by Nicholas Katz (cf. [42], whose introduction gives a very clear overview of the problems involved, and also [43], [46] and [44]).

Properties of Picard–Fuchs differential equations

The derivation sketched above shows that the Picard–Fuchs differential equation of the function $t \mapsto \int_{[Y]} \omega_t$ depends only on the form ω_t and not on the homology class $[Y]$ over which it is integrated, and this is just as it should be: a linear differential equation of order b has a precisely b-dimensional space of solutions at a generic point (Cauchy), and by varying $[Y]$ over the b-dimensional space $H_r(X_t;\mathbb{Z}) \otimes \mathbb{C}$ we obtain all of these solutions. This has a very important consequence. For any linear differential equation of order b, say defined with respect to a parameter t ranging over $\mathbb{C}$ minus a finite set S of singular points, we can choose a basis of b solutions at some non-singular point t_0 and analytically continue them around any path in $\mathbb{C} \smallsetminus S$. If we choose a closed path, then the b analytically continued solutions again belong to the space of solutions at t_0, so they are linear combinations of the original solutions. This gives a $b \times b$ matrix depending only on the homotopy class of the closed path and hence a homomorphism from $\pi_1(\mathbb{C} \smallsetminus S, t_0)$ to $GL(b,\mathbb{C})$, called the *monodromy representation*. In the case of the Picard–Fuchs differential equation of a period function, the basis of solutions can be chosen by integrating over a $\mathbb{Z}$-basis of $H_r(X_{t_0};\mathbb{Z})$, and it follows that the monodromy representation of the differential equation that we have found is always integral, i.e., with respect to a suitable basis of the space of solutions at a given point it takes values in the subgroup $GL(b,\mathbb{Z})$ of $GL(b,\mathbb{C})$. However, we should warn the reader explicitly that the minimal differential equation satisfied by a period function may *not* have integral monodromy, because this equation may arise from a lower-dimensional subspace of the cohomology (we will see examples below) and this piece may not be defined over $\mathbb{Q}$. In such a case, the monodromy matrices of the Picard–Fuchs differential equation (in a suitable basis) may have entries with values in a number field rather than in $\mathbb{Q}$.

As well as this integrality (or near-integrality) of the monodromy, Picard–Fuchs differential equations have several other special properties among the class of all linear differential equations with polynomial coefficients:

- They have only regular singular points. We recall the definition: if we write the equation as $y^{(n)} + a_1(x)y^{(n-1)} + \cdots + a_n(x)y = 0$, where $y^{(i)} = d^iy/dx^i$ and $a_1(x), \dots, a_n(x)$ are rational functions, then a point $x_0 \in \mathbb{C}$ is called a *singular point* of the equation if some $a_i(x)$ has a pole at x_0 and is *regular* if the pole of every a_i at x_0 has order at most i. The same definitions apply also to $x_0 = \infty \in \mathbb{P}^1(\mathbb{C})$ after making the change of variables $x \mapsto 1/x$ in the differential equation.
- Their local monodromy around any singular point (i.e., the matrix representing the analytically continued solutions in terms of the original ones when we analytically continue a basis of solutions near the singular point around a small loop circling it) is quasiunipotent (all its eigenvalues are roots of unity).
- The numerators and denominators of the Taylor coefficients of a solution of the differential equation around a singular point have at most exponential growth, i.e., they satisfy the same integrality or near-integrality as we saw for the Apéry sequences $\{A_n\}$ and $\{B_n\}$ at the beginning of this paper.

It is believed that these properties characterize Picard–Fuchs differential equations. For a detailed discussion and more precise conjectures, see Simpson's article [65].

3 Examples

In this section we give several examples of Picard–Fuchs differential equations.

Example 1: The Legendre elliptic curve

Consider the elliptic curve given by the Legendre equation

$$E_t : \quad y^2 = x\,(x-1)\,(x-t)\,, \tag{3.1}$$

where t is a complex parameter. On this curve, like on any elliptic curve given by a Weierstrass equation over $\mathbb{C}$, there is a unique (up to a constant multiple) holomorphic 1-form, given by

$$\omega_t = \frac{dx}{y} = \frac{dx}{\sqrt{x(x-1)(x-t)}}\,.$$

This form is automatically closed because it is a holomorphic 1-form and our variety has complex dimension 1. With a little bit of experimentation one finds an exact form that is a combination of partial derivatives of ω_t with respect to t, namely

$$\begin{aligned} -2\,d\Big(\frac{x^{1/2}\,(x-1)^{1/2}}{(x-t)^{3/2}}\Big) &= \frac{(x-t)^2 + 2(2t-1)(x-t) + 3t(t-1)}{x^{1/2}\,(x-1)^{1/2}\,(x-t)^{5/2}}\,dx \\ &= \Big(1 + 4(2t-1)\,\frac{\partial}{\partial t} + 4t(t-1)\,\frac{\partial^2}{\partial t^2}\Big)\,\omega_t\,, \end{aligned} \tag{3.2}$$

and from this it follows that the integral $P(t) = \int_\gamma \omega_t$ for any closed curve γ on E_t [2] satisfies the Legendre differential equation

$$\Big(1 + 4(2t-1)\,\frac{d}{dt} + 4t(t-1)\,\frac{d^2}{dt^2}\Big)\,P(t) = 0\,.$$

The integral basis of the space of solutions that we discussed at the end of Section 2 can be seen clearly in this example. Suppose for concreteness that t is a real number between 0 and 1. As a basis of the rank 2 group $H_1(E_t;\mathbb{Z})$ we take the classes $[C_i]$ of the two curves C_1 and C_0 on E_t given by the double cover $y = \pm\sqrt{x(x-1)(x-t)}$

2. Note that both $\partial\omega_t/\partial t$ and $\partial^2\omega_t/\partial t^2$ have poles at $x = t$, but are meromorphic 1-forms of the second kind (vanishing residues), in accordance with the discussion in the previous section.

of the real interval $[1,\infty)$ and the double cover $y = \pm i\sqrt{|x|(|x|+1)(|x|+t)}$ of the real interval $(-\infty,0]$, respectively. Then

$$\int_{C_1}\omega_t = 2\int_1^\infty \frac{dx}{\sqrt{x(x-1)(x-t)}} = 2\sum_{n=0}^\infty \binom{-1/2}{n}(-t)^n \int_1^\infty \frac{dx}{x^{n+1}\sqrt{x-1}}$$
$$= 2\pi \sum_{n=0}^\infty \binom{-1/2}{n}^2 t^n = 2\pi\, F(\tfrac12,\tfrac12;1;t) \tag{3.3}$$

by a term-by-term expansion-and-integration calculation exactly similar to the one used for Euler's equation (2.1), while a similar but messier computation gives

$$\int_{C_0}\omega_t = \frac{2}{i}\sum_{n=0}^\infty \binom{-1/2}{n}^2 \Bigl(\log\bigl(\frac{t}{16}\bigr) + 4\bigl(\frac{1}{n+1}+\cdots+\frac{1}{2n}\bigr)\Bigr)\, t^n \tag{3.4}$$

for the integral of ω_t over the other basis element. These two functions of t are then a $\mathbb{Z}$-basis of the canonical $\mathbb{Z}$-lattice of solutions of the differential equation whose existence was explained in the last section, and the monodromy matrices obtained by analytically extending them around a closed loop in $\mathbb{C}\smallsetminus\{0,1\}$ lie in the subgroup $SL(2,\mathbb{Z})$ of $GL(2,\mathbb{C})$. This statement is closely related to the modular interpretation of the Legendre family that will be discussed in Section 5.

Example 2: The Dwork quintic pencil

Our next example, which we will see again several times later in this paper, is the subvariety $\mathcal{Q}_\psi \subset \mathbb{P}^4(\mathbb{C})$ given by the homogeneous equation

$$\mathcal{Q}_\psi:\quad x_1^5+x_2^5+x_3^5+x_4^5+x_5^5 = \psi\, x_1x_2x_3x_4x_5\,, \tag{3.5}$$

where ψ is a complex parameter. For $\psi\neq 0,\,5$ this is a smooth quintic hypersurface in $\mathbb{P}^4$ and as such a Calabi–Yau 3-fold. The family $\{\mathcal{Q}_\psi\}_\psi$ is very famous because it was the starting point of the whole field of "mirror symmetry" (to which we will return in Section 8) in the famous 1991 paper [14] by Candelas, de la Ossa, Green and Parkes, in which its Picard–Fuchs differential equation was related to the problem of counting rational curves on generic quintic hypersurfaces in $\mathbb{P}^4(\mathbb{C})$.

To say that $\mathcal{Q}_\psi$ is a Calabi–Yau 3-fold means that there is a nowhere vanishing holomorphic 3-form Ω on it. Using non-homogeneous coordinates $(w:x:y:z:1)$ rather than $(x_1:\cdots:x_5)$ on $\mathbb{P}^4(\mathbb{C})$, we can give this form explicitly by

$$\Omega = \frac{dx\,dy\,dz}{\partial F/\partial w}\Big|_{F=0} = \frac{dx\,dy\,dz}{xyz-5\psi^{-1}w^4}\Big|_{F=0}, \tag{3.6}$$

where $F = F_\psi(w,x,y,z) = wxyz - \psi^{-1}(w^5+x^5+y^5+z^5+1)$. (A similar formula would apply to any smooth hypersurface $F=0$ of degree n in $\mathbb{P}^{n-1}(\mathbb{C})$, such a hypersurface always being a Calabi–Yau manifold.) To find the corresponding Picard–Fuchs differential equation, we compute the integral of this form over a suitably chosen 3-cycle in $\mathcal{Q}_\psi$ and then, imitating Euler's derivation of the differential equation (2.1), find the differential equation satisfied by this integral. The 3-cycle we choose is the deformed 3-torus given (for $|\psi| > 5$) by

$$T: \quad |x| = 1, \; |y| = 1, \; |z| = 1, \; w = w(x,y,z) \qquad (|\psi| \text{ large}),$$

where $w(x,y,z)$ denotes the "small" solution of the equation $F_\psi(w,x,y,z) = 0$ (i.e., the one given asymptotically by $w \approx (1 + x^5 + y^5 + z^5)/\psi xyz$, as opposed to the four "large" solutions $w \approx i^\nu \sqrt[4]{\psi xyz}$ with $\nu \in \mathbb{Z}/4\mathbb{Z}$). From the Lagrange inversion formula we obtain the two power series expansions

$$w = \sum_{n=0}^{\infty} \frac{\binom{5n}{n}}{4n+1} \frac{(1 + x^5 + y^5 + z^5)^{4n+1}}{(\psi xyz)^{5n+1}},$$

$$\frac{1}{1 - 5w^4/\psi xyz} = \sum_{n=0}^{\infty} \binom{5n}{n} \frac{(1 + x^5 + y^5 + z^5)^{4n}}{(\psi xyz)^{5n}},$$

and combining the second of these with the Cauchy integral formula we obtain

$$\begin{aligned} \frac{1}{(2\pi i)^3} \int_T \Omega &= \frac{1}{(2\pi i)^3} \int_T \frac{1}{1 - 5w^4/\psi xyz} \frac{dx}{x} \frac{dy}{y} \frac{dz}{z} \\ &= \sum_{n=0}^{\infty} \binom{5n}{n} \text{c.t.}\Big(\frac{(1 + x^5 + y^5 + z^5)^{4n}}{(\psi xyz)^{5n}} \Big) \qquad (\text{"c.t." as in (1.11)}) \\ &= \sum_{n=0}^{\infty} \binom{5n}{n} \frac{(4n)!}{n!^4} \psi^{-5n} = \sum_{n=0}^{\infty} \frac{(5n)!}{n!^5} \psi^{-5n}. \end{aligned} \tag{3.7}$$

This formula, in which the sum on the right satisfies a hypergeometric differential equation (see the next subsection), is the celebrated result of Candelas et al.

We make two further comments in connection with this example. First of all, the differential equation satisfied by the function (3.7) has order only 4, whereas the third Betti number of $\mathcal{Q}_\psi$ is the much larger number 204. In fact, there is an obvious abelian group of order 125 (given by multiplying the coordinates x_i in (3.5) by 5th roots of unity), and the invariant part of $H^3(\mathcal{Q}_\psi)$ under the action of this group has dimension 4. In other words, our differential equation corresponds to a "natural piece" of the cohomology group rather than to the whole group, a phenomenon that we already mentioned in the previous section. Such "natural pieces" are precisely Grothendieck's motives, to which we will come back in Section 6.

Secondly, there is the question of the integrality of the monodromy. Here the monodromy group is in fact integral (even though the differential equation corresponds to only part of the full cohomology group) and a complete description of it was given in [14], which also gave the transition matrices between the "Frobenius bases" of solutions of the hypergeometric equation at 0 and ∞ and the integral base. We will discuss this in more detail in Example 2 of Section 7.

Example 3: Hypergeometric functions

An extremely important class of differential equations are the ones satisfied by the Euler–Gauss hypergeometric function (2.2) and its generalization

$$F(a_1,\dots,a_r;b_1,\dots,b_s;t) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_r)_n}{(b_1)_n \cdots (b_s)_n\, n!}\, t^n\,, \tag{3.8}$$

sometimes also denoted ${}_rF_s\left(\begin{matrix}a_1,\dots,a_r\\ b_1,\dots,b_s\end{matrix}\,\middle|\, t\right)$. We will be most interested in the case when $s = r-1$ and all the a's and b's are rational numbers, a typical example being the function appearing in (3.7), which can be written $F(\frac15,\frac25,\frac35,\frac45;1,1,1;(5/\psi)^5)$. Under these assumptions the hypergeometric differential equation is always a Picard–Fuchs equation. (An explicit expression of the function (3.8) in this case as a period integral is given by its classical expression as the integral over $[0,1]^{d-1}$ of a monomial in $x_1,\dots,x_{d-1},1-x_1,\dots,1-x_{d-1}$, and $1-x_1\cdots x_{d-1}t$.) Here the condition $r = s-1$ is equivalent to the condition that the three singular points $x = 0,\,1,\,\infty$ of the differential equation are regular, and the rationality condition says that the local monodromy matrices at these singularities are quasiunipotent.

If we write the expansion (3.8) as $\sum u_n t^n$, then we see from the definition of the "ascending Pochhammer symbol" $(a)_n$ as $a(a+1)\cdots(a+n-1)$ that the coefficients u_n satisfy the two-term recursion

$$b(n)\,u_{n+1} = a(n)\,u_n$$

with polynomials $a(x) = \prod_{i=1}^{r}(x+a_i)$ and $b(x) = \prod_{i=1}^{s+1}(x+b_i)$ $(b_{s+1}=1)$. This translates into the differential equation $\mathcal{L}F = 0$ satisfied by (3.8), where $\mathcal{L}$ is the hypergeometric differential operator $\mathcal{L} = a(D)t - b(D-1)$, with $D = t\,d/dt$ as in (1.8). The fact that the recursion has only two terms corresponds to the fact that the differential equation has only three singular points at 0, 1 and ∞; three-term recursions like the ones (1.1) and (1.2) satisfied by the Apéry sequences correspond to differential equations with four singularities (two of them again at 0 and ∞), and more generally Picard–Fuchs differential equations with singularities at 0, ∞, and ℓ finite points typically correspond to recursions with $\ell+1$ terms (or "length ℓ").

Example 4: Algebraic functions

Any algebraic function, i.e., any function $y(x)$ satisfying a polynomial equation $P(x,y(x)) = 0$ with complex coefficients, satisfies a linear differential equation with polynomial coefficients. Indeed, all derivatives of y belong to the function field $K = \mathbb{C}(x,y)/(P(x,y)=0)$, as we see from the formula $\frac{dy}{dx} = -\frac{\partial P/\partial x}{\partial P/\partial y}$ and induction, and since this field has dimension d over $\mathbb{C}(x)$, where d is the degree of P with respect to y, we see that the derivatives $y, y', \dots, y^{(d)}$ must be linearly dependent over $\mathbb{C}[x]$. This differential equation always has regular singularities and is in fact always a Picard–Fuchs differential equation, because evaluating an algebraic function at a point is just the special case $r = 0$ of integrating an algebraic r-form over an r-dimensional manifold. The monodromy group is given by permutation matrices.

Sometimes hypergeometric functions are algebraic, and these cases are especially interesting. The criterion for an Euler–Gauss hypergeometric function (2.2) to be algebraic was found by Schwarz in the 19th century, and the corresponding criterion for the general case (3.8) by Frits Beukers and Gert Heckman in 1989 [10]. A particularly nice case is that of hypergeometric functions

$$F(t) = F_{\mathbf{c},\mathbf{d}}(t) = \sum_{n=0}^{\infty} \frac{(c_1 n)! \cdots (c_p n)!}{(d_1 n)! \cdots (d_q n)!}\, t^n \tag{3.9}$$

involving only factorials, like (3.7). Here Villegas found that the Beukers-Heckman criterion is equivalent to the three conditions $q = p + 1$, $\sum_i c_i = \sum_j d_j$, and $F_{\mathbf{c},\mathbf{d}}(t) \in \mathbb{Z}[[t]]$. A simple example is the binomial coefficient series

$$B_{M,N}(t) = \sum_{n=0}^{\infty} \binom{Mn}{Nn} t^n = \sum_{n=0}^{\infty} \frac{(Mn)!}{(Nn)!\,((M-N)n)!}\, t^n \tag{3.10}$$

($M \geq N \geq 0$), which we will discuss in §7. Three more complicated examples are

$$\sum_{n=0}^{\infty} \frac{(6n)!\,n!}{(3n)!\,(2n)!^2}\, t^n, \quad \sum_{n=0}^{\infty} \frac{(10n)!\,n!}{(5n)!\,(4n)!\,(2n)!}\, t^n, \quad \sum_{n=0}^{\infty} \frac{(30n)!\,n!}{(15n)!\,(10n)!\,(6n)!}\, t^n. \tag{3.11}$$

The integrality of the coefficients of the last of these series (equivalent to the statement that the periodic and integer-valued function $[30x]+[x]-[15x]-[10x]-[6x]$ is non-negative) is a famous discovery of Chebyshev, who used it to prove the weaker version $c\,\frac{x}{\log x} < \pi(x) < \frac{6c}{5}\,\frac{x}{\log x}$ ($c = \log(30^{-1/30}2^{1/2}3^{1/3}5^{1/5}) = 0.921\cdots$) of the prime number theorem almost half a century before the full theorem was proved by Hadamard and de la Vallée Poussin. The first two of the functions (3.11) have degree 6 and 30 over $\mathbb{Q}(t)$, respectively, whereas the expected degree of the last one, according to Villegas, is a whopping 483840.

4 Differential equations and modular forms

Modular forms are a wonderful mathematical theory because they give an especially clear link between complex analysis and arithmetic. (The same holds also for higher-dimensional modular objects, a prime example being the Hilbert modular surfaces and Hilbert modular forms studied by Hirzebruch.) Specifically, on the one side they have an arithmetic nature associated with words like

Hecke theory, eigenvalues, L-functions, Galois representations, ...

and on the other side an analytic nature associated with words like periods, special values, differential equations, ...

They thus fit particularly well into the subject matter of this paper, and many of our examples – including of course the Apéry numbers with which we started – have a modular nature.

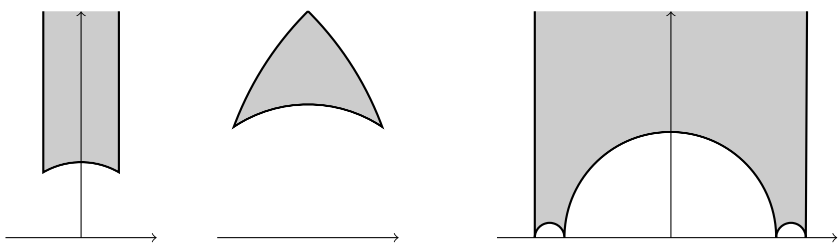
In this section we recall the definitions and main properties of modular forms, with many examples, and say something about their arithmetic side (Hecke theory and L-functions), but only briefly since we will return to this in the next section. Our main goal will be to explain and sketch the proof of the following

Key Fact. A modular form of positive integral weight k, written as a function of a modular function on the same group, satisfies a linear differential equation of order $k+1$ with algebraic coefficients.

We refer to [13] for more details, examples, and applications of all of these topics.

Modular forms and modular functions

We denote by $\mathbb{H}$ the complex upper half-plane $\{\tau \in \mathbb{C} \mid \Im(\tau) > 0\}$, on which the group $SL(2,\mathbb{R})$ acts in the usual way by Möbius transformations: $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\tau = \frac{a\tau+b}{c\tau+d}$. Let Γ be a discrete subgroup of $SL(2,\mathbb{R})$ (Fuchsian group) for which the quotient $\mathbb{H}/\Gamma$ has finite area. We can visualize this quotient by choosing a fundamental domain for this action, as illustrated in the figure (which shows the fundamental domains for the standard modular group $\Gamma_1 = SL(2,\mathbb{Z})$, for a so-called triangle group, and for the non-arithmetic group that will be discussed at the end of this section), but the important object is the quotient $\mathbb{H}/\Gamma$, not the fundamental domain. This quotient is a Riemann surface that is either compact (as in the middle picture) or else can be compactified by the addition of finitely many "cusps" (one in the first picture, three in the third) and then becomes a projective curve $X_\Gamma = \overline{\mathbb{H}/\Gamma}$. A *modular function* on Γ is then a meromorphic function on X_Γ, i.e., a Γ-invariant meromorphic function in $\mathbb{H}$ together with appropriate growth properties at the cusps if $\mathbb{H}/\Gamma$ is non-compact, whereas a *modular form* of (integral) weight k on Γ is a holomorphic function in $\mathbb{H}$ satisfying the more general transformation equation $f\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^k f(\tau)$ for all $\tau \in \mathbb{H}$ and $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, again together with a restriction on the growth of $f(\tau)$ at the cusps if $\mathbb{H}/\Gamma$ is non-compact. This restriction is most easily stated in terms of



the Fourier expansion of f. If Γ has cusps, then without loss of generality we can assume that one of them is at infinity and that the stabilizer of ∞ in Γ is generated by the matrix $T = \left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ (which pictorially means that the part of the fundamental domain above a certain height is a strip of width 1); then $f(\tau)$ has a Fourier expansion $f(\tau) = \sum_{n\in\mathbb{Z}} a(n)e^{2\pi in\tau}$ and the growth assumption is that the Fourier coefficients $a(n) = a_f(n)$ are bounded by a polynomial in n. This (non-obviously) implies that $a(n)$ vanishes for $n < 0$, so we can also think of this Fourier expansion as the Taylor expansion $\sum a(n)q^n$ of f with respect to the local parameter $q = q^{2\pi i\tau}$ of X_Γ at infinity. The modular functions that we consider will also always be holomorphic in $\mathbb{H}$, so they also have expansions $\sum a(n)q^n$, but now possibly with non-zero coefficients $a(n)$ for finitely many negative values of n and with the weaker growth property $a(n) = \mathrm{O}(C^{\sqrt{n}})$ rather than $a(n) = \mathrm{O}(n^C)$.

In the rest of this section we give examples of modular forms and modular functions, concentrating mostly on the full modular group Γ_1. Our first and very important example is the *Dedekind eta-function*

$$\eta(\tau) = q^{1/24} \prod_{n=1}^{\infty}(1-q^n) = q^{1/24}\left(1 - q - q^2 + q^5 + q^7 - \cdots\right). \tag{4.1}$$

This is not quite a modular form as defined above, but instead a *modular form of weight* 1/2 *with non-trivial multiplier system* on Γ_1. We omit the exact definition of these words, but here they mean that $\eta(\tau)$ satisfies the transformation properties $\eta(\tau+1) = e^{\pi i/12}\eta(\tau)$ and $\eta(-1/\tau) = \sqrt{\tau/i}\,\eta(\tau)$ with respect to the generators $T = \left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $S = \left(\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right)$ of Γ_1. It follows that the 24th power

$$\Delta(\tau) = q\prod_{n=1}^{\infty}(1-q^n)^{24} = q - 24q^2 + 252q^3 - \cdots - 6048q^6 + \cdots \tag{4.2}$$

of $\eta(\tau)$, the so-called *discriminant function*, is a true modular form of weight 12. We also have the *weight k Eisenstein series $E_k(\tau)$* on Γ_1 of any even weight $k \geq 4$, the first three of which are given by

$$\begin{aligned}
E_4(\tau) &= 1 + 240\sum_{n=1}^{\infty}\frac{n^3q^n}{1-q^n} = 1 + 240q + 2160q^2 + 6720q^3 + \cdots, \qquad (4.3)\\
E_6(\tau) &= 1 - 504\sum_{n=1}^{\infty}\frac{n^5q^n}{1-q^n} = 1 - 504q - 16632q^2 - 122976q^3 - \cdots,\\
E_8(\tau) &= 1 + 480\sum_{n=1}^{\infty}\frac{n^7q^n}{1-q^n} = 1 + 480q + 61920q^2 + 1050240q^3 + \cdots.
\end{aligned}$$

One shows fairly easily that the ring $M_*(\Gamma_1)$ of all modular forms on Γ_1 is the free algebra on two generators E_4 and E_6, and this immediately implies identities like $E_8 = E_4^2$ or $1728\Delta = E_4^3 - E_6^2$ that would be completely mysterious from an elementary point of view, showing at a very simple level the power of modularity.

We also mention the weight 2 Eisenstein series

$$E_2(\tau) = \frac{\Delta'(\tau)}{\Delta(\tau)} = 1 - 24\sum_{n=1}^{\infty}\frac{n\,q^n}{1-q^n} = 1 - 24q - 72q^2 - 96q^3 - \cdots .$$

(Here and from now on $f'(\tau)$ for a holomorphic function f in $\mathbb{H}$ will denote the derivative of f with respect to $2\pi i\tau$; this is convenient as it preserves the rationality or integrality of the Fourier expansion of f.) It is not a modular form but a so-called *quasimodular form* on Γ_1, meaning that it satisfies the modified transformation property $E_2(\frac{a\tau+b}{c\tau+d}) = (c\tau+d)^2E_2(\tau) - \frac{6i}{\pi}c(c\tau+d)$ for $\tau\in\mathbb{H}$ and $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\Gamma_1$. The ring $\widetilde{M}_*(\Gamma_1)$ of all quasimodular forms (we omit the definition) on Γ_1 is generated freely by E_2, E_4 and E_6 and is closed under differentiation, with $E_2' = \frac{1}{12}(E_2^2 - E_4)$, $E_4' = \frac{1}{3}(E_2E_4 - E_6)$, $E_6' = \frac{1}{2}(E_2E_6 - E_4^2)$ (Ramanujan's formulas).

As the basic example of a modular function we have the *modular j-invariant*

$$j(\tau) = \frac{E_4(\tau)^3}{\Delta(\tau)} = q^{-1} + 744 + 196884\,q + 21493760\,q^2 + \cdots , \tag{4.4}$$

which is invariant under Γ_1 because both E_4^3 and Δ are modular forms of weight 12. This function gives a holomorphic isomorphism between $\mathbb{H}/\Gamma_1$ and $\mathbb{C}$ and between the compactification $X_{\Gamma_1} = \mathbb{H}/\Gamma_1 \cup \{\infty\}$ and $\mathbb{P}^1(\mathbb{C})$. Such a modular function (on any Fuchsian group of genus 0) is called a *Hauptmodul*.

One can also give explicit examples of modular functions and modular forms for other Fuchsian groups Γ. The most important for arithmetic purposes are the so-called "congruence subgroups" of Γ_1 such as the group $\Gamma_0(N)$, which is defined for any integer $N \geq 1$ as the set of matrices $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right) \in \Gamma_1$ with c divisible by N, or the principal congruence subgroup $\Gamma(N)$, defined as the set of matrices in Γ_1 that are congruent to the identity modulo N. For instance, on $\Gamma_0(2)$ we have the modular form of weight 2

$$E_{2,2}(\tau) = 2E_2(2\tau) - E_2(\tau) = 1 + 24q + 24q^2 + 96q^3 + 24q^4 + \cdots , \tag{4.5}$$

and on the principal congruence group $\Gamma(2)$ (whose associated modular curve $X(2) = X_{\Gamma(2)}$ again has genus 0) the Hauptmodul

$$\lambda(\tau) = 16\frac{\eta(\tau/2)^8\eta(2\tau)^{16}}{\eta(\tau)^{24}} = 1 - \frac{\eta(\tau/2)^{16}\eta(2\tau)^8}{\eta(\tau)^{24}} = \left(\frac{\vartheta_2(\tau)}{\vartheta_3(\tau)}\right)^4 . \tag{4.6}$$

Here ϑ_2, ϑ_3 are the Jacobi theta functions

$$\vartheta_2(\tau) = \sum_{n\in\mathbb{Z}} q^{(n+\frac{1}{2})^2/2}, \qquad \vartheta_3(\tau) = \sum_{n\in\mathbb{Z}} q^{n^2/2} \tag{4.7}$$

which are modular forms of weight $1/2$, again with a multiplier system that we omit, for $\Gamma(2)$. Further examples will be given in the next subsection when we discuss Hecke eigenforms.

Modular forms with multiplicative Fourier coefficients

The observant reader may have noticed, as Ramanujan did in 1916, that the coefficient -6048 of q^6 in (4.2) is the product of the coefficients -24 and $+252$ of q^2 and q^3, respectively. This is a special case of the more general property that the coefficients of Δ are *multiplicative*, meaning that the coefficient of q^{mn} is the product of those of q^m and q^n whenever m and n are coprime, as Ramanujan also observed and as Mordell proved the following year. This was later generalized by Hecke to all k by showing that the space $M_k(\Gamma_1)$ is uniquely spanned by modular forms having this same multiplicative property. These are called *Hecke eigenforms* (or simply *Hecke forms*) because the multiplicativity is equivalent to the statement that the form is a simultaneous eigenvector of an infinite collection of operators on $M_k(\Gamma_1)$ called Hecke operators, whose definition we omit. The Eisenstein series $\frac{1}{240}E_4$, $-\frac{1}{504}E_6$ and $\frac{1}{480}E_8$ also have multiplicative Fourier coefficients and are also Hecke eigenforms, but for reasons that we will see in the next section they are much less interesting than Δ, which is a cusp form. (A *cusp form* of weight k is a modular form of weight k that is bounded in the upper half-plane by $\Im(\tau)^{-k/2}$.) Hecke's result shows that the space $S_k(\Gamma_1)$ of weight k cusp forms on Γ_1 has a unique basis consisting of Hecke forms, the next example after Δ being the form

$$\Delta_{16}(\tau) = E_4(\tau)\,\Delta(\tau) = q + 216q^2 - 3348q^3 \cdots - 723168q^6 + \cdots.$$

of weight 16. Here, as for $k = 12$, the coefficients of the eigenform belong to $\mathbb{Z}$, but we should warn the reader that for weight 24 and all weights $k \geq 28$ the space of cusp forms on Γ_1 has dimension $d > 1$ and the Hecke cusp forms may (and conjecturally always do) have coefficients in a number field of degree d over $\mathbb{Q}$ rather than in $\mathbb{Q}$ itself. For instance, the two Hecke cusp forms of weight 24 have coefficients in $\mathbb{Q}(\sqrt{144169})$, as Hecke himself showed.

A similar, though more complicated, statement ("theory of newforms") is true also for the congruence subgroups $\Gamma_0(N)$. In particular, if the space $S_k(\Gamma_0(N))$ of cusp forms of weight k on $\Gamma_0(N)$ happens to be 1-dimensional, then its unique generator is (if properly normalized) automatically a Hecke form, which we will then denote by $f_{k,N}$. Three cases besides Ramanujan's original example $f_{12,1}(\tau) = \Delta(\tau) = \eta^{24}(\tau)$ where this happens, and a fourth eigenform with integral coefficients, are the cusp forms

$$f_{2,11}(\tau) = \eta(\tau)^2\,\eta(11\tau)^2 = q - 2q^2 - q^3 + 2q^4 + q^5 + \cdots, \tag{4.8}$$

$$f_{4,9}(\tau) = \eta(3\tau)^8 = q - 8q^4 + 20q^7 - 70q^{13} + 64q^{16} + \cdots, \tag{4.9}$$

$$f_{4,8}(\tau) = \eta(2\tau)^4\,\eta(4\tau)^4 = q - 4q^3 - 2q^5 + 24q^7 - 11q^9 - \cdots, \tag{4.10}$$

$$f_{4,25}(\tau) = \eta(5\tau)^4 \sum_{i=0}^{4} a_i\,\eta(\tau)^{4-i}\,\eta(25\tau)^i = q + q^2 + 7q^3 - 7q^4 + \cdots, \tag{4.11}$$

where in the last line $a = (1, 5, 20, 25, 25)$. Each of these will reappear later in this paper. Note that each of these forms has been expressed in terms of the Dedekind eta-function, but for most Hecke forms this is not possible.

Hecke eigenforms (in particular, cuspidal Hecke eigenforms) are the most important objects in the theory of modular forms from the arithmetic point of view, and the whole modern theory of automorphic forms via the representation theory of adelic groups (Jacquet-Langlands theory) can be seen as a vast generalization of the theory of Hecke operators. We mention here two fundamental (and related) properties of Hecke eigenforms, both of which will play a role later. The first is that the multiplicativity of the Fourier coefficients $a_n = a_n(f)$ of a Hecke form f translates into the fact that the corresponding L-series $L(f,s) = \sum_{n\ge1} a_n(f)n^{-s}$ has an Euler product, and in fact a strengthening of this multiplicativity property (also already observed by Ramanujan for the coefficients of Δ in 1916 and proved by Mordell in 1917) says that this Euler product has only quadratic factors and more specifically is of the form

$$L(f,s) = \prod_{p \text{ prime}} \frac{1}{1 - a_f(p)\, p^{-s} + \chi(p)\, p^{k-1-2s}} \tag{4.12}$$

for some Dirichlet character χ. The other is that a Hecke form has two associated "periods" $\omega_\pm(f)$ such that the value of $L(f,s)$ at any integral argument s between 0 and the weight k is an algebraic multiple of a power of π times either $\omega_+(f)$ or $\omega_-(f)$. (This statement makes sense because the analytic properties of cusp forms imply that the L-function $L(f,s)$ of any cusp form f has an analytic continuation to all s, as well as satisfying a functional equation, so that we can talk about its values even for arguments outside of the domain of absolute convergence of its defining series or Euler product representation.) These are typical properties of motivic L-functions, as we will see in Section 6, but there they are only conjectural in general, whereas in the case of modular forms they are theorems.

Modular forms satisfy differential equations

Modular forms and quasimodular forms always satisfy non-linear differential equations of order 3 with constant coefficients as functions of the variable τ in the upper half-plane. The reason is simply that the ring of all quasimodular forms on any Fuchsian group Γ has transcendence degree 3 (for instance, for Γ_1 it is the free algebra on E_2, E_4 and E_6) and is closed under differentiation, so if f belongs to this ring then there must be a polynomial relation among f, f', f'' and f'''. (A nice example is the *Chazy equation* $E_2''' - E_2E_2'' + \frac{3}{2}{E_2'}^2 = 0$ satisfied by $E_2(\tau)$.) However, these non-linear differential equations do not have many applications in arithmetic precisely because they are non-linear.[3] Much more useful – and surprisingly little known, even to specialists, although it was the starting point for the whole theory in its early years in

3. An exception is a theorem proved by Villegas and myself [71] which says that, modulo the Birch–Swinnerton-Dyer conjecture, a prime p of the form $9m+1$ is a sum of two rational cubes if and only if $p|A_{3m}$, where $\sum A_n x^n/n!$ is the Taylor expansion (in suitable coordinates) of $\eta(\tau)$ at the point $\tau = e^{2\pi i/3}$ in $\mathbb{H}$. The non-linear differential equation of the power series is

the late 19th and early 20th centuries – is the "key fact" stated that at the beginning of the section, which says that a modular form of positive integral weight satisfies a *linear* differential equation (of order one greater than the weight) if it is expressed in terms of a modular function of τ rather than in terms of τ itself.

So let $t(\tau)$ be a (meromorphic) modular function for some Γ and $f(\tau)$ a (meromorphic) modular form of positive integral weight k on the same group. We want to show that if we express $f(\tau)$ as $\varphi(t(\tau))$, then the function $\varphi(t)$ satisfies a linear differential equation of order $k+1$ with algebraic coefficients. But of course we *cannot* express f this way globally, since $t(\tau)$ is invariant under substitutions $\tau \mapsto \gamma\tau$ with $\gamma \in \Gamma$ while $f(\tau)$ is not. However, we can do so locally, say in a small neighborhood of a cusp or of a point of $\mathbb{H}$, and then the many-valuedness of φ when we leave this neighborhood and then come back to it is precisely the monodromy representation that we want from the solution of a linear differential equation. This observation is in fact the key to the proof of our assertion, because the $k+1$ linearly independent functions $\tau^i f(\tau)$ $(i = 0, \ldots, k)$ form a basis for the space of solutions of the differential equation satisfied by φ.

More explicitly, let $\vec{f} : \mathbb{H} \to \mathbb{C}$ be the (column) vector-valued function with entries $\tau^i f(\tau)$ (in reverse order). From the equation $f(\gamma\tau) = (c\tau+d)^k f(\tau)$ we get

$$\vec{f}(\gamma\tau) = \begin{pmatrix} (a\tau+b)^k f(\tau) \\ \vdots \\ (c\tau+d)^k f(\tau) \end{pmatrix} = \begin{pmatrix} a^k & \cdots & b^k \\ & \ddots & \\ c^k & \cdots & d^k \end{pmatrix} \begin{pmatrix} \tau^k f(\tau) \\ \vdots \\ f(\tau) \end{pmatrix} = S^k(\gamma)\,\vec{f}(\tau)$$

for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, where $S^k(\gamma) \in GL(k+1, C)$ is the kth symmetric power of γ. The point is that this matrix is independent of τ, so we can differentiate the equation $\vec{f} \circ \gamma = S^k(\gamma)\vec{f}$ without obtaining any extra terms, as we would have if we simply differentiated $(c\tau+d)^k f(\tau)$. However, the derivative of $\gamma\tau$ is $(c\tau+d)^{-2}$, so the differentiated equation would have the form $\vec{f}' \circ \gamma = (c\tau+d)^2\, S^k(\gamma)\vec{f}'$, which again contains a multiplicative factor depending on τ. Since we want to go to higher derivatives, we cannot iterate this procedure. But we actually want to differentiate with respect to $t = t(\tau)$, and since $d_t := d/dt$ equals $t'(\tau)^{-1} d/d\tau$ and $t'(\tau)$ is a (possibly meromorphic) modular form of weight 2 and therefore also acquires a factor $(c\tau+d)^2$ when we replace τ by $\gamma\tau$, everything is all right after all: we have $(d_t\vec{f}) \circ \gamma = S^k(\gamma)(d_t\vec{f})$ and by induction $(d_t^i\vec{f}) \circ \gamma = S^k(\gamma)(d_t^i\vec{f})$ for all $i \geq 0$. Writing down these equations for $i = 0, 1, \ldots, k+1$ and noting that $k+2$ vectors of length $k+1$ are linearly dependent, we get an identity of the form $\sum_{i=0}^{k+1} m_i\, d_t^i \vec{f} = 0$ where each $m_i = m_i(\tau)$ is the determinant of a $(k+1) \times (k+1)$ matrix that transforms under $\tau \mapsto \gamma\tau$ by multiplication on the left by $S^k(\gamma)$, so that its determinant m_i is a mod-

---

then equivalent to a non-linear recursion for the coefficients A_n, but even here there is a more convenient method of calculating these numbers based on a different parametrization using the expansions of integral weight modular forms as hypergeometric series with respect to a Hauptmodul.

ular function of τ. Since every such function is an algebraic function of the chosen modular function $t(\tau)$, we obtain our desired differential equation. (A more detailed exposition of this proof, and two other elementary proofs of the "key fact", are given on pages 61–62 of [13].)

Note that this differential equation has algebraic coefficients of t in general, but if $\mathbb{H}/\Gamma$ has genus 0 and $t(\tau)$ is a Hauptmodul, then every modular function on Γ is a rational function of t and therefore the equation, after multiplying through by a common denominator, in fact has polynomial coefficients, implying that we have a linear recursion for the coefficients of any power series solution. We also see from the proof that the monodromy group is the kth symmetric power $S^k(\Gamma) \subset GL(k+1,\mathbb{C})$ of the original Fuchsian group Γ and in particular is integral if Γ is a subgroup of the full modular group Γ_1. On the other hand, in the last section we saw that Picard–Fuchs differential equations also have integral (or nearly integral) monodromy representations. This is not a coincidence, since the differential equations associated to subgroups of $SL(2,\mathbb{Z})$ are the Picard–Fuchs equations associated to families of elliptic curves. We will see explicitly how this works in Example 1 of the next section.

The "key fact" is illustrated by a classical result of H.A. Schwarz (see[72], Chapter III) that the monodromy group of the hypergeometric differential equation of the function (2.2) for special values of the parameters a, b and c (with each of $1-c$, $c-a-b$ and $a-b$ equal to 0 or to the reciprocal of an integer) is a triangle group in $SL(2,\mathbb{R})$, i.e., a Fuchsian group Γ whose fundamental domain is either a hyperbolic triangle as pictured at the beginning of this section or else a union of two such triangles. In this case one can take the parameter t in (2.2) to be a modular function $t(\tau)$ (in fact, a Hauptmodul) for Γ, the function $f(\tau) = F(a,b;c;t(\tau))$ to be a modular form of weight 1 on Γ, and the second solution of the differential equation to be $\tau f(\tau)$, the (many-valued) map $t \mapsto \tau$ formed by the quotient of these two solutions then being the classical Schwarz map. We will give many further examples, of a more arithmetic nature, in the next section.

5 Examples

In this section we illustrate the theorem discussed above by describing a number of triples "Fuchsian group – modular form – modular function" and the associated differential equations.

Example 1: Jacobi theta series

Our first example is classical. We take as our Fuchsian group the principal congruence subgroup $\Gamma = \Gamma(2)$ of Γ_1, as our modular form f the square of the Jacobi theta function ϑ_3 defined in (4.7), and as our modular function t the Hauptmodul λ defined in (4.6). Here $k = 1$, so $\vartheta_3(\tau)^2$ should satisfy a differential equation of order 2 with respect to $\lambda(\tau)$, and indeed

$$\vartheta_3(\tau)^2 = \sum_{k=0}^{\infty} \binom{-1/2}{k}^2 \lambda(\tau)^k \tag{5.1}$$

for $|\lambda(\tau)| \le 1$. This is the same hypergeometric function $F(\frac{1}{2}, \frac{1}{2}; 1; t)$ that we saw in Section 3 when computing the Picard–Fuchs differential equation of the Legendre elliptic curve (3.1), and indeed the $\mathbb{Z}$-basis $(\int_{C_0} \omega_t, \int_{C_1} \omega_t)$ of solutions that we computed in equations (3.4) and (3.3) is (up to a factor 2π) the same as the $\mathbb{Z}$-basis $(\tau f(\tau), f(\tau))$ coming from the proof in the last section, as we can check numerically by inverting the q-expansion of $\lambda(\tau)$ and taking logarithms:

$$\begin{aligned}
t = \lambda(\tau) &= 16\,q^{1/2}\,(1 - 8\,q^{1/2} + 44\,q - 192\,q^{3/2} + \cdots),\\
q^{1/2} &= \frac{t}{16}\Big(1 + \frac{t}{2} + \frac{21\,t^2}{64} + \frac{31\,t^3}{128} + \cdots\Big),\\
\pi i \tau &= \log(q^{1/2}) = \log\Big(\frac{t}{16}\Big) + \frac{t}{2} + \frac{13\,t^2}{64} + \frac{23\,t^3}{192} + \cdots\\
&= \log\Big(\frac{t}{16}\Big) + \frac{4\sum_{n=1}^{\infty}\binom{-1/2}{n}^2 \big(\frac{1}{n+1} + \cdots + \frac{1}{2n}\big)\,t^n}{\sum_{n=0}^{\infty}\binom{-1/2}{n}^2 t^n} = \pi i\,\frac{\int_{C_0}\omega_t}{\int_{C_1}\omega_t}\,.
\end{aligned}$$

Example 2: The function E_4

Another classical example, this time for the full modular group Γ_1, is an identity of Fricke and Klein expressing the Eisenstein series (4.3) in terms of the modular invariant (4.4) as

$$E_4(\tau) = 1 + \frac{60}{j(\tau)} + \frac{39780}{j(\tau)^2} + \cdots = F\Big(\frac{1}{12}, \frac{5}{12}; 1; \frac{1728}{j(\tau)}\Big)^4,$$

where $F(a, b; c; x)$ once again denotes the hypergeometric function. This is indeed the solution of a differential equation of order five (one more than the weight of E_4), since the fourth symmetric power of a two-dimensional space is five-dimensional, but in fact we see in this example that the 4th root $\sqrt[4]{E_4(\tau)}$ satisfies a second order differential equation, even though it is not a modular form (or even a holomorphic function in the upper half-plane) and therefore is not strictly covered by the statement of the "key fact" as given in the previous section. Actually this behavior is generic: if $f(\tau)$ is any modular form of integral weight $k > 0$, then its kth root, expressed in terms of a modular function on the same group, always satisfies a differential equation of order 2, and the equation of order $k + 1$ satisfied by f itself is the one derived from this by the kth symmetric power operation.

As a related example we mention the expansion

$$t(\tau) = \frac{1}{864}\Big(1 - \frac{E_6(\tau)}{E_4(\tau)^{3/2}}\Big) \quad \Rightarrow \quad \sqrt[4]{E_4(\tau)} = \sum_{n=0}^{\infty} \frac{(6n)!}{(3n)!(2n)!n!}\, t(\tau)^n, \tag{5.2}$$

in which the coefficients $\frac{(6n)!}{(3n)!(2n)!n!}$ are a solution of the Apéry-like recursion (1.3) with $(A,B,C) = (432,60,0)$; this is another one of the "miraculous" cases when this recursion has an integral solution, but is not contained in the list (1.4) because there we excluded "degenerate" (= hypergeometric!) cases where $C = 0$.

Example 3: The Apéry numbers

Here we take

$$t(\tau) = \frac{\eta(\tau)^{12}\,\eta(6\tau)^{12}}{\eta(2\tau)^{12}\,\eta(3\tau)^{12}}, \qquad f(\tau) = \frac{\eta(2\tau)^{7}\,\eta(3\tau)^{7}}{\eta(\tau)^{5}\,\eta(6\tau)^{5}}, \tag{5.3}$$

which are a modular function and a modular form of weight 2, respectively, on the congruence group $\Gamma_0(6)$. Then $f(\tau) = \mathcal{A}(t(\tau))$, where $\mathcal{A}(t) = 1 + 5t + 73t^2 + \cdots$ is the generating function for the Apéry numbers for $\zeta(3)$ as discussed in Section 1. This is just Beukers's identity (1.7), but now written in standard modular notation.

Examples 4–18: Apéry-like numbers

The generating series of the Apéry numbers for $\zeta(2)$ defined by his recursion (1.2) also has a modular parametrization, again due to Beukers: here we take for Γ the group $\Gamma_1(5)$ of matrices congruent to $\left(\begin{smallmatrix}1 & *\\ 0 & 1\end{smallmatrix}\right)$ modulo 5, for $t(\tau)$ the 5th power of Ramanujan's modular function

$$r(\tau) = q^{1/5}\prod_{n=1}^{\infty}(1-q^n)^{\left(\frac{n}{5}\right)} = \cfrac{q^{1/5}}{1+\cfrac{q}{1+\cfrac{q^2}{1+\cdots}}}$$

(in which the equality of the two expressions on the right, without the factor $q^{1/5}$, is sometimes called "the most beautiful formula in mathematics"), which is a Hauptmodul for $\Gamma_1(5)$, and for $f(\tau)$ the function $\eta(5\tau)^{5/2}/t(\tau)^{1/2}\eta(\tau)^{1/2}$, which is a modular form (Eisenstein series) of weight 1 on the same group.

There are similar modular parametrizations for the other six cases listed in (1.4) when the recursion (1.3) has integral solutions. All are given explicitly in [74]. (See also [69].) Apart from these seven, there exist eight further "degenerate" triples (A,B,C) (up to rescaling) for which the generating series of the solution of (1.3) has a modular parametrization: four "hypergeometric" cases with $C = 0$ and $(A,B) = (16,4)$, $(27,6)$, $(64,12)$ or $(432,60)$ and four "Legendrian" ones with $C = A^2/4$ and $(A,B) = (32,12)$, $(54,21)$, $(128,52)$ or $(864,372)$. The modular parametrization for the case $(A,B,C) = (432,60,0)$ was already given in (5.2). Another example, corresponding to $(A,B,C) = (27,6,0)$, is the differential equation satisfied by the generating series $\sum \frac{(3n)!}{n!^3}\,t^n$: on the one hand this is the Picard–Fuchs differential equation of the family of plane cubic curves $x_1^3 + x_2^3 + x_3^3 = \psi x_1x_2x_3$ (with $t = (3\psi)^{-3}$) by a calculation identical to the one given in Example 2 of Section 3 but with "5" replaced by "3" everywhere, and on the other hand the series can be parametrized by the modular function t and modular form f of weight 1 given by

$$t(\tau) = \frac{\eta(3\tau)^{12}}{\eta(\tau)^{12} + 27\,\eta(3\tau)^{12}}, \qquad f(\tau) = \sum_{a,\,b\in\mathbb{Z}} q^{a^2+ab+b^2}.$$

Examples 19–34: Mirrors of Fano 3-folds

In connection with the predictions of mirror symmetry, Vasily Golyshev found a specific collection of 17 families of K3 surfaces, corresponding to the 17 smooth Fano 3-folds of Picard rank 1 (details will be given in Section 8), each of which has a period with Taylor coefficients given by a formula of the form (1.11) for some Laurent polynomial L and each of which admits a modular parametrization. One of these families is the Beukers-Peters family (1.9) with the modular parametrization (5.3) (which is why we numbered this subsection "Examples 19–34" rather than "Examples 19–35"!). Another is the family of quartic surfaces $x_1^4 + x_2^4 + x_3^4 + x_4^4 = \psi x_1x_2x_3x_4$, whose period is given by the hypergeometric series $\sum \frac{(4n)!}{n!^4}\, t^n$ (with $t = (4\psi)^{-4}$) by the same calculation as the ones for the corresponding families with "4" replaced by "5" or "3" and has a modular parametrization given by taking $t(\tau) = \frac{\Delta(\tau)\Delta(2\tau)}{(\Delta(\tau)+64\Delta(2\tau))^2}$ and $f(\tau)$ the Eisenstein series of weight 2 on $\Gamma_0(2)$ defined in (4.5).

Example 35: An integrality enigma

Our last example is of a somewhat different nature, and serves both as an illustration of the arithmetic subtleties involved in the relation between modular forms and differential equations and as our first specific application of the mathematics of Friedrich Hirzebruch.

In connection with the theory of Teichmüller curves, Irene Bouw and Martin-Möller [12] studied a Picard–Fuchs differential equation associated to a specific family of genus 2 curves over the projective line defined over the real quadratic field $\mathbb{Q}[\sqrt{17}]$ and showed that its power series solution, which begins

$$\varphi(t) = 1 + \tfrac{81-15\sqrt{17}}{16}\, t + \tfrac{4845-1155\sqrt{17}}{64}\, t^2 + \tfrac{3200225-775495\sqrt{17}}{2048}\, t^3 + \cdots,$$

has integral coefficients (apart from a power of 2 in the denominators that can be removed by rescaling t), even though the recursion defining these coefficients (which here has length 3 rather than 2 as in the Apéry cases) begins $(n+1)^2 u_{n+1} = \cdots$, so that a priori the denominator of the nth coefficient could be as large as $n!^2$. This example, which was taken up again by Möller and myself in [55], has the mysterious property that it *has* a modular parametrization, for a specific cofinite Fuchsian group Γ_{17} (the one whose fundamental domain is depicted in the third of the pictures at the beginning of Section 4), but that this parametrization, unlike the parametrization (1.7) used by Beukers to explain the integrality of the Apéry numbers, does *not* imply the corresponding integrality here in any obvious way. (The proof of integrality given by Bouw and Möller was completely different and used p-adic techniques.) The reason is that the group Γ_{17} is not an arithmetic one and therefore, although the power series φ can be parametrized as $\varphi(t(\tau)) = f(\tau)$ for an explicitly computable Hauptmodul t and modular form f of weight 1, the q-expansions of t and f do not have integral coefficients and hence the argument that implied the integrality of the A_n as a consequence of the parametrization (5.3) fails. In fact, the q-expansions of t and f do not even have algebraic coefficients, but instead belong to the power series ring $\mathbb{Q}[\sqrt{17}][[Aq]]$ for some real constant $A = -7.483708229911735369\cdots$ that we eventually recognized (first numerically and then theoretically) as $-2\,(3+\sqrt{17})\,(\frac{5-\sqrt{17}}{2})^{(\sqrt{17}-1)/4}$, which is transcendental by the Gelfond-Schneider theorem. What's more, even the power series in Aq, although its coefficients are now algebraic, has infinitely many primes in its denominators. The solution of the mystery turned out to be that one had to embed the base curve of the Bouw-Möller family (Teichmüller curve) into the Hilbert modular surface for $\mathbb{Q}[\sqrt{17}]$, and thus also to embed the non-arithmetic group Γ_{17} into the corresponding Hilbert modular group, which is arithmetic. The q-expansions of t and f can then be understood using Hirzebruch's description of the geometry of Hilbert modular surfaces near their cusps [38], and the integrality follows.

6 Differential equations and arithmetic

One of the great developments in 20th century mathematics was the discovery by Artin, Weil, Dwork, Grothendieck, Deligne and many other mathematicians of deep links between number theory and topology. This connection starts with a relation between counting solutions of algebraic equations in finite fields and the topology of the corresponding algebraic variety over $\mathbb{C}$, but then extends to many further topics like global L-functions, variations of Hodge structures and periods, etc. that give rise to the interconnections referred to in the title of this paper.

From point counting to cohomology

We begin by giving a very abbreviated account (omitting all technicalities and occasionally oversimplified) of the passage from point-counting to topology. Let X be a smooth projective variety defined over $\mathbb{Q}$. Such a variety is given as a subspace of some projective space by equations with rational coefficients. If we multiply by a suitable integer, we can take the defining equations to have coefficients in $\mathbb{Z}$ and can then reduce them modulo any prime p, leading to a variety $X_p = X/\mathbb{F}_p$ defined over the finite field $\mathbb{F}_p$ that is again smooth for all but finitely many p. (The remaining "bad" primes will be ignored in our simplified discussion here.) We then have as a basic invariant the number $\#X_p(\mathbb{F}_p)$ of solutions of the defining equations with the variables taking their values in the field of p elements, or more generally the number $\#X_p(\mathbb{F}_{p^n})$ of solutions over the finite field $\mathbb{F}_{p^n}$ for any integer $n \geq 1$. These numbers for a given p can be put together in the form of the *local zeta-function*

$$Z(X/\mathbb{F}_p, T) = \exp\left(\sum_{n=1}^{\infty} \#X_p(\mathbb{F}_{p^n})\, \frac{T^n}{n}\right). \tag{6.1}$$

As the three simplest examples, for X equal to a point, the projective line, or an elliptic curve one finds that

$$Z(\{\text{pt.}\}/\mathbb{F}_p, T) = \frac{1}{1-T}, \qquad Z(\mathbb{P}^1/\mathbb{F}_p, T) = \frac{1}{(1-T)(1-pT)},$$
$$Z(E/\mathbb{F}_p, T) = \frac{1 - a_p(E)T + pT^2}{(1-T)(1-pT)} \qquad (a_p(E) \in \mathbb{Z}),$$

where in the last case $a_p(E)$ can be calculated in terms of Legendre symbols by

$$a_p(E) = p + 1 - \#E_p(\mathbb{F}_p) = -\sum_{x \in \mathbb{Z}/p\mathbb{Z}} \left(\frac{x^3 + Ax + B}{p}\right) \tag{6.2}$$

if $p \neq 2$ and E is given by the Weierstrass equation $y^2 = x^3 + Ax + B$ with A, $B \in \mathbb{Z}$, as one sees by noting that the number of solutions of $y^2 \equiv N \pmod{p}$ equals $1 + \left(\frac{N}{p}\right)$ for any $N \in \mathbb{Z}$. More generally, the right-hand side of (6.1), which *a priori* is just a power series in T with coefficients in $\mathbb{Q}$, is in fact a rational function and has integral coefficients, as was proved in Emil Artin's thesis (1923) for some hyperelliptic curves, by F.K. Schmidt in 1931 for arbitrary smooth curves, and by Dwork in the 50's for varieties of arbitrary dimension. For instance, if X is a curve of genus g then $Z(X/\mathbb{F}_p, T)$ has the form $\frac{P(T)}{(1-T)(1-pT)}$ where $P(T) \in \mathbb{Z}[T]$ is a polynomial of degree $2g$ with all roots of absolute value $p^{-1/2}$, as was proved by Deuring and Hasse for elliptic curves and by André Weil in 1949 for curves of arbitrary genus. We can write this expression in the form $\frac{P_1(T)}{P_0(T)P_2(T)}$ where $P_i(T) \in \mathbb{Z}[T]$ is a polynomial of degree $b_i(X)$, the ith Betti number of X (equal to 1, $2g$, 1 for $i = 0, 1, 2$) and where the roots of $P_i(T)$ have absolute value $p^{-i/2}$, and this and other examples led Weil to conjecture that in general the local zeta function of a d-dimensional smooth projective variety should have the form

$$Z(X/\mathbb{F}_p, T) = \prod_{i=0}^{2d} P_i(X/\mathbb{F}_p, T)^{(-1)^{i-1}} \tag{6.3}$$

where each $P_i(X/\mathbb{F}_p, T)$ is a polynomial of degree $b_i(X)$ with integral coefficients and with all roots of absolute value $p^{-i/2}$ ("Riemann hypothesis"), and further that it should be possible to prove this by finding an appropriate cohomology theory that could apply to the variety X_p over the finite field $\mathbb{F}_p$. These statements imply a very deep link between arithmetic and topology, including in particular the statement that the dimensions of all the homology and cohomology groups of the complex manifold $X(\mathbb{C})$ can be read off from the cardinalities of the finite sets $X(\mathbb{F}_{p^n})$ for even a single (good) prime p.

Weil's proposal for a cohomology theory for varieties over $\mathbb{F}_p$ was realized through the work of Alexander Grothendieck, Michael Artin and others, except that there was not just one such cohomology theory, but infinitely many. More specifically, for a smooth projective variety X defined over any field K and for any integer n prime to the characteristic of K one can define finite "(geometric) étale cohomology groups" $H^i_{\text{ét}}(\overline{X};\mathbb{Z}/n\mathbb{Z})$ via the étale coverings of $\overline{X}$ of degree n (étale coverings, the analogue in the algebraic context of unramified coverings in topology, are maps between algebraic varieties that induce an isomorphism of tangent spaces at every point), where $\overline{X}$ (usually denoted $X \otimes_K \overline{K}$) means X thought of as a variety over $\overline{K}$ together with the natural action of $\mathrm{Gal}(\overline{K}/K)$. The ℓ-adic cohomology group $H^i(\overline{X};\mathbb{Q}_\ell)$ is then defined for any prime ℓ different from the characteristic of K as the inverse limit of $H^i_{\text{ét}}(\overline{X};\mathbb{Z}/\ell^k\mathbb{Z})$ as $k \to \infty$, tensored with $\mathbb{Q}_\ell$, the field of ℓ-adic numbers, and by construction carries an action of the Galois group of $\overline{K}$ over K. In the case $K = \mathbb{F}_p$ this Galois group contains (and is topologically generated by) the Frobenius element Fr_p, defined on $\overline{\mathbb{F}}_p$ by the formula $x \mapsto x^p$ and on $X_p(\overline{\mathbb{F}}_p)$ by applying the same formula to every coordinate. The fixed points of the nth power of Fr_p on $X_p(\overline{\mathbb{F}}_p)$ are precisely the points of X over the finite field $\mathbb{F}_{p^n}$, and then by the analogue of the classical Lefschetz trace formula, proved in this context by Grothendieck, one gets

$$\# X_p(\mathbb{F}_{p^n}) = \# X_p(\overline{\mathbb{F}}_p)^{\mathrm{Fr}_p^n} = \sum_{i=0}^{2d} (-1)^i \operatorname{tr}\big((\mathrm{Fr}_p^*)^n,\, H^i(\overline{X_p};\mathbb{Q}_\ell)\big)\,.$$

This formula translates after a short calculation into the formula (6.3), but with the polynomial $P_i(X/\mathbb{F}_p, T)$ replaced by the characteristic polynomial

$$P_{i,\ell}(X/\mathbb{F}_p, T) = \det\big(1 - \mathrm{Fr}_p^*\, T,\, H^i(\overline{X_p}, \mathbb{Q}_\ell)\big)$$

of the action of Frobenius on the ℓ-adic cohomology. This is *a priori* a polynomial with coefficients in $\mathbb{Q}_\ell$ and depending on ℓ, but in fact it belongs to $\mathbb{Z}[T]$ and is independent of ℓ, as follows (in this case, for smooth projective varieties – for general varieties it is still only conjectural!) from Deligne's proof of the "Riemann hypothesis" part of Weil's conjectures, because if the ith polynomial in the alternating product in (6.3) has roots of absolute value $p^{-i/2}$ then there can be no cancellation among the factors and we can read off each P_i separately from the left-hand side, which does not depend on ℓ.

Global zeta functions

The above considerations were all local and would have applied to any smooth projective variety defined over $\mathbb{F}_p$, not just to the reduction $X_p = X/\mathbb{F}_p$ of a "global" variety X defined over $\mathbb{Q}$. In the global situation the ℓ-adic cohomology groups $H^i(\overline{X};\mathbb{Q}_\ell)$ are isomorphic as $\mathbb{Q}_\ell$-vector spaces to $H^i(X(\mathbb{C});\mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ for all ℓ ("comparison theorem") and also to $H^i(\overline{X_p};\mathbb{Q}_\ell)$ for all (good) primes $p \neq \ell$ ("base change"). For such primes the element Fr_p corresponds to a well-defined conjugacy class (also denoted Fr_p) in the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on $H^i(\overline{X};\mathbb{Q}_\ell)$ whose characteristic polynomial is equal to $P_i(X/\mathbb{F}_p, T)$. The facts that all of the local zeta functions $Z(X/\mathbb{F}_p, T)$ come from the same variety $X/\mathbb{Q}$ and that all of the polynomials $P(X/\mathbb{F}_p, T)$ come from a single cohomology group $H^i(\overline{X}; \mathbb{Q}_\ell)$ are reflected in the expected analytic properties of the global zeta and L-functions associated to X. Specifically, one defines the *Hasse-Weil zeta function* of X by the formula

$$\zeta(X/\mathbb{Q}, s) = \prod_p Z(X/\mathbb{F}_p,\, p^{-s}) \qquad (\Re(s) \gg 0),$$

which by virtue of (6.3) is the alternating product of the *global L-functions*

$$L_i(X/\mathbb{Q}, s) = \prod_p P_i(X/\mathbb{F}_p,\, p^{-s})^{-1} \qquad (\Re(s) \gg 0),$$

where in both cases the product is over all primes p but the description of the Euler factors at "bad" or ramified primes (those where the reduction of X modulo p is no longer a smooth variety over $\overline{\mathbb{F}}_p$) is different from the one for "good" or unramified primes. (For instance, the degree of $P_i(X/\mathbb{F}_p, T)$ in T for bad primes is strictly less than the ith Betti number of X.) The basic conjecture, which can perhaps be considered the single most outstanding open problem in arithmetic algebraic geometry, is that the L-function $L_i(X/\mathbb{Q}, s)$, initially defined only for s in some right half-plane, has a holomorphic (or sometimes just meromorphic, but with specified and very simple pole behavior) continuation to all complex values of s and satisfies a functional equation with respect to the symmetry $s \mapsto i + 1 - s$ in which the necessary gamma-factor is given by an explicit recipe (due to Serre) involving the Hodge decomposition of $H^i(X(\mathbb{C}); \mathbb{C})$. The function $L_i(X/\mathbb{Q}, s)$ is also expected to satisfy a Riemann hypothesis (all zeros on the line $\Re(s) = \frac{i+1}{2}$), but this is not known even if X is a point and $i = 0$, when it reduces to the usual Riemann hypothesis. Finally, there is a marvelous conjecture of Deligne [23] relating the special values of $L_i(X/\mathbb{Q}, s)$ at "critical" values of s (those integers for which neither s nor $i + 1 - s$ is a pole of the gamma-product occurring in the functional equation of L_i) to the periods of X obtained by comparing the Betti and de Rham rational subspaces of $H^i(X(\mathbb{C}); \mathbb{C})$.

For points, projective spaces, Grassmannians and a few other varieties ("Tate motives"), the Hasse–Weil zeta functions are multiplicative combinations of shifted Riemann zeta functions and the meromorphic continuation and functional equation are therefore known, but in almost all other cases they are conjectural. In 1955 Taniyama observed that the Euler factors of the L-functions $L_1(E/\mathbb{Q}, s)$ of elliptic curves E over $\mathbb{Q}$ and the L-functions (4.12) of Hecke cusp forms of weight 2 had exactly the same form, and raised the question whether they sometimes coincided, and in 1969 Weil showed that the expected analytic properties of Hasse–Weil zeta functions would imply that the L-function of any elliptic curve over $\mathbb{Q}$ is in fact equal to the L-function of a weight 2 cusp form. The conjecture that this aways holds became particularly famous after it was shown by the work of Frey, Serre and Ribet that it would imply Fermat's Last Theorem, but it remained open until the spectacularly difficult proof by Andrew Wiles in 1994 of the special cases needed for the Fermat theorem and its extension to the general case by Breuil, Conrad, Diamond and Taylor during the next few years. The Taniyama–Weil conjecture is now a special case of the "Langlands program," which predicts that the L-functions coming from algebraic varieties will always coincide with L-functions of appropriate automorphic forms, from which the desired analytic properties would follow. But this is only known in isolated cases, and the analytic continuation and functional equation of the L-functions associated even to H^1 of curves of genus bigger than 1, let alone to arbitrary cohomology groups of varieties of higher dimension, remain conjectural even today. Surprisingly, however, these properties can be verified numerically, because there is a method, observed by several people and worked out in detail both theoretically and in the form of a software package by Tim Dokchitser [24], to calculate a Dirichlet series with an assumed functional equation to arbitrarily high precision even outside its domain of convergence and at the same time to test the functional equation numerically. This method is self-verifying in the sense that it involves a free parameter and that if the final result of the calculation turns out to be independent of the choice of this parameter then one has convincing evidence both of the correctness of the presumed functional equation and of the accuracy of the numerical evaluation, and in all of the many cases that have been tested the predicted analytic properties and special values were verified.

The idea of a motive

We have seen that one can associate to a (smooth, projective) algebraic variety $X/\mathbb{Q}$ several different kinds of cohomology groups: the Betti cohomology $H_B^*(X) = H^*(X(\mathbb{C}), \mathbb{Q})$, the algebraic de Rham cohomology $H_{dR}^*(X)$ defined in terms of differential forms, and the ℓ-adic cohomology groups $H^*(\overline{X}; \mathbb{Q}_\ell)$ on which the Galois group of $\overline{\mathbb{Q}}$ over $\mathbb{Q}$ acts, and that these are interrelated in many ways: the complexifications (tensor product over $\mathbb{Q}$ with $\mathbb{C}$) of the Betti and de Rham cohomology are canonically identified with one another and with the complex cohomology group $H^*(X(\mathbb{C}); \mathbb{C})$, which in turn has a Hodge decomposition as the direct sum of complex subspaces $H^{p,q}(X)$, the ℓ-adic cohomology groups are isomorphic as $\mathbb{Q}_\ell$-vector spaces with the tensor product of $H_B^*(X)$ with $\mathbb{Q}_\ell$ and are related among each other by the fact that the characteristic polynomial of the Frobenius element Fr_p for a (good) prime $p \neq \ell$ is independent of ℓ. The coefficients of the transition matrix coming from the passage between the $\mathbb{Q}$-bases of $H^*(X(\mathbb{C}); \mathbb{C})$ coming from Betti and de Rham cohomology are the periods of X, which are arithmetically interesting numbers if X is fixed and give rise to the arithmetically interesting Picard–Fuchs differential equation if X varies algebraically in a family, while the ℓ-adic Galois representations provide the deep arithmetic information that leads from point counting over finite fields to the global L-functions with their mysterious and still mostly conjectural analytic properties. The word "motive" is used to describe a purely linear-algebra structure that has all of these properties ($\mathbb{Q}$-vector spaces labelled by the names "Betti" and "de Rham" whose complexifications are isomorphic and have a Hodge decomposition and families of ℓ-adic Galois representations that satisfy the above-mentioned compatibilities) but is not explicitly required to come from the cohomology of any specific algebraic variety. The idea is due to Grothendieck, who gave a concrete way to produce such objects as a piece of a cohomology group cut out by algebraic correspondences. (More precisely, if X is a variety defined over $\mathbb{Q}$ then any correspondence $Z \subset X \times X$ defined over $\mathbb{Q}$ induces an endomorphism of each of the cohomology groups associated to X, and if one has a $\mathbb{Q}$-linear combination P of such induced maps that is a projector, i.e., that satisfies $P^2 = P$, then the image of P gives a collection of subspaces of the Betti, de Rham, and ℓ-adic cohomology groups that has all of the above-named properties.) More generally, the same is true of any "natural piece" of the cohomology of X, meaning a collection of $\mathbb{Q}$-subspaces of the Betti and de Rham cohomology whose complexifications are equal and that are the direct sums of their intersections with the Hodge spaces $H^{p,q}(X)$, and of $\mathbb{Q}_\ell$-subspaces of the ℓ-adic cohomology groups that correspond to these under the comparison maps and are stable under the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. If the Hodge and Tate conjectures are true, then these two classes of motives coincide.

In the early years there was no clear way to establish the existence of a well-defined category of motives having all the desired properties, and there were (and still are) mathematicians who deprecated the whole subject as mere castles in the air lacking both definitions and theorems. That situation has changed over the years and there are now well-defined theories of motives and motivic cohomology due to the work of Tate, Deligne, Bloch, Beilinson, Voevodsky, André, Nori and many others, even though the full theory, in the sense of having a well-defined abelian category of motives in full generality satisfying all the expected properties, is still not in its final shape. The fact that there are rival and not necessarily equivalent candidates for the "right" definition is not in itself a problem, just as there are many ways to define cohomology groups (singular, de Rham, Čech, ...) that are all valid and useful in various settings. But even if there were no general definitions or theorems at all, there would still be plenty of perfectly well-defined examples, either as pieces of actual cohomology groups (like the 4-dimensional piece of the 204-dimensional $H^3(\mathcal{Q}_\psi)$ defined as its invariant part under a group action) or else by various specific constructions, like polylogarithmic motives or the motives discussed in Examples 1 and 2 below. There are two main points to be made here. The first is that, even though conjecturally there are no other motives than the Grothendieck ones (pieces of cohomology groups cut out by geometrically defined correspondences), the realization of a given motive in this form is not unique or in any sense canonical and it is extremely useful to think of it as an object "in its own right" that can be studied and used independently of any such realization. The second is that, even if certain expected properties are still conjectural and no proof is in sight, they can nevertheless often be used to make concrete predictions that can be tested numerically or proved by classical methods but that might be hard to discover without the motivic way of thinking. We will give illustrations of the first point in the two examples below (modular and hypergeometric motives) and of the second in the three examples treated in the next section.

Example 1: Motives associated to modular forms

To any Hecke cusp form f (= cusp form with multiplicative Fourier coefficients, as discussed in §4) of weight k there is an associated 2-dimensional motive M_f of weight $k-1$ and Hodge type (dimensions of the pieces of the Hodge decomposition) $h^{k-1,0} = h^{0,k-1} = 1$. For $k = 2$ this follows from the work of Eichler and Shimura, who showed that L_1 of the modular curve $X_0(N)$ (the algebraic curve over $\mathbb{Q}$ whose complex version is the compactification of $\mathbb{H}/\Gamma_0(N)$) for any $N \geq 1$ is the product of the L-functions (4.12) of Hecke forms of weight 2 and level N, with each factor $L(f,s)$ being the L-series of a 2-dimensional subspace of $H^1(X_0(N)/\mathbb{Q})$ defined as the intersection of the kernels of $T_p - a_p(f)$ for all (or sufficiently many) primes $p \nmid N$, where T_p denotes the pth Hecke operator (a correspondence of degree $(p+1, p+1)$ from $X_0(N)$ to itself) and $a_p(f)$ the coefficient of q^p in the Fourier development of f. For $k > 2$ the motivic nature of f was proved by Deligne, who showed that for each $p \nmid N$ the coefficient $a_p(f)$ coincides with the trace of the Frobenius Fr_p on a particular 2-dimensional subspace (again cut out by the $T_p - a_p(f)$) of the $(k-1)$st cohomology group of the associated $(k-1)$-dimensional Kuga variety (a compactification of the total space of the fibre bundle over $\mathbb{H}/\Gamma_0(N)$ whose fibre over any point is the $(k-2)$nd symmetric power of the corresponding elliptic curve). Ramanujan's conjecture $|a_p(\Delta)| \leq p^{11/2}$ (or more generally $|a_p(f)| \leq 2p^{(k-1)/2}$ for any Hecke cusp form of weight k) followed by combining this result with Deligne's later proof of the "Riemann hypothesis" part of the Weil conjectures, and has therefore sometimes been referred to by Serre as the "theorem of Deligne and Deligne." An explicit construction of M_f as a Grothendieck motive was given by Scholl [62].

The comparison of the Betti and de Rham $\mathbb{Q}$-bases of the 2-dimensional complex realization of M_f gives rise to periods $\omega_\pm(f) \in \mathbb{R}$ that are related to the values of the L-series $L(f,s)$ at $s = 1, 2, \ldots, k-1$, in accordance with Deligne's general conjecture on special motivic L-values. For example, these two periods for the cusp form $\Delta \in S_{12}(SL(2,\mathbb{Z}))$, in a suitable normalization, have the numerical values $\omega_+ = 0.046346\cdots$, $\omega_- = 0.045751\cdots$ and are related to the special values of the completed L-function $\hat{L}(\Delta, s) := (2\pi)^{-s}\Gamma(s)L(\Delta, s) = \hat{L}(\Delta, 12-s)$ by

s	6	6 ± 1	6 ± 2	6 ± 3	6 ± 4	6 ± 5
$\hat{L}(\Delta,s)$	$\frac{\omega_+}{30}$	$\frac{\omega_-}{28}$	$\frac{\omega_+}{24}$	$\frac{\omega_-}{18}$	$\frac{2\omega_+}{25}$	$\frac{90\omega_-}{691}$

As a concrete example, the motive M_f for the form $f = f_{2,11}$ defined in (4.8) is given by H^1 of the elliptic curve $E_{11}/\mathbb{Q}$ with Weierstrass equation $y^2 - y = x^3 - x^2$, meaning that the number of solutions of this equation in $\mathbb{F}_p^2$ equals $p - a_p(f)$ for all p. But it is also equal to H^1 of the modular curve $X_0(11)/\mathbb{Q}$ by the Eichler-Shimura theory, and the elliptic curves $X_0(11)$ and E_{11} are isogenous but not isomorphic. Already this simple example illustrates the point made above that a motive should be thought of "in its own right" and not as a specific subspace of the cohomology of some specific variety, since there are typically many realizations of the same motive and we do not necessarily know or need an explicit geometric correspondence between them. A more striking example is given by the Hecke form

$$f = f_{4,10} = q + 2q^2 - 8q^3 + 4q^4 + 5q^5 - 16q^6 + \cdots$$

of weight 4 and level 10. An old paper of Ron Livné [51] showed that the L-function of f is a factor of the Hasse-Weil zeta function of the smooth 7-dimensional variety W_{10} defined as the set of points $(x_1 : \cdots : x_{10})$ in $\mathbb{P}^9(\mathbb{C})$ satisfying $\sum x_i = \sum x_i^3 = 0$ (more specifically: the number of points of W_{10} over $\mathbb{F}_p$ for $p \nmid 10$ equals $-p^2 a_p(f)$ plus a polynomial in p, so the Hasse-Weil zeta function is a multiplicative combination of shifts of $\zeta(s)$ and of $L(f,s)$), and a recent paper of Matthias Schütt [64] shows that the *same* modular form f also occurs in the zeta-function of a certain Calabi–Yau threefold $\widehat{W}^3$. In both cases the result is established purely arithmetically (by counting points modulo p for small p and using theorems of Faltings and Serre to deduce that if the desired equality of Frobenius traces is true for sufficiently many p, then it is always true), without exhibiting any explicit correspondence between the algebraic variety W_{10} or $\widehat{W}^3$ and any variety having a modular parametrization. There are many more examples in the literature, a particularly nice collection being given by the modularity theorem [32] (previously the modularity conjecture) for "rigid" Calabi–Yau threefolds (those with $b_3 = 2$) defined over $\mathbb{Q}$, of which Schütt's example is a special case. We refer to Noriko Yui's survey article [73] and Christian Meyer's book [53] for more detailed discussions.

Example 2: Hypergeometric motives

The starting point here is an old observation of Deuring. Consider the Legendre elliptic curve (3.1) for some rational value of t and reduce modulo a prime p not dividing the numerator or denominator of t or $1 - t$. Then from (6.2) we find that the integer $a_p(E_t) = p + 1 - \#\overline{E_t}(\mathbb{F}_p)$ is given modulo p by

$$\begin{aligned} a_p(E_t) = -\sum_{x \pmod p} \left(\frac{x(x-1)(x-t)}{p}\right) &\equiv -\sum_{x \pmod p} \big(x(x-1)(x-t)\big)^{(p-1)/2} \\ &\equiv \text{Coefficient of } x^{p-1} \text{ in } \big(x(x-1)(x-t)\big)^{(p-1)/2} \pmod p \end{aligned}$$

because the sum of x^m over $x \in \mathbb{Z}/p\mathbb{Z}$ equals -1 for $m = p - 1$ and 0 for other values of m between $\frac{p-1}{2}$ and $\frac{3(p-1)}{2}$. Calculating by the binomial theorem, we find

$$(-1)^{(p-1)/2} a_p(E_t) \equiv \sum_{n=0}^{(p-1)/2} \binom{(p-1)/2}{n}^2 t^n \equiv \sum_{n=0}^{(p-1)/2} \binom{-1/2}{n}^2 t^n \pmod p,$$

which (as was later observed by Igusa) is just a truncated version of the hypergeometric series in (3.3) giving the period function for the Legendre family as t varies. This observation led to a vast development, begun by Dwork, continued by Katz, and now being systematically developed by Villegas and his collaborators [59]. To any pair $\mathbf{a} = (a_1, \ldots, a_r)$, $\mathbf{b} = (b_1, \ldots, b_r)$ of tuples of rational numbers of the same length r and to each rational number t one associates an r-dimensional motive $M(\mathbf{a},\mathbf{b};t)$ whose complex realization in the case $b_r = 1$ has periods coming from the hypergeometric function (3.8) with $r = s+1$ and whose L-function is defined by an Euler product whose Euler factors are given by a beautiful explicit formula in terms of Gauss and Jacobi sums at the "unramified" primes (those not dividing the denominator of any a_i or b_i or of t, $\frac{1}{t}$ or of $\frac{1}{1-t}$) that can also be written in terms of truncated p-adic hypergeometric series. (There is also a much more complicated recipe, discussed in detail in [59], for the Euler factors at the ramified primes.) These motives are well-defined objects and can be realized geometrically as part of the cohomology of some variety, as was shown by Katz [45] (and more explicitly in [9], where varieties are constructed whose number of points is given by the p-adic hypergeometric function when we are in the situation (3.9), when $M(\mathbf{a},\mathbf{b};t)$ is defined over $\mathbb{Q}$). But the real point is that these are intrinsically defined motives, whose complex and ℓ-adic parts can be written down directly in terms of the defining data $(\mathbf{a},\mathbf{b},t)$ without needing the geometric realizations (which are neither canonical nor particularly natural).

The prototypical and motivating example of a hypergeometric motive is the one associated to $\mathbf{a} = (\frac{1}{5},\frac{2}{5},\frac{3}{5},\frac{4}{5})$ and $\mathbf{b} = (1,1,1,1)$, which is the motive given by the same 4-dimensional piece of the 204-dimensional third cohomology group of the Dwork quintic $\mathcal{Q}_\psi$ (with $t = (5/\psi)^5$) that led to the Picard–Fuchs differential equation and period function $F(\mathbf{a};\mathbf{b};t) = \sum \frac{(5n)!}{5^{5n}n!^5}t^n$ that we found in Example 2 of §3. A detailed discussion of this case and of the way that one can obtain the L-function by first counting points on $\mathcal{Q}_\psi$ and then subtracting the contributions from the unwanted 200-dimensional part of the cohomology, is given in the paper [15] by Candelas, de la Ossa, and Villegas, where an explicit expression for the number of points in terms of p-adic gamma functions is derived (penultimate formula on p. 46 of [15]). To quote Candelas: "The fact that you can count the numbers of $\mathbb{F}_p$-rational points using periods is very interesting." Here and in [59] the L-functions are calculated, and their analytic continuation and functional equation (which are in general still conjectural) verified numerically, for many rational values of t. These degree 4 L-functions should have an automorphic meaning as the spinor L-functions of certain Siegel modular forms of degree 2 and weight 3.

7 Examples

In this section we will describe in detail three examples, all of which I heard about from Vasily Golyshev, where a motivic argument suggested a concrete mathematical statement that could then be checked numerically and/or theoretically.

Example 1. Apéry numbers of fractional index

Our first example has to do with the interpolation to non-integral values of n of the Apéry numbers A_n. Golyshev pointed out that Apéry's hypergeometric closed formula (1.5), slightly rewritten, makes sense also for $n \notin \mathbb{Z}$ and, on the basis of motivic considerations that will be indicated below, predicted that the real number $A_{-1/2}$ should be a simple multiple of a special value of the L-series of the modular form $f_{4,8}$ defined in equation (4.10). This turned out to be true (Proposition 2 below), and in fact we learned subsequently from Wadim Zudilin that an equivalent result (but as a purely hypergeometric formula for the L-value, not from the point of view of interpolating the A_n) had also just been proven by Rogers, Wan and Zucker (eq. (32) of [60], where many other examples of similar type are given). However, the main point here is not the proof of the identity, which is a nice illustration of the connection between modular forms and period integrals but is not particularly difficult, but rather the power of the way of thought that made it possible to predict that such an identity had to hold in the first place.

We define a number A_x for any $x \in \mathbb{C}$ by the absolutely convergent series

$$A_x = \sum_{k=0}^{\infty} \binom{x}{k}^2 \binom{x+k}{k}^2 . \tag{7.1}$$

For $x = n \in \mathbb{Z}_{\geq 0}$ this series terminates at $k = n$ and agrees with Apéry's formula (1.5) for the Apéry numbers $\{A_n\} = \{1, 5, 73, \ldots\}$, so (7.1) gives a natural interpolation of these numbers to arbitrary complex arguments. A small surprise here was that this interpolation does *not* satisfy the original recursion (1.1) of the Apéry numbers, but only a modified version of it:

Proposition 7.1 *The sum* (7.1) *defines a holomorphic function which for all* $x \in \mathbb{C}$ *satisfies the symmetry property* $A_x = A_{-x-1}$ *and the functional equation*

$$(x+1)^3 A_{x+1} - (34x^3 + 51x^2 + 27x + 5)\, A_x + x^3 A_{x-1} = \frac{8}{\pi^2}\,(2x+1)\, \sin^2 \pi x\, .$$

This will be proved below. (I mention here that Golyshev and I have now found a way different from (7.1) to interpolate $\{A_n\}$ to complex values that satisfies the original recursion (1.1), based on the method of Frobenius limits that will be discussed in Section 9. This will be presented in a later paper.)

Now we consider $A_{-1/2}$. The series (7.1) converges too slowly to be used directly, but by convergence acceleration techniques one can calculate the value

$$A_{-1/2} = 1.118636387164187068349619257525640916794857551529361191 48 \cdots$$

and verify numerically that it satisfies the statement of the next proposition.

Proposition 7.2 *The value of the function* A_n *at its point of symmetry is given by*

$$A_{-1/2} = \frac{16}{\pi^2}\, L(f_{4,8}, 2)\, , \tag{7.2}$$

where $f_{4,8}(\tau)$ *is the normalized Hecke eigenform in* $S_4(\Gamma_0(8))$ *defined in* (4.10).

This too will be proved below, but first we explain what lay behind Golyshev's prediction that an identity of this sort was to be expected. At first sight it looks very strange, since the modular interpretation of the numbers A_n that we have seen so far has to do with the modular form f of weight 2 on $\Gamma_0(6)$ occurring in (5.3), while (7.2) involves the completely different modular form $f_{4,8}$ of weight 4 on $\Gamma_0(8)$. The connection between them occurs through the congruence

$$A_{(p-1)/2} \equiv \gamma_p \pmod{p} \qquad (p > 2 \text{ prime}) \tag{7.3}$$

proved by Beukers in [7], where γ_n denotes the coefficient of q^n in $f_{4,8}(\tau)$. In an earlier paper [6] he had showed that the numbers A_{mp^r-1} have a p-adic limit as $r \to \infty$ for any prime p and any positive integer m, and formal group methods that Stienstra and he developed in [66] give a kind of fusion of these two results, namely that the p-adic limit of $A_{(mp^r-1)/2}/A_{(mp^{r-1}-1)/2}$ exists for any odd prime p and odd positive integer m and that its value u_p is independent of m and related to γ_p by $\gamma_p = u_p + p^3/u_p$. All of these results show that there is a deep connection between the Apéry numbers A_n and the L-series of $f_{4,8}$ as defined in (4.12) (here with χ the trivial Dirichlet character modulo 2). Already in [7], in connection with the congruence (7.3), Beukers had written "Although we do not know all the details yet, this congruence must arise from the interplay between the numbers A_n and the ζ-function of a certain algebraic threefold." In fact the connection, as suggested by the factor $\frac{1}{2}$ in the index of A_n in (7.3), is with the double covering W of the Beukers-Peters family of $K3$ surfaces given by $w^2 = L(x,y,z)$, where $L(x,y,z)$ is the Laurent polynomial (1.10) defining this family, as we can test numerically by verifying the formula

$$\sum_{x,y,z \in \mathbb{F}_p^\times} \left(\frac{L(x,y,z)}{p}\right) = -p - \gamma_p \qquad (p > 2 \text{ prime}),$$

in which the left-hand side (up to sign and an additive term $(p-1)^3$) gives the number of points $(x,y,z,w) \in \mathbb{F}_p^{\times 3} \times \mathbb{F}_p$ lying on W. This point-counting identity says that the L-function of the cusp form $f_{4,8}$ is part of the Hasse-Weil zeta function of W, so if one believes the motivic philosophy then that means that the Galois representation, and thus the motive, of $f_{4,8}$ is contained in that of W and hence (according to the Tate conjecture) that there must be an algebraic correspondence over $\mathbb{Q}$ between the Kuga variety over $X_0(8)$ in which the former motive lives and the variety W. Of course the Tate conjecture is not known in this generality, but the existence of the predicted correspondence can in principle always be verified by an actual construction (and in the case under consideration is presumably given by a birational map found by van Straten and mentioned on p. 170 of [53]), and even in cases where this cannot be carried out we can still formulate and test the identity of periods that it implies, which here is precisely (7.2)

We end the subsection by giving proofs of the two propositions.

Proof of Proposition 1. Denote the kth summand in (7.1) by $\alpha_k(x)$. The asymptotic formula $\binom{a}{k} \sim \frac{(-1)^k}{\Gamma(-a)} k^{-a-1}$ $(k \to \infty)$ and the duplication formula of the gamma function give $\alpha_k(x) \sim \sin^2(\pi x)/(\pi k)^2 = \mathrm{O}(k^{-2})$ as $k \to \infty$, so the series (7.1) converges absolutely and locally uniformly and hence defines a holomorphic function in the entire complex plane. The symmetry under $x \mapsto -x-1$ is obvious since each term in (7.1) has this property. Finally, to prove the recursion, we observe that by induction on K we have

$$\sum_{k=0}^{K} \left((x+1)^3 \alpha_k(x+1) - (34x^3+51x^2+27x+5)\alpha_k(x) + x^3\alpha_k(x-1)\right)$$
$$= 4\left(K(2K+1)(2x+1) - (2x+1)^3\right)\alpha_K(x)\,,$$

and the limiting value of this as $K \to \infty$ has the value claimed because of the asymptotic formula for $\alpha_K(x)$. (This calculation is just a rewriting of the standard proof by the method of telescoping series of Apéry's original recursion formula.) □

Proof of Proposition 2. Define a function $B(t)$ for $|t| \le 1$ by $B(t) = \sum_{k=0}^{\infty} \binom{-1/2}{k}^2 t^k$, the hypergeometric series occurring in the period integral (3.3). Then we have

$$A_{-1/2} = \sum_{k=0}^{\infty} \binom{-1/2}{k}^4 = \frac{1}{2\pi i} \oint_{|t|=1} B(t)\,B(1/t)\,\frac{dt}{t}\,.$$

If we set $t = \lambda(\tau)$ with $\tau \in \mathbb{H}$, then from the modular parametrization (5.1) we obtain $B(t) = \vartheta_3(\tau)^2$. Using the modular transformation properties

$$\frac{1}{\lambda(\tau)} = \lambda\Big(\frac{\tau}{1-\tau}\Big)\,, \qquad \vartheta_3\Big(\frac{\tau}{1-\tau}\Big)^2 = (1-\tau)\,\vartheta_2(\tau)^2$$

and the modular form identity

$$\frac{1}{2\pi i}\,\vartheta_3(\tau)^2\,\vartheta_2(\tau)^2\,\frac{\lambda'(\tau)}{\lambda(\tau)} = 2\,f(\tau/4)\,,$$

with $f = f_{4,8}$, we then find the integral representation

$$A_{-1/2} = 2\int_0^2 (1-\tau)\,f(\tau/4)\,d\tau\,,$$

where the integral is taken along the hyperbolic geodesic from 0 to 2 (= Euclidean semicircle with center 1 and radius 1), which is mapped by λ isomorphically to the unit circle. Since f is a cusp form, we can replace this path of integration by the difference of the two vertical lines from 0 to $i\infty$ and from 2 to $i\infty$, and since $f(\tau+\frac{1}{2}) = -f(\tau)$ (because f has a q-expansion containing only odd powers of q), this gives finally

$$A_{-1/2} = 2\Big(\int_0^{\infty} - \int_2^{\infty}\Big)(1-\tau)\,f(\tau/4)\,d\tau = -4\int_0^{\infty} \tau\,f(\tau/4)\,d\tau\,,$$

which is equivalent to (7.2) by the standard integral representation of $L(f,s)$. □

Example 2. Periods of the mirror quintic family

Our next example, again involving the periods of a cusp form of weight 4, is connected with the Dwork quintic $\mathcal{Q}_\psi$ studied in Example 2 of Section 3. There we gave the first calculation (3.7) of an explicit period on $\mathcal{Q}_\psi$, following [14], finding that the integral of the algebraic differential form (3.6) over an appropriate 3-cycle equals $(2\pi)^3\Phi(\psi^{-5})$, where $\Phi(t)$ denotes the hypergeometric function $\sum_{n=0}^{\infty}\frac{(5n)!}{n!^5}t^n$. This means that $(2\pi i)^3\Phi(t)$ is part of the $\mathbb{Z}$-lattice in the 4-dimensional solution space of the corresponding hypergeometric differential equation obtained by integrating Ω over a basis of $H_3(\mathcal{Q}_\psi;\mathbb{Z})$, as discussed in Section 2. A natural question, here or for any other differential equation with regular singular points, is to give the complete transition matrices between the basis of the solution space obtained by local expansions at each singular point and an integral basis of this lattice. In the case at hand there are three singular points $t = 0$ ("point of maximum unipotent monodromy"), $t = \infty$ ("orbifold point"), and $t = 5^{-5}$ ("conifold point"). At $t = 0$ we have the "Frobenius basis" $\{\Phi_i(t)\}_{0\le i\le 3}$ (cf. (9.4) below), where each $\Phi_i(t)$ is a polynomial of degree i in $\log t$ with coefficients in $\mathbb{Q}[[t]]$ and with leading term $\Phi(t)(\log t)^i/i!$, and at $t = \infty$ we have the basis given by choosing any four of the five functions $\sum_{5\nmid n}\frac{\Gamma(n)^5}{\Gamma(5n)}\zeta^n t^{-n/5}$ ($\zeta^5 = 1$), which sum to 0. Candelas et al. gave explicit formulas for the 4×4 transition matrices between each of these and the integral basis, and a generalization applying to all hypergeometric differential equations was given by Golyshev and Mellit in [30]. But for the conifold point only seven of the entries of the corresponding transition matrix (namely, those of its last row and column in appropriate bases) are known in closed form, and the remaining nine only numerically to high precision (calculations by Albrecht Klemm, Emanuel Scheidegger, and myself). Golyshev told us that among the remaining entries of this matrix one should find both periods of the cuspidal eigenform $f_{4,25}$ of weight 4 and level 25 given in (4.11). We checked this prediction, and also our own further prediction that the "quasiperiods" as well as the periods of $f_{4,25}$ should appear in the transition matrix that we had already calculated numerically, and indeed simple rational multiples or rational linear combinations of all four numbers appeared, at least numerically to very high precision [49].

Where did these predictions come from? In the final example of the last section we discussed how point counting on the Dwork quintic $\mathcal{Q}_\psi$ for generic ψ leads to a polynomial $P(T) = P(M(\frac{1}{5},\frac{2}{5},\frac{3}{5},\frac{4}{5};1,1,1,1;(5/\psi)^5)/\mathbb{F}_p,T)$ of degree 4 for every good prime p. At the singular point $\psi = 5$, the degree of this polynomial drops by 1 and it factors as $(1-(\frac{p}{5})pT)(1-c_pT+p^3T^2)$, where c_p is the coefficient of q^p in the cusp form $f_{4,25}$, as was shown 30 years ago by Chad Schoen [62] by computing the values in question for sufficiently many primes and then invoking the results of Faltings and Serre to deduce their equality in general. This says that the 2-dimensional motive associated to $f_{4,25}$ is contained in the third cohomology group of $\mathcal{Q}_5$ (more precisely, that the corresponding ℓ-adic Galois representation occurs in $H^3(\overline{\mathcal{Q}_5};\mathbb{Q}_\ell)$). Hence, if one believes the Tate conjecture, there should be an actual correspondence, defined over $\mathbb{Q}$, between the associated geometric objects, and this in turn implies that the two periods of $f_{4,25}$ must show up in the period matrix of the Dwork pencil at the point $\psi = 5$ as claimed. A similar degeneration at the singular fiber of the generic fourth degree point-counting polynomial into a linear factor and a quadratic factor associated to a modular form of weight 4 has been found by Villegas [70] for all 14 hypergeometric families of Calabi–Yau threefold, e.g., the L-series of the cusp form (4.9) occurs in the L-series of the motive $M((\frac{1}{4}, \frac{1}{3}, \frac{2}{3}, \frac{3}{4}); (1,1,1,1); 1)$. One can then predict that the periods and quasiperiods of these cusp forms will appear in the transition matrices between the relevant bases of the spaces of solutions of the corresponding hypergeometric differential equations.

Example 3. Hypergeometric algebraic units

The last example is of a somewhat different nature. In Example 4 of Section 3 we discussed hypergeometric functions $F(t)$ of the form (3.9) that are algebraic, giving Villegas's criterion for this and also the examples (3.10) and (3.11). Here Golyshev predicted, based on an argument about extensions of motives that I will not reproduce, that the power series $Q(t) = \exp\Big(\int \frac{F(t)}{t}\,dt\Big) = t\exp\Big(\sum_{n>0} a_n \frac{t^n}{n}\Big)$, where a_n denotes the coefficient of t^n in $F(t)$, must always be an algebraic function in the field $\mathbb{Q}(t, F(t))$, and in fact always an algebraic unit over $\mathbb{Z}[1/t]$. (This implies in particular that the value of $Q(t)$ if one substitutes for t the reciprocal of any integer bigger than the inverse of the radius of convergence is an algebraic unit in $\overline{\mathbb{Q}}$.) Specifically, he asked me whether I could prove this for the special case of the binomial series (3.10), and this turned out not to be too hard, as shown in Proposition 3 and its proof below.[4] (Strangely enough, precisely this question had appeared quite recently in various contexts in physics, e.g., in [19] and [25], as I was informed by Yan Soibelman.) I also checked Golyshev's prediction for the first two power series in (3.11) (Proposition 4 below), but in view of the huge degree I was not able to do the same for the third example. Spencer Bloch sketched to me a proof of the algebraicity of $Q(t)$ whenever the curve defined by the algebraic hypergeometric function $F(t)$ is rational (as happens for $B_{M,2}(t)$ for all M and also for $F_{(6,1),(3,2,2)}(t)$; see below), but as far as I know there is no proof yet for the general case.

Proposition 7.3 *The function defined for $|t| < 1$ by the power series expansion*

$$Q_{M,N}(t) = t\,\exp\bigg(\sum_{n=1}^{\infty} \binom{Mn}{Nn} \frac{t^n}{n}\bigg) \quad \in \mathbb{Q}[[t]]$$

is algebraic for all $M > N > 0$, and is a unit over $\mathbb{Z}[1/t]$.

Proof. For t small, the polynomial $P_t(u) = t(1+u)^M - u^N$ of degree M has N "small" roots $u_1(t), \ldots, u_N(t)$ near the circle $|u| = |t|^{1/N}$ and $K := M - N$ "large"

4. A much more general result was given by Maxim Kontsevich in his article 'Noncommutative identities' (arXiv:1109.2469) written on the occasion of my $3 \cdot 4 \cdot 5$ birthday.

roots $u_{N+1}(t),\dots,u_M(t)$ near the circle $|u|=|t|^{-1/K}$. The function (3.10) is then given explicitly as an algebraic function of t by

$$B_{M,N}(t)=\sum_{i=1}^{N}\frac{1+u_i(t)}{N-Ku_i(t)}\,, \tag{7.4}$$

as one can see by using the Cauchy residue theorem twice to write

$$\begin{aligned}B_{M,N}(t)&=\sum_{n=0}^{\infty}\left(\frac{1}{2\pi i}\int_{|u|=1}\frac{(1+u)^{Mn}\,du}{u^{Nn+1}}\right)t^n\\&=-\frac{1}{2\pi i}\int_{|u|=1}\frac{u^{N-1}\,du}{P_t(u)}=-\sum_{i=1}^{N}\frac{u_i(t)^{N-1}}{P_t'(u_i(t))}\,.\end{aligned}$$

Differentiating the equation $P_t(u_i(t))=0$ with respect to t, we find that the ith summand in (7.4) equals $tu_i'(t)/u_i(t)$, and then dividing by t, integrating and exponentiating we get the formula

$$Q_{M,N}(t)=(-1)^{N-1}\prod_{i=1}^{N}u_i(t) \tag{7.5}$$

for $Q_{M,N}(t)$ in terms of the roots $u_i(t)$. This proves the proposition since the polynomial $t^{-1}P_t(u)$ is monic of degree M over $\mathbb{Z}[1/t]$ and therefore each $u_i(t)$ is an algebraic unit over this ring. □

As an example of the proposition, the function $Q_{5,2}(t)$ is a root of the equation $(Q+1)^{10}t^2-Q(Q+1)^5(Q^2-5Q+1)t+Q^7=0$, which has degree $10=\binom{5}{2}$ in Q. In general the degree of the algebraic function $B_{M,N}(t)$ (or $Q_{M,N}(t)$) over $\mathbb{Q}(t)$ equals $\binom{M}{N}$ for M and N coprime, because its conjugates are given by replacing $u_1,\dots,u_N$ in (7.4) (or (7.5)) by any subset of $\{u_1,\dots,u_M\}$ of cardinality N, while if $(M,N)=d>1$ the degree equals $\binom{M/d}{N/d}^d$. Another remark is that the algebraic curve defined by the algebraic functions $B_{M,N}(t)$ or $Q_{M,N}(t)$ has a rational parametrization, and hence has genus 0, if $N=1$ or $N=2$. Indeed, for $N=1$ equation (7.4) simplifies to $B_{M,1}(t)=\frac{1+u(t)}{1-Ku(t)}$, where $u(t)$ is the solution of $u=t(1+u)^M$ in $t+t^2\mathbb{Q}[[t]]$ (this formula also follows from the Lagrange inversion formula) and equation (7.5) becomes simply $Q_{M,1}(t)=u$, so that we have the rational parametrization $t=\frac{u}{(1+u)^M}$, $B_{M,1}(t)=\frac{1+u}{1-Ku}$, $Q_{M,1}(t)=u$. If $N=2$ we denote by $u=u_1(t)$ and $v=u_2(t)$ the solutions of $t=\frac{u^2}{(1+u)^M}=\frac{v^2}{(1+v)^M}$ with $u\sim\sqrt{t}$, $v\sim-\sqrt{t}$ and set $\frac{1+u}{1+v}=x^2$. Then $\frac{u}{v}=-x^M$ and we can solve to get

$$\begin{aligned}u&=\frac{x^M-x^{M-2}}{1+x^{M-2}}\,,\quad v=\frac{1-x^2}{x^2+x^M}\,,\quad t=\frac{x^{2M-4}(1-x^2)^2(1+x^{M-2})^{M-2}}{(1+x^M)^M}\,,\\B_{M,2}(t)&=\frac{1+x^M}{2(1+x^M)+Mx^{M-2}(1-x^2)}+\frac{1+x^M}{2(1+x^M)-M(1-x^2)}\,.\end{aligned}$$

When we divide this by t, integrate, and exponentiate, there is a huge cancellation and the formula for $Q_{M,2}(t)$ is much simpler, with two surprising factorizations:

$$Q_{M,2}(t) \;=\; \frac{x^{M-4}(1-x^2)^2}{(1+x^{M-2})^2}\,, \qquad 1+Q_{M,2}(t) \;=\; \frac{(1+x^{M-4})(1+x^M)}{(1+x^{M-2})^2}\,.$$

Finally, we verify Golyshev's prediction for the first two series in (3.11).

Proposition 7.4 *Each of the two power series*

$$t\,\exp\Bigl(\sum_{n=1}^{\infty}\frac{(6n)!\,n!}{(3n)!\,(2n)!^2}\,\frac{t^n}{n}\Bigr),\quad t\,\exp\Bigl(\sum_{n=1}^{\infty}\frac{(10n)!\,n!}{(5n)!\,(4n)!\,(2n)!}\,\frac{t^n}{n}\Bigr) \tag{7.6}$$

is algebraic, and is a unit over the ring $\mathbb{Z}[1/t]$.

Proof. The proof is purely computational, using the first terms of each power series to guess the algebraic equation and then verifying that it satisfies the correct differential equation, so we content ourselves with describing the structure of the equations of the hypergeometric series $F(t) = F_{\mathbf{c},\mathbf{d}}(t)$ and the corresponding unit $Q(t)$ in each case. The degrees of $F(t)$ and $Q(t)$ over $\mathbb{Q}(t)$ are 6 and 30, as already mentioned in §3, but in each case one of the conjugates of F is $-F$ and the functions $F(t)^2$ and $Q(t)+Q(t)^{-1}$ therefore have degree only 3 and 15, respectively. For instance, in the first case the equation satisfied by $G = (1-108t)F_{(6,1),(3,2,2)}(t)^2$ over $\mathbb{Q}(t)$ is $G(4G-3)^2 = (1-216t)^2$ and that of $H = Q(t)+2+Q(t)^{-1}$ is $H^3t^2-(H^2-27H+108)t+1 = 0$. In this case the curve has genus 0 and can be given parametrically by $Q\bigl(\frac{u(1-u)^2(3-u)^2(4-u)}{432}\bigr) = \frac{u(1-u)}{(3-u)(4-u)}$. In the second case the equation satisfied by $H = Q+2+Q^{-1}$ over $\mathbb{Q}(t)$ has the form $t^6H^{15}-t^5H^{14}+432t^5H^{13}-(14500t-184)t^4H^{12}+\cdots+65t(3125t-9)^2 = 0$, where the intermediate coefficients are complicated and have been omitted. $\square$

8 Differential equations and mirror symmetry

The usual description of mirror symmetry involves two Calabi–Yau varieties (or more properly families of Calabi–Yau varieties) and relates the Gromov–Witten invariants ("A-side") of one of them to the Picard–Fuchs equation ("B-side") of the other. A particularly clean class of examples, which is the only one we will talk about in this paper, starts with the family of Calabi–Yau varieties arising as anticanonical divisors of a Fano manifold of arbitrary dimension. In this section we will discuss how this works, starting with the definition and examples of Fano manifolds and then explaining the conjectured "mirror" correspondence and describing a large number of examples for which it is known. In a few words, Gromov–Witten theory associates to any complex symplectic manifold a collection of invariants defined by counting holomorphic maps of curves into the variety with prescribed homological data (like the genus and number of marked points on the curves, the homology class in the target variety to which it maps, and constraints on the images of the marked points). The genus 0 Gromov–Witten invariants are used to construct a "quantum cohomology ring" that is a deformation of the usual cohomology ring of the variety, and from this structure in turn one constructs an explicit linear differential equation of finite order called the quantum differential equation. The mirror correspondence is then characterized by the statement that the quantum differential equation of the Fano variety is the Laplace transform of the Picard–Fuchs differential equation of its mirror variety, meaning in particular that the unique power series solution of the quantum differential equation (the so-called "quantum period") is given by $\sum A_n z^n/n!$, where $\sum A_n t^n$ is the power series solution of the Picard–Fuchs equation on the mirror side. We will describe this correspondence in detail for the case of the 17 rank one Fano 3-folds, for one of which the numbers A_n are precisely the Apéry numbers with which we began this article.

We will give only brief descriptions of Gromov–Witten theory, mirror symmetry, quantum cohomology and quantum differential equations, referring the reader to the expositions in [20], [36], [28], [47], [52] and [56] (in roughly increasing order of difficulty).

Fano manifolds and their mirrors

We recall that a *Fano n-fold* is by definition a smooth n-dimensional complex manifold F whose anticanonical class $-K$ is ample. (For a topologist this would be expressed as the positivity of the first Chern class $c_1 = c_1(F)$, since $K = -c_1$.) Examples in all dimensions are given by complex projective spaces (or more generally Grassmannians and flag varieties) and their products. The only Fano curve is the projective line $\mathbb{P}^1(\mathbb{C})$. There are precisely ten Fano surfaces, otherwise known as del Pezzo surfaces, namely $\mathbb{P}^1 \times \mathbb{P}^1$ and dP_n (the blow-up of the projective plane $\mathbb{P}^2$ in $9-n$ points in general position) for $1 \le n \le 9$. Notice that these are actually families of surfaces, since dP_n for $n \le 4$ has a positive-dimensional moduli space corresponding to the positions of the points in the plane that are blown up. In dimension 3 there are exactly 105 Fano manifolds up to deformation, as was proved in the early 1980's by Mori and Mukai by a very subtle analysis (and as an extremely impressive illustration of the power of Mori's theory of extremal rays). The Fano 3-folds F with Picard rank ρ (which here is equal to the second Betti number $\dim H^2(F,\mathbb{Q})$) equal to 1 had been classified a few years earlier by Iskovskikh, who showed that there were precisely 17 of them (all but one of which had already been known to Fano). This will be reviewed below. In dimension 4 thousands of examples are known but there is no complete classification.

If F is a Fano n-fold, then the adjunction formula implies that its anticanonical divisors (divisors whose homology class is the Poincaré dual of $-K$) are Calabi–Yau manifolds. According to the mirror symmetry philosophy – the experts assure me that it would be premature to call it a well-defined theory – there should be a mirror dual family $X = \{X_t\}_{t\in S}$ of Calabi–Yau $(n-1)$-folds whose associated Picard–Fuchs differential equation is the Laplace transform of the quantum differential equation of F, which we will describe in a moment. In the case $n = 3$ in which we will be most interested, this will be a family of K3 surfaces of Picard number $20 - \rho$, where ρ is the Picard number of F, which can range from 1 to 6. The dimension of the moduli space of anticanonical divisors in F is also $20 - \rho$, while the dimension of the base space S of the mirror family is equal to ρ. For instance, the moduli space of quartic hypersurfaces in $\mathbb{P}^3$ has dimension $\binom{7}{3} - 4^2 = 19$, and the mirror dual is the quartic analogue in $\mathbb{P}^3$ of the quintic pencil (3.5) in $\mathbb{P}^4$ studied in Section 3. In the $\rho = 1$ cases that we will be especially interested in, the base space S of the mirror family is always $\mathbb{P}^1(\mathbb{C})$ (or more correctly, since the periods here always have a modular parametrization, a moduli curve of genus 0 that has been identified with $\mathbb{P}^1(\mathbb{C})$ by choosing a Hauptmodul).

We next have to describe the mirror of F, which should be a family of $(n-1)$-dimensional Calabi–Yau manifolds. In our case this family will always be given by a Landau–Ginzburg model, i.e., there is a Laurent polynomial L in n variables and the Calabi–Yau manifolds are the fibres of the map $L : \mathbb{C}^{*n} \to \mathbb{C}$. The relation between F and its mirror family can be described at many levels, e.g., as an isomorphism between the derived category of coherent sheaves on F and an appropriate Fukaya category on the Landau–Ginzburg side. We will use a more elementary description in terms of the differential equations associated to both objects. On the Fano side this is the quantum differential equation, which we now recall, and on the mirror side it is the Picard–Fuchs equation that we have been studying throughout the paper.

The quantum differential equation associated to F is defined in terms of its ("small") quantum cohomology ring. We will not give the complete definitions, since they play no role for us; a short description in the rank 1 case is given in Section 4 of [31] and more detailed expositions can be found in [16] and in the references listed at the beginning of the section. Very briefly and very roughly, the quantum cohomology ring is the vector space $H^*(F;\mathbb{Q}) \otimes \mathbb{Q}[z]$ (at least in the cases with $\rho = 1$; if $\rho > 1$ then one has to take z to be a multi-variable of length ρ) equipped with an associative multiplication $\star$ extending the usual cup product (the specialization to $z = 0$) that is defined in terms of the genus 0 Gromov–Witten invariants of F (the counting functions of rational curves in F intersecting divisors with given homology classes and having a given image in $H_2(F;\mathbb{Z})$). This data can be encoded in the form of a first-order vector-valued differential equation on $H^*(F;\mathbb{Q})[z]$ or in terms of an ordinary differential equation with respect to z. The quantum period $G_F(z)$, which is the unique solution in $\mathbb{Q}[[z]]$ of this differential equation, is then given by $\sum_{n=0}^{\infty} a_n z^n$, where a_n can be thought of as some kind of "volume" of the moduli space $\mathcal{M} = \mathcal{M}_{0,1,n}(F)$ of morphisms f of anticanonical degree n from a genus 0 curve with one marked point x_0 to F. (Somewhat more precisely, $a_n = \int_{\mathcal{M}} \psi^{n-2} \mathrm{ev}^*([\mathrm{pt}])$, where ψ is the first Chern class of the line bundle on $\mathcal{M}$ whose fibre at $[f]$ is given by the cotangent bundle of the curve at $f(x_0)$ and $\mathrm{ev} : \mathcal{M} \to F$ is the evaluation map $f \mapsto f(x_0)$.) The duality between the Fano variety F and its mirror $L : \mathbb{C}^{*n} \to \mathbb{C}$ is then summarized in the equality $a_n = A_n/n!$, where A_n is the constant term of L^n. Since the generating series $\sum A_n t^n$, as we have seen, is then a period of the family $\{X_t = L^{-1}(t)\}_t$ and is a solution of the associated Picard–Fuchs differential equation, the relationship between the two sides can also be expressed by saying that the quantum differential equation is the Laplace transform of the Picard–Fuchs differential equation.

In [29] Golyshev made a precise conjecture giving the mirror duals for the 17 Fano 3-folds of the Iskovskikh classification, and all cases of this were proved in that paper and in the subsequent article [58] by Przyjalkowski. We will describe these results in the next subsection. The corresponding results for the remaining 88 Fano 3-folds of the Mori-Mukai classification were conjectured by Tom Coates, Alessio Corti, Sergey Galkin and Alexander Kasprzyk together with Golyshev and proved in all cases in their (physically and mathematically) huge article [16], which gives new explicit descriptions of the Fano varieties as well as a Laurent polynomial defining the mirror family in every case. For example, opening [16] at random on page 220, one finds that the Fano variety $F = \mathrm{MM}_{3\text{–}19}$ (= the 19th of the Fano 3-folds with Picard number $\rho = 3$ in the Mori-Mukai classification) has quantum period given by

$$G_F(z) = e^{-2z} \sum_{m \ge \ell \ge 0} \binom{2m}{m} \frac{z^{m+2\ell}}{\ell!^3 (m-\ell)!} .$$

If we write this as $\sum A_n z^n/n!$, where $(A_0, A_1, \dots) = (1, 0, 2, 12, 54, 240, 1280, \dots)$, then A_n is the integer defined by the constant-term formula (1.11) with

$$L = L(x,y,z) = xz + x + y + z + \frac{1}{x} + \frac{1}{yz} + \frac{1}{xyz} ,$$

and the mirror family of F is given by the Landau–Ginzburg model $t\,L(x,y,z) = 1$. Similar results are given for 738 Fano 4-folds in [17].

The 17 rank one Fano 3-folds and their mirrors

As already said above, there are exactly 17 families of Fano 3-folds with Picard number $\rho = 1$, as classified by Iskovskikh in 1977–78. The families are labeled by two numerical invariants, the *index* $d = [H^2(F;\mathbb{Z}) : \mathbb{Z}\,c_1]$, where $c_1 = c_1(F) \in H^2(F;\mathbb{Z})$ is the first Chern class of the tangent bundle of F, and the *level* N, defined as $\langle c_1^3, [F]\rangle/2d^2$, which is always a positive integer. The 17 possible pairs (d,N) are given by the table

d	1	2	3	4
N	1, …, 9, 11	1, …, 5	3	2

Each of the corresponding Fano varieties has a name and an algebraic description. For instance, the projective space $\mathbb{P}^3$ corresponds to the last entry $(d,N) = (4,2)$ of the table, while the Fano 3-fold corresponding to $(d,N) = (1,6)$, which is the one related to the Apéry numbers, is called V_{12} and is defined by starting from the Fano 10-fold $G(10,5)$ (orthogonal Grassmannian of 5-dimensional isotropic subspaces in $\mathbb{C}^{10}$) and then taking generic hyperplane sections 7 times in a row to reduce the dimension to 3.

We now give the description of the mirrors of these varieties as found by Golyshev, though in a somewhat modified form taken from [31]. This description is completely modular. For every integer $N \geq 1$ we have the congruence group $\Gamma_0(N)$ and corresponding modular curve $X_0(N) = X_{\Gamma_0(N)}$, as described in Section 4. This curve has an interpretation as the moduli space of ordered pairs (E, E') of elliptic curves together with a cyclic isogeny of degree N from E to E'. (In terms of $\tau \in \mathbb{H}$ these are given by $E = \mathbb{C}/(\mathbb{Z}\tau + \mathbb{Z})$ and $E' = \mathbb{C}/(N\mathbb{Z}\tau + \mathbb{Z})$, with the obvious map.) There is an involution W_N on $X_0(N)$ corresponding to interchanging E and E', given in terms of the parameter τ by $\tau \mapsto -1/N\tau$ (Fricke involution), so we can consider the quotient $X_0^*(N) = X_0(N)/W_N$, which is the moduli space of unordered pairs of cyclically N-isogenous elliptic curves. To the point of $X_0^*(N)$ corresponding to such a pair $\{E, E'\}$ we can associate the Kummer surface obtained by dividing the abelian surface $E \times E'$ by the involution $(u, u') \mapsto (-u, -u')$ and blowing up the 16 resulting singularities. These Kummer surfaces generically have Picard number 19 as desired, because there are 3 linearly independent algebraic cycles in $E \times E'$ given by the classes of E and E' and the graph of the isogeny between them, and 16 further linearly independent cycles on the Kummer surface coming from the blow-ups of the singularities (exceptional curves).

Next we must single out special values of N, as well as of an auxiliary integer d corresponding to a covering of $X_0^*(N)$. For every $N > 1$, there is a modular form $F_N(\tau)$ defined as the unique Eisenstein series of weight 2 that equals 1 at ∞ (i.e., has a q-expansion beginning $1 + \mathrm{O}(q)$), is anti-invariant under W_N (meaning that $F_N(-1/N\tau) = -N\tau^2 F_N(\tau)$), and vanishes at all of the cusps of $X_0(N)$ other than 0 and ∞. For $N = 1$ this definition makes no sense, since all modular forms of level 1 are invariant under W_1. In this case we set $F_1 = \sqrt{E_4}$ (which is not quite a modular form, or even a well-defined function in $\mathbb{H}$, but works well anyway; cf. Example 2 of Section 5). We then consider the cases when $X_0^*(N)$ has genus 0 and a Hauptmodul t_N for $\Gamma_0^*(N)$ is given by the formula

$$t_N(\tau) = \left(\frac{\eta(\tau)^2\,\eta(N\tau)^2}{F_N(\tau)}\right)^{\frac{12}{N+1}},$$

the strange-looking exponent $\frac{12}{N+1}$ being forced by the requirement that $t_N(\tau)$ has an expansion at ∞ beginning $q + \mathrm{O}(q^2)$. It turns out that there are precisely ten values of N for which this happens, namely $N = 1, \ldots, 9$ and 11, and precisely 17 pairs (d, N) for which also the Hauptmodul $t_N(\tau)$ has a dth root, these being exactly the 17 pairs of the Iskovskikh classification. The result conjectured by Golyshev and proved by him and Przyjalkowski is that the corresponding families of K3 surfaces, with periods $\Phi_{N,d}(t) \in \mathbb{Q}[[t]]$ defined by the modular parametrization $F_N(\tau) = \Phi_{N,d}(t_N(\tau)^{1/d})$, are the mirror duals in the sense explained above of the Fano 3-folds of the Iskovskikh classification. Each of the 17 power series $\Phi_{N,d}(t)$ has Taylor coefficients A_n given by the formula (1.11) for a suitable Laurent polynomial L, and the differential equations that they satisfy are the ones that were discussed in Examples 3 and 19–34 of Section 5.

9 Differential equations and topology

In this final main section of the paper we describe a conjecture due to Galkin, Golyshev and Iritani relating the asymptotic behavior at infinity of the solutions of the quantum differential equation of a Fano manifold with the so-called Gamma class of this manifold. Its subject matter is thus a direct continuation of that of Section 8, but we have put it into a section of its own because there is no direct relation to mirror symmetry (both sides of the conjecture involve the Fano variety, not its mirror manifold) and also because it makes a more direct connection with the word "topology" in the title of the paper. Specifically, the Gamma class is a universally defined multiplicative characteristic class in the sense of Hirzebruch, so that this section also provides another link between our topic and his work.

The Gamma class

We begin by recalling Hirzebruch's definition of multiplicative characteristic classes [37], one of his most beautiful and fruitful discoveries. Let $f(x)$ be a power series with constant term 1 and coefficients in a ring K (say $\mathbb{Q}$ or $\mathbb{R}$). Then for any complex vector bundle E over a base space X we define a characteristic class $\chi_f(E) \in H^*(X;K)$ by the formula

$$\chi_f(E) = \prod_j f(\alpha_j)\,, \tag{9.1}$$

where the α_j are formal degree 2 cohomology classes such that the total Chern class $c(E) \in H^*(K)$ factors as $\prod(1+\alpha_j)$. The α_j can be interpreted topologically to some extent by finding a pullback of E under some map $X' \to X$ that splits as a sum of line bundles and thinking of the pull-backs of the α_j as the first Chern classes of these, but this is not necessary: if we simply multiply out the power series on the right-hand side of (9.1), then the degree $2d$ part of the expansion obtained is a symmetric homogeneous polynomial of degree d in the α's and hence a weighted homogeneous polynomial in their elementary symmetric polynomials $c_i(E)$, with coefficients in K, and therefore belongs to $H^{2d}(X;K)$. If X is a smooth complex manifold and we take for E its tangent bundle, we write simply $\chi_f(X)$ instead of $\chi_f(T(X))$. Hirzebruch showed that one can obtain important topological invariants of X by evaluating $\chi_f(X)$ on the fundamental class of X for suitable power series f (genera) or by multiplying it by the Chern character of some bundle over X and then evaluating on $[X]$ (as in the Hirzebruch–Riemann–Roch theorem). The most important examples here were those associated to the three power series $f(x) = \frac{x}{\tanh x}$, giving the Hirzebruch L-class and the signature theorem, $f(x) = \frac{x}{1-e^{-x}}$, giving the Todd genus, and $f(x) = \frac{x/2}{\sinh(x/2)}$, giving the $\widehat{A}$-genus. Note that the last two of these power series differ only by a factor $e^{x/2}$, so that the corresponding characteristic classes differ only by a factor $e^{c_1(E)/2}$.

If one now looks at the last of the above power series and remembers Euler's formula $\frac{\pi x}{\sin \pi x} = \Gamma(1+x)\Gamma(1-x)$, then it is natural to introduce the *Gamma class*

$\widehat{\Gamma}_X \in H^*(X;\mathbb{R})$ associated to the power series $f(x) = \Gamma(1+x)$. Euler's formula then implies that the $\widehat{A}$-class of X (or its Todd class, up to a factor $e^{c_1(X)/2}$) factors as the product of $\widehat{\Gamma}_X(-1)$ and its complex conjugate, where the "jth Tate twist" $\xi(j)$ of a cohomology class $\xi \in H^{\mathrm{ev}}(X)$ is defined by multiplying its degree d part by $(2\pi i)^{jd}$ for all d. Thus we can think of the gamma class of X as some sort of a square-root of its Todd class, and the authors of the paper [26] in which the Gamma Conjecture is formulated describe the conjecture as a kind of square-root of the index theorem.

Since $\Gamma(1+x)$ has an expansion beginning

$$\begin{aligned}\Gamma(1+x) &= \exp\Big(-\gamma x + \sum_{n\geq 2} \frac{(-1)^n \zeta(n)}{n}\, x^n\Big)\\ &= 1 - \gamma x + \frac{\gamma^2+\zeta(2)}{2}\, x^2 - \frac{\gamma^3 + 3\gamma\zeta(2) + 2\zeta(3)}{6}\, x^3 + \cdots,\end{aligned}$$

we have

$$\begin{aligned}\widehat{\Gamma}_X = 1 \,-\, \gamma\, c_1 &+ \Big(-\zeta(2)\, c_2 + \frac{\zeta(2)+\gamma^2}{2}\, c_1^2\Big)\\ &+ \Big(-\zeta(3)\, c_3 + (\zeta(3) + \gamma\,\zeta(2))\, c_1 c_2 \,-\, \frac{2\zeta(3) + 3\gamma\,\zeta(2) + \gamma^3}{6}\, c_1^3\Big) + \cdots,\end{aligned}$$

where $c_i = c_i(X) \in H^{2i}(X)$ are the Chern classes of X. This formula simplifies a lot if we introduce the *modified gamma class* $\widehat{\Gamma}_X^0$, defined by

$$\widehat{\Gamma}_X = \Gamma(1+c_1)\, \widehat{\Gamma}_X^0, \tag{9.2}$$

in which case it reduces to

$$\widehat{\Gamma}_X^0 = 1 - \zeta(2)\, c_2 + \zeta(3)\,(c_1c_2 - c_3) + \cdots. \tag{9.3}$$

Note that $\widehat{\Gamma}_X = \widehat{\Gamma}_X^0$ if X is a Calabi–Yau manifold, since then $c_1(X) = 0$. In any case, there is a characteristic appearance of the number $\zeta(3)$ in any formula involving threefolds and the Gamma class. Such formulas have played a role in string theory in recent years, the process not having been entirely painless since certain formulas that were thought to have been established turned out to be wrong until they were corrected by incorporating the Gamma class.

The Gamma Conjecture for Fano varieties

We can now formulate the Gamma Conjecture for Fano varieties (actually one of two "Gamma Conjectures" stated in [26], but we will not discuss the other). We concentrate mainly on the case of Fano 3-folds F of Picard rank one, for which the cohomology ring $H^*(F;\mathbb{Q})$ is simply $\mathbb{Q}[c_1]/(c_1^4 = 0)$. The relationship between the quantum cohomology of F and its quantum differential equation is such that the 4-dimensional space of solutions of the latter can be canonically identified with the dual space $H_*(F;\mathbb{C})$ of the cohomology ring of F, so any linear functional κ assigning to each element Ψ of the solution space a complex number $\kappa(\Psi)$ can be thought of as an element of $H^*(F;\mathbb{C})$. In particular, since all solutions of the quantum differential equation grow at infinity like the sum of a multiple of the holomorphic solution $G_F(z)$ and a term of exponentially lower order, we can take κ to be the asymptotic limit functional $\kappa(\Psi) = \lim\limits_{z\to\infty} \frac{\Psi(z)}{G_F(z)}$. The Gamma Conjecture then says that the cohomology class of F corresponding to this functional is the Gamma class of F.

This statement can be made more explicit in the cases where $H^*(F)$ is generated by c_1 by using the Frobenius basis of solutions.[5] These are the four functions $\Psi_i(z)$ ($0 \le i \le 3$), where $\Psi_i(z) \in \mathbb{Q}[[z]][\log z]$ is defined for all $i \ge 0$ by the expansion

$$\sum_{n=0}^{\infty} a_n(\varepsilon)\, z^{n+\varepsilon} = \sum_{i=0}^{\infty} \Psi_i(z)\, \varepsilon^i\,, \tag{9.4}$$

with $a_n(\varepsilon) \in \mathbb{Q}(\varepsilon)$ defined by the same recursion as that satisfied by the coefficients a_n of the quantum period $G_F(z) = \sum a_n z^n$ itself, but with n replaced by $n+\varepsilon$ and with initial conditions $a_0(\varepsilon) = 1$, $a_n(\varepsilon) = 0$ for $n < 0$. For instance, in the Apéry case $F = V_{12}$ one has

$$\begin{aligned} \Psi_0(z) &= G_F(z) = 1 + 5z + \frac{73z^2}{2} + \frac{1445z^3}{6} + \cdots\,, \\ \Psi_1(z) &= G_F(z)\, \log z + 7z + \frac{201z^2}{4} + \frac{10733z^3}{36} + \cdots\,, \end{aligned}$$

and in general $\Psi_i(z) = G_F(z)(\log z)^i/i!$+(lower order terms) for all $i \ge 0$. The *ith Frobenius limit* κ_i is then defined as the limit of $\Psi_i(z)/G_F(z)$ as $z \to \infty$. (Thus κ_0 is always 1, since $\Psi_0 = G_F$.) In terms of these numbers, the Gamma Conjecture says simply that the Gamma class of F equals $\sum_{i=0}^{3} \kappa_i c_1^i$. The Gamma class is easily computed by purely topological considerations, so this gives an explicit prediction for the values of the Frobenius limits, e.g. $\kappa_1 = -\gamma$, $\kappa_2 = \frac{\gamma^2-3\zeta(2)}{2}$ and $\kappa_3 = \frac{-\gamma^3+9\gamma\zeta(2)+15\zeta(3)}{6}$ in the Apéry case. These formulas simplify to $\kappa_1^0 = 0$, $\kappa_2^0 = -2\zeta(2)$ and $\kappa_3^0 = \frac{17}{6}\zeta(3)$ if we replace the Frobenius limits by the corresponding limits for the Laplace-transformed differential equation satisfied by the generating function of the original Apéry numbers, in which case the Gamma Conjecture becomes $\hat{\Gamma}_F^0 = \sum_{i=0}^{3} \kappa_i^0 c_1^i$.

The Gamma Conjecture was proved by its authors for projective spaces, toric manifolds, and certain toric complete intersections and Grassmannians, and in [31] for all of the Fano 3-folds with $\rho = 1$ (some cases of which were already known previously by work of Dubrovin and others). Actually, two methods of proof were given in [31]. The first is combinatorial and proceeds by giving explicit formulas for the coefficients of the power-series parts of the Frobenius solution $\Psi_i(z)$, involving the harmonic numbers $1 + \frac{1}{2} + \cdots + \frac{1}{n}$ and the nth partial sums of $\zeta(k)$ for $2 \le k \le i$, while

5. The insight that the Gamma class gives the discrepancy between the Frobenius and the integral basis of the solution space of a Picard–Fuchs differential equation, already mentioned in §7 in the hypergeometric case, is due to Katzarkov–Kontsevich–Pantev [48] and Iritani [41].

the second is based on the modular parametrizations of the power series involved, and more specifically on the properties of Eichler integrals of weight 4 Eisenstein series. The first method works cleanly in all hypergeometric cases (which includes 10 of the 17 Iskovskikh cases), but is messy in general and was only worked out in detail in [31] for the case $F = V_{12}$ corresponding to the Apéry numbers. However, it has the advantage of working in the two cases $(d, N) = (1, 1)$ or $(2, 1)$ of the Iskovskikh classification for which the modular proof fails because $F_1(\tau)$ is not a modular form, and also of being potentially applicable in higher-dimensional situations, where modularity is almost never available. The modular proof is much smoother and works in a completely uniform way in all 15 cases to which it applies.

These calculations contained one nice surprise. The Frobenius functions $\Psi_i(z)$ exist even for $i > 3$, even though they are then no longer solutions of the differential equation satisfied by the quantum period, and the Frobenius limits κ_i and κ_i^0 are therefore well-defined real numbers also for these i. In the course of the calculations with Golyshev, I calculated them numerically to high precision for $i \le 11$ and looked whether they could also be written, like the numbers κ_2^0 and κ_3^0, as polynomials in Riemann zeta-values with rational coefficients. This turned out to be true for i up to 10, but false for κ_{11}^0, which was instead a rational linear combination of products of zeta values and of the *multiple zeta value*

$$\zeta(3,5,3) = \sum_{0<\ell<m<n} \frac{1}{\ell^3 m^5 n^3} = 0.002630072587647\cdots .$$

Multiple zeta values are old friends of mine and I was very pleased to see one show up here, but I could not understand why the first appearance was only in weight 11, rather than in weight 8 (the first case where not all multiple zeta values are expressible in terms of Riemann zeta values). I showed Golyshev my numerical discovery and proffered the conjecture that this must be an extremely deep fact and that it would probably take many decades until anybody could explain why things changed only at the value 11. This turned out to be one of the least accurate conjectures I had ever made, since he gave the answer within seconds rather than decades: it had to be related to the fact that the corresponding Fano variety V_{12} can be “unsectioned” seven times (cf. the geometric description of this variety as given in the previous section) to give a Fano 10-fold, but cannot be unsectioned an eighth time without introducing singularities. Since the Gamma conjecture is supposed to be true for Fano varieties of any dimension and only involves Riemann zeta-values, this explained the phenomenon that I had found numerically, and indeed in the subsequent weeks Golyshev was able to compute the Gamma classes of $G(10,5)$ and its successive sections and to verify that the numbers obtained were the same as the ones that I had found on my computer. This also provides a numerical verification of the Gamma conjecture for a number of Fano varieties of dimension going up to 10 (though not a proof since the higher Frobenius limits have only been evaluated numerically and not proved). In any case, the nature of the numbers κ_i^0 for $i \le 10$ can now be considered to be understood, but the appearance of $\zeta(3,5,3)$ in κ_{11}^0 remains a mystery requiring new insights.

10 Miscellaneous examples, open questions, and remarks

In this section we describe a few miscellaneous topics that belong to our subject but did not fit naturally into any of the main subjects treated so far, and also mention a couple of open questions suggested by the results discussed here.

Two further connections with mirror symmetry

We start by describing two specific results that relate between mirror symmetry or Gromov–Witten theory to differential equations and therefore could in principle have been included in Section 8, but that are of an entirely different nature from the material there.

The first is a rather odd statement that was needed by Aleksey Zinger for his proof of the Bershadsky–Cecotti–Ooguri–Vafa mirror symmetry predictions for the genus 1 Gromov–Witten invariants of a quintic 3-fold in $\mathbb{P}^4$ (or more generally of a hypersurface of degree d in $\mathbb{P}^{d-1}$ for any $d > 0$), and that we proved in our joint paper [77]. Let $\mathcal{P}$ be the group of power series in t with coefficients in $\mathbb{Q}(\varepsilon)$ that have constant term 1 and no pole at $\varepsilon = 0$. Then the hypergeometric deformation of the power series $\sum \frac{(5n)!}{n!^5}\, t^n$ given by

$$\sum_{n=0}^{\infty} \frac{\prod_{r=1}^{5n}(r+5\varepsilon)}{\prod_{r=1}^{n}\big((r+\varepsilon)^5-\varepsilon^5\big)}\, t^n$$

is a fixed point of the 5th power of the non-linear map

$$F(t,\varepsilon) \;\mapsto\; \left(1+\frac{t}{\varepsilon}\,\frac{\partial}{\partial t}\right)\frac{F(t,\varepsilon)}{F(t,0)}$$

from $\mathcal{P}$ to itself, and similarly with “5” replaced by any positive integer d.

The second example comes from the paper [3] by Marco Bertola, Boris Dubrovin, and Di Yang in which the authors find power series satisfying linear differential equations whose coefficients are defined by integrals over suitable moduli spaces, but now with the summation being over genera rather than over the degrees of maps from a genus 0 curve to a target space as in the case of the quantum period that we discussed in Section 8. One of their series begins

$$\sum_{n=0}^{\infty} c_n\, t^n = 1 - \frac{161}{2^{10}3^5}\, t + \frac{26605753}{2^{23}3^{12}5^2}\, t^2 + \cdots,$$

where c_n is defined by an integral over a moduli space (more precisely, up to a simple factor it is the integral over the moduli space $\overline{\mathcal{M}}_{5n,1}$ of stable 1-pointed curves of genus $5n$ of the product of ψ^{12n-2} with a so-called Witten 5-spin class) and satisfies the four-term recursion relation

$$\begin{aligned}&80352000\, n(5n-1)(5n-2)(5n-4)c_n\\ &+25(2592000n^4-16588800n^3+39118320n^2-39189168n+14092603)\, c_{n-1}\\ &\qquad\qquad+20\,(4500n^2-18900n+19739)\, c_{n-2}+c_{n-3} \;=\; 0\,.\end{aligned}$$

When I saw these numbers, which decay roughly like $1/n!^2$, I naturally asked whether they might share with the coefficients of the quantum periods discussed in Section 8 the property that when they are multiplied by $n!^2$, or possibly by the product of two Pochhammer symbols, they become integers (which would then potentially be the coefficients of the power series solution to some Picard–Fuchs differential equation). This indeed turned out to be the case, but somewhat surprisingly in two different ways: Yang and I found a formula showing that the numbers $a_n := (2^{10}3^5 5^4)^n (\frac{3}{5})_n (\frac{4}{5})_n c_n$ are integers of exponential growth (and hence can be expected to have a generating series that is a period, although we have not succeeded in finding it), and Dubrovin and Yang found that the numbers $b_n := (2^{12}3^5 5^4)^n (\frac{2}{5})_n (-\frac{1}{10})_n c_n$ are *also* integral and that in this case the generating function $\sum b_n t^n$ is not only of Picard–Fuchs type, but is actually algebraic! So this is a very mysterious example from both the mirror symmetry point of view and from the point of view of the elementary number-theoretical (divisibility) properties of numbers defined by recursions with polynomial coefficients.

Some open questions

We next list a few questions, some well known and some less so, that are suggested by the results and observations discussed in the main body of the paper.

1. How can one recognize whether a given differential equation is geometric in origin, i.e., whether it can arise as the Picard–Fuchs equation of the periods in some algebraic family? In particular, if $\{A_n\}$ is a sequence of integers of at most exponential growth satisfying a linear recursion of finite length with polynomial coefficients, is it always of Picard–Fuchs type? This question was already mentioned in Section 2, but is so basic to our theme that it seems worth emphasizing.
2. In a related direction, given a sequence $\{A_n\}$ of integers as above, how can we recognize whether they can be defined as the constant terms of the powers of a Laurent polynomial, as was the case for the Apéry numbers and for all of the Picard–Fuchs equations discussed in connection with mirror symmetry? In some cases one can in fact *exclude* the existence of such a representation, because an equation like (1.11) implies certain obvious congruences like $A_p \equiv A_1 \pmod{p}$ (by Fermat's little theorem) as well as much less obvious ones such as the Lucas-like congruences given in [61] and [54], and if these fail for a given sequence $\{A_n\}$ then there can be no representation of this type. As an example, the integrality of the solutions of the Bouw-Möller recursion discussed in the final example of Section 5, where non-elementary proofs using p-adic analysis or the theory of Hilbert modular forms were described, *cannot* be proved in an elementary way by a formula like (1.11), because already the Fermat-like congruence for primes splitting in $\mathbb{Q}(\sqrt{17})$ fails. But in many cases one knows that there is a formula for the sequence of coefficients as constant terms of powers of some polynomial L, and the problem of finding this polynomial algorithmically remains. Note, by the way, that the problem is only to *find* the polynomial L, not to prove that it works,

since once one has a candidate it is an elementary procedure to find the recursion for the constant terms of its powers, and if this recursion and its initial values agree with those for the A_n, then the required identity is true.

As well as asking about the existence of a Laurent polynomial producing a given sequence of constant terms, one can ask about its uniqueness. It is known that different polynomials can give the same sequence of numbers, if they are obtained from one another by a sequence of so-called mutations, and several authors (e.g., Galkin and Usnich [27]) have studied the question whether the converse of this statement is also true. This is not known and seems very hard, but in any case the question of having a criterion for the *existence* of a Laurent polynomial, or equivalently of a Landau–Ginzburg model, seems even more fundamental than the question of its *uniqueness*.

3. Again in a related direction, given a sequence of rational numbers defined by a recursion with polynomial coefficients, is there any criterion to determine whether this sequence can be multiplied by a quotient of products of Pochhammer symbols to obtain a new sequence that is integral (perhaps up to a factor M^n for some fixed $M \in \mathbb{N}$) and has exponential growth? In this case, and if the answer to the first question is positive, one would have a relation to periods and to algebraic geometry. Also interesting is the extent to which this modified sequence is unique. The Bertola-Dubrovin-Yang example described above shows that the answer to this latter question is not completely trivial.
4. The Picard–Fuchs equations associated to families of elliptic curves or families of K3 surfaces are always modular (at least in practice; I do not know whether there is any theorem to this effect), but for families of higher-dimensional Calabi–Yau varieties, like the mirror quintic family with period $\sum \frac{(5n)!}{n!^5} t^n$, this is known not to be true. (The proof is easy: the differential equation satisfied by a modular form of weight k with respect to a modular function as independent variable is always the kth symmetric power of a second order equation, as mentioned in Example 2 in Section 5, but the differential equation satisfied by $\sum \frac{(5n)!}{n!^5} t^n$ is not a symmetric cube.) Is there nevertheless some non-trivial relation between the periods in this case and any kind of automorphic functions or forms? The Bouw-Möller equation discussed at the end of Section 5 is an example of a differential equation that cannot be parametrized directly by modular forms on an arithmetic group (of finite index of $SL(2, \mathbb{Z})$), but which nevertheless, as we saw, embeds into a higher-dimensional modular variety (in this case, a Hilbert modular surface) in which such a parametrization exists. So one can at least wonder whether there can be any kind of correspondence between the Calabi–Yau varieties and some automorphic varieties that relates the periods of the former to automorphic quantities.
5. The final question concerns the integrality of Gromov–Witten invariants. If the mirror story is to be true, then the quantum period $\sum a_n z^n$ of a Fano variety as discussed in Section 8 must have coefficients a_n with denominator at most $n!$, since they are supposed to be given by $a_n = A_n/n!$ with A_n defined as in (1.11) for some Laurent polynomial L. Can one prove this integrality directly from Gromov–Witten theory? That the denominator is *at least* $n!$ is to be expected, since a_n

is defined as the evaluation of a certain cohomology class on the fundamental class of the moduli space $X_{0,1,n}$ and this moduli space is a stack some of whose points can have a stabilizer of order as large as $n!$ (namely, the points given by composing a degree 1 map from $\mathbb{P}^1$ to F with a degree n map from $\mathbb{P}^1$ to itself, where the Galois group of the latter can be the full symmetric group on n letters). Apparently no geometric argument is known showing that $n!a_n$ is always integral, but there seems to be a possibility of showing at least that it cannot have more than exponential growth, e.g., (in the Fano 3-fold case) that it is a divisor of l.c.m.$\{1^3, \dots, n^3\}$ rather than merely of $n!^3$, which is all that one gets directly from the recursive formula.

Higher dimensions

Throughout this paper we have concentrated on ordinary differential equations, whose solutions are functions of a single complex variable, so that in the geometric situation we were studying families defined over a curve (and indeed almost always over $\mathbb{P}^1(\mathbb{C})$, since then the associated differential equations have polynomial rather than algebraic coefficients and the coefficients satisfy a recursion with polynomial coefficients). One exception was in Section 8, where the quantum differential equations associated to Fano 3-folds with Picard rank $\rho > 1$ involve ρ arguments, and similarly for the Picard–Fuchs differential equations on the mirror sides since ρ is the dimension of the corresponding family, but even here we used this higher-dimensional system of differential equations to construct a power series in a single variable, the quantum period, whose coefficients still satisfied a recursion with polynomial coefficients. Of course higher-dimensional situations are also extremely interesting, but much harder to study. I would like to end the paper by saying a few words about them, especially because, as mentioned in the opening paragraph of the paper, this was a subject that was of great interest to Hirzebruch and that he studied actively in the last decades of his life. For reasons of both space and competence I will say only a few words about it here.

A particularly beautiful result of Hirzebruch's is his proportionality principle, which says that the Chern numbers of the compact quotient of a bounded symmetric domain X by a properly discontinuous and fixed-point-free group action are proportional to the Chern classes of the compact dual X' (a compact algebraic manifold into which X is naturally embedded). If X is the complex 2-ball, then $X' = \mathbb{P}^2(\mathbb{C})$, which has Chern numbers $c_1^2 = 9$, $c_2 = 3$, so any compact quotient B_2/Γ, where Γ is a group acting freely, has Chern numbers satisfying $c_1^2 = 3c_2$. Conversely it was proved by Miyaoka and Yau that one has $c_1^2 \leq 3c_2$ for any compact complex surface of general type, with equality only for quotients of the 2-ball. One is particularly interested in examples of non-arithmetic groups acting freely and with compact quotients. There are three main sources of these: certain groups generated by complex reflections, surfaces obtained as coverings of $\mathbb{P}^2(\mathbb{C})$ ramified along lines, and finally monodromy groups of generalized (Appell-Lauricella) hypergeometric functions. Hirzebruch made an intensive study of the second class [1], deriving in particular from the Miyaoka-Yau inequality an inequality concerning the combinatorics of line arrangements in the plane that was stronger than anything that had been obtained by elementary methods and was later applied to prove the so-called "bounded negativity conjecture" about such line configurations. There are many links between the three classes, as one can read in detail in [2], in the book [22] of which it is a review, and in the book [68] that, as mentioned in the introduction, was originally an outgrowth of a course that Hirzebruch gave on the subject in 1996. Of particular interest in connection with the present article are the discussions of the classical hypergeometric differential equations in Chapter 2 and of the Appell hypergeometric functions and their associated monodromy groups in Chapter 7 of [68]. I say no more here except that the whole field is still very active, a very recent example being a new construction by Martin Deraux [23] of non-arithmetic lattices via coverings of line arrangements.

References

[1] Gottfried Barthel, Friedrich Hirzebruch and Thomas Höfer, *Geradenkonfigurationen und algebraische Flächen*. Aspects of Mathematics **D4**, Vieweg, Braunschweig (1987).

[2] Paula Beazley Cohen and Friedrich Hirzebruch, Book review: *Commensurabilities among lattices in $PU(1,n)$*, by Pierre Deligne and G. Daniel Mostow. *Bull. AMS* **32** (1995), 88–105.

[3] Marco Bertola, Boris Dubrovin and Di Yang, Simple Lie algebras and topological ODEs. To appear in IMRN, 36 pages. arXiv:1508.03750 (2015).

[4] Frits Beukers, A note on the irrationality proof of $\zeta(2)$ and $\zeta(3)$. *Bull. London Math. Soc.* **11** (1979), 268–272.

[5] Frits Beukers, Irrationality proofs using modular forms. In *Journées arithmétiques de Besançon (Besançon, 1985)*, Astérisque **147–148** (1987), 271–283, 345.

[6] Frits Beukers, Some Congruences for the Apéry Numbers. *J. Number Theory* **21** (1985), 141–155.

[7] Frits Beukers, Another congruence for the Apéry Numbers. *J. Number Theory* **25** (1987), 201–210.

[8] Frits Beukers, On B. Dwork's accessory parameter problem. *Math. Z.* **241** (2002), 425–444.

[9] Frits Beukers, Henri Cohen and Anton Mellit, Finite hypergeometric functions. *Pure Appl. Math. Q.* **11** (2015), 559–589.

[10] Frits Beukers and Gert Heckman, Monodromy for the hypergeometric function ${}_nF_{n-1}$. *Invent. Math.* **95** (1989), 325–354.

[11] Frits Beukers and Chris Peters, A family of K3 surfaces and $\zeta(3)$. *J. Reine Angew. Math.* **351** (1984) 42–54.

[12] Irene Bouw and Martin Möller, Differential equations associated with non-arithmetic Fuchsian groups. *J. London Math. Soc.* **81** (2010), 65–90.

[13] Jan Bruinier, Günter Harder, Gerard van der Geer and Don Zagier, *The 1–2–3 of Modular Forms: Lectures at a Summer School in Nordfjordeid, Norway*, (ed. K. Ranestad), Universitext, Springer-Verlag, Berlin–Heidelberg–New York 2008.

[14] Philip Candelas, Xenia de la Ossa, Paul Green and Linda Parkes, A pair of Calabi–Yau manifolds as an exactly soluble superconformal theory. *Nuclear Phys. B* **359** (1991), 21–74.

[15] Philip Candelas, Xenia de la Ossa and Fernando Rodriguez Villegas, Calabi–Yau manifolds over finite fields, I. Preprint arXiv:hep-th/0012233 (2000), 76 pages; II. in *Calabi-Yau varieties and mirror symmetry*, *Fields Inst. Commun.* **38** (2003), 121–157.

[16] Tom Coates, Alessio Corti, Sergey Galkin and Alexander Kasprzyk, Quantum periods for 3-dimensional Fano manifolds. *Geometry & Topology* **20** (2016) 103–256.

[17] Tom Coates, Alexander Kasprzyk and Tom Prince, Four-dimensional Fano toric complete intersections. *Proc. R. Soc. A* **471** (2015), id. 20140704.

[18] Paula Cohen and Jürgen Wolfart, Modular emeddings for some nonarithmetic Fuchsian groups. *Acta Arith.* **56** (1990) 93–110.

[19] Clay Córdova and Shu-Heng Shao, Asymptotics of ground state degeneracies in quiver quantum mechanics. *Comm. in Number Theory and Physics* **10** (2016) 339–371.

[20] David Cox and Sheldon Katz, *Mirror Symmetry and Algebraic Geometry*. Mathematical Surveys and Monographs **68**, AMS, Providence 1999.

[21] Pierre Deligne, Valeurs de fonctions L et périodes d'intégrales. In *Automorphic forms, Representations, and L-functions*, Proc. Symp. Pure Math. **33**, AMS (1979), 313–346.

[22] Pierre Deligne and G. Daniel Mostow, Commensurabilities among lattices in PU(1,n). Annals of Mathematics Studies **132**, Princeton University Press (1993), viii+183 pp.

[23] Martin Deraux, Non-arithmetic lattices and the Klein quartic. *J. reine angew. Math.* (2017), 24 pages.

[24] Tim Dokchitser, Computing special values of motivic L-functions. *Exper. Math.* **13** (2004), 137–149.

[25] Dmitry Galakhov, Pietro Longhi, Tom Mainiero, Gregory Moore and Andrew Neitzke, Wild wall crossing and BPS giants. *JHEP* **1311** (2013), 46–152.

[26] Sergey Galkin, Vasily Golyshev and Hiroshi Iritani, Gamma classes and quantum cohomology of Fano manifolds: Gamma Conjectures. *Duke Math. J.* **165** (2016), 2005–2077.

[27] Sergey Galkin and Alexandr Usnich, Mutation of polynomials. Preprint (2010), 7 pages. IPMU 10-0100.

[28] Alexander Givental, A tutorial on quantum cohomology. In: *Symplectic geometry and topology (Park City, UT, 1997)*, IAS/Park City Math. Ser. **7** (1999) 231–264.

[29] Vasily Golyshev, Classification problems and mirror duality. *London Math. Soc. Lecture Note Ser.* **338** (2007), 88–121.

[30] Vasily Golyshev and Anton Mellit, Gamma structures and Gauss's contiguity. *J. Geom. Phys.* **78** (2014), 12–18.

[31] Vasily Golyshev and Don Zagier, Proof of the Gamma Conjecture for Fano 3-folds of Picard rank one. *Izvest. Russ. Acad. Sci., Math. Series* **80** (2016).

[32] Fernando Gouvea and Noriko Yui, Rigid Calabi–Yau threefolds over $\mathbb{Q}$ are modular. *Expo. Math.* **29** (2011), 142–149.

[33] Jeremy Gray, *Linear Differential Equations and Group Theory from Riemann to Poincaré*. Birkhäuser, Boston–Basel–Stuttgart 1986.

[34] Phillip Griffiths and Joseph Harris, *Principles of Algebraic Geometry.* John Wiley & Sons, New York–Chichester–Brisbane–Toronto 1978.

[35] Alexander Grothendieck, On the de Rham cohomology of algebraic varieties. *Pub. math. de l'I.H.É.S.* **29** (1966), 95–103.

[36] Martin Guest, *From Quantum Cohomology to Integrable Systems* (Oxford Graduate Texts in Mathematics; vol. 15). Oxford University Press, Oxford 2008.

[37] Friedrich Hirzebruch. *Neue topologische Methoden in der algebraischen Geometrie.* Springer-Verlag, Berlin 1956.

[38] Friedrich Hirzebruch, Hilbert modular surfaces. *L'Enseignement mathématique* **19** (1973), 183–281.

[39] Christian Houzel, *La Géométrie algébrique : Recherches historiques.* Aspects of Mathematics **18**, Vieweg, Braunschweig (1992).

[40] Wilfred Hulsbergen, *Conjectures in Arithmetic Algebraic Geometry.* Blanchard, Paris 2003.

[41] Hiroshi Iritani, An integral structure in quantum cohomology and mirror symmetry for toric orbifolds. *Advances in Mathematics* **222** (2009), 1016–1079.

[42] Nicholas Katz, On the differential equations satisfied by period matrices. *Pub. math de l'I.H.É.S.* **35** (1968), 71–106.

[43] Nicholas Katz, On the intersection matrix of a hypersurface. *Ann. scient. Éc. Norm. Sup.* **2** (1969), 583–598.

[44] Nicholas Katz, The regularity theorem in algebraic geometry. *Actes, Congrès Intern. math.* **1** (1070), 437–443.

[45] Nicholas Katz, *Exponential Sums and Differential Equations.* 1990.

[46] Nicholas Katz and Tadao Oda, On the differentiation of De Rham cohomology classes with respect to parameters. *J. Math. Kyoto Univ.* **8** (1968) 199–213.

[47] Sheldon Katz, *Enumerative Geometry and String Theory.* Student Mathematical Library, IAS/Park City Mathematical Subseries **32**, AMS, Providence 2006.

[48] Ludmil Katzarkov, Maxim Kontsevich, and Tony Pantev, Hodge theoretic aspects of mirror symmetry. *Proc. Sympos. Pure Math.* **78** (2008), 87–174.

[49] Albrecht Klemm, Emanuel Scheidegger and Don Zagier, Periods and quasiperiods of modular forms and *D*-brane masses for the mirror quintic. In preparation.

[50] Maxim Kontsevich and Don Zagier, Periods. In *Mathematics Unlimited – 2001 and Beyond* (B. Engquist and W. Schmid, eds.), Springer, Berlin-Heidelberg-New York 2001, 771–808.

[51] Ron Livné, Cubic exponential sums and Galois representations. In *Current Trends in Arithmetical Algebraic Geometry*, *Contemp. Math.* **67** (1987), 247–261.

[52] Yuri Manin, *Frobenius Manifolds, Quantum Cohomology, and Moduli Spaces.* AMS Colloquium Publications **47**, AMS, Providence 1999.

[53] Christian Meyer, *Modular Calabi–Yau Threefolds.* Fields Institute Monographs, **22**, AMS, Providence, 2005.

[54] Anton Mellit and Masha Vlasenko, Dwork's congruences for the constant terms of powers of a Laurent polynomial. *Int. J. Number Theory* **12** (2016), 313–321.

[55] Martin Möller and Don Zagier, Modular embeddings of Teichmüller curves. *Comp. Math.* **152** (2016) 2269–2349.

[56] Rahul Pandharipande, Rational curves on hypersurfaces (after A. Givental). In *Séminaire Bourbaki 1997–1998*, exposé 848, *Astérisque* **252** (1998), 307–340.

[57] Alf van der Poorten, A proof that Euler missed ... Apéry's proof of the irrationality of $\zeta(3)$. An informal report. *Mathematical Intelligencer* **1** (1979), 195–203.

[58] V. V. Przyjalkowski, Quantum cohomology of smooth complete intersections in weighted projective spaces and in singular toric varieties. *Rossiĭskaya Akademiya Nauk. Matematicheskiĭ Sbornik* **198** (2007), 107–122.

[59] David Roberts, Fernando Rodriguez Villegas, Mark Watkins, *Hypergeometric Motives.* In preparation.

[60] Matt Rogers, J. Wan, I. J. Zucker, Moments of elliptic integrals and critical L-values. *Ramanujan J.* **37** (2015) 113–130.

[61] Kira Samol and Duco van Straten, Dwork congruences and reflexive polytopes. *Ann. Math. Qué.* **39** (2015), 185–203.

[62] Chad Schoen, On the geometry of a special determinantal hypersurface associated to the Mumford-Horrocks vector bundle. *J. Reine Angew. Math.* **364** (1986) 85–111.

[63] Anthony Scholl, Motives for modular forms. *Invent. Math.* **100** (1990) 419–430.

[64] Matthias Schütt, New examples of modular rigid Calabi–Yau threefolds. *Collect. Math.* **55** (2004), 219–228.

[65] Carlos Simpson, Transcendental aspects of the Riemann-Hilbert correspondence. *Illinois J. of Math.* **34** (1990), 368–391.

[66] Jan Stienstra and Frits Beukers, On the Picard–Fuchs equation and the formal Brauer group of certain elliptic K3-surfaces. *Math. Annalen* **271** (1985), 269–304.

[67] Burt Totaro, Euler and algebraic geometry. *Bull. AMS* **44** (2007) 541–559.

[68] Paula Tretkoff, *Complex Ball Quotients and Line Arrangements in the Projective Plane.* Princeton University Press, Princeton and Oxford 2016.

[69] Helena Verrill, Picard–Fuchs equations of some families of elliptic curves. In *Proceedings on Moonshine and Related Topics (Montreal, 1999)*, CRM Proc. Lecture Notes **30**, AMS (2001), 253–268.

[70] Fernando Rodriguez Villegas, Hypergeometric families of Calabi–Yau manifolds. In *Calabi–Yau varieties and mirror symmetry (Toronto, ON, 2001)*, Fields Inst. Commun. **38**, Amer. Math. Soc., Providence, RI, 2003, 223–231.

[71] Fernando Rodriguez Villegas and Don Zagier, Which primes are sums of two cubes? In *Number Theory (Proceedings of the Third Conference of the Canadian Number Theory Association)*, ed. K. Dilcher, CMS Conference Proceedings **15** (1995), 295–306,

[72] Masaaki Yoshida, *Hypergeometric Functions, My Love.* Aspects of Mathematics **32**, Vieweg, Braunschweig 1997.

[73] Noriko Yui, Modularity of Calabi–Yau varieties: 2011 and beyond. In *Arithmetic and geometry of K3 surfaces and Calabi–Yau threefolds*, Fields Inst. Commun. **67** (2013), 101–139.

[74] Don Zagier, Integral solutions of Apéry-like recurrence equations. In *Groups and Symmetries: From the Neolithic Scots to John McKay.* CRM Proceedings and Lecture Notes 47 (2009), Amer. Math. Society, 349–366.

[75] Don Zagier. Lines on the Dwork quintic pencil and its higher degree analogues. *J. Differential Geometry* **97** (2014), 177–189.

[76] Don Zagier, Life and work of Friedrich Hirzebruch. *Jahresbericht der Deutschen Mathematiker-Vereinigung* **117** (2015), 93–132.

[77] Don Zagier and Alexei Zinger, Some properties of hypergeometric series associated with mirror symmetry. In *Modular Forms and String Duality*, Fields Institute Communications Vol. **54** (2008), 163–177.

[78] Wadim Zudilin, Arithmetic hypergeometric series. *Russian Math. Surveys* **66** (2011) 369–420.

[79] Wadim Zudilin, A hypergeometric version of the modularity of rigid Calabi–Yau manifolds. In preparation.

Don Zagier
Max-Planck-Institut für Mathematik,
Vivatsgasse 7, 53111 Bonn, Germany
dbz@mpim-bonn.mpg.de