Hypergeometric Functions, How Special Are They?

Frits Beukers

Section 1. Introduction

In the world of standard functions, the hypergeometric functions take a prominent position in mathematics, both pure and applied, and in many branches of science. They were introduced by Euler as power series expansions of the form

$$1 + \frac{a \cdot b}{c \cdot 1} z + \frac{a(a+1)b(b+1)}{c(c+1) \cdot 1 \cdot 2} z^2 + \cdots ,$$

where a, b, c are rational parameters. By specialization of the parameters, Euler obtained the various classical functions that were around at that time. For example, taking $b = c = 1$ gives us Newton's binomial series for $(1-z)^{-a}$ and taking $a = b = 1/2, c = 3/2$ gives us $\arcsin(\sqrt{z})/\sqrt{z}$. Finally, taking all parameters equal to 1 recovers the ordinary geometric series, which more or less explains the name hypergeometric series that was given by Euler to his series. Hypergeometric functions also include functions that were entirely new in Euler's time. For example, taking $a = b = 1/2, c = 1$, one obtains the function

$$\frac{2}{\pi} \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-zx^2)}},$$

a so-called elliptic integral of the first kind. It is a period of the family of elliptic curves $y^2 = (1-x^2)(1-zx^2)$ parameterized by z and is one of the most often quoted functions in algebraic geometry. Euler also found the hypergeometric equation, which is the second-order linear differential equation that is satisfied by hypergeometric series. It reads

$$z(z-1)f'' + ((a+b+1)z - c)f' + abf = 0.$$

Frits Beukers is a professor of mathematics at Utrecht University. His email address is f.beukers@uu.nl.

This is a differential equation that occurs in a multitude of branches of mathematics, mathematical physics, and applied sciences.

Years later, Gauss studied hypergeometric functions not only as values of Euler's hypergeometric series but also as solutions of the hypergeometric equation throughout the complex plane, an approach that was entirely new at that time. In this way Gauss very soon became aware of the problem of their multivaluedness, known nowadays as the monodromy problem. He published only part of his work in 1812 [11, p. 123]. A sequel that describes the behavior in the complex plane was found in Gauss's *Nachlass*; see [11, p. 207]. Because of Gauss's work, the functions given by Euler's hypergeometric series are now often called Gauss hypergeometric functions.

The next major contribution came from Riemann. In the article [19] from 1857 he gave a complete description of the monodromy group for Gauss's hypergeometric function. The monodromy group of a linear differential equation in the complex plane characterizes the behavior of the analytic continuation of its solutions. In this way hypergeometric functions became an important testing ground for Riemann's fundamentally new ideas on analytic continuation. Riemann's work was taken up by H. A. Schwarz, Felix Klein, and others, and hypergeometric functions made their appearance in the early days of algebraic geometry and modular forms. In his 1893 lectures, Klein [17] gave an extensive exposition of Riemann's ideas and their consequences. For a more concise overview see [2].

Simultaneously, by the end of the nineteenth century, people had introduced many generalizations of Gauss hypergeometric functions by increasing the number of parameters or the number of variables or both. With a few exceptions, the ensuing

study of them was for a large part of a descriptive nature, meaning: for each new class the system of corresponding (partial) differential equations had to be determined, a basis of solutions around various special points was to be found, together with their domain of convergence. In this process, the conceptual understanding of these generalizations lagged far behind the well-developed ideas around Gauss's hypergeometric function. Thus a large part of the subject of generalized hypergeometric functions became relegated to the domain of special functions, which to many people means many formulas but no mathematical depth. Fortunately, developments by the end of the twentieth century turned this image around somewhat, and it is my hope that in this century the turnaround will be complete. It is the purpose of this article to introduce generalized hypergeometric functions in one and several variables and hint at some simple, almost combinatorial, structures that underlie them. We do this by looking at hypergeometric functions that are at the same time algebraic. The structure of this article is as follows:

- Section 2: A number of random-looking but relevant examples;
- Sections 3, 4, 5: Gauss hypergeometric functions and examples of their generalizations;
- Section 6: A-hypergeometric functions, a unified way of looking at all the previous examples;
- Section 7: An example of a result that holds for general A-hypergeometric systems;
- Section 8: A short discussion on monodromy.

Section 2. Some Peculiar Examples

The solution of the general fifth-degree equation has a notorious history, and it is known that it cannot be achieved by the repeated use of taking radicals. However, it does not mean that the equation is hard to solve. It is well known that, after some transformations involving radicals, the fifth-degree equation can be reduced to a three-term equation of the form

$$zx^5 - x + 1 = 0,$$

where z is a parameter and x the unknown. Such forms are called Bring-Jerrard forms. An exercise in Lagrange inversion shows that a solution is given by the power series

$$\sum_{n \geq 0} \binom{5n}{n} \frac{z^n}{4n+1}.$$

Lagrange inversion
Let $f(x)$ be a power series in x with $f(0) = 0$, $f'(0) \neq 0$. Then Lagrange inversion tells us that a solution to $f(x) = z$ in x is given by the power series

$$x = \sum_{k \geq 1} \frac{z^k}{k!} \left(\frac{d}{dx} \right)^{k-1} \left(\frac{x}{f(x)} \right)^k \bigg|_{x=0}$$

in powers of z. Change the fifth-degree equation to $z(x+1)^5 - x = 0$ and apply the inversion with $f(x) = x/(x+1)^5$ to obtain our solution of the fifth-degree equation.

Another exercise with a surprising answer is when we take $f(x) = xe^{-x}$.

This turns out to be a hypergeometric function. To see this, let us introduce the symbol $(x)_n = x(x+1)\cdots(x+n-1)$, the so-called *Pochhammer symbol* or *shifted factorial*. In particular, $(1)_n = n!$. Note that we can rewrite the coefficients of our fifth-degree solution as

$$\binom{5n}{n} \frac{1}{4n+1} = \frac{5^{5n}}{4^{4n}} \times \frac{(1/5)_n(2/5)_n(3/5)_n(4/5)_n}{(1/2)_n(3/4)_n(5/4)_n n!}.$$

Denote

$$\phi(z) = \sum_{n \geq 0} \frac{(1/5)_n(2/5)_n(3/5)_n(4/5)_n}{(1/2)_n(3/4)_n(5/4)_n n!} z^n.$$

Then the solution to our fifth-degree equation can be written as $\phi(5^5 z/4^4)$. The function ϕ is an example of a generalized hypergeometric function.

Our second example comes from a beautiful observation by Fernando Rodriguez-Villegas [20]. In his work on estimates for the prime counting function $\pi(x)$, Chebyshev used arguments that amount to studying prime factors of the numbers

$$u_n = \frac{(30n)!n!}{(15n)!(10n)!(6n)!}.$$

Rodriguez-Villegas observed that the numbers u_n are integers and that the generating function $u(z) = \sum_{n \geq 0} u_n z^n$ is an algebraic function in z. This means that $u(z)$ satisfies a nontrivial polynomial equation with polynomials in z as coefficients. In this example, the minimal degree of the equation turns out to be 483840. The proof of this observation is to rewrite $u(z)$ as $\psi(2^{14}3^9 5^5 z)$ (see Figure 1) and then use a result of Heckman and the author in [6] on generalized hypergeometric functions, which implies that $\psi(z)$ is an algebraic function.

A third example involves two-variable functions. Consider the polynomial

$$\Delta = 1 + 4x + 4y + 18xy - 27x^2y^2$$

and the algebraic function $g(x, y)$ defined by the cubic equation

$$g^3 - g^2 - (3xy - x - y)g - xy(x+y+1) = 0$$

$$\psi(z) = \sum_{n \geq 0} \frac{(1/30)_n(7/30)_n(11/30)_n(13/30)_n(17/30)_n(19/30)_n(23/30)_n(29/30)_n}{(1/5)_n(1/3)_n(2/5)_n(1/2)_n(3/5)_n(2/3)_n(4/5)_n n!} z^n,$$

<div align="right">Figure 1.</div>

and $g(0,0) = 1$. Then

$$\sqrt{(g - 3xy)/\Delta} = \sum_{m,n \geq 0} \frac{(1/2)_{2m-n}(1/2)_{2n-m}}{m!n!} x^m y^n.$$

The two-variable series is a so-called Horn series of type G_3. Horn series have the property that the Pochhammer symbols that occur may have negative indices. In that case one must use the more general definition $(a)_k = \Gamma(a + k)/\Gamma(a)$. For negative values of k, this amounts to $(a)_k = 1/(a - 1)(a - 2) \cdots (a - |k|)$.

These three examples may have reaffirmed your impression that we are working with special functions indeed. However, the goal of this article is to explain Theorem 5, which describes a general combinatorial criterion to detect instances such as the ones above. In fact, the two-variable example was found that way. I believe that this is only part of a much wider circle of ideas that underlie the domain of hypergeometric functions in one and more variables. In the next section I will be more systematic and introduce hypergeometric functions, together with a number of their properties.

Section 3. Gauss Hypergeometric Function
Choose three parameters a, b, c, usually rational numbers, and consider the power series

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

in the complex variable z. It is well defined as long as c is not an integer ≤ 0. This is Euler's series from the introduction of this article. It is a simple exercise to show that its radius of convergence is equal to 1 whenever $a, b \notin \mathbb{Z}_{\leq 0}$. In the case when a or b is in $\mathbb{Z}_{\leq 0}$, the infinite series becomes a polynomial.

Let us abbreviate the notation $_2F_1(a, b, c|z)$ by $f(z)$ for the moment, and let θ be the differential operator $z\frac{d}{dz}$. A simple calculation using the factorial structure of the coefficients of $f(z)$ shows that

$$z(\theta + a)(\theta + b)f = \theta(\theta + c - 1)f.$$

After replacing θ by $z\frac{d}{dz}$ again and an expansion, we find the second-order differential equation

$$z(z - 1)f'' + ((a + b + 1)z - c)f' + abf = 0,$$

known as the *hypergeometric differential equation*. A second solution around $z = 0$, when $c \notin \mathbb{Z}$, is given by the series expansion

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

Together with the above series they form a basis of the vector space of solutions of the hypergeometric equation around the origin of the complex plane.

Beginning with Gauss, mathematicians such as Kummer and Goursat discovered many relations that exist between hypergeometric functions with their argument z replaced by certain rational functions in z. An example is

$$_2F_1(a, b, a + b + 1/2|4z - 4z^2)$$
$$= \, _2F_1(2a, 2b, a + b + 1/2|z),$$

and for an account I refer to my overview [2]. It was through Riemann's work that many such identities found a conceptual basis with proofs that are only a few lines long.

As an application of Riemann's work on analytic continuation in geometrical form, H. A. Schwarz in 1873 gave a classification of all hypergeometric functions which are at the same time algebraic; see [21]. This resulted in the famous Schwarz list. For example, it turns out that $_2F_1(19/60, 49/60, 4/5|z)$ is an algebraic function of degree 720. Its Galois group—and at the same time monodromy group—G has the property that G modulo its center is the alternating group A_5. For the full list I refer to Schwarz's original paper and, for a more recent account, [7] or [2]. In these lists one has to realize that only irreducible hypergeometric equations are considered, i.e., equations whose differential operator does not factor in the ring $\mathbb{C}(z)[d/dz]$ of differential operators. Here is a simple criterion.

Theorem 1. *The hypergeometric equation is irreducible if and only if the sets $\{a, b\}$ and $\{0, c\}$ are disjoint when considered modulo $\mathbb{Z}$.*

From now on we will be interested only in irreducible (systems of) differential equations.

Section 4. Higher-Order Hypergeometric Functions
There is a generalization of the hypergeometric equation to higher-order equations for which one could also compute the analytic continuation, albeit in a preliminary way. This was realized notably

through the work of Clausen and later in the article [26] by J. Thomae from 1870.

Let $k \in \mathbb{Z}_{\geq 2}$ and take two sets of parameters $\mathbf{a} = \{a_1, \ldots, a_k\}$ and $\mathbf{b} = \{b_1, \ldots, b_k\}$, where we take $b_k = 1$ by default. The *hypergeometric function of order k* is defined by

$$_kF_{k-1}(\mathbf{a}, \mathbf{b}|z) = \sum_{n \geq 0} \frac{(a_1)_n \cdots (a_k)_n}{(b_1)_n \cdots (b_{k-1})_n n!} z^n.$$

Sometimes it is also referred to as the Clausen-Thomae function. Again, the radius of convergence is 1, and it satisfies a differential equation of order k given by

$$z(\theta + a_1) \cdots (\theta + a_k)f$$
$$= (\theta + b_1 - 1) \cdots (\theta + b_{k-1} - 1)\theta f.$$

The derivation goes along in the same way as in the Gauss case. This time it is better to leave it in this form. Again, the equation is irreducible if and only if the sets $\mathbf{a}$ and $\mathbf{b}$ are disjoint modulo $\mathbb{Z}$.

In 1989 Heckman and I [6] succeeded in extending Schwarz's list of algebraic Gauss hypergeometric functions to the case of higher-order hypergeometric functions. As a byproduct of [6] there is a simple criterion to decide whether or not a given hypergeometric function is algebraic. It is called the *interlacing criterion*. We say that two sets of k real numbers $\{a_1, \ldots, a_k\}$ and $\{b_1, \ldots, b_k\}$ in the interval $[0, 1)$ *interlac* if they are disjoint and their elements occur alternately in increasing order. For example, the white set and black set pictured below are interlacing.

$$0 \vdash\!\!-\!\bullet\!-\!\circ\!-\!\bullet\!-\!\circ\!-\!\bullet\!-\!\circ\!-\!\bullet\!-\!\circ\!-\!\dashv 1$$
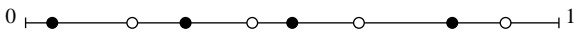
More generally, two sets of real numbers $\{a_1, \ldots, a_k\}$ and $\{b_1, \ldots, b_k\}$ are said to interlace modulo $\mathbb{Z}$ if the sets $\{a_1 - \lfloor a_1 \rfloor, \ldots, a_k - \lfloor a_k \rfloor\}$ and $\{b_1 - \lfloor b_1 \rfloor, \ldots, b_k - \lfloor b_k \rfloor\}$ interlace on $[0, 1)$. The interlacing criterion reads as follows.

Theorem 2 (Beukers, Heckman (1989)). *Let $a_1, \ldots, a_k, b_1, \ldots, b_k \in \mathbb{Q}$ be the parameters of an irreducible generalized equation (we have $b_k = 1$ by default). Let D be their common denominator. Then the solutions of the hypergeometric equation are algebraic functions in z if and only if the sets*

$$\{ra_1, \ldots, ra_k\} \quad \text{and} \quad \{rb_1, \ldots, rb_k\}$$

interlace modulo $\mathbb{Z}$ for all integers r with $\gcd(r, D) = 1$ and $1 \leq r < D$.

As an application recall the function $\psi(z)$ from the introduction. We recognize it as a generalized hypergeometric function with parameters

$$\mathbf{a} = (1/30, 7/30, 11/30, 13/30, 17/30, 19/30, 23/30, 29/30)$$

and

$$\mathbf{b} = (1/5, 1/3, 2/5, 1/2, 3/5, 2/3, 4/5, 1).$$

Notice that these two sets interlace modulo $\mathbb{Z}$. The interlacing criterion tells us that we should also look at the sets $r\mathbf{a}$ and $r\mathbf{b}$ for all integers r relatively prime to 120. However, the numerators of the elements of $\mathbf{a}$ form a complete set of numbers between 1 and 30 that are relatively prime to 30. Consequently, $r\mathbf{a} \equiv \mathbf{a} \pmod{\mathbb{Z}}$ for all r relatively prime to 30. Similarly, $r\mathbf{b} \equiv \mathbf{b} \pmod{\mathbb{Z}}$ for all r relatively prime to 30. Therefore, interlacing $r\mathbf{a}$ and $r\mathbf{b}$ modulo $\mathbb{Z}$ is equivalent to interlacing of $\mathbf{a}$ and $\mathbf{b}$, which we have already verified. The conclusion is that $\psi(z)$ is an algebraic function. Unfortunately the interlacing criterion gives no information about the degree of this function over the field $\mathbb{C}(z)$. For this we must invoke the other parts of [6], which tell us that the Galois group in this case is $W(E_8)$, the Weyl group of the root system E_8. This group has order 696729600, and a careful analysis shows that the degree of $\psi(z)$ is 483840, a divisor of $|W(E_8)|$.

Section 5. Several Variables

By the end of the nineteenth century several authors had introduced several-variable versions of hypergeometric functions. The best known are the Appell functions F_1, F_2, F_3, F_4. Their basic properties are described at length in the book [1] by Appell and Kampé de Fériet from 1926. We will not go into these functions extensively, but recall the definition of Appell's F_1:

$$F_1(a, b, b', c|x, y) = \sum_{m,n \geq 0} \frac{(a)_{m+n}(b)_m(b')_n}{(c)_{m+n}m!n!} x^m y^n.$$

Of course a, b, b', c are parameters. The other Appell functions are variations on this type of series. They all satisfy systems of partial differential equations of order 2. Since these equations look rather unappealing, I will not quote them here. Another example is the Horn G_3 function, alluded to in the beginning:

$$G_3(a, b|x, y) = \sum_{m,n \geq 0} \frac{(a)_{2m-n}(b)_{2n-m}}{m!n!} x^m y^n.$$

Horn G_3 is one among a list of fourteen Horn series, which includes Appell's series. They arise from the philosophy that a two-variable power series $\sum_{m,n \geq 0} A(m, n)x^m y^n$ can be considered hypergeometric if the ratios $P(m, n) = A(m, n)/A(m, n-1)$ and $Q(m, n) = A(m, n)/A(m-1, n)$ are rational functions in m, n, together with the compatibility conditions $P(m, n)Q(m-1, n) = P(m, n-1)Q(m, n)$. If, in addition, one assumes that the numerator and denominator of P, Q have degree 2, one arrives at the ten Horn series and the four Appell series. The degree 2 condition corresponds

to the requirement that the partial differential equations for these functions have order 2.

Of course one can increase the number of variables. We then get the so-called Lauricella functions F_A, F_B, F_C, F_D. We give the series expansion for the three-variable version of Lauricella F_D introduced by G. Lauricella in 1893 [18]:

$$F_D(a, b, b', b'', \gamma | x, y, z)$$
$$= \sum_{m,n,p \geq 0} \frac{(a)_{m+n+p}(b)_m(b')_n(b'')_p}{(c)_{m+n+p} m! n! p!} x^m y^n z^p.$$

It extends Appell's F_1 to three variables, and it is probably not hard for the reader to figure out what the n-variable version is. I mention this class in particular, since the structure of its monodromy group has been the subject of an extensive study that started with E. Picard in the nineteenth century and was completed by Deligne and Mostow around 1980; see [9]. In this paper Deligne and Mostow study discrete and arithmetic symmetry groups of the complex hyperbolic ball. Using Lauricella F_D functions, they discovered a number of groups that were hitherto unknown.

Apart from increasing the number of variables, one could also increase the order of the partial differential equations that define these functions. For example, in the book by Appell and Kampé de Fériet [1, Part I, Ch. IX], we find a section on two-variable functions that are the analogue of the one-variable higher-order hypergeometric functions. Then one is again obliged to study the dimension of the solution space, systems of differential equations, etc. This is what I refer to as a descriptive activity, and one might wonder where the end is.

Section 6. A-hypergeometric Functions

Fortunately there is a way to treat all previously mentioned examples of hypergeometric functions in a unified way. It was discovered by Gel'fand, Kapranov, and Zelevinsky by the end of the 1980s in a series of papers [12], [13], [14], [15] and is known under the name of A-hypergeometric functions. In honor of their discoverers, they are also referred to as GKZ-hypergeometric functions. My preference is the name A-hypergeometric function. Around the same time, B. Dwork developed a general theory of hypergeometric functions that has many parallels with A-hypergeometric functions and that culminated in his book [10]. Unfortunately, the book is hard to read because of its cumbersome notation and the interference of p-adic considerations in which Dwork was mostly interested.

In what follows I give a crash course on A-hypergeometric functions, concentrating only on those aspects that are immediately required for this article. Many other important issues will not be touched upon. For a more complete account, one might consult the overviews [25], [3], or the book [22], which is very computationally flavored.

The A-hypergeometric approach starts with a finite set A (hence the name) of N lattice points $\mathbf{a}_1, \ldots, \mathbf{a}_N$ in $\mathbb{Z}^r$ such that the rth coordinate of each point is 1 and the $\mathbb{Z}$-span of the $\mathbf{a}_i$ equals $\mathbb{Z}^r$. The $r \times N$-matrix with $\mathbf{a}_1, \ldots, \mathbf{a}_N$ as columns will also be denoted by A. We call it the A-matrix. There is also a vector $\boldsymbol{\alpha} \in \mathbb{Q}^r$ of parameters. The data $A, \boldsymbol{\alpha}$ are the combinatorial data that determine a system of hypergeometric differential equations. We arrive at them in the following way. Let $L \in \mathbb{Z}^N$ be the lattice of relations,

$$(l_1, \ldots, l_N) \in L \iff l_1 \mathbf{a}_1 + \cdots + l_N \mathbf{a}_N = \mathbf{0}.$$

Its rank equals $d := N - r$. Choose a basis of L and let B be the $d \times N$ matrix with these basis vectors as rows. We call B a B-matrix. In the literature the transpose of B is usually introduced under the name of Gale dual. Notice that $A.B^t = O$, the $r \times d$ zero matrix. Denote the columns of B by $\mathbf{b}_1, \ldots, \mathbf{b}_N$. Our differential equations and their solutions live in a space of N complex variables $v_1, \ldots, v_N$. Choose $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)$ such that $\gamma_1 \mathbf{a}_1 + \cdots + \gamma_N \mathbf{a}_N = \boldsymbol{\alpha}$ and consider the formal Laurent series

$$\Psi(B, \boldsymbol{\gamma} | \mathbf{v}) = \sum_{\mathbf{m} \in \mathbb{Z}^d} \prod_{j=1}^{N} \frac{v_j^{\mathbf{b}_j \cdot \mathbf{m} + \gamma_j}}{\Gamma(\mathbf{b}_j \cdot \mathbf{m} + \gamma_j + 1)}$$

in the variables $v_1, \ldots, v_N$. Note that, given A and $\boldsymbol{\alpha}$, it is independent of the particular choice of B, but it does depend on the choice of $\boldsymbol{\gamma}$. The system of A-hypergeometric functions is the system of all partial differential equations in the derivations $\partial/\partial v_i$ and coefficients in $\mathbb{C}(v_1, \ldots, v_N)$ that annihilate this formal series. There is a very standard way to write down these operators, but I refrain from writing them here. It turns out that we hardly need them. What is important is that there is a freedom of choice in the solutions $\boldsymbol{\gamma}$ to $\gamma_1 \mathbf{a}_1 + \cdots + \gamma_N \mathbf{a}_N = \boldsymbol{\alpha}$ up till shifts in the vector space generated by L. However, the system of differential equations is in general independent of this choice. We denote the system of A-hypergeometric equations by $H_A(\boldsymbol{\alpha})$.

Let us see how this works in the case of the Gauss hypergeometric function. Consider

$$f_{\text{Gauss}} = \sum_{n \geq 0} \frac{\Gamma(n+a)\Gamma(n+b)}{\Gamma(n+c)\Gamma(n+1)} z^n,$$

which equals the Gauss hypergeometric series up to a constant factor. We use the Γ-function identity $\Gamma(z)\Gamma(1-z) = \pi/\sin(\pi z)$ to rewrite this as a series proportional to

$$\sum_{n \geq 0} \frac{z^n}{\Gamma(-n-a+1)\Gamma(-n-b+1)\Gamma(n+c)\Gamma(n+1)}.$$

This is beginning to look like our series Ψ. First we note that we might as well sum over all $n \in \mathbb{Z}$,

simply because $1/\Gamma(n+1)$ is zero whenever n is a negative integer. As a last step, replace z^n by $v_1^{-n-a}v_2^{-n-b}v_3^{n+c-1}v_4^n$ to get

$$f_A = \sum_{n \in \mathbb{Z}} \frac{v_1^{-n-a}v_2^{-n-b}v_3^{n+c-1}v_4^n}{\Gamma(-n-a+1)\Gamma(-n-b+1)\Gamma(n+c)\Gamma(n+1)},$$

which has the shape of Ψ defined above with $B = (-1,-1,1,1)$ and $\boldsymbol{\gamma} = (-a,-b,c-1,0)$. The interpretation is that f_A is a series in four variables v_i whose restriction to $v_1 = v_2 = v_3 = 1, v_4 = z$ reproduces the Gauss hypergeometric series. Similarly, the restriction of the A-hypergeometric equations to this line recovers the classical hypergeometric equation. One of the innovations of this approach is that in Ψ we have freedom of choice in the parameter $\boldsymbol{\gamma}$. We can shift it by rational multiples of $(-1,-1,1,1)$ and still have a formal series satisfying the same differential equations. As an exercise one can consider $\boldsymbol{\gamma}' = (c-a-1, c-b-1, 0, 1-c)$, which equals $\boldsymbol{\gamma}+(1-c)(-1,-1,1,1)$. Then write the corresponding $\Psi(B, \boldsymbol{\gamma}'|\mathbf{v})$, set $v_1 = v_2 = v_3 = 1, v_4 = z$, and carry out the above procedure in reverse direction. We get the second series solution

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

of the Gauss hypergeometric equation. Similarly, shifting $\boldsymbol{\gamma}$ such that the first or second component becomes zero reproduces the two basic solutions of the Gauss hypergeometric equation around $z = \infty$.

Thus we can identify the Gauss hypergeometric equation with an A-hypergeometric system with parameters $N = 4, d = 1$ and a B-matrix $(-1,-1,1,1)$. Clearly, $r = N - d = 3$, and an $r \times N$ A-matrix which has B as a kernel can be given by

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

The columns are vectors in $\mathbb{R}^3$ but with the last coordinate $x_3 = 1$. The corresponding points, together with their convex hull, can be pictured as

2F1
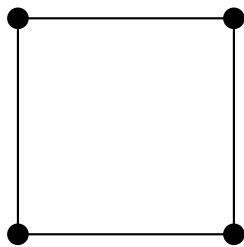
in the plane given by $x_3 = 1$.

The same point of view works for every classical hypergeometric series we have seen so far. As an example take the series expansion of Appell's F_1 given earlier. There one easily sees that $d = 2$
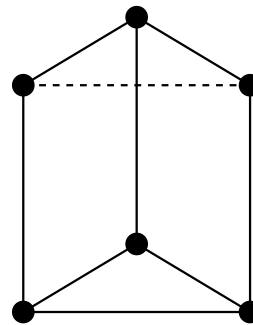
(the number of variables), $N = 6$ (the number of Γ-factors), and the B-matrix is

$$\begin{pmatrix} -1 & -1 & 0 & 1 & 1 & 0 \\ -1 & 0 & -1 & 1 & 0 & 1 \end{pmatrix}.$$

From $r = N - d = 4$ and this B-matrix one deduces the following possible A-matrix:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Here is a picture of the corresponding points in the hyperplane $x_4 = 1$:

F1

By playing with the parameters $\boldsymbol{\gamma}$ in the corresponding formal series $\Psi(B, \boldsymbol{\gamma}|\mathbf{v})$, one easily obtains all the local series expansions for solutions of the Appell F_1-system, which are obtained with far more effort in the classical literature.

We now present some of the striking results on A-hypergeometric equations. Given the set $A \subset \mathbb{Z}^r$, we can construct its convex hull, which we denote by $Q(A)$. This is a polytope of dimension $r-1$ which lies in the affine hyperplane $x_r = 1$. We also introduce the positive cone $C(A)$, which is the $\mathbb{R}_{\geq 0}$-span of the vectors from A. We say that the A-hypergeometric system is *resonant* if $\boldsymbol{\alpha} + \mathbb{Z}^r$ and the boundary of $C(A)$ have a nontrivial intersection. We say that the system is *nonresonant* if the intersection is empty.

Theorem 3 (GKZ). *A nonresonant A-hypergeometric system is irreducible.*

The converse of the statement is almost true; see for example [24] or [4]. This irreducibility result, together with its converse, implies all the irreducibility statements that have been compiled over the years for individual systems of hypergeometric equations. Another striking theorem is the following.

Theorem 4 (GKZ). *Around any nonsingular point the analytic solution space of an A-hypergeometric system has finite dimension. Moreover, if the system*

*is nonresonant, the dimension equals* $(r-1)!$ *times the Euclidean volume of* $Q(A)$.

The normalization for the volume of $Q(A)$ is that which assigns volume 1 to the simplex with vertices $\mathbf{0}, \mathbf{e}_1, \ldots, \mathbf{e}_{r-1}$ in $\mathbb{R}^{r-1}$ (the $\mathbf{e}_i$ are the standard basis vectors). The dimension of the solution space is also called the holonomic rank of the system. The theorem, together with the pictures above, tells us immediately that the holonomic rank of the Gauss hypergeometric equation is 2 and, for Appell F_1, it is 3.

Another classic result of GKZ gives a one-to-one correspondence between local series expansions of certain bases of solutions with the regular triangulations of A. Since we do not need it here, I refer instead to the overviews [25], [3].

Section 7. Algebraic Solutions

In this section we extend the interlacing criterion for one-variable equations to the case of A-hypergeometric systems. We make the additional assumption that the sets A are assumed to be *normal*. That is, $C(A) \cap \mathbb{Z}^r$ is equal to the $\mathbb{Z}_{\geq 0}$-span of A. All classical hypergeometric systems satisfy this condition. We will also assume that we are in the *nonresonant* situation.

Let us now consider the set $K(A, \boldsymbol{\alpha}) = C(A) \cap (\boldsymbol{\alpha} + \mathbb{Z}^r)$. Recall that $C(A)$ is the positive cone spanned by the elements of A. We shall call a point $\mathbf{p}$ in this set an *apex-point* if $\mathbf{p} - \mathbf{q}$ does not lie in $C(A)$ for any other $\mathbf{q} \in K(A, \boldsymbol{\alpha})$. Loosely speaking, a point $\mathbf{p}$ is an apex point if it cannot be seen from any other point in $K(A, \boldsymbol{\alpha})$ in a direction in $C(A)$. Below is a two-dimensional sketch. The grey area represents the cone $C(A)$, and the dots form the shifted lattice $\boldsymbol{\alpha} + \mathbb{Z}^2$. The two white dots are the apex points in this example.


C(A)
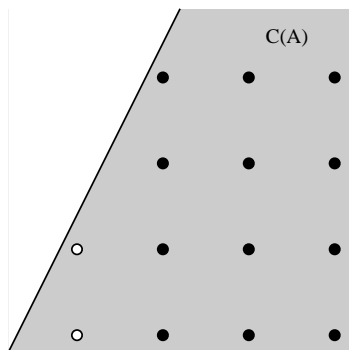
It is an easy lemma to show that the number of apex points is at most the rank of the hypergeometric system. We say that the set of apex points is *maximal* if their number equals the rank. We are now in a position to state the main theorem of this article, which we call the *apex point criterion*.

Theorem 5 (Beukers, 2010). *Consider a nonresonant A-hypergeometric system* $H_A(\boldsymbol{\alpha})$ *and assume that A is normal. Suppose the parameters* $\alpha_1, \ldots, \alpha_r$ *are rational with common denominator D. Then the solution space of the system* $H_A(\boldsymbol{\alpha})$ *contains a nontrivial algebraic function if and only if the apex sets of* $C(A) \cap (r\boldsymbol{\alpha} + \mathbb{Z}^r)$ *are maximal for all integers r with* $0 < r < D$ *and* $\gcd(r, D) = 1$.

Because of the irreducibility of $H_A(\boldsymbol{\alpha})$, the occurrence of *one* nontrivial algebraic solution is equivalent to *all* solutions being algebraic. It is a nontrivial exercise to show directly that this condition is equivalent to the interlacing condition we had earlier in the one-variable case.

To illustrate the apex point criterion, we go to the Horn G_3-system, which has the advantage that $A \subset \mathbb{Z}^2$, so we can draw pictures. Recall the series $G_3(a, b|x, y)$. A possible set A consists of the points $(0, 1), (1, 1), (2, 1), (3, 1)$. Below is a picture, together with the positive cone $C(A)$. The darker grey area represents the location of the apex points.
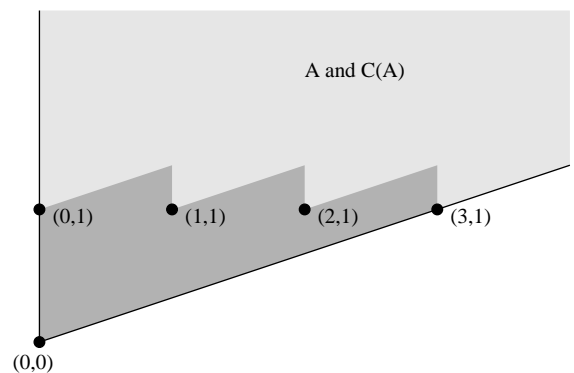

The set $Q(A)$ is the line segment between $(0, 1)$ and $(3, 1)$. It is the union of three unit intervals, so its volume is 3, which is the rank of the G_3-system. As a fine point, I mention that in the classical literature the rank is 4. This is because there is a classical spurious solution $x^\rho y^\sigma$ with $\rho = -(2a + b)/3, \sigma = -(a + 2b)/3$. However, the system in the A-hypergeometric version of G_3 contains more equations and does not allow this solution. So we get rank 3.
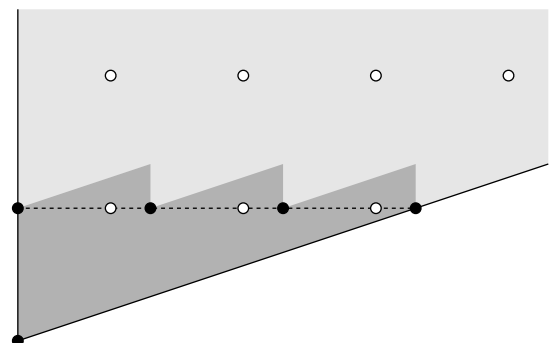
The parameters a, b are represented in this picture by the vector $(-2a - b, -a - b)$. Let us first take the case where a, b are rational and $a + b$ is an integer. Thus the second coordinate of $(-2a - b, -a - b)$ is an integer. All points with integral second coordinate not on the boundary of $C(A)$ and in the domain of apex points are pictured below by the dashed line. In addition, the white dots represent an example of a shifted lattice $(-2a - b, -a - b) + \mathbb{Z}^2$.

So one easily sees that the number of apex points is maximal in this case. Of course, the apex points of $r(-2a - b, -a - b) + \mathbb{Z}^2$ are also on this line for any integer r. Hence the conditions of Theorem 5 are fulfilled, and we conclude that $G_3(a, b|x, y)$ is an algebraic function in x, y if a, b are nonintegral rationals with $a + b \in \mathbb{Z}$. Our initial example $a = b = 1/2$ is a particular case of this.

There is another choice that satisfies the apex point criterion. Take $a = 1/2, b = 1/3$ and $a = 1/2, b = 2/3$. The corresponding shifted lattices are depicted by the white and grey dots.
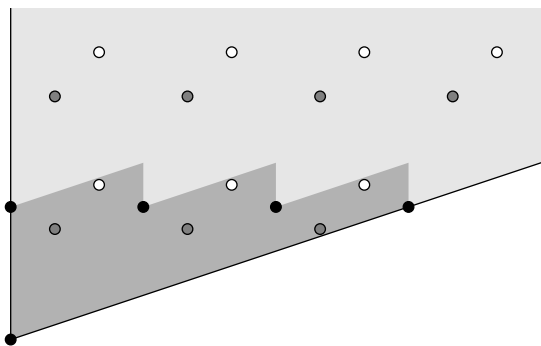
In both cases we see that the number of apex points is maximal, and we can conclude that $G_3(1/2, 1/3|x, y)$ and $G_3(1/2, 2/3|x, y)$ are algebraic. Essentially, up to shifts by integers and interchange of a, b, one can show that these are the only possibilities for $G_3(a, b|x, y)$ to be algebraic. An analysis such as this has been carried out by Esther Bod in her Ph.D. thesis [7]. She extended Schwarz's list of algebraic hypergeometric functions to all two-variable Appell and Horn functions and the several-variable Lauricella systems. Before her work, other authors had already made such an extension to Appell's functions and Lauricella F_D but with entirely different methods.

Section 8. Monodromy

I will finish this article with a few remarks on analytic continuation and monodromy. Consider an A-hypergeometric system $H_A(\boldsymbol{\alpha})$ and the vector space of solutions around a point $P \in \mathbb{C}^N$. Take a closed loop **c** beginning and ending in P and continue the solutions analytically along **c**. After returning in P, the solutions may not have returned

to their original value. Instead the solution space underwent a linear transformation. The group generated by these transformations is called the monodromy group of $H_A(\boldsymbol{\alpha})$. Of course, if all solutions of the system are algebraic, the monodromy group is finite. With a bit more effort one can show that the converse is also true.

Unfortunately, determination of the monodromy group for general A-hypergeometric systems is still an open problem. In particular, Theorem 5 cannot be proven by using monodromy calculations. Thus, in order to prove Theorem 5, we need to resort to other methods. Fortunately there is a key in the form of a conjecture.

Conjecture 6 (Grothendieck). *Consider a finite system $\mathcal{L}$ of linear partial differential equations in the independent variables $z_1, \ldots, z_N$ and with coefficients in $\mathbb{Z}[z_1, \ldots, z_N]$. Suppose that $\mathcal{L}$ has finite rank R. Then the solution space of $\mathcal{L}$ consists of algebraic functions if and only if, for almost all primes p, the system L reduced modulo p has R polynomial solutions which are linearly independent over $\mathbb{F}_p[z_1^p, \ldots, z_N^p]$.*

Some remarks are in order. By almost all primes we mean all primes minus a finite set of them. The notation $\mathbb{F}_p$ stands for the integers modulo p, which is a finite field. In characteristic p the partial derivates of elements in $\mathbb{F}_p[z_1^p, \ldots, z_N^p]$ are zero and hence should be considered as constants. Roughly speaking, Grothendieck's conjecture comes down to saying that all solutions of a linear system of partial differential equations are algebraic if and only if almost every reduction mod p has a basis consisting of polynomial solutions modulo p. It would be great if this theorem could be proven, but it has not—not even in the case of ordinary ($N = 1$) differential equations.

However, one particular instance has been proven. In a beautiful paper, Nick Katz [16] has shown that the conjecture is true for systems that are part of a Gauss-Manin system, that is, a system of differential equations associated with periods of a family of algebraic varieties. For $\boldsymbol{\alpha} \in \mathbb{Q}^r$ the A-hypergeometric system $H_A(\boldsymbol{\alpha})$ can be shown to be of the required form, and thus Grothendieck's conjecture is true in this case. The proof of Theorem 5 now consists of showing that the apex point criterion is equivalent to the statement that the solution set of $H_A(\boldsymbol{\alpha})$ modulo p contains the maximal number of polynomial solutions for almost all p.

This proof sketch may also explain why we cannot say anything about the degrees of the algebraic functions involved or the order of the monodromy group. It is clear that at some point one would like to be able to determine monodromy for general A-hypergeometric systems. Of course,

in many papers that deal with special instances of several-variable hypergeometric functions, the determination of the monodromy group does play a crucial role. The drawback of this approach is that, for every new set A, one must perform a nontrivial amount of work to determine the singular locus of the system and the fundamental group of its complement and then find some way to determine analytic continuation of so-called Euler integrals defining the A-hypergeometric functions. It is my hope that eventually one can circumvent these intricate considerations and instead use the conceptual simplicity of A-hypergeometric systems to describe their monodromy, for example, through constructions in combinatorial algebra. If this could be possible, the turnaround announced in the introduction of this article might be complete.

Acknowledgments

References

[1] P. Appell and J. Kampé de Fériet, *Fonctions hypergéométriques et hypersphériques: Polynomes d'Hermite*, Paris: Gauthier-Villars, 1926.

[2] F. Beukers, Gauss' hypergeometric function, in: *Arithmetic and Geometry around Hypergeometric Functions*, Progr. Math., 260, Birkhäuser, Basel, 2007, pp. 23–42. Also accessible at `http://www.staff.science.uu.nl/~beuke106/GaussHF.pdf`

[3] _____, Notes on A-hypergeometric functions, in: *Arithmetic and Galois Theories of Differential Equations*, Séminaires et Congrès, Société mathématique de France, vol. 23; 2011, pp. 25–61. Also accessible at `http://www.staff.science.uu.nl/~beuke106/AHGcourse.pdf`

[4] _____, Irreducibility of A-hypergeometric systems, *Indag. Math (N.S.)* **21** (2011), 30–39.

[5] _____, Algebraic A-hypergeometric functions, *Invent. Math.* **180** (2010), 589–610.

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

[7] E. Bod, Algebraicity of the Appell-Lauricella and Horn hypergeometric functions, *J. Differential Equations* **252** (2012), 541–566.

[8] W. Couwenberg, G. Heckman, and E. Looijenga, Geometric structures on the complement of a projective arrangement, *Publ. Math. Inst. Hautes Études Sci.* **101** (2005), 69–161.

[9] P. Deligne and G. D. Mostow, Monodromy of hypergeometric functions and nonlattice integral monodromy, *Publ. Math. Inst. Hautes Études Sci.* **63** (1986), 5–89.

[10] B. Dwork, *Generalized Hypergeometric Functions*. Oxford Mathematical Monographs, Oxford University Press, New York, 1990.

[11] C. F. Gauss, Gesammelte Werke, Bd. III, 1866.

[12] I. M. Gelfand, M. I. Graev, and A. V. Zelevinsky, Holonomic systems of equations and series of hypergeometric type, *Doklady Akad. Nauk SSSR* **295** (1987), 14–19 (in Russian).

[13] I. M. Gelfand, A. V. Zelevinsky, and M. M. Kapranov, Equations of hypergeometric type and Newton polytopes, *Doklady Akad. Nauk SSSR* **300** (1988), 529–534 (in Russian).

[14] _____, Hypergeometric functions and toral manifolds, *Functional Analysis and its Applications* **23** (1989), 94–106. Correction in ibid. **27** (1993), 295.

[15] _____, Generalized Euler integrals and A-hypergeometric functions, *Adv. in Math.* **84** (1990), 255–271.

[16] N. M. Katz, A conjecture in the arithmetic theory of differential equations, *Bull. Soc. Math. France* **110** (1982), 203–239; corrections ibid., 347–348.

[17] F. Klein, *Vorlesungen über die hypergeometrische Funktion*, Grundlehren der Mathematischen Wissenschaften, vol. 34, Springer Verlag, Berlin, 1933.

[18] G. Lauricella, Sulla funzioni ipergeometriche a più variabili, *Rend. Circ. Math. Palermo* **7** (1893), 111–158.

[19] B. Riemann, Beiträge zur Theorie der durch die Gauss'sche Reihe $F(\alpha, \beta, \gamma, x)$ darstellbare Funktionen, *Mathematische Werke*, Artikel IV.

[20] F. Rodriguez-Villegas, Integral ratios of factorials and algebraic hypergeometric functions, *Oberwolfach Reports*, July 2005, math/0701362. Also accessible at `http://www.math.utexas.edu/users/villegas/publications/oberwolfach-05.pdf`

[21] H. A. Schwarz, Über diejenigen Fälle, in welchen die Gaussische hypergeometrische Reihe eine algebraische Funktion ihres vierten Elementes darstellt, *J. Reine Ange. Math.* **75** (1873), 292–335.

[22] M. Saito, B. Sturmfels, and N. Takayama, *Gröbner Deformations of Hypergeometric Differential Equations*, Springer, New York, 2000.

[23] G. C. Shepard and J. A. Todd, Finite unitary reflection groups, *Canadian Math. J.* **6** (1954), 274–304.

[24] M. Schulze and U. Walther, Resonance equals reducibility for A-hypergeometric systems, *Algebra Number Theory* **6** (2012), 527–537.

[25] J. Stienstra, GKZ hypergeometric structures, in: *Arithmetic and Geometry around Hypergeometric Functions*, Progr. Math., vol. 260, Birkhäuser, Basel, 2007, pp. 313–371.

[26] J. Thomae, Ueber die höheren hypergeometrischen Reihen, insbesondere über die Reihe: $1 + \frac{a_0 a_1 a_2}{1 \cdot b_1 b_2} x + \frac{a_0(a_0+1)a_1(a_1+1)a_2(a_2+1)}{1 \cdot 2 \cdot b_1(b_1+1)b_2(b_2+1)} x^2 + \cdots$, *Math. Ann.* **2** (1870), 427–441.