Apéry’s theorem. Thirty years after

Wadim Zudilin
Max-Planck-Institut für Mathematik
Vivatsgasse 7, D-53111 Bonn, Germany
e-mail: wzudilin@mpim-bonn.mpg.de

Abstract

We present a new, rather elementary, proof of the irrationality of $\zeta(3)$ based on some recent ‘hypergeometric’ ideas of Yu. Nesterenko, T. Rivoal, and K. Ball, as well as on the Gosper–Zeilberger algorithm of creative telescoping.

1 Introduction

A question of arithmetic nature of the values of Riemann’s zeta function

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}$$

at odd integral points $s = 3, 5, 7, \ldots$ looks like a challenge for number theorists. An expected answer ‘each odd zeta value is transcendental’ is still far from being resolved. We only dispose of a particular information on the irrationality of odd zeta values, namely:

- $\zeta(3)$ is irrational (R. Apéry [1], 1978);
- infinitely many of the numbers $\zeta(3), \zeta(5), \zeta(7), \ldots$ are irrational (T. Rivoal [11], 2000);
- at least one of the four numbers $\zeta(5), \zeta(7), \zeta(9), \zeta(11)$ is irrational (this author [16], [20], 2001);
- each set $\zeta(s+2), \zeta(s+4), \ldots, \zeta(8s-3), \zeta(8s-1)$ with odd $s > 1$ contains at least one irrational number (this author [17], 2002).

Key words and phrases: Zeta value, hypergeometric series, Apéry’s theorem, creative telescoping.

AMS (MOS) Subject Classifications: Primary 11J82, 33C20; Secondary 11B37, 11M06.

All these results have a classical hypergeometric origin, and we refer the reader roused the curiosity of this terminology to the works [6], [13], [14], [18], [20] for details. The aim of this note is to prove Apéry’s famous result by ‘elementary means’.

**Theorem** (Apéry [1]). The number $\zeta(3)$ is irrational.

The idea of the following proof is due to T. Rivoal [12], [13], who combines approaches of L. Gutnik [5] and Yu. Nesterenko [7], and of K. Ball; our contribution here is to make a use of the Gosper–Zeilberger algorithm of creative telescoping in the most elementary manner.

## 2 Gutnik’s series

In what follows, we denote by $D_n$ the least common multiple of the numbers $1, 2, \ldots, n$ (and take $D_0 = 1$ for completeness). The prime number theorem (whose proof, by the way, depends on the behavior of $\zeta(s)$ in a neighborhood of $s = 1$) implies $D_n^{1/n} \to e$ as $n \to \infty$.

Our starting point is repetition of [7, Section 1] (which in turn originates from the construction in [5]). For each integer $n = 0, 1, 2, \ldots$, define the rational function

$$R_n(t) := \frac{(t - 1) \cdots (t - n)}{(t + 1) \cdots (t + n)}.$$

**Lemma 1** (cf. [7, Lemma 1]). The equality

$$r_n := -\sum_{t=1}^{\infty} R_n'(t) = u_n \zeta(3) - v_n$$

holds, where $u_n \in \mathbb{Z}$ and $D_n^3 v_n \in \mathbb{Z}$.

**Proof.** Taking square of the partial-fraction decomposition

$$\frac{(t - 1) \cdots (t - n)}{t(t + 1) \cdots (t + n)} = \sum_{k=0}^{n} \frac{(-1)^{n-k}(n+k)}{t+k} t + k$$

with a help of the relation

$$\frac{1}{t+k} \cdot \frac{1}{t+l} = \frac{1}{l-k} \cdot \left( \frac{1}{t+k} - \frac{1}{t+l} \right) \quad \text{for} \quad k \neq l,$$
we arrive at the formula

\[ R_n(t) = \sum_{k=0}^{n} \left( \frac{A_{2k}^{(n)}}{(t+k)^2} + \frac{A_{1k}^{(n)}}{t+k} \right), \]

with \( A_{jk} = A_{jk}^{(n)} \) satisfying the inclusions

\[ A_{2k} = \binom{n+k}{n}^2 \binom{n}{k}^2 \in \mathbb{Z} \quad \text{and} \quad D_n A_{1k} \in \mathbb{Z}, \quad k = 0, 1, \ldots, n. \]  

(2)

Furthermore,

\[ \sum_{k=0}^{n} A_{1k} = \sum_{k=0}^{n} \text{Res}_{t=-k} R_n(t) = -\text{Res}_{t=\infty} R_n(t) = 0 \]

since \( R_n(t) = O(t^{-2}) \) as \( t \to \infty \); hence the quantity

\[ r_n = \sum_{l=1}^{\infty} \sum_{k=0}^{n} \left( \frac{2A_{2k}}{(t+k)^3} + \frac{A_{1k}}{(t+k)^2} \right) = \sum_{l=1}^{\infty} \sum_{k=0}^{n} \left( \frac{2A_{2k}}{l^3} + \frac{A_{1k}}{l^2} \right) \]

has the desired form (1), with

\[ u_n = 2 \sum_{k=0}^{n} A_{2k} \quad \text{and} \quad v_n = 2 \sum_{k=0}^{n} A_{2k} \sum_{l=1}^{k} \frac{1}{l^3} + \sum_{k=0}^{n} A_{1k} \sum_{l=1}^{k} \frac{1}{l^2}. \]  

(3)

Finally, using the inclusions (2) and

\[ D_j^n \cdot \sum_{l=1}^{k} \frac{1}{U} \in \mathbb{Z} \quad \text{for} \quad k = 0, 1, \ldots, n, \quad j = 2, 3, \]

we deduce that \( u_n \in \mathbb{Z} \) and \( D_3^n v_n \in \mathbb{Z} \) as required.

Since

\[ R_0(t) = \frac{1}{t^2} \quad \text{and} \quad R_1(t) = \frac{1}{t^2} + \frac{4}{(t+1)^2} - \frac{4}{t} + \frac{4}{t+1}, \]

in accordance with formulae (3) we find that

\[ r_0 = 2\zeta(3) \quad \text{and} \quad r_1 = 10\zeta(3) - 12. \]  

(4)
3 Creative telescoping

Now, with a help of the Gosper–Zeilberger algorithm of creative telescoping [8, Chapter 6] we get the rational function $S_n(t) := s_n(t)R_n(t)$, where

$$s_n(t) := 4(2n+1)(-2t^2 + t + (2n+1)^2),$$ \hspace{1cm} (5)

satisfying the following property.

**Lemma 2.** For each $n = 1, 2, \ldots$, the following identity is true:

$$(n+1)^3R_{n+1}(t) - (2n+1)(17n^2 + 17n + 5)R_n(t) + n^3R_{n-1}(t) = S_n(t+1) - S_n(t).$$ \hspace{1cm} (6)

‘One-line’ proof. Divide both sides of (6) by $R_n(t)$ and verify numerically the resulted identity

$$(n+1)^3 \left( \frac{t-n-1}{t+n+1} \right)^2 - (2n+1)(17n^2 + 17n + 5) + n^3 \left( \frac{t+n}{t-n} \right)^2 = s_n(t+1) \left( \frac{t^2}{(t-n)(t+n+1)} \right)^2 - s_n(t),$$

where $s_n(t)$ is given in (5).

**Lemma 3.** The quantity (1) satisfies the difference equation

$$(n+1)^3r_{n+1} - (2n+1)(17n^2 + 17n + 5)r_n + n^3r_{n-1} = 0$$ \hspace{1cm} (7)

for $n = 1, 2, \ldots$.

**Proof.** Since $R_n'(t) = O(t^{-3})$ and $S_n'(t) = O(t^{-2})$, differentiating identity (6) and summing the result over $t = 1, 2, \ldots$ we arrive at the equality

$$(n+1)^3r_{n+1} - (2n+1)(17n^2 + 17n + 5)r_n + n^3r_{n-1} = S_n'(1),$$

because the sum on the right-hand side telescopes. It remains to note that, for $n \geq 1$, both functions $R_n(t)$ and $S_n'(t) = s_n(t)R_n(t)$ have zero of multiplicity 2 at $t = 1$. Thus $S_n'(1) = 0$ for $n = 1, 2, \ldots$ and we obtain the desired recurrence (7) for the quantity (1).
4 Ball’s series

Consider another rational function

\[ \tilde{R}_n(t) := n!^2 (2t + n) \frac{(t - 1) \cdots (t - n) \cdot (t + n + 1) \cdots (t + 2n)}{(t(t + 1) \cdots (t + n))^4} \tag{8} \]

and the corresponding hypergeometric series

\[ \tilde{r}_n := \sum_{t=1}^{\infty} \tilde{R}_n(t) \tag{9} \]

proposed by K. Ball.

Lemma 4 (cf. [3, the second proof of Lemma 3]). For each \( n = 0, 1, 2, \ldots \), we have the inequality

\[ 0 < \tilde{r}_n < 20(n + 1)^4 (\sqrt{2} - 1)^{4n}. \tag{10} \]

Proof. Since \( \tilde{R}_n(t) = 0 \) for \( t = 1, 2, \ldots, n \) and \( \tilde{R}_n(t) > 0 \) for \( t > n \), we deduce that \( \tilde{r}_n > 0 \).

With a help of elementary inequality

\[ \frac{1}{m} \cdot \frac{(m + 1)^m}{m^{m-1}} = \left( 1 + \frac{1}{m} \right)^m < e < \left( 1 + \frac{1}{m} \right)^{m+1} = \frac{1}{m} \cdot \frac{(m + 1)^{m+1}}{m^m} \]

implying \( (m + 1)^m / m^{m-1} < em < (m + 1)^{m+1} / m^m \) for \( m = 1, 2, \ldots \), we deduce that

\[ e^{-n} \frac{(m + n)^{m+n-1}}{m^{m-1}} < m(m + 1) \cdots (m + n - 1) < e^{-n} \frac{(m + n)^{m+n}}{m^m}. \]

Therefore, for integers \( t \geq n + 1 \),

\[ \tilde{R}_n(t) \cdot \frac{(t + n)^5}{(2t + n)(t + 2n)} = n!^2 \cdot \frac{(t - 1) \cdots (t - n) \cdot (t + n) \cdots (t + 2n - 1)}{(t(t + 1) \cdots (t + n - 1))^4} \]

\[ < (n + 1)^{2(n+1)} \cdot \frac{t^{5t-4}(t + 2n)^{t+2n}}{(t-n)^{t-n}(t+n)^{5(t+n)-4}} \]
and, as a consequence,
\[
\tilde{R}_n(t) \cdot \frac{t^4(t + n)}{(2t + n)(t + 2n)(n + 1)^2} < (n + 1)^2 \cdot \frac{t^{5t}(t + 2n)^{t+2n}}{(t-n)^{t-n}(t + n)^5(t+n)}
= \left(1 + \frac{1}{n}\right)^{2n} \cdot e^{nf(t/n)} < e^2 \cdot \left(\sup_{\tau > 1} e^{f(\tau)}\right)^n,
\]
where
\[
f(\tau) := \log \frac{\tau^5(\tau + 2)^{\tau+2}}{(\tau - 1)^{\tau-1}((\tau + 1)^5(\tau+1))}.
\]
The unique (real) solution \(\tau_0\) of the equation
\[
f'(\tau) = \log \frac{\tau^5(\tau + 2)}{(\tau - 1)\tau+1(\tau + 1)^5} = 0
\]
in the region \(\tau > 1\) is the zero of the polynomial
\[
\tau^5(\tau + 2) - (\tau - 1)(\tau + 1)^5 = -\left(\tau + \frac{1}{2}\right) \left(2\left(\tau + \frac{1}{2}\right)^4 - 5\left(\tau + \frac{1}{2}\right)^2 - \frac{7}{8}\right);\]
it can be determined explicitly:
\[
\tau_0 = -\frac{1}{2} + \sqrt{\frac{5}{4} + \sqrt{2}}.
\]
Thus,
\[
\sup_{\tau > 1} f(\tau) = f(\tau_0)
= f(\tau_0) - \tau_0 f'(\tau_0) = 2 \log(\tau_0 + 2) + \log(\tau_0 - 1) - 5 \log(\tau_0 + 1)
= 4 \log(\sqrt{2} - 1)
\]
and we continue the estimate (11) as follows:
\[
\tilde{R}_n(t) \cdot \frac{t^4(t + n)}{(2t + n)(t + 2n)} < e^2(n + 1)^2(\sqrt{2} - 1)^4n,
\]
Finally, we apply the inequality (12) to deduce the required estimate (10):

\[
\tilde{r}_n = \sum_{t=n+1}^{\infty} \tilde{R}_n(t) < e^2(n + 1)^2(\sqrt{2} - 1)^{4n} \sum_{t=n+1}^{\infty} \frac{(2t + n)(t + 2n)}{t^4(t + n)}
\]

\[
< e^2(n + 1)^2(\sqrt{2} - 1)^{4n} \sum_{t=n+1}^{\infty} \left( \frac{2}{t^5} + \frac{5n}{t^4} + \frac{2n^2}{t^3} \right)
\]

\[
\leq e^2(n + 1)^2(\sqrt{2} - 1)^{4n}
\]

\[
< 20(n + 1)^4(\sqrt{2} - 1)^{4n}.
\]

5 Coincidence of Gutnik’s and Ball’s series

Applying the algorithm of creative telescoping, this time with data (8), we obtain the certificate

\[
\tilde{S}_n(t) := \frac{\tilde{R}_n(t)}{(2t + n)(t + 2n - 1)(t + 2n)} \cdot (-t^6 - (8n - 1)t^5 + (4n^2 + 27n + 5)t^4
\]

\[
+ 2n(67n^2 + 71n + 15)t^3 + (358n^4 + 339n^3 + 76n^2 - 7n - 3)t^2
\]

\[
+ (384n^5 + 396n^4 + 97n^3 - 29n^2 - 17n - 2)t
\]

\[
+ n(153n^5 + 183n^4 + 50n^3 - 30n^2 - 22n - 4)) \quad (13)
\]

Lemma 5. For each \( n = 1, 2, \ldots \), the identity

\[
(n+1)^3\tilde{R}_{n+1}(t) - (2n+1)(17n^2+17n+5)\tilde{R}_n(t) + n^3\tilde{R}_{n-1}(t) = \tilde{S}_n(t+1) - \tilde{S}_n(t)
\]

holds.

‘One-line’ proof. Divide both sides of (14) by \( \tilde{R}_n(t) \) and verify the resulted identity.

Lemma 6. The quantity (9) satisfies the difference equation (7) for \( n = 1, 2, \ldots \).

Proof. Since \( \tilde{R}_n(t) = O(t^{-5}) \) and \( \tilde{S}_n(t) = O(t^{-2}) \) as \( t \to \infty \) for \( n \geq 1 \), summation of equalities (14) over \( t = 1, 2, \ldots \) yields the relation

\[
(n + 1)^3\tilde{r}_{n+1} - (2n + 1)(17n^2 + 17n + 5)\tilde{r}_n + n^3\tilde{r}_{n-1} = -\tilde{S}_n(1).
\]

It remains to note that, for \( n \geq 1 \), both functions (8) and (13) have zero at \( t = 1 \). Thus \( \tilde{S}_n(1) = 0 \) for \( n = 1, 2, \ldots \) and we obtain the desired recurrence (7) for the quantity (9).
Lemma 7. For each \( n = 0, 1, 2, \ldots \), the quantities (1) and (9) coincide.

Proof. Since both \( r_n \) and \( \tilde{r}_n \) satisfy the same second-order difference equation (7), we have to verify that \( r_0 = \tilde{r}_0 \) and \( r_1 = \tilde{r}_1 \). Direct calculations show that

\[
\tilde{R}_0(t) = \frac{2}{t^2}, \quad \tilde{R}_1(t) = -\frac{2}{t^4} + \frac{2}{(t+1)^4} + \frac{5}{t^3} + \frac{5}{(t+1)^3} - \frac{5}{t^2} + \frac{5}{(t+1)^2},
\]

hence \( \tilde{r}_0 = 2\zeta(3) \) and \( \tilde{r}_1 = 10\zeta(3) - 12 \), and comparison of this result with (4) yields the desired coincidence.

6 Proof of Apéry’s theorem

Suppose, on the contrary, that \( \zeta(3) = p/q \), where \( p \) and \( q \) are positive integers. Then, using the ‘trivial’ bound \( D_n < 3^n \) valid for \( n \geq n_0 \), we deduce that, for each index \( n \geq n_0 \), the integer \( qD_n^3r_n = D_n^3u_np - D_n^3v_nq \) satisfies the estimate

\[
0 < qD_n^3r_n < 20q(n+1)^43^{3n}(\sqrt{2} - 1)^{4n};
\]

this is not possible since \( 3^3(\sqrt{2} - 1)^4 = 0.7948 \ldots < 1 \) and the right-hand side of (15) is less than 1 for \( n \) sufficiently large which is a contradiction.

7 Concluding remarks

In spite of its elementary argument, our proof of Apéry’s theorem does not look much simpler than the original (also elementary) Apéry’s proof well-explained in A. van der Poorten’s informal report [9], or (almost elementary) Beukers’s proof [4] using Legendre polynomials and multiple integrals. In fact, Poincaré’s theorem on asymptotic behavior of solutions of difference equations allows one to avoid using Ball’s series (9) and the analysis in Sections 4 and 5: Dividing both sides of (7) by \( n^3 \) we see that the solutions of (7) are, in a certain sense, ‘close’ to the solutions of

\[
r_{n+1} - 34r_n + r_n = 0, \quad n = 0, 1, 2, \ldots ,
\]

which are \( r_n = c_1\lambda_1^n + c_2\lambda_2^n \) with \( \lambda_1 = (\sqrt{2} - 1)^4 \), \( \lambda_2 = (\sqrt{2} + 1)^4 \) and \( c_1, c_2 \) arbitrary (complex) constants. This ‘closeness’ is the subject of Poincaré’s
Apéry’s theorem. Thirty years after

theorem, which states that for any non-trivial solution \( r_n \) (and we are in this case, since \( r_0 \) and \( r_1 \) are non-zero according to (4)) of (7),

\[
\lim_{n \to \infty} |r_n|^{1/n} \in \{\lambda_1, \lambda_2\}.
\]

Already naive estimating the terms in (1) gives a bound of the form \( |r_n| < Cn \) with \( C > 0 \) a constant, hence only \( |r_n|^{1/n} \to \lambda_1 = (\sqrt{2} - 1)^4 \) remains possible. With this asymptotic estimate one easily concludes with the irrationality of \( \zeta(3) \) as we do in Section 6.

The fact that \( \tilde{r}_n = \tilde{u}_n \zeta(3) - \tilde{v}_n \) with \( D_n \tilde{u}_n, D_n^4 \tilde{v}_n \in \mathbb{Z} \) was first discovered by K. Ball; the proof follows lines of the proof of Lemma 1 and the vanishing of the coefficients of \( \zeta(4) \) and \( \zeta(2) \) is due to the well-poised origin of the hypergeometric series (9). Recently, C. Krattenthaler and T. Rivoal [6, Section 16] showed the coincidence of Gutnik’s and Ball’s series without use of the difference equation (7) but applying directly certain hypergeometric transformations to the coefficients \( u_n \) and \( v_n \) of the linear forms (1). This proof resembles a classical (unfortunately, less elementary) recipe: Lemma 7 can be proved by specialisation of Bailey’s identity [2, Section 6.3, formula (2)]

\[
\begin{align*}
\Gamma\left(1 + \frac{1}{2}a\right) & \Gamma\left(1 + a - b\right) \Gamma\left(1 + a - c\right) \Gamma\left(1 + a - d\right) \Gamma\left(1 + a - e\right) \Gamma\left(1 + a - f\right) \\
\times & \Gamma(1 + a - c - d) \Gamma(1 + a - e - f) \\
\times & \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\Gamma(b + t) \Gamma(c + t) \Gamma(d + t) \Gamma(1 + a - e - f + t)}{\Gamma(1 + a - e + t) \Gamma(1 + a - f + t)} dt,
\end{align*}
\]

provided that the very-well-poised hypergeometric series on the left-hand side converges. Namely, taking \( a = 3n + 2 \) and \( b = c = d = e = f = n + 1 \) in (16) we obtain Ball’s sequence (9) on the left and Apéry’s sequence (1) on the right (for the last fact see [7, Lemma 2]). The general identity (16) can be put forward for explaining on how the permutation group from [10] for linear forms in 1 and \( \zeta(3) \) appears (see [20, Sections 4 and 5] for details). On the other hand, the hypergeometric machinery developed in [6] could lead to further novelties on the arithmetic nature of odd zeta values.

We would like to point out that our way of deducing the recursion (7) for the sequence \( r_n \) easily extends to showing that the coefficients \( u_n \) and \( v_n \)
satisfy (7) as well. Indeed, if we multiply both sides of (6) by \((t + k)^2\), substitute \(t = -k\) and sum over all integers \(k\), then we arrive at the difference equation (7) for the sequence \(u_n\); as for the sequence \(v_n\), it follows from the equality \(v_n = u_n\zeta(3) - r_n\). This approach slightly differs from those used in [9, Section 8] and [15, Section 13] although it is based on the same algorithm of creative telescoping. This algorithm and the above scheme allow us [18], [19] to obtain Apéry-like difference equations for \(\zeta(4)\) and Catalan’s constant \(G = \sum_{n=0}^{\infty} (-1)^n(2n + 1)^{-2}\).

Acknowledgement. The work was supported by the Hausdorff Center for Mathematics (Bonn).

References


Apéry’s theorem. Thirty years after


[16] W. Zudilin, One of the numbers \( \zeta(5), \zeta(7), \zeta(9), \zeta(11) \) is irrational, Uspekhi Mat. Nauk [Russian Math. Surveys], 56:4 (2001), 149–150.


