EXPERIMENTAL METHODS IN NUMBER THEORY AND COMBINATORICS

By

ROBERT DOUGHERTY-BLISS

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

Doron Zeilberger

And approved by

New Brunswick, New Jersey

February 27, 2024

ABSTRACT OF THE DISSERTATION

Experimental Methods in Number Theory and Combinatorics

By ROBERT DOUGHERTY-BLISS

Dissertation Director:

Doron Zeilberger

Some results in experimental mathematics are presented. In particular, new primality tests based on the theory of linear recurrences; the confirmation of some conjectures by Manuel Kauers and Christoph Koutschan; some new proofs of famous summation identities; a connection to Gosper summability of factorials and bell numbers; the creation of new, decidable diophantine equations; experiments related to Beukers' proof of the irrationality of Apéry's constant; and an enumeration of certain types of restricted Dyck paths.

Contents

List of Figures

Acknowledgements

This thesis would not have been possible without the support of many people. To begin, I thank my family for their support: My mom Tina Dougherty, my dad Drew Bliss, my grandpop Bill Dougherty, my grandmon Patrica Shipley, my grandfather Robert Bliss, and in particular my Nana, Nancy Bliss.

I have been mentored by many excellent teachers. In roughly chronological order: David Caldwell, Iona Wynter-Parks, Philip Tiu, Lynn Gieger, John Merkel, John Nardo, Brian Patterson, Mary Gardner, Michele Benzi, and many more than I can list here. A heartfelt thank you to all of these educators.

My work has been supported by many administrative and support staff who have made work more tolerable, At Rutgers, I offer my most sincere thanks to Katie Guarino, who was always a pleasure to see around the department.

My day-to-day spirit was often supported by my friends and students. In particular I have the deepest affection for Leonidas Daskalakis, Charles Kenney, and Natasha Ter-Saakov, but it has been a great pleasure to share knowledge and laughs with everyone at Rutgers.

The vast majority of this text consists of articles I previously wrote with collaborators. Each one of them—Doron Zeilberger, Manuel Kauers, Christoph Koutschan, AJ Bu, and Charles Kenney—generously gave their permission to reuse our material.

I must express my gratitude to the citizens of South Carolina and New Jersey, who provided robust public schools where I began my education and worked while completing this thesis. This gratitude extends to all citizens of the United States, who helped fund my education through various grants and aid programs. For those hoping for a more corporate kind of patriotism, I also thank the Coca-Cola Foundation for funding a generous scholarship I received at Oglethorpe University. (Shareholders: I promise I am buying your delicious beverages as fast as I can to pay you back.)

Finally, the greatest force exerted upon this thesis was provided by my advisor, Doron Zeilberger. He has provided guidance, encouragement, and a commitment to mentoring without peer. His insight, dedication, and inimitable style radiates from every word which follows.

Chapter 1

Introduction

This thesis contains results across a variety of fields, chiefly combinatorics, number theory, and the theory of linear recurrences and summation. The chapters are essentially independent and either self-contained or accompanied by enough references that an interested reader could follow along. Pick your favorite parts and ignore the rest. The remainder of this chapter is an introduction to experimental mathematics as I have seen it practiced at Rutgers and elsewhere, targeted at the uninitiated but interested mathematician.

The theme which underlies this thesis is an embrace of the experimental approach. Roughly speaking, this means incorporating an aggressive interrogation of data within the process of doing mathematics. The problems about things that you can touch and manipulate easily on a computer, the methods focus on writing programs to generate and analyze data, and the results are sometimes conjectures rather than proven theorems.

Of course all mathematicians make conjectures from data, but the difference here is one of *degrees*. There is a collection of methods and techniques that experimental mathematicians seem to make use of more than others, all related to the idea of guessing. To clarify the point, the following sections are a collection of brief case studies where the experimental approach seems to shine through.

1.1 Origami

In 2021, Natalya Ter–Saakov asked me for help solving a graph theory problem related to the theory of origami folding [Hul+22]. It is simple enough to state without any knowledge of origami.

Definition 1. An *assignment array* of length $2n$ is a circular array of length $2n$ which contains -1's and 1's whose entries sum to ± 2. The *all-equal-angles origami flip graph* A_{2n} consists of one vertex for every assignment array of length $2n$, and edges joining two arrays if one can be obtained from the other by negating

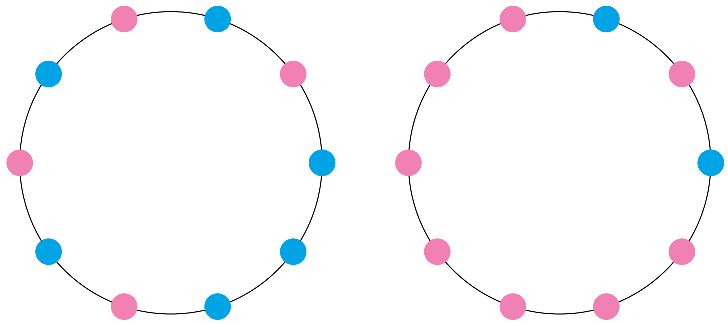


Figure 1.1: Left: A valid array of size 10. Right: An invalid array of size 10.

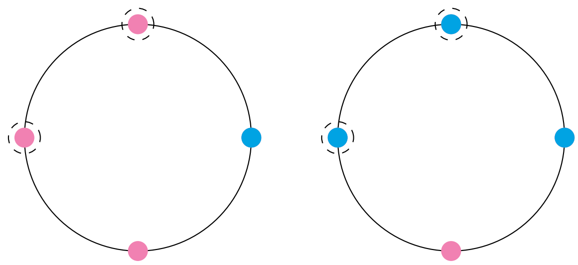


Figure 1.2: Two adjacent vertices in A_4 and their flipped bits.

two adjacent entries.

Ter-Saakov's goal was to determine the number of edges in A_{2n}. She and her collaborators had a program that could compute a few terms of the sequence, which starts as follows:

$$2, 16, 84, 400, 1820, 8064, 35112, 151008, 643500, 2722720, 11454872, 47969376, 200107544.$$

The group was unable to find a closed-form expression on their own, and unfortunately the sequence was not in the OEIS [OEI24]. (The sequence divided by 2 was entered into the OEIS after the end of this story.)

I was able to find a closed form expression for their sequence in a few minutes. I typed the numbers into Maple and passed them to Zeilberger's `FindRec.txt` package [Zei24a] as follows:

```
> x := [2, 16, 84, 400, 1820, 8064, 35112, 151008,
        643500, 2722720, 11454872, 47969376, 200107544]:
> Findrec(x, n, N, 3);
                2 (3 n + 1) (2 n - 1)
              - --------------------- + N
                     n (3 n - 2)
```

This told me that the sequence $a(n)$, the number of edges in A_{2n}, seemed to satisfy the recurrence

$$a(n+1) = \frac{2(3n+1)(2n-1)}{n(3n-2)} a(n). \tag{1.1}$$

From here it was a simple matter to unroll the recurrence and conjecture the correct answer

$$a(n) = \frac{(n+1)(3n-2)}{2n-1} \binom{2n}{n-1}.$$

Ter-Saakov, inspired by my formula, quickly found a proof using probabilistic methods.

The method to conjecture the recurrence (1.1) is not particularly complicated. Combinatorial quantities can often be expressed as the product and quotient of factorials, binomial coefficients, and polynomials. If $a(n)$ is of this form, then the term ratio $a(n+1)/a(n)$ will be some fixed rational function in n. So, as an educated guess, we write

$$\frac{a(n+1)}{a(n)} = \frac{c_d n^d + c_{d-1} n^{d-1} + \cdots + c_0}{b_m n^m + b_{m-1} n^{m-1} + \cdots + b_0}, \tag{1.2}$$

make the numerator and denominator of the rational function have a reasonable degree like 4, and plug in some values of the sequence $a(n)$ that we know from computation. Cross multiplication and equating coefficients leads to a system of linear equations in the coefficients of the rational function, and solving this gives the conjectured expression for $a(n+1)/a(n)$. This is how one could find a recurrence satisfied by *hypergeometric* sequences, which are precisely those which satisfy an equation of the form (1.2). Zeilberger's `Findrec` works on a more general class of sequences which are called *D-finite* [Kau23].

The key point is that recurrences like (1.1), while common, can be arbitrarily difficult to find by hand. Many sequences satisfy simple recurrences, but many more satisfy very complicated ones, and even more satisfy *no* recurrence of this form. Rather than trying to find a needle in one of infinitely many haystacks, it is better to run automated analyses.

1.2 Permutations and the OEIS

The most famous experimental method to solve a problem is "consult the OEIS," the authoritative sequence collection. This has led to some really stupendous discoveries.

For instance, in 2019, Lara Pudwell was studying the following types of permutations [Pud20].

Definition 2. A permutation π on $\{1,2,\ldots,n\}$ is *alternating* provided that $\pi(1) > \pi(2) < \pi(3) > \cdots$, and so on. A permutation π *contains the pattern 123* provided that there are integers $i < j < k$ such that $\pi(i) < \pi(j) < \pi(k)$. The number of 123 patterns contained in π is the number of such $i < j < k$ triples.

Pudwell discovered that the maximum number of 123 patterns contained in a permutation of length $n+3$ is A168380. This happens to have the closed form

$$\frac{(n+1)(2n^2+4n+3-3(-1)^n)}{12}.$$

More surprisingly, A168380 came up eight years earlier in an unrelated fashion. Alonso del Arte pointed out in the OEIS that the first eight terms of the sequence (prepended by a zero) happen to equal the "atomic numbers of the augmented alkaline earth group in Charles Janet's spiral periodic table." Or, as Pudwell puts it, "the quasi-polynomial sequence 2, 4, 12, 20, 38, 56, 88, ... [gives] the atomic numbers of helium, beryllium, magnesium, calcium, and more" [Pud20]. The OEIS is great for finding these kind of unexpected connections.

Another success occurred in 2022. Christoph Koutschan and Manuel Kauers used lattice reduction techniques to build a cutting edge recurrence guesser, then ran it on every sequence in the OEIS. One promising result was a conjectured recurrence for A189281:

$$a(n) = \{\pi \in S_n \mid \pi(i+2) - \pi(i) \neq 2,\ 1 \leq i < n-1\}.$$

Like most sequences counting permutations, it is difficult to compute terms of $a(n)$ from the definition. According to a comment in the OEIS, in 2012 it took 78 computer hours to compute $a(34)$, 147 to compute $a(35)$, and no one bothered to compute beyond that. The method of "classical guessing," as implemented in Zeilberger's `FindRec.txt`, could not find a recurrence satisfied by the available terms in 2012. With their enhanced approach, Koutschan and Kauers conjectured (essentially) the following recurrence:

$$\begin{aligned}
&((-1+n)^2 na(n))/4 + (n(-16+38n+11n^2)a(1+n))/16+\\
&\quad (3/2+(139n)/16+(29n^2)/8+(3n^3)/16)a(2+n)+\\
&\quad (-21/4-(51n)/4-(79n^2)/16-(5n^3)/8)a(3+n)+\\
&\quad (-15/2-n/8+(5n^2)/4+n^3/8)a(4+n)+\\
&\quad (603/4+(307n)/4+(49n^2)/4+(11n^3)/16)a(5+n)+\\
&\quad (-41-(533n)/16-(49n^2)/8-(5n^3)/16)a(6+n)+\\
&\quad (-911/2-161n-(303n^2)/16-(3n^3)/4)a(7+n)+\\
&\quad (-363-(417n)/4-(37n^2)/4-n^3/4)a(8+n)+\\
&\quad (-993/4-53n-(11n^2)/4)a(9+n)+(-130-(93n)/4-n^2)a(10+n)+\\
&\quad (-71/4-2n)a(11+n)+(-10-n)a(12+n)+a(13+n)=0.
\end{aligned}$$

Inspired by the potential for faster computation of $a(n)$, independent checkers worked hard to confirm that the recurrence agrees with $a(n)$ up to $n = 300$, though it is still officially an open problem to prove that this recurrence is correct.

1.3 RIES

A lesser-known cousin of the OEIS is RIES, Robert Munafo's *RILYBOT Inverse Equation Solver* [Mun]. Given any floating point approximation to a real number, RIES tries to guess an equation that it might satisfy. This has many silly applications—for example, this thesis will be published in $\lfloor e^5 + 4^{2e} \rfloor$ CE, approximately $5e^{\sqrt{e}}$ years after I was born—but also some serious ones.

The primary use is to identify constants which come up in experimental work. A nice example occurred in a Math Stack Exchange question [EHE]. Does the series

$$\sum_{k\geq 1} (-1)^k \frac{\zeta(2k)}{2^{2k-1}},$$

where $\zeta(s) = \sum_{k\geq 1} k^{-s}$ is the Riemann zeta function, have a closed form? Summing the first 400 terms of this series yields the approximation

$$0.71268857495964775560916908658601926830.$$

If we pass this to RIES then we obtain many potential equations, but one with an unusually good fit (output edited for brevity):

```
$ ries 0.712688574959647755609169086586019268 3

   Your target value: T = 0.712688574959648        mrob.com/ries

         1/x^2 = 2                      for x = T - 0.00558179  {44}

        -ln(x) = 1/3                    for x = T + 0.00384274  {57}

     1/cospi(x) = -phi                  for x = T - 0.000617977 {60}

       sinpi(x) = atan2(1)              for x = T - 0.000230344 {57}

                                  ...

       x+1-pi/2 = pi/(e^pi-1)           for x = T - 1.11022e-16 {127}

  (Stopping now because best match is within   1e-15 of target value.)
```

RIES conjectured the closed form

$$\sum_{k\geq 1}(-1)^k\frac{\zeta(2k)}{2^{2k-1}}=\frac{\pi}{e^{\pi}-1}+\frac{\pi}{2}-1. \tag{1.3}$$

With the confidence given by a convincing guess, this is easy to establish with Euler's identity

$$\zeta(2k)=(-1)^{k+1}\frac{(2\pi)^{2k}B_{2k}}{2(2k)!}$$

relating $\zeta(2k)$ to π and the $2k$th Bernoulli number B_{2k}, and the generating function for the Bernoulli numbers themselves.

See the tutorial on constant identification by Stoutemyer for more information [Sto23].

1.4 Conclusion

We have still not exhausted the experimental toolkit. We have not mentioned the integer relation algorithm PSQL [FBA99], famously used to discover the following formula for π [BBP97]:

$$\pi=\sum_{k\geq 0}\frac{1}{16^k}\left(\frac{4}{8k+1}-\frac{2}{8k+4}-\frac{1}{8k+5}-\frac{1}{8k+6}\right).$$

We have not mentioned *gfun*, a package which guesses and manipulates generating functions [SZ94]. And we have not mentioned any of the Mathematica packages for experimental work developed at the Research Institute for Symbolic Computation [RIS].

Nontheless, hopefully this has communicated some of what experimental mathematics is about: The aggressive interrogation of data as a regular part of mathematical research.

Chapter 2

Integral Recurrences from A to Z

The following article appeared in *The American Mathematical Monthly* [Dou22]. It was written over the winter of 2020 as a distant companion piece to a joint paper with Doron Zeilberger and Christoph Koutschan about Beukers integrals [DKZ22]. I thought of it as something to hand to a sharp undergraduate student as a stepping stone, somewhat towards number theory but mostly towards computer algebra. We sometimes teach undergraduates calculus as if it were 1755 [Eul55], and I wanted to show students this is not the case anymore.

This is chiefly an expository work, but I include it here because I have some small updates to the references. In particular, I have mentioned where readers can learn more about the Almkvist–Zeilberger algorithm itself, as well as modern implementations.

I do not remember what the biblical quote was supposed to convey.

2.1 Introduction

Behold, I will stand before thee there upon the rock in Horeb; and thou shalt smite the rock and there shall come water out of it, that the people may drink.

— Exodus 17:6

You have been up all night working out the ingenious solution to your latest problem. Your answer depends on the integral sequence

$$I(n) = \int_{-\infty}^{\infty} \frac{x^{2n}}{(x^2+1)^{n+1}}\, dx,$$

which you desperately need to evaluate. You know that you could break out special functions, contour integrals, or some other method, but you would really just like a quick answer without much fuss.

You run to download the file EKHAD from

https://sites.math.rutgers.edu/~zeilberg/tokhniot/EKHAD

and read it into Maple with "`read EKHAD;`". You type the command

```
AZd(x^(2 * n) / (x^2 + 1)^(n + 1), x, n, N);
```

and hardly a second has passed when Maple produces the following:

```
-2 n - 1 + (2 n + 2) N, -x.
```

You cry out in joy, for the *Almkvist–Zeilberger algorithm* has told you that your integrand satisfies the "recurrence"

$$(-2n-1+(2n+2)N)\frac{x^{2n}}{(x^2+1)^{n+1}} = -\frac{d}{dx}x\frac{x^{2n}}{(x^2+1)^{n+1}},$$

where N is the shift operator defined by $Nf_n(x) = f_{n+1}(x)$. Integrating this equation on $(-\infty,\infty)$ gives the identity

$$(-2n-1+(2n+2)N)I(n) = 0,$$

which would traditionally be written as

$$I(n+1) = \frac{2n+1}{2(n+1)}I(n).$$

Using the initial condition $I(0) = \pi$, you crank out the first few terms of the sequence:

$$\pi, \frac{\pi}{2}, \frac{3\pi}{8}, \frac{5\pi}{16}, \frac{35\pi}{128}, \frac{63\pi}{256}, \frac{231\pi}{1024}, \frac{429\pi}{2048}, \frac{6435\pi}{32768}, \frac{12155\pi}{65536}, \frac{46189\pi}{262144}.$$

The denominators look like powers of 2. After some experimentation, you let $\pi = 1$ and multiply by 4^n. This produces some *integers*:

$$1, 2, 6, 20, 70, 252, 924, 3432, 12870, 48620, 184756.$$

You visit the On-Line Encyclopedia of Integer Sequences [OEI24] at https://oeis.org, type in your integers, and receive word that they are the *central binomial coefficients* $\binom{2n}{n}$. You have just conjectured that

$$I(n) = \int_{-\infty}^{\infty} \frac{x^{2n}}{(x^2+1)^{n+1}}\,dx = \frac{\pi}{4^n}\binom{2n}{n}.$$

To wrap up, you note that the final expression satisfies the same initial condition. You check the recurrence by typing the commands

```
L := Pi * binomial(2 * (n + 1), n + 1) / 4^(n + 1):
R := Pi * binomial(2 * n, n) / 4^n:
simplify(convert(L / R, factorial));
```

and observing the output

```
2 n + 1
-------.
2 n + 2
```

This is a rigorous proof requiring minimal effort on your part. Such is a normal case study of the *AZ* algorithm.

In general, we often want to understand the sequence of *definite* integrals

$$I(n) = \int F_n(x)\,dx.$$

Perhaps we would like to compute the first twenty or so terms to see what $I(n)$ looks like. Sometimes we can ask a computer to churn these out, but other times $F_n(x)$ is so complicated that even our electronic friends would struggle to keep up for large n. What we need is an *efficient algorithm* to compute the terms of $I(n)$. We need a *recurrence*.

There are plenty of ad hoc methods to find a recurrence for $I(n)$. You could integrate by parts or differentiate under the integral sign, for example. But these all require ingenuity, insight, and hard work. As Sir Alfred Whitehead once remarked, such ingenuity is overrated. No one wants to work hard—we want answers!

The *painless* way to discover these recurrences for large classes of integrals is the *Almkvist–Zeilberger* algorithm. This is the direct analog of the celebrated Wilf–Zeilberger method of automatic definite summation, but it has received less attention than its discrete counterpart. Our goal here is to explore the Almkvist–Zeilberger algorithm with a few case studies, leaving the door open for more experimentation.

2.2 A quick start guide to the AZ algorithm

The Wilf–Zeilberger method of definite summation was a breakthrough in automatic summation. Roughly, the Wilf–Zeilberger method can automatically prove (and semi-automatically discover) most commonly occurring summation identities of the form

$$S(n) = \sum_k f(n,k) = RHS(n).$$

One piece of the puzzle is that, whenever $f(n,k)$ is a "suitable" function, it satisfies a special type of inhomogeneous linear recurrence with polynomial coefficients in n. Specifically, there exists a nonnegative integer d and polynomials $p_j(n)$ such that

$$\sum_{j=0}^{d} p_j(n) f(n+j,k) = G(n,k+1) - G(n,k),$$

where $G(n,k)$ is some function with $G(n,\pm\infty) = 0$. Summing over k yields the recurrence

$$\sum_{j=0}^{d} p_j(n) S(n+j) = G(n,\infty) - G(n,-\infty) = 0.$$

This method, also known as creative telescoping, has been (rightly) advertised from here to the Moon and back. See the article [Tef04], the book [PWZ97], the lecture notes [Zei95], and the lively *Monthly* article [Nem+97].

The Almkvist–Zeilberger algorithm is to definite integrals what the Wilf–Zeilberger method is to definite sums. The input to the algorithm is a "suitable" function $F_n(x)$ with a discrete parameter n. The output is a nonnegative integer d, polynomials $p_k(n)$, and a rational function $R(x)$ such that

$$\sum_{k=0}^{d} p_k(n) F_{n+k}(x) = \frac{d}{dx} R(n,x) F_n(x).$$

The left-hand side is independent of x except for the $F_n(x)$, so integrating this equation on $[0,1]$, say, gives

$$\sum_{k=0}^{d} p_k(n) \int_0^1 F_{n+k}(x) = R(n,1)F_n(1) - R(n,0)F_n(0).$$

If $F_n(0) = F_n(1) = 0$ and $R(n,x)$ is well-behaved, then $I(n) = \int_0^1 F_n(x)$ satisfies

$$\sum_{k=0}^{d} p_k(n) I(n+k) = 0,$$

meaning that we have discovered a recurrence for the sequence of integrals $I(n)$. The only thing to verify is that $F_n(x)$ is "suitable," and that $R(n,x)$ is well-behaved on the region of integration.

What functions are "suitable"? The requirement is that $F_n(x)$ is *hypergeometric* in n and x, meaning that there exist fixed rational functions $R_1(n,x)$ and $R_2(n,x)$ such that

$$F_{n+1}(x)/F_n(x) = R_1(n,x)$$
$$F_n'(x)/F_n(x) = R_2(n,x).$$

This is all that the algorithm needs to produce its identity.

The version of the Almkvist–Zeilberger algorithm that we will use is implemented in the procedure `AZd(f, x, n, N)` in the Maple package `EKHAD` referenced in the introduction. It takes an expression f in the continuous variable x and discrete parameter n. The symbol N stands for the "shift" operator N on the set of sequences by

$$Na(n) = a(n+1).$$

For example, the Fibonacci numbers $F(n)$ satisfy

$$(N^2 - N - 1)F(n) = 0.$$

I have deliberately chosen to not sketch the inner workings of the Alkmvist–Zeilberger algorithm or detail every modern improvement or implementation. To learn the details, see the original article [AZ90]. To learn about some improvements to the theory, and where to find efficient multivariate implementations, see [AZ06; Abl21].

Let us get on to the case studies.

2.3 Factorials

Let us begin humbly, by evaluating an integral that we already know.

Proposition 1. *For each integer $n \geq 0$,*

$$I(n) = \int_0^\infty e^{-x} x^n \, dx = n!.$$

Proof. Typing the command

```
AZd(exp(-x) * x^n, x, n, N);
```

into Maple produces:

```
N - n - 1, -x.
```

That is, the Almkvist–Zeilberger algorithm has told us that

$$(N-(n+1))f_n(x) = -\frac{d}{dx}e^{-x}x^{n+1}.$$

Since the antiderivative of the right-hand side vanishes for $x=0$ and $x=\infty$, integrating on $[0,\infty)$ gives

$$(N-(n+1))I(n) = 0,$$

and since $I(0)=1$, we have $I(n)=n!$. $\square$

2.4 "A Complicated Integral"

This is from Section 3.8 of [BM04].

Proposition 2.

$$I(n) = \int_0^\infty \frac{x^n}{(x+1)^{n+r+1}}\,dx = \left[r\binom{r+n}{n}\right]^{-1}.$$

Proof. Typing the command

```
AZd(x^n / (x + 1)^(n + r + 1), x, n, N);
```

into Maple produces:

```
(n + 1) + (-n - r - 1) N, x.
```

And for $r>0$, integrating the implied identity

$$((n+1)-(n+r+1)N)\frac{x^n}{(x+1)^{n+r+1}} = \frac{d}{dx}x\frac{x^n}{(x+1)^{n+r+1}}$$

yields

$$((n+1)-(n+r+1)N)I(n) = 0.$$

The sequence $(r\binom{r+n}{n})^{-1}$ satisfies the same recurrence and initial condition (check!). $\square$

2.5 Central binomial coefficients

Proposition 3. *The integral sequence*

$$I(n) = \int_0^1 (x(1-x))^n\,dx$$

satisfies

$$(N-\frac{n+1}{2(2n+3)})I(n)=0.$$

Proof. Typing the command

```
AZd((x * (1 - x))^n, x, n, N);
```

into Maple produces:

```
n + 1 + (-4 n - 6) N, (-1 + 2 x) (-1 + x) x.
```

Integrating the implied identity

$$(N-\frac{n+1}{2(2n+3)})(x(1-x))^n=\frac{d}{dx}(2x-1)(x-1)x(x(1-x))^n,$$

on $[0,1]$ yields the result, since the antiderivative of the right-hand side vanishes at $x=0$ and $x=1$. □

The recurrence implies that $I(n)$ begins as follows:

$$1/6,\ 1/30,\ 1/140,\ 1/630,\ 1/2772,\ 1/12012,\ 1/51480,\ldots.$$

Corollary 1.

$$I(n)=\frac{1}{(2n+1)\binom{2n}{n}}.$$

Proof. Both sequences satisfy the same recurrence and initial condition (check!). □

Integral evaluation is an area full of unintended consequences. Here we have an example, since one way to *try* and evaluate $I(n)$ is by applying the binomial theorem to the integrand:

$$\begin{aligned}I(n)&=\int_0^1 x^n(1-x)^n\,dx\\&=\int_0^1\sum_{k=0}^{n}\binom{n}{k}(-1)^k x^{n+k}\,dx\\&=\sum_{k=0}^{n}\binom{n}{k}\frac{(-1)^k}{n+k+1}.\end{aligned}$$

This remaining sum is complicated, but we can pair it with our previous corollary to get another.

Corollary 2.

$$\sum_{k=0}^{n}\binom{n}{k}\frac{(-1)^k}{n+k+1}=\frac{1}{(2n+1)\binom{2n}{n}}.$$

2.6 Irrationality

Our final case study is a slightly more complicated sequence of integrals. We will not derive a closed form, but to compensate we will uncover a wealth of other information.

Proposition 4. *The integral sequence*

$$I(n) = \int_0^1 (x(1-x))^n e^{-x}\, dx$$

satisfies

$$(N^2 + 2(2n+3)(n+2)N - (n+1)(n+2))I(n) = 0.$$

Proof. Let $f_n(x)$ be the integrand. The Almkvist–Zeilberger algorithm produces the "calculus exercise"

$$\begin{aligned}(N^2+&2(2n+3)(n+2)N-(n+1)(n+2))f_n(x)\\ &= \frac{d}{dx}(-2nx^3 - x^4 + 3nx^2 - 2x^3 - nx + 5x^2 - 2x)f_n(x),\end{aligned}$$

and integrating this proves the proposition. □

This recurrence is too complicated to solve in a reasonable way. However, it does help us produce the following initial terms:

$$-1+\frac{3}{e},\ 14-\frac{38}{e},\ -426+\frac{1158}{e},\ 24024-\frac{65304}{e},\ \ldots.$$

This data is very suggestive! It seems that

$$I(n) = a_n + b_n e^{-1}$$

for some *integers* a_n and b_n. We can check this immediately with the recurrence: if $I(n) = a_n + b_n e^{-1}$ and $I(n+1) = a_{n+1} + b_{n+1}e^{-1}$, then

$$\begin{aligned}I(n+2) &= -2(2n+3)(n+2)I(n+1) + (n+1)(n+2)I(n)\\ &= -2(2n+3)(n+2)(a_{n+1} + b_{n+1}e^{-1}) + (n+1)(n+2)(a_n + b_n e^{-1})\\ &= a_{n+2} + b_{n+2}e^{-1},\end{aligned}$$

where we take

$$\begin{aligned}
a_{n+2} &= -2(2n+3)(n+2)a_{n+1} + (n+1)(n+2)a_n \\
b_{n+2} &= -2(2n+3)(n+2)b_{n+1} + (n+1)(n+2)b_n.
\end{aligned}$$

That is, a_n and b_n are sequences of integers which satisfy the *same recurrence* that $I(n)$ satisfies, only the initial conditions are different:

$$\begin{aligned}
a_1 = -1 \quad a_2 = 14 \\
b_1 = 3 \quad b_2 = -38.
\end{aligned}$$

Better yet, note that

$$\begin{aligned}
-\frac{a_4}{b_4} &= \frac{24024}{65304} \\
&= 0.36787945\ldots \\
&\approx e^{-1}.
\end{aligned}$$

That is, $-a_n/b_n$ seems to be a good approximation to e^{-1}!

To see why this is, we must go back to the initial integral. For $0 \le x \le 1$, we have $x(1-x) \le 1/4$, therefore

$$0 \le I(n) = \int_0^1 e^{-x}(x(1-x))^n \, dx \le \frac{1}{4^n} \int_0^1 e^{-x} \, dx,$$

which shows that $I(n)$ goes to zero exponentially quickly. Therefore

$$|a_n + b_n e^{-1}| \to 0$$

exponentially quickly, meaning that $-a_n/b_n$ gives an exponentially-good rational approximation of e^{-1}. To double check, we can use our recurrence to compute a_{20}/b_{20}:

$$\begin{aligned}
-\frac{a_{20}}{b_{20}} &= \frac{4932941648663833516994295346011418332399206400000}{1340912564441170249019237618446466016434749440000} \\
&= 0.3678794411714423215955237701614608674\ldots \\
&\approx e^{-1}.
\end{aligned}$$

Better still, this remarkable approximation $-a_n/b_n \approx e^{-1}$ is *too good to be true* in the following sense.

Proposition 5. *Let α be a real number. If there exist sequences of integers a_n and b_n such that $|b_n| \to \infty$, the ratio a_n/b_n is not eventually constant, and*

$$|\alpha - \frac{a_n}{b_n}| \leq \frac{C}{|b_n|^{1+\delta}}$$

for some positive constants C and δ, then α is irrational.

Proof. If $\alpha = a/b$ is rational, then

$$|\alpha - \frac{a_n}{b_n}| = \frac{|(b_n a - b a_n)/b|}{|b_n|}.$$

The "numerator" $b_n a - b a_n$ is a nonzero integer infinitely often, and so

$$|\alpha - \frac{a_n}{b_n}| \geq \frac{D}{|b_n|}$$

for some positive constant D and infinitely many n. But the inequality

$$\frac{D}{|b_n|} \leq \frac{C}{|b_n|^{1+\delta}}$$

is impossible if $|b_n| \to \infty$. □

This fact together with our approximation $-a_n/b_n \approx e^{-1}$ gives us a proof that e is irrational.

Proposition 6. *e is irrational with $\delta = 1$.*

Proof. Let a_n and b_n be the approximating sequences induced by

$$I(n) = \int_0^1 e^{-x}(x(1-x))^n \, dx.$$

We have

$$|a_n + b_n e^{-1}| \leq \frac{1}{4^n} \int_0^1 e^{-x} = \frac{C}{4^n}.$$

The sequence b_n satisfies the recurrence

$$(N^2 + 2(2n+3)(n+2)N - (n+1)(n+2))b_n = 0.$$

It turns out—see [WZ85]—that this reveals considerable asymptotic information about b_n. In particular, if we rewrite the recurrence as a polynomial in n, the leading coefficient is $4N - 1$. The only solution to $4N - 1 = 0$

is $N = 1/4$, and this implies that $1/4^n \leq C' \frac{1}{|b^n|}$ for some constant C'. Thus

$$|a_n + b_n e^{-1}| \leq \frac{C'}{|b_n|},$$

or

$$|\frac{a_n}{b_n} + e^{-1}| \leq \frac{C'}{|b_n|^{1+\delta}},$$

where $\delta = 1$. The claim follows from the previous proposition. □

It is a cruel irony that almost every real is irrational, yet we are often helpless to prove that any naturally occurring constant such as $e\pi$, $e+\pi$, or γ (the Euler–Mascheroni constant) is irrational. The "constructive irrationality" method we have just used gives us a *possible framework* to approach irrationality: find an approximation, check that $\delta > 0$. Because irrationality results are so difficult, any hint or direction is worth investigating.

This constructive style of proof was made very famous by Roger Apéry [Apé79; Poo79]. In 1978, during an infamous talk at Marseille, Apéry proved that

$$\zeta(2) = \sum_{k \geq 1} \frac{1}{k^2} \quad \text{and} \quad \zeta(3) = \sum_{k \geq 1} \frac{1}{k^3}$$

are irrational. Euler knew that $\zeta(2) = \pi^2/6$, so the irrationality of $\zeta(2)$ was no big deal. The irrationality of $\zeta(3)$ was stunning.

Apéry's proof relied on writing down (seemingly at random) the recurrence

$$(n+2)^3 N^2 - (2n+3)(17n^2 + 51n + 39)N + (n+1)^3 = 0.$$

and choosing two solutions a_n and b_n with different initial conditions, just as we did for e^{-1}. After considerable checking, it turned out that a_n/b_n converged to $\zeta(3)$ sufficiently quickly to prove its irrationality with $\delta \approx 0.080529$.

The proof left the audience with many questions. Where did that recurrence come from? Why did those particular solutions work? Would Apéry's argument generalize to other constants, such as $\zeta(5)$?

Shortly after Apéry's proof, Frits Beukers [Beu79] cleared up some of the mystery when he elegantly *reproved* proved that $\zeta(2)$ and $\zeta(3)$ are irrational by considering integrals of the form

$$\int_0^1 \int_0^1 \frac{x^n y^n (1-x)^n (1-y)^n}{1-xy} \, dx \, dy$$

and

$$\int_0^1 \int_0^1 \int_0^1 \frac{x^n y^n z^n (1-x)^n (1-y)^n (1-z)^n}{(1-(1-xy)z)^{n+1}} \, dx \, dy \, dz,$$

respectively, which is similar to what we have done here.

Despite the efforts of many experts, it remains unclear how to generalize *either* Apéry's or Beukers' arguments to prove that any odd-zeta value other than $\zeta(3)$ is irrational. Simply put, we do not properly understand how $\zeta(3)$ is related to the ratio a_n/b_n of distinct solutions to a single recurrence, so we cannot construct similar recurrences for $\zeta(5)$ and beyond. The relevant term here is *Apéry limit* [CS21].

2.7 Irrationality Measures

Proving that e is irrational is an easy exercise, but our constructive proof gives more: a *quantitative measure* on the irrationality of e. For any real α, the *irrationality measure* of α, denoted $\mu(\alpha)$ is defined to be the smallest real μ such that

$$\left| \alpha - \frac{p}{q} \right| > \frac{1}{q^{\mu+\varepsilon}}$$

holds for any $\varepsilon > 0$ and for all integers p and q with q sufficiently large. If no such μ exists, then we set $\mu(\alpha) = \infty$.

Irrationality measures are intimately tied to constructive irrationality proofs. If we can construct a sequence of rationals a_n/b_n such that

$$|\alpha - \frac{a_n}{b_n}| \leq \frac{C}{b_n^{1+\delta}},$$

then $\mu(\alpha) \geq 1 + \delta$. In many cases—see [Poo79]—this also implies the upper bound $\mu(\alpha) \leq 1 + \frac{1}{\delta}$. Our proof happens to be one of these cases, and we get both $\mu(e) \geq 1 + 1 = 2$ and $\mu(e) \leq 1 + \frac{1}{1} = 2$, which implies $\mu(e) = 2$.

Though we have used integrals to construct our approximations, irrationality measures are also tightly linked with simple continued fractions. See Section 11.3 of the classic book *Pi and the AGM* by the great masters Jonathan and Peter Borwein for another proof of the irrationality measure of e, along with much more discussion about irrationality measures in general [BB87].

It is unusual that we know $\mu(e)$ exactly. Irrationality measures fall into three "regions":

α is ...	$\mu(\alpha)$
rational	1
algebraic with degree > 1	2
transcendental	≥ 2

So $\mu(e) \geq 2$ is automatic by its transcendence, but $\mu(e) = 2$ is a surprise. Normally the best we can do is give an upper bound on $\mu(\alpha)$ for specific transcendental α.

In fact, since it is so hard to establish irrationality, we have invented a *new game*: finding better and better upper bounds for the irrationality measure of famous constants. If you can find the best upper bound for $\mu(\pi)$, or $\mu(\zeta(3))$, then you get to hold the world record for a few months until someone beats you. For example, the current "world record" upper bound on $\mu(\pi)$ is held by Zeilberger and Zudilin [ZZ20], who showed that

$$\mu(\pi) \leq 7.103205334137\ldots.$$

Ignoring technical details, their proof is very similar to ours. The basic idea is to find a rapidly-decaying sequence of integrals $I(n)$ such that $I(n) = a_n + \pi b_n$ for integers a_n and b_n, then show that a_n and b_n have nice asymptotic properties.

To find their approximating sequences, Zeilberger and Zudilin tweaked integrals similar to the integrals Beukers used in his $\zeta(3)$ proof. They added parameters to the integrands and performed an exhaustive computer search to find those parameters which gave the empirically best upper bound, then went back and checked the details. This computer search method continues to provide possible avenues for constructive irrationality proofs; see [DKZ22] and [ZZ21].

2.8 Conclusions

It is too late for us to become famous proving that $\zeta(3)$ is irrational. We should be content just to have some new tools to play with. We should certainly show students how to use the AZ algorithm.

But you never know—one day you might just plug the right integrand into the Almkvist–Zeilberger algorithm to prove that

(FAMOUS CONSTANT)

is irrational. Until then, have fun!

Chapter 3

Hardinian arrays

The On-Line Encyclopedia of Integer sequences [OEI24] contains over 350,000 sequences and perhaps tens of thousands of conjectures about them. Here we resolve some of these conjectures related to a family of sequences due to R.H. Hardin.

For any positive integer r, let $H_r(n,k)$ be the number of $n \times k$ arrays which obey the following rules:

- The entry in position $(1,1)$ is 0, and the entry in position (n,k) is $\max(n,k) - r - 1$.
- The entry in position (i,j) must equal or be one more than each of the entries in positions $(i-1,j)$, $(i,j-1)$, and $(i-1,j-1)$.
- The entry in position (i,j) must be within r of $\max(i,j) - 1$.

We call these arrangements of numbers Hardinian arrays. For $r = 1,2,3$, they are counted by the tables A253026, A253223, and A253004, respectively. Below is an example for $r = 1$, $n = 6$, and $k = 5$.

$$\begin{bmatrix} 0 & 1 & 2 & 2 & 3 \\ 1 & 1 & 2 & 2 & 3 \\ 2 & 2 & 2 & 3 & 3 \\ 3 & 3 & 3 & 3 & 4 \\ 4 & 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 & 4 \end{bmatrix}$$

Hardin noticed several interesting patterns. For example, for every fixed r and k, the sequence $H_r(n,k)$ seems to be a polynomial in n of degree r for sufficiently large n. He also conjectured an evaluation of the diagonal for $r = 1$, namely

$$H_1(n,n) = \frac{1}{3}(4^{n-1} - 1).$$

More recently, Kauers and Koutschan [KK23] performed an automated search for sequences in the OEIS which satisfy linear recurrences with polynomial coefficients. Hardin happened to submit the diagonal of $r = 2$ as its own sequence, which led Kauers and Koutschan to conjecture a recurrence for $f(n) = H_2(n,n)$, namely

$$\begin{aligned}
&32(n+1)(2n+1)^2(1575n^6 + 21285n^5 + 117954n^4 + 343020n^3 + 551943n^2 + 465785n + 161046)f(n)\\
&- 8(121275n^9 + 1933470n^8 + 13267683n^7 + 51280818n^6 + 122556360n^5 + 186866686n^4\\
&\quad + 180574335n^3 + 105734340n^2 + 33718283n + 4443102)f(n+1)\\
&+ 2(294525n^9 + 4763070n^8 + 33170868n^7 + 130145646n^6 + 315713355n^5 + 488415476n^4\\
&\quad + 478464380n^3 + 283626704n^2 + 91378536n + 12137328)f(n+2)\\
&+ (294525n^9 + 4668570n^8 + 31877118n^7 + 122735586n^6 + 292620525n^5 + 445804136n^4\\
&\quad + 431097970n^3 + 252913504n^2 + 80866406n + 10688508)f(n+3)\\
&- (121275n^9 + 1961820n^8 + 13655808n^7 + 53503836n^6 + 129484209n^5 + 199650088n^4\\
&\quad + 194784258n^3 + 114948300n^2 + 36871922n + 4877748)f(n+4)\\
&+ 2(2n+7)(1575n^6 + 11835n^5 + 35154n^4 + 52554n^3 + 41382n^2 + 16118n + 2428)(n+3)^2 f(n+5) = 0.
\end{aligned}$$

Our main results are that many of these conjectures are correct. In Section 3.1 we will prove Hardin's conjectured closed form for $H_1(n,n)$ and extend this to a closed form for the rectangular case $H_1(n,k)$. In Section 3.2 we will prove that the conjectured recurrence of Kauers and Koutschan for $H_2(n,n)$ is correct, and in fact that *every* $H_r(n,n)$ satisfies such a recurrence. We will provide rigorous asymptotic estimates of $H_2(n,n)$ and conjecture asymptotic estimates for $H_r(n,n)$ when $r \geq 3$.

3.1 The case $r = 1$

This case can be settled by an elementary combinatorial argument. Let us first consider the diagonal and confirm the closed form representation conjectured by Hardin. In the following proof we index our arrays beginning from 0 rather than 1.

Theorem 1. *$H_1(n,n) = \frac{1}{3}(4^{n-1} - 1)$ for all $n \geq 1$.*

Proof. Consider a valid $n \times n$ array. Above the upper diagonal, draw a dividing path between row entries which are equal to their king-distance and less than their king-distance. Draw the same path below the diagonal, but make it with respect to columns. See Figure 3.1 for an example.

By the monotonicty rule, the upper path can only move down and to the right. Further, if the first entry to its right in row i is (i,j), then the first entry to its right in row $i+1$ is either $(i+1,j)$ or $(i+1,j+1)$. Thus the upper-path essentially consists of two kinds of steps: down and right-down. The situation is mirrored in the lower path.

If the upper path does not divide row i just after the row's entry on the main diagonal, then the row is determined from the diagonal to the right endpoint. Entries between the diagonal and the path equal their king-distance, entries after the path equal one less than their king-distance, and the diagonal must equal i as its king-distance is i and to its right is an $i+1$. The analogous statement is true for the lower path with respect to columns. Thus every entry is determined except for when both paths divide the ith row and column just after the diagonal. In fact, the *first* time this happens, the diagonal entry is still determined, as one of the entries above or to the left of the diagonal entry equals i.

In summary, the only entries not determined by these paths are the diagonal entries which both paths are adjacent to, except the first one and last one (by rule). If one path first touches the diagonal at position i, and the other at position $j>i$, then there are $n-j-2$ diagonal entries not determined. Of these entries, we may choose at most one to be the first less than its king-distance. After this choice all later entries must do the same. Thus each such pair of paths generates $n-j-1$ valid arrays.

If $C(k)$ is the number of paths which are first adjacent to the diagonal at position k, then

$$H_1(n,n) = 2\sum_{j=0}^{n-1}\sum_{i=0}^{j-1} C(i)C(j)(n-j-1) + \sum_{j=0}^{n-1} C(j)^2(n-j-1).$$

Because each path essentially has two steps to choose from, both of them moving one step closer to their end, we have $C(k)=2^{k-1}$ if $k>0$ and $C(0)=1$. Evaluating the above summations and simplifying produces $H_1(n,n) = (4^{n-1}-1)/3$. □

The double-path idea used in the proof above extends to the case of rectangular Hardinian arrays. The closed form expression for $H_1(n,k)$ shown next confirms conjectures stated by Hardin for $H_1(n,1)$, $H_1(n,2)$, $\ldots$, $H_1(n,7)$.

Theorem 2. $H_1(n,k) = 4^{k-1}(n-k)+\frac{1}{3}(4^{k-1}-1)$ *for all* $n \geq k \geq 1$.

Proof. Draw the same paths indicated in the proof of Theorem 1. See Figure 3.2 for an example.

A lower path now is either adjacent to the diagonal at some point or not. The number of valid arrays where the lower path is adjacent to the diagonal at some point is $H_1(k,k)$. All other pairs of paths contribute

$$\begin{bmatrix} 0 & 1 & 1 & 2 & 3 \\ 1 & 1 & 2 & 2 & 3 \\ 1 & 1 & 2 & 2 & 3 \\ 2 & 2 & 2 & * & 3 \\ 3 & 3 & 3 & 3 & 3 \end{bmatrix}$$

Figure 3.1: A generic 5×5 matrix with two specific paths as constructed in the combinatorial proof of Theorem 1. Every entry is determined by the paths except the one labeled $*$, which may be 3 or 2.

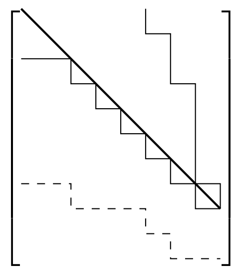


Figure 3.2: The generic picture for paths in the proof of Theorem 2. The lower two paths are examples of the two possible cases.

only one valid array. There are $n-k$ possible ending positions for a lower path which is never adjacent to the diagonal and 2^{k-1} paths originating from each. Thus this case contributes $(n-k)4^{k-1}$ valid arrays. Together this yields $H_1(n,k) = 4^{k-1}(n-k) + H_1(k,k)$. $\square$

As these combinatorial arguments do not seem to extend to $r > 1$, we give some alternative proofs of Theorem 1. They all rely on the theorem of Gessel and Viennot [Kra15, Theorem 10.13.1], which translates the counting problem into a determinant evaluation problem. We will evaluate the determinant in three different ways. The following notation will be used.

Definition 3. 1. For each positive integer n, let $M(n)$ be the $n \times n$ matrix of binomial coefficients

$$\left\{ \binom{u+v}{u} \right\}_{0 \le u,v < n}.$$

Observe that rows and columns are indexed starting from zero.

2. For any $n \times n$ matrix A, any distinct row indices $i_1, i_2, \ldots, i_r \in \{0, \ldots, n-1\}$ and distinct column indices $j_1, j_2, \ldots, j_r \in \{0, \ldots, n-1\}$, let $A^{j_1,j_2,\ldots,j_r}_{i_1,i_2,\ldots,i_r}$ be the $(n-r) \times (n-r)$ matrix obtained from A by deleting rows $i_1, \ldots, i_k$ and columns $j_1, \ldots, j_k$.

$$\begin{bmatrix} 0 & 1 & 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 2 & 2 & 3 & 4 & 5 \\ 2 & 2 & 2 & 2 & 3 & 4 & 5 \\ 2 & 2 & 3 & 3 & 3 & 4 & 5 \\ 3 & 3 & 3 & 3 & 3 & 4 & 5 \\ 4 & 4 & 4 & 4 & 4 & 4 & 5 \\ 5 & 5 & 5 & 5 & 5 & 5 & 5 \end{bmatrix}$$

Figure 3.3: The contiguous regions of a Hardinian array are separated by a tuple of nonintersecting lattice walks starting on the left and ending a the top.

3. For every $n \geq 1$, define

$$\Delta(n) = \det M(n)$$
$$\Delta(n)^{j_1,j_2,\ldots,j_r}_{i_1,i_2,\ldots,i_r} = \det M(n)^{j_1,j_2,\ldots,j_r}_{i_1,i_2,\ldots,i_r}.$$

Lemma 1. $\Delta(n) = 1$ *for all* n.

Proof. Observe that $M(n) = AB$ where A is the matrix whose entry at (u,v) is $\binom{u}{v}$ and B is the matrix whose entry at (u,v) is $\binom{v}{u}$. This follows from Vandermonde's identity $\binom{u+v}{v} = \sum_k \binom{u}{k}\binom{v}{k}$. As A and B are triangular matrices with 1's on the diagonal, the claim follows from $\Delta(n) = \det(M(n)) = \det(A)\det(B)$. □

The key observation is that the valid $n \times n$ arrays can be partitioned into contiguous regions, as shown in Figure 3.3. There is a region for 0, a region for 1, a region for 2, and so on. In the $n \times n$ case, the region corresponding to k is obtained by beginning at the lowest occurrence of k in the first column, moving as far right as possible while only passing k's, and moving up when stuck. For an $n \times n$ Hardinian array this process always terminates in the first row.

Proposition 7. $H_1(n,n) = \displaystyle\sum_{i=0}^{n-2}\sum_{j=0}^{n-2} \Delta(n-1)^j_i$ *for all* $n \geq 1$.

Proof. The $n-1$ contiguous regions in a Hardinian array of size $n \times n$ are separated by $n-2$ nonintersecting lattice paths. These paths begin on one of the $n-1$ edges between entries in the first column and end on one of the $n-1$ edges between entries in the first row, using only steps to the right ($\rightarrow$) and upwards ($\uparrow$). Each Hardinian array corresponds to exactly one such set of paths.

In the other direction, each such set of paths corresponds to a Hardinian array. Given such a set, assign the induced regions the values 0, 1, $\ldots n-2$ in order from the top-left to the bottom-right. The top left will contain a 0, the bottom right will contain an $n-2$, and adjacent entries differ by no more than 1. To see that the king-distance rule is not violated, note that it is not violated at the entries before the boundaries on the first column and first row—because at most one entry does not have a path just before it—and that these points have the largest king-distance of any entry reached using the available steps.

It follows that the number of Hardinian arrays of size $n \times n$ equals the number of sets of nonintersecting lattice paths we have described. If we label the possible starting and ending positions $0, 1, \ldots, n-2$, then there are altogether $\binom{u+v}{v}$ paths from u to v, for any u and v.

Consider the set of paths where i is the unique unchosen startpoint and j the unique unchosen endpoint. In this case the kth path ($k = 0, \ldots, n-3$) starts at $k + [i \le k]$ and ends at $k + [j \le k]$. By the theorem of Gessel and Viennot, the number of such sets of paths is the determinant of the $(n-2) \times (n-2)$ matrix whose entry at position (u, v) is $\binom{u+v+[i \le u]+[j \le v]}{v+[j \le v]}$. This determinant equals $\Delta(n-1)_i^j$. It follows that $H_1(n,n)$ is the sum of $\Delta(n-1)_i^j$ over all possible rows i and columns j. $\square$

The proposition reduces the enumeration problem to the problem of evaluating a sum of determinants. This can be done as follows.

Second proof of Theorem 1. Let $\tilde{M}(n)$ be the $(n+1) \times (n+1)$ matrix obtained from $M(n)$ by first attaching an additional row $1, -1, 1, -1, \ldots$ at the top and then an additional column $0, -1, 1, -1, 1, \ldots$ at the left, e.g.,

$$\tilde{M}(5) = \begin{vmatrix} 0 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 4 & 5 \\ -1 & 1 & 3 & 6 & 10 & 15 \\ 1 & 1 & 4 & 10 & 20 & 35 \\ -1 & 1 & 5 & 15 & 35 & 70 \end{vmatrix}.$$

By expanding along the first row and then along the first column, we have $\det \tilde{M}(n) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \Delta(n)_i^j$.

It remains to determine the determinant of $\tilde{M}(n)$.

Subtract the $(n-2)$nd row from the $(n-1)$st, then the $(n-3)$rd row from the $(n-2)$nd, and so on, and analogously for the columns, e.g.,

$$\begin{vmatrix} 0 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 4 & 5 \\ -1 & 1 & 3 & 6 & 10 & 15 \\ 1 & 1 & 4 & 10 & 20 & 35 \\ -1 & 1 & 5 & 15 & 35 & 70 \end{vmatrix} = \begin{vmatrix} 0 & 1 & -2 & 2 & -2 & 2 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 1 & 1 & 1 \\ -2 & 0 & 1 & 2 & 3 & 4 \\ 2 & 0 & 1 & 3 & 6 & 10 \\ -2 & 0 & 1 & 4 & 10 & 20 \end{vmatrix}$$

In general, the proposed row and column operations replace the entry $\binom{u}{v}$ by

$$\binom{u}{v} - \binom{u-1}{v} - (\binom{u}{v-1} - \binom{u-1}{v-1}) = \binom{u-1}{v-1}.$$

Now expand along the second row (or column) to obtain

$$\det \tilde{M}(n) = \Delta(n-1) + 4 \det \tilde{M}(n-1) = 4 \det \tilde{M}(n-1) + 1$$

for every n. Together with the initial value $\det \tilde{M}(1) = 1$, it follows by induction that $\det \tilde{M}(n) = \frac{1}{3}(4^n - 1)$. In view of Prop. 7, Theorem 1 follows by replacing n by $n-1$. $\square$

Third proof of Theorem 1. This proof uses computer algebra, in the spirit of an approach proposed by Zeilberger [Zei07]. Because of $\Delta(n) = 1$ and Cramer's rule, $(-1)^{i+j}\Delta(n)_i^j$ is the entry of $M(n)^{-1}$ at position (i, j). For $n \geq 1$ and $i, j = 0, \ldots, n-1$, define

$$c(n,i,j) = (-1)^{i+j} \sum_{\ell=0}^{n-1} \binom{i}{\ell}\binom{j}{\ell}.$$

Using symbolic summation algorithms (as implemented, e.g., in Koutschan's package [Kou10]), it can be easily shown that

$$\sum_{k=0}^{n-1} \binom{i+k}{k} c(n,k,j) = \delta_{i,j}$$

for all $n \geq 1$ and all $i, j \geq 0$. Therefore, $c(n,i,j)$ is the entry at (i, j) of $M(n)^{-1}$, and thus equal to $(-1)^{i+j}\Delta(n)_i^j$.

Applying summation algorithms once more, we can prove that the sum $s(n) = \sum_{i,j}(-1)^{i+j}c(n,i,j)$ satisfies the recurrence

$$s(n+2) = 5s(n+1) - 4s(n)$$

for all $n \geq 1$. Together with the initial values $s(1) = 1$ and $s(2) = 5$, the claimed closed form expression now follows again by induction. $\square$

While the sum $\Delta(n)_i^j = \sum_{\ell=0}^{n-1} \binom{i}{\ell}\binom{j}{\ell}$ does not have a hypergeometric closed form, it does simplify in the special case $j = n-1$, where it turns out to be equal to $\binom{n-1}{i}$. Taking the knowledge of this special case for granted, we can give a fourth proof of Theorem 1.

Fourth proof of Theorem 1. Dodgson's identity (cf. Prop. 10 of Krattenthaler's tutorial on evaluating determinants [Kra99]) says that

$$\det(A)\det(A_{i,n-1}^{j,n-1}) = \det(A_i^j)\det(A_{n-1}^{n-1}) - \det(A_i^{n-1})\det(A_{n-1}^j)$$

for every $n \times n$ matrix A. (Actually, Krattenthaler states the equation for $i = j = 0$, but it is easily seen that it holds for arbitrary i and j, because we can multiply A with suitable permutation matrices from the left and the right in order to reduce to the case $i = j = 0$.)

Consider $A = M(n)$ and observe that $A_{n-1}^{n-1} = M(n-1)$. Then, because of $\Delta(n) = \Delta(n-1) = 1$ it follows that

$$\Delta(n-1)_i^j = \Delta(n)_i^j - \Delta(n)_i^{n-1}\Delta(n)_{n-1}^j.$$

Using $\Delta(n)_i^{n-1} = \binom{n-1}{i}$ and $\Delta(n)_{n-1}^j = \binom{n-1}{j}$, it follows that

$$\Delta(n)_i^j = \Delta(n-1)_i^j + \binom{n-1}{i}\binom{n-1}{j}.$$

Summing over all i and j gives

$$s(n) = s(n-1) + 4^{n-1},$$

and with $s(1) = 1$, the claim follows again by induction. $\square$

3.2 The case $r \geq 2$

Via the theorem of Gessel and Viennot, we also have access to the sequences $H_r(n,n)$ for $r > 1$. The argument is the same as for $r = 1$, except that now a Hardinian array of size $n \times n$ consists of $n - r$ contiguous regions, separated by $n - r - 1$ nonintersecting lattice paths, whose start points and end points are taken from the set $\{0,\dots,n-2\}$. According to Gessel and Viennot, $\Delta(n-1)^{j_1,\dots,j_r}_{i_1,\dots,i_r}$ is the number of sets of $n-r-1$ nonintersecting lattice walks whose start points are $\{0,\dots,n-2\}\setminus\{i_1,\dots,i_r\}$ and whose end points are $\{0,\dots,n-2\}\setminus\{j_1,\dots,j_r\}$.

In order to deal with these determinants, it helps to observe that Dodgson's identity quoted in the fourth proof of Theorem 1 is a special case of a more general identity due to Jacobi [Jac33; RT07; Abe14]: For an $n\times n$ matrix A and two choices $0\le i_1<i_2<\dots<i_r<n$ and $0\le j_1<j_2<\dots<j_r<n$ of indices, form the $r\times r$ matrix B whose entry at (u,v) is defined as $\det(A^{j_v}_{i_u})$. Then Jacobi's identity says

$$\det(A)^{r-1}\det(A^{j_1,\dots,j_r}_{i_1,\dots,i_r})=\det(B).$$

For example, for $r=2$ we obtain

$$\det(A)\det(A^{j_1,j_2}_{i_1,i_2})=\begin{vmatrix}\det(A^{j_2}_{i_2}) & \det(A^{j_1}_{i_2})\\ \det(A^{j_2}_{i_1}) & \det(A^{j_1}_{i_1})\end{vmatrix}=\det(A^{j_1}_{i_1})\det(A^{j_2}_{i_2})-\det(A^{j_2}_{i_1})\det(A^{j_1}_{i_2}),$$

and setting $i_2=j_2=n-1$ gives Dodgson's version.

Theorem 3. *For every $r\ge 2$, the sequence $H_r(n,n)$ is D-finite. In particular, the sequences A253217 ($r=2$) and A252998 ($r=3$) are D-finite.*

Proof. For $A=M(n)$, Jacobi's identity implies

$$\Delta(n)^{j_1,\dots,j_r}_{i_1,\dots,i_r}=\begin{vmatrix}\Delta(n)^{j_1}_{i_1} & \cdots & \Delta(n)^{j_r}_{i_1}\\ \vdots & \ddots & \vdots\\ \Delta(n)^{j_1}_{i_r} & \cdots & \Delta(n)^{j_r}_{i_r}\end{vmatrix}$$

For every fixed r, the determinant on the right is D-finite because it depends polynomially on quantities which we have recognized in the previous section as being D-finite. It follows that the left hand side is D-finite, and consequently,

$$H_r(n,n)=\sum_{0\le i_1<\dots<i_r\le n-2}\ \sum_{0\le j_1<\dots<j_r\le n-2}\Delta(n-1)^{j_1,\dots,j_r}_{i_1,\dots,i_r}$$

is D-finite, too. □

Theorem 3 is not quite enough to confirm the correctness of the recurrence equation Kauers and Koutschan obtained for $H_2(n,n)$ via guessing [KK23]. The theorem only implies that the sequence satisfies *some* recurrence. In order to explicitly construct a recurrence, we have to evaluate the two 6-fold sums

$$\begin{aligned}
S_1(n) &= \sum_{i_1\geq 0}\sum_{i_2>i_1}\sum_{j_1\geq 0}\sum_{j_2>j_1}\sum_{u=0}^{n}\sum_{v=0}^{n}\binom{u}{i_1}\binom{u}{j_1}\binom{v}{i_2}\binom{v}{j_2}\\
&= \sum_{u=0}^{n}\sum_{v=0}^{n}\underbrace{\Bigg(\sum_{i_1\geq 0}\sum_{i_2>i_1}\binom{u}{i_1}\binom{v}{i_2}\Bigg)}_{=:s(u,v)}\underbrace{\Bigg(\sum_{j_1\geq 0}\sum_{j_2>j_1}\binom{u}{j_1}\binom{v}{j_2}\Bigg)}_{=s(u,v)} \text{ and}\\
S_2(n) &= \sum_{i_1\geq 0}\sum_{i_2>i_1}\sum_{j_1\geq 0}\sum_{j_2>j_1}\sum_{u=0}^{n}\sum_{v=0}^{n}\binom{u}{i_1}\binom{u}{j_2}\binom{v}{i_2}\binom{v}{j_1}\\
&= \sum_{u=0}^{n}\sum_{v=0}^{n}\underbrace{\Bigg(\sum_{i_1\geq 0}\sum_{i_2>i_1}\binom{u}{i_1}\binom{v}{i_2}\Bigg)}_{=s(u,v)}\underbrace{\Bigg(\sum_{j_1\geq 0}\sum_{j_2>j_1}\binom{v}{j_1}\binom{u}{j_2}\Bigg)}_{=s(v,u)}.
\end{aligned}$$

It seems best to do this using generating functions. We have

$$\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}s_1(u,v)x^u y^v = \frac{y}{(1-x-y)(1-2y)}.$$

The generating functions of $s(u,v)^2$ and $s(u,v)s(v,u)$ can be expressed as Hadamard products. As explained in [Bos+17], Hadamard products can be rephrased as residues, and residues can be computed via creative telescoping [Zei90a]. Using Koutschan's implementation [Kou10], it is easy to prove

$$\begin{aligned}
\frac{y}{(1-x-y)(1-2y)}\odot_{x,y}\frac{y}{(1-x-y)(1-2y)} &= \frac{y}{2x+2y-1}\left(\frac{1}{\sqrt{x^2-2x(y+1)+(y-1)^2}}+\frac{2}{4y-1}\right)\\
\frac{y}{(1-x-y)(1-2y)}\odot_{x,y}\frac{x}{(1-x-y)(1-2x)} &= \frac{1}{2(2x+2y-1)}\left(\frac{x+y-1}{\sqrt{x^2-2x(y+1)+(y-1)^2}}+1\right),
\end{aligned}$$

respectively. Summing u from 0 to n and v from 0 to m amounts to multiplying these series by $\frac{1}{(1-x)(1-y)}$, and setting m to n amounts to taking the diagonals of the resulting bivariate series:

$$\begin{aligned}
&\operatorname{diag}\frac{1}{(1-x)(1-y)}\frac{y}{2x+2y-1}\left(\frac{1}{\sqrt{x^2-2x(y+1)+(y-1)^2}}+\frac{2}{4y-1}\right),\\
&\operatorname{diag}\frac{1}{(1-x)(1-y)}\frac{1}{2(2x+2y-1)}\left(\frac{x+y-1}{\sqrt{x^2-2x(y+1)+(y-1)^2}}+1\right),
\end{aligned}$$

respectively. As diagonals can also be rephrased as residues (cf. again [Bos+17] for a detailed discussion), we can apply creative telescoping to obtain linear differential operators annihilating these series. Their least common left multiple is an annihilator of the generating function of $H_2(n,n)$.

In the end, we obtained a linear differential operator of order 10 with polynomial coefficients of degree 43. With this certified operator at hand, we can prove that the guessed recurrence of Kauers and Koutschan is correct.

In principle, we could derive a recurrence for $H_r(n,n)$ for any $r \geq 2$ in the same way, but already for $r = 3$ the computations become too costly. We can however use the formula

$$H_r(n,n) = \sum_{0 \leq i_1 < \cdots < i_r \leq n-2} \; \sum_{0 \leq j_1 < \cdots < j_r \leq n-2} \Delta(n-1)^{j_1,\ldots,j_r}_{i_1,\ldots,i_r}$$

to compute some more terms of the sequences. In order to do this efficiently, we can recycle the idea of the second proof of Theorem 1 and translate some of the summation signs into additional rows and columns of the determinant. For example, for $r = 3$ we have

$$H_r(n,n) = \sum_{i=0}^{n-2} \sum_{j=0}^{n-2} |\det(A_{i,j})|$$

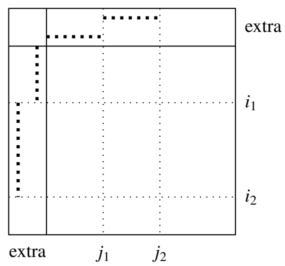
where $A_{i,j}$ is the matrix obtained from $M(n-1)$ by removing the ith row and the jth column and adding a row with alternating signs in the column range $0 \ldots j-1$ followed by zeros and an additional row with zeros in the column range $0 \ldots j-1$ followed by alternating signs; and similarly two additional columns. For example, for $n = 8, i = 4, j = 5$ we have

$$A_{i,j} = \begin{array}{cl} \left(\begin{array}{cc|cccc:cc} 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 & -1 & 1 & 0 & 0 \\ \hline 0 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 & 3 & 4 & 6 & 7 \\ 0 & -1 & 1 & 3 & 6 & 10 & 21 & 28 \\ \hdashline -1 & 0 & 1 & 5 & 15 & 35 & 126 & 210 \\ 1 & 0 & 1 & 6 & 21 & 56 & 252 & 462 \\ -1 & 0 & 1 & 7 & 28 & 84 & 462 & 924 \end{array}\right) & \begin{array}{l} \text{extra} \\ \text{rows} \\ \\ \\ i\text{th row} \\ \text{deleted} \\ \\ \\ \end{array} \\ \begin{array}{cc} \text{extra columns} & \qquad\qquad j\text{th column deleted} \end{array} & \end{array}$$

With this optimiziation, it is not difficult to compute the first 100 terms, and using these, the technique of [KK22] is able to guess a convincing recurrence equation of order 9 and degree 36. It is not reproduced here.

For $r = 4$, we explicitly delete two rows and columns and add two rows and columns with alternating signs, as shown in Figure 3.4 on the left. This allows us to reduce the original 8-fold sum to a 4-fold sum. A 4-fold sum is also sufficient for $r = 5$, where we can even eliminate six summations by adding extra rows and columns, as shown in Figure 3.4 on the right. By computing the sums over all these determinants, we were able to determine the first ≈ 65 terms of the sequences $H_4(n,n)$ and $H_5(n,n)$. Unfortunately, these terms were not sufficient to find a recurrence by guessing.

However, the terms are enough to obtain convincing conjectured expressions for their asymptotics. We



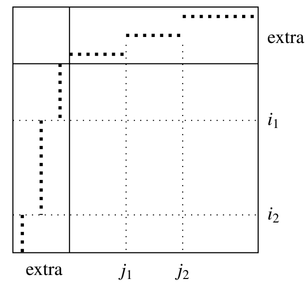


Figure 3.4: Left: the computation of $\sum_{i_1<i_2<i_3<i_4}\sum_{j_1<j_2<j_3<j_4}\Delta(n-1)^{j_1,j_2,j_3,j_4}_{i_1,i_2,i_3,i_4}$ is equivalent to the computation of the sum over i_1,i_2 and j_1,j_2 of the determinants constructed as shown in the figure. Light dots indicated omitted rows and columns; strong dots indicate regions filled with alternating signs.
Right: the computation of $\sum_{i_1<i_2<i_3<i_4<i_5}\sum_{j_1<j_2<j_3<j_4<j_5}\Delta(n-1)^{j_1,j_2,j_3,j_4,j_5}_{i_1,i_2,i_3,i_4,i_5}$ is equivalent to the computation of the sum over i_1,i_2 and j_1,j_2 of the determinants constructed as shown in the figure.

obtained the following conjectures:

r	asymptotics	remark
0	1	trivial
1	$\frac{1}{2^2 3}\,4^n$	by Theorem 1
2	$\frac{1}{2^2 3^4 \pi}\,16^n\,n^{-1}$	from the proven recurrence
3	$\frac{1}{2^2 3^9 \pi}\,64^n\,n^{-3}$	from the guessed recurrence
4	$\frac{2^2}{3^{16}\pi^2}\,256^n\,n^{-6}$	from the first 70 terms
5	$\frac{2^4}{3^{23}\pi^2}\,1024^n\,n^{-10}$	from the first 70 terms

Altogether, it seems that for every $r \geq 0$, we have

$$H_r(n,n) \sim c\,2^{2rn} n^{-\binom{r}{2}} \qquad (n \to \infty)$$

for some constant c that can be expressed as a power product of 2, 3, and π.

At least for specific values of r, it might be possible to prove these conjectured asymptotic formulas using the powerful techniques of analytic combinatorics in several variables [PW13; Mel21]. However, in order to invoke these techniques, we would need to know more about the bivariate sequences $H_r(n,k)$. Unfortunately, while we found an explicit expression for $H_1(n,k)$, we were not able to show that $H_r(n,k)$ is D-finite as a bivariate sequence in n and k for any $r \geq 2$, although we suspect it to be.

Chapter 4

Lots and Lots of Perrin-Type Primality Tests and Their Pseudo-Primes

The following article describes a set of primality tests based on the theory of linear recurrences, as well as some results about their pseudoprimes. The tests are a generalization of Fermat's primality test: A number n is *probably* prime if $a^n \equiv a \pmod{n}$ for many integers a.

A composite integer n is a *psuedoprime* of a probable primality test if it passes mistakenly. Fermat's test is particularly bad. For example, the smallest pseudoprime with $a = 2$ is $n = 341$:

$$2^{341} \equiv 2 \pmod{341},$$

yet

$$341 = 31 \cdot 11.$$

The obvious fix here is to try multiple a. For example, with $a = 3$ we have

$$3^{341} \equiv 168 \pmod{341},$$

so 341 is correctly discarded as composite as long as you try $a = 2$ and $a = 3$. Unfortunately this strategy is doomed to fail. The famous *Carmichael numbers* are composite integers n such that $a^n \equiv a \pmod{n}$ for *all* a. The smallest example is $n = 561$, but there are infinitely many of them and they are not particularly rare.

Fermat's primality test can be extended by using linear recurrences with constant coefficients. In particular, any monic linear recurrence with integer coefficients induces a sequence $a(n)$ and an integer e which satisfies $a(p) \equiv e \pmod{p}$ for all primes p. These are the tests the following article describes.

4.1 How it all started thanks to Vince Vatter

It all started when we came across Vince Vatter's delightful article [Vat22], where he gave a cute combinatorial proof of the following fact that goes back to Raoul Perrin [Per99] (See also A001608, A013998, [Ste96], and [Wik].)

Proposition 8. *Let $P(n)$ be the integer sequence defined by*

$$\begin{aligned}&P(1) = 1 \qquad P(2) = 2 \qquad P(3) = 3\\&P(n) = P(n-2) + P(n-3).\end{aligned}$$

Then p divides $P(p)$ for every prime p.

To contrast with Vatter, and to provide a hint of things to come, let us see an algebraic proof of this fact.

Proof. It is a simple linear algebra problem to show that $P(n) = x^n + y^n + z^n$ where x, y, and z are the distinct roots of $t^3 - t - 1$. Note that $x + y + z = 0$. For any prime p, we have

$$\begin{aligned}0 &= (x+y+z)^p\\&= x^p + y^p + z^p + \sum_{\substack{i+j+k=p\\ i,j,k<p}} \frac{p!}{i!j!k!} x^i y^j z^k\\&= P(p) + f(x,y,z).\end{aligned}$$

The term $f(x,y,z)$ is a symmetric, integer-coefficient polynomial in x, y, and z such that every coefficient is divisible by p. By the theory of symmetric polynomials it follows that $f(x,y,z)$ is an integer divisible by p, and so $P(p) \equiv 0 \pmod{p}$. $\square$

Vatter proved that $P(n)$ is the number of circular words of length n in the alphabet $\{0,1\}$ that avoid the subwords $\{000, 11\}$. We can divide these words into equivalence classes based on shifts. If $n = p$ is prime, then all the p circular shifts are distinct except for possibly the constant words, since otherwise there would be a non-trivial period. The constant words 0^p and 1^p can't avoid both 00 and 111, so the number of such words is p times the number of equivalence classes. This gives Perrin's theorem.

When we saw Vatter's proof we got excited. Vatter's argument transforms verbatim to counting circular words in any finite alphabet which avoid any finite set of patterns. More than twenty years ago one of us (DZ) wrote a paper, in collaboration with his then PhD student, Anne Edlin [EZ00], that computes the (rational) generating function in any such scenario, hence this is a cheap way to manufacture lots and lots of Perrin-style primality tests. This inspired us to write our first Maple package, `PerrinVV.txt`, available from `https://sites.math.rutgers.edu/~zeilberg/tokhniot/PerrinVV.txt`.

4.2 An even better way to manufacture Perrin-style Primality tests

After the initial excitement we had an epiphany: There is an easier way to generate primality tests! (It turns out it was already made, in 1990, by Stanley Gurak [Gur90].) Take any polynomial $Q(x)$ with integer coefficients and constant term 1, and write it as

$$Q(x) = 1 - e_1 x + e_2 x^2 - \cdots + (-1)^k e_k x^k$$

and factor it over the complex numbers as

$$Q(x) = (1-\alpha_1 x)(1-\alpha_2 x)\cdots(1-\alpha_k x).$$

Note that $e_1, e_2, \ldots$ are the elementary symmetric functions in $\alpha_1, \ldots, \alpha_k$. Thanks to Newton's identities, the sequence

$$a(n) = \alpha_1^n + \alpha_2^n + \cdots + \alpha_k^n \quad ,$$

is an integer sequence. We claim that each such integer sequence engenders a *Perrin-style* primality test, namely $a(p) \equiv e_1 \pmod{p}$. To see this, note that

$$(\alpha_1 + \cdots + \alpha_k)^p = a(p) + pA(p),$$

where

$$A(p) = \sum_{\substack{i_1+i_2+\cdots+i_k=p \\ i_1,i_2,\ldots i_k<p}} \frac{(p-1)!}{i_1!\cdots i_k!} \alpha_1^{i_1}\cdots\alpha_k^{i_k}$$

is a symmetric polynomial in the α_i with integer coefficients. The fundamental theorem of symmetric functions [Mac95] implies that $A(p)$ is an integer. Fermat's little theorem then gives

$$a(p) \equiv (\alpha_1 + \cdots + \alpha_k)^p = e_1^p \equiv e_1 \pmod{p}$$

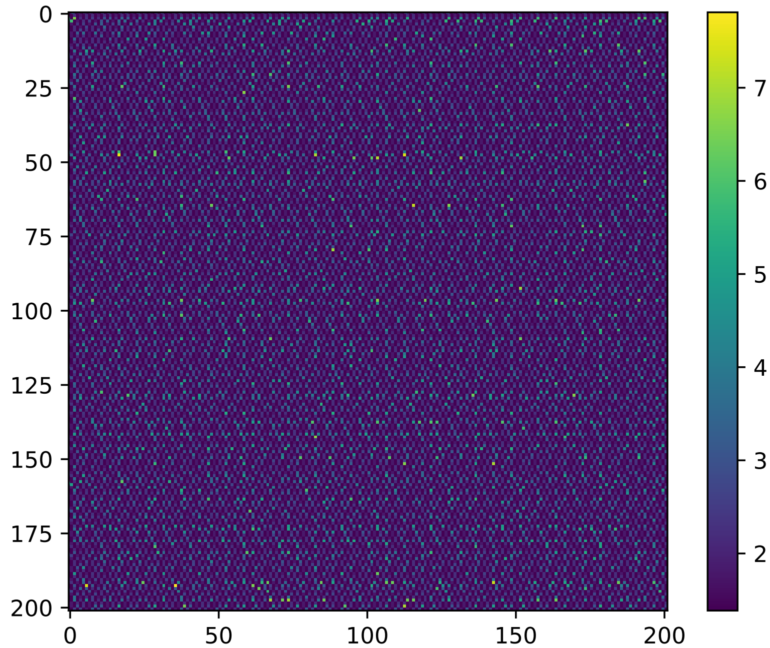


Figure 4.1: Log-heatmap of the first pseudoprime of $x^2 - ax - b$.

So this is an even easier way to manufacture lots and lots of Perrin-style primality tests, and we can let the computer search for those that have as few small pseudo-primes as possible.

This is implemented in the Maple package `Perrin.txt`, available from https://sites.math.rutgers.edu/~zeilberg/tokhniot/Perrin.txt. See the front of this article https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/perrin.html.

for many such primality tests, inspired by this more general method (first suggested by Stanley Gurak [Gur90]).

4.3 The DB-Z primality test

Let

$$\sum_{n=0}^{\infty} a(n)x^n := \frac{-3x^4 - 5x^2 - 6x + 7}{-4x^7 - x^4 - x^2 - x + 1} \quad ,$$

or equivalently, define the integer sequence $a(n)$ by by

$$a(1) = 1\,, a(2) = 3\,, a(3) = 4\,, a(4) = 11\,, a(5) = 16\,, a(6) = 30\,, a(7) = 78,$$

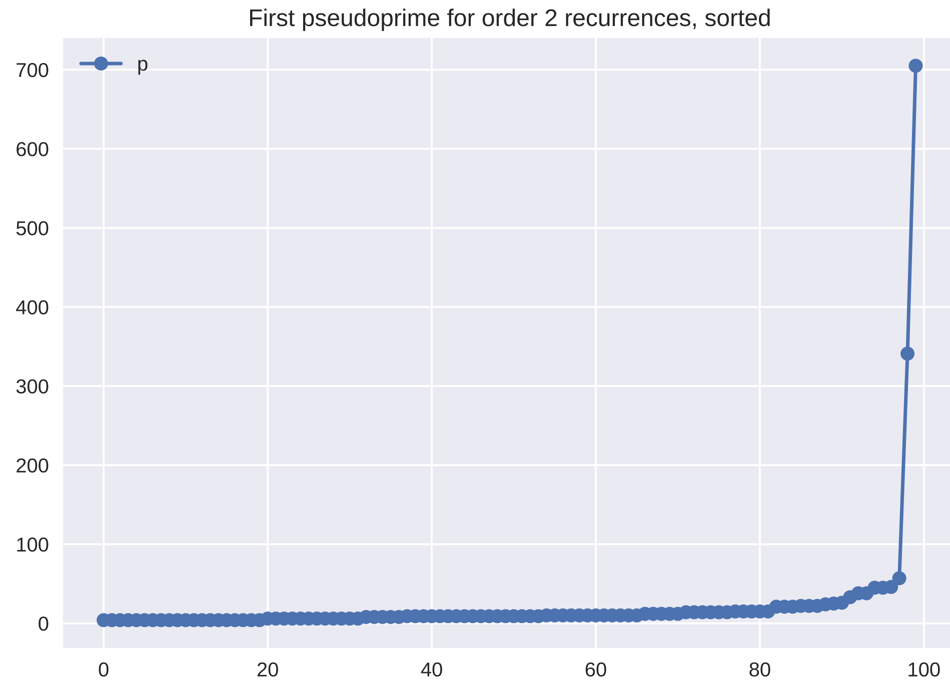


Figure 4.2: Sorted first pseudoprime of $x^2 - ax - b$

$$a(n) = a(n-1) + a(n-2) + a(n-4) + 4a(n-7)\ (for \quad n > 7).$$

Then $a(p) \equiv 1 \pmod{p}$ for all p. then if p is prime, we have $a(p) \equiv 1 \pmod{p}$.

Manuel Kauers kindly informed us that the seven smallest DB-Z pseudo-primes are as follows:

- $1531398 = 2 \cdot 3 \cdot 11 \cdot 23203$
- $114009582 = 2 \cdot 3 \cdot 17 \cdot 1117741$
- $940084647 = 3 \cdot 47 \cdot 643 \cdot 10369$
- $4206644978 = 2 \cdot 97 \cdot 859 \cdot 25243$
- $7962908038 = 2 \cdot 191 \cdot 709 \cdot 29401$
- $20293639091 = 11 \cdot 3547 \cdot 520123$
- $41947594698 = 2 \cdot 3 \cdot 19 \cdot 523 \cdot 703559$

At the time this was a difficult computational challenge. For meeting it, we donated 100 dollars to the OEIS in Kauers' honor.

4.4 The DB-Kauers primality test

After the first version of this paper was written, with the help of Manuel Kauers, we discovered an even better primality test.

The DB-Kauers primality test

Let

$$\sum_{n=0}^{\infty} a(n)\, x^n := \frac{-9x^5-16x^4-10x+6}{-3x^6-9x^5-8x^4-2x+1} \quad ,$$

or equivalently, let $a(n)$ be the integer sequence defined by

$$a(1)=2\,,\, a(2)=4\,,\, a(3)=8\,,\, a(4)=48\,,\, a(5)=157\,,\, a(6)=382\,,$$

$$a(n) \,=\, 2a(n-1) \,+\, 8\,a(n-4) \,+\, 9\,a(n-5) \,+\, a(n-6) \;(for \quad n>6).$$

Then $a(p) \equiv 2 \pmod{p}$ for all prime p.

The smallest pseudoprime happens to be $2,260,550,373 = 3 \cdot 103 \cdot 107 \cdot 68371$.

4.5 Perrin-Style Primality Tests with Explicit Infinite Families of Pseudo-Primes

We are particularly proud of the next primality test, featuring the Companion Pell numbers (see https://oeis.org/A002203). These numbers have been studied extensively, but as far as we know using them as a primality test is new. It is not a very good one, but the novelty is that it has an explicit, doubly-infinite set of pseudo-primes.

The Companion Pell Numbers Primality Test Let

$$\sum_{n=0}^{\infty} a(n)\, x^n := \frac{2-2x}{-x^2-2x+1} \quad ,$$

or equivalently,

$$a(1)=2\,,\, a(2)=6 \quad , \quad a(n) \,=\, a(n-1) \,+\, 2a(n-2) \quad (for \quad n>2) \quad ,$$

then if p is a prime, we have

$$a(p) \equiv 2\,(mod\; p) \quad .$$

Theorem 4. *Every element of*

$$\{2^i \cdot 3^j \mid i \geq 3, j \geq 0\}$$

is a Companion-Pell pseudoprime. In other words, $a(2^i \cdot 3^j) \equiv 2 \pmod{2^i 3^j}$ *for* $i \geq 3$ *and* $j \geq 0$.

Proof. With $\alpha_1 = 1 + \sqrt{2}$ and $\alpha_2 = 1 - \sqrt{2}$, we have the Binet formula $a(n) = \alpha_1^n + \alpha_2^n$. This and the product $\alpha_1 \alpha_2 = -1$ lead to the recurrences

$$a(2n) = a(n)^2 + 2(-1)^{n+1}$$
$$a(3n) = a(n)^3 + 3(-1)^{n+1} a(n).$$

Define the sequence $b(n) = a(n) - 2$. We will show by induction that $b(2^i 3^j)$ is divisible by $2^i 3^j$ if $i \geq 3$ and $j \geq 0$.

Suppose that n is even, i.e., that $i \geq 1$. Then the above identities yield

$$b(2n) = b(n)(b(n) + 4).$$

It is easy to prove that $b(n)$ is even for all n, so this identity shows that if an even n divides $b(n)$, then $2n$ divides $b(2n)$. Therefore if $2^i 3^j$ divides $b(2^i 3^j)$ with $i \geq 1$, then $2^{i+1} 3^j$ divides $b(2^{i+1} 3^j)$.

The second recurrence implies

$$b(3n) + 2 = (b(n) + 2)^3 + 3(-1)^{n+1}(b(n) + 2).$$

Again, if n is even, i.e. $i \geq 1$, then this gives

$$b(3n) = b(n)(b(n)^2 + 6b(n) + 12).$$

It is easy to show that $b(n)$ is divisible by 3 if $i \geq 1$ and $j \geq 1$. Therefore, if n divides $b(n)$ with $i \geq 1$ and $j \geq 1$, then $3n$ divides $b(n)$.

Considering the above remarks, we have shown that if $2^i 3^j$ divides $b(2^i 3^j)$, then $2^{i+1} 3^j$ divides $b(2^{i+1} 3^j)$ and $2^i 3^{j+1}$ divides $b(2^i 3^{j+1})$. Since 24 divides $b(24)$, the theorem follows by induction. $\square$

We now state without proofs (except for Theorem 4, where we give a sketch) a few other primality tests that have explicit infinite families of pseudoprimes.

Theorem 5. *Let*

$$\sum_{n=0}^{\infty} a(n)x^n := \frac{2-x}{-2x^2-x+1} \quad ,$$

or equivalently,

$$a(1)=1\,,a(2)=5 \quad , \quad a(n) \,=\, a(n-1) \,+\, 2a(n-2) \quad (for \quad n>2) \quad ,$$

then if p is a prime, we have

$$a(p) \equiv 1\,(mod\ p) \quad .$$

Furthermore, $\{2^i\,|i \geq 2\}$ *are all pseudo-primes, in other words*

$$a(2^i) \equiv 1\,(mod\ 2^i) \quad , \quad i \geq 2.$$

Theorem 6. *Let*

$$\sum_{n=0}^{\infty} a(n)x^n := \frac{2-2x}{-2x^2-2x+1} \quad ,$$

or equivalently,

$$a(1)=2\,,a(2)=8 \quad , \quad a(n) \,=\, 2a(n-1) \,+\, 2a(n-2) \quad (for \quad n>2) \quad ,$$

then if p is a prime, we have

$$a(p) \equiv 2\,(mod\ p) \quad .$$

Furthermore, the following infinite families are all pseudo-primes:

$\{3^i\,|i \geq 2\}$, $\{2 \cdot 3^i\,|i \geq 1\}$, $\{11 \cdot 81^i\,|i \geq 1\}$, $\{23 \cdot 3^{5i}\,|i \geq 1\}$, $\{29 \cdot 3^{4+12i}\,|i \geq 0\}$, $\{31 \cdot 3^{16i}\,|i \geq 1\}$.

Theorem 7. *Let*

$$\sum_{n=0}^{\infty} a(n)x^n := \frac{2x^2+3}{2x^3+2x^2+1} \quad ,$$

or equivalently,

$$a(1)=0\,,a(2)=-4\,,a(3)=-6 \quad , \quad a(n) \,=\, -2a(n-2) \,-\, 2a(n-3) \quad (for \quad n>2) \quad ,$$

then if p is a prime, we have

$$a(p) \equiv 0\,(mod\ p) \quad .$$

Furthermore, the following infinite families are all pseudo-primes:

$$\{2^i \,|\, i \geq 2\} \quad , \quad \{3 \cdot 2^{4i} \,|\, i \geq 2\} \quad , \quad \{11 \cdot 2^{18i} \,|\, i \geq 2\} \quad , \quad \{13 \cdot 2^{17+20i} \,|\, i \geq 2\} \, .$$

Sketch of proof We use the *C-finite ansatz* [Zei13]. Let

$$b(n) = a(2n) - a(n)^2 \quad ,$$

then it follows from the C-finite anzatz that $b(n)$ satisfies *some* recurrence, that turns out to be

$$b(1) = -4 \, , b(2) = -8 \, , b(3) = -40 \quad , \quad b(n) \, = \, 2b(n-1) + 4b(n-3) \;\; (for \quad n > 3)$$

We now define

$$c(n) := \frac{b(n)}{2^{\lfloor n/2 \rfloor}} \quad ,$$

and once again it follows that $c(n)$ satisfies the recurrence,

$$c(1) = -4 \, , c(2) = -4 \, , c(3) = -20, \, c(4) = -24, \, c(5) = -56, \, c(6) = -76$$

$$c(n) \, = \, 2c(n-2) + 4c(n-4) + 2c(n-6) \;\; (for \quad n > 6) \quad .$$

Note that $c(n)$ are manifestly **integers**. Going back to $a(n)$ we have the recurrence

$$a(2n) = a(n)^2 + 2^{\lfloor n/2 \rfloor} c(n) \quad ,$$

and it follows by induction that $a(2^i)/2^i$ are all integers. A similar argument goes for the other infinite families claimed.

Theorem 8. *Let*

$$\sum_{n=0}^{\infty} a(n) x^n \, := \, \frac{-2x^2 - 2x + 3}{-x^3 - 2x^2 - x + 1} \quad ,$$

or equivalently,

$$a(1) = 1 \, , a(2) = 5 \, , a(3) = 10 \quad , \quad a(n) \, = \, a(n-1) \, + \, 2a(n-2) + a(n-3) \quad (for \quad n > 2) \quad ,$$

then if p is a prime, we have

$$a(p) \equiv 1 \, (mod \; p) \quad .$$

Furthermore, the following infinite families are all pseudo-primes:

$$\{3^i \,|\, i \geq 2\} \quad , \quad \{5 \cdot 3^{6+10i} \,|\, i \geq 0\} \quad , \quad \{5 \cdot 3^{8+10i} \,|\, i \geq 0\} \quad , \quad \{7 \cdot 3^{4+6i} \,|\, i \geq 0\} \, ,$$

We found 9 other such primality tests, with infinite explicit families of presodoprimes, that can be viewed by typing

`PDB(x);` ,

in the Maple package `Perrin.txt`.

For fast computations and explorations using C programs, readers are welcome to explore the GitHub repository `https://github.com/rwbogl/pt`.

Acknowledgment: Many thanks to Manuel Kauers for his computational prowess, and to the referee for a helpful remark.

Chapter 5

Creating decidable diophantine equations

5.1 Preface

In 1970, 23-year-old Yuri Matiyasevich, standing on the shoulders of Julia Robinson, Martin Davis, and Hilary Putnam, shocked the world of mathematics by showing that David Hilbert's dream of finding an algorithm that inputs *any* polynomial

$$P(x_1,\ldots,x_n)$$

with integer coefficients, and outputs **true** or **false** if $P=0$ has, or does not have, solutions in integers, can never come to be.

Of course, for *specfic* equations, and even, many specific infinite families, one can often decide, but there is no *magic bullet* that, can decide all of them.

For example, Pythagoras got very upset when Hippasus of Metapontum discovered that the diophantine equation

$$x^2-2y^2=0 \quad ,$$

has **no solution**, and Hippasus (probably) gave a fully rigorous proof (**not** the usual one but a geometrical version of the **reduction formula** if (x,y) $x>y>0$ is a solution so is $(y,2x-y$ and since $(1,0)$ is not a solution **qed**. Going backwards, using the fact that if (x,y) is a solution of *Pell's equation*

$$x^2-2y^2\;=\;\pm 1$$

then so is $(x+2y,x+y)$, we can prove that there are **infinitely many** solutions. A little more effort will show that these are the only ones (see [Zei14]).

Much harder is the fact proved by Sir Andrew Wiles [Wil95], that for every $n > 2$ the diophantine equation

$$x^n + y^n = z^n \quad ,$$

has no solution.

On the positive side, Noam Elkies [Elk88] famously proved that

$$A^4 + B^4 + C^4 = D^4 \quad ,$$

has *infinitely* many solutions.

Wouldn't it be nice to be able to manufacture, *at will*, many examples of diophantine equations for which we can explicitly construct *all solutions*?

There is a cheap way to do this. As most of us know, the triple

$$a = A^2 - B^2 \quad , \quad , b = 2AB \quad , \quad c = A^2 + B^2 \quad ,$$

satisfies

$$a^2 + b^2 = c^2$$

A little more challenging is to prove that all solutions of $a^2 + b^2 = c^2$ with $\gcd(a,b) = 1$ are of this form. But this is true, and the same idea applies more generally.

Take any $m+1$ polynomials in m variables with integer coefficients

$$P_i(a_1, \ldots, a_m) \quad , \quad 1 \leq i \leq m+1$$

and define

$$X_i = P_i(a_1, \ldots, a_m) \quad , \quad 1 \leq i \leq m+1 \quad ,$$

using, e.g **Gröbner bases** (the Buchberger algorithm) we can *eliminate* $a_1, \ldots, a_m$ and get a polynomial equation (with integer coefficients)

$$Q(X_1, \ldots, X_m) \,=\, 0 \quad ,$$

that has a parametric solution as above. Alas, in general this leads to monster equations and unlike the case with Pythagorean triples, it is not clear that there aren't other solutions.

So it would be nice to be able to generate, in a systematic way, many examples of simple diophantine equations for which we know infinitely many solutions, and to be able to prove that these are all of them. It would also be nice to manufacture not too complicated, but non-trivial, diophantine equations for which we can conclusively prove that there aren't any solutions.

5.2 Diophantine equations from recurrences

In Matiyasevich's proof (we use the versions in [JM91; Mat93]) a central role is played by *Pell's equation* and the fact ([Mat93], pp. 19-20) that two consecutive terms $x = a_b(n), y = a_b(n+1)$ of the sequence of integers defined by the **second-order** linear recurrence

$$a_b(0) = 0 \quad , \quad a_b(1) = 1 \quad , \quad a_b(n+2) = b a_b(n+1) - a_b(n) \quad ,$$

satisfy the diophantine equation

$$x^2 - bxy + y^2 = 1 \quad ,$$

and conversely if $x > y$ satisfies it, then there must be an n such that $x = a_b(n+1)$, $y = a_b(n)$. This gave us the idea to consider higher-order recurrences.

If F_n is the nth Fibonacci number, then taking the determinant of the well-known matrix identity

$$\begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n ,$$

yields Cassini's identity $F_{n+1}F_{n-1} - F_n^2 = (-1)^n$. If we square this and apply the Fibonacci recurrence, then we obtain $P(F_{n-1}, F_n) = 1$ for some polynomial P. From another perspective, starting with the Fibonacci numbers we created a polynomial diophantine equation $P(x,y) = 1$ with infinitely many solutions. Our goal is to repeat this for a wider class of recurrences.

Consider the linear recurrence

$$a(n) = c_1 a(n-1) + \cdots + c_d a(n-d). \tag{5.1}$$

Our recipe has two parts. First, the matrix

$$B = \begin{bmatrix} c_1 & c_2 & \cdots & c_d \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

satisfies the "forward identity"

$$\begin{bmatrix} a(n+1) \\ a(n) \\ \vdots \\ a(n-d+2) \end{bmatrix} = B \begin{bmatrix} a(n) \\ a(n-1) \\ \vdots \\ a(n-d+1) \end{bmatrix}$$

for any sequence which satisfies (5.1). Second, (5.1) has d "fundamental solutions" which form a basis for all solutions. They are the solutions whose initial conditions are all zero except for a single entry, which is instead one. By luck, the columns of the identity matrix are exactly the initial conditions of these fundamental solutions. If we call the fundamental solutions $e_0(n), e_1(n), \ldots, e_{d-1}(n)$, then

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & & \cdots & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} e_0(d-1) & e_1(d-1) & \cdots & e_{d-1}(d-1) \\ e_0(d-2) & e_1(d-2) & \cdots & e_{d-1}(d-2) \\ e_0(d-3) & e_1(d-3) & \cdots & e_{d-1}(d-3) \\ & \cdots & & \\ e_0(0) & e_1(0) & \cdots & e_{d-1}(0) \end{bmatrix}.$$

This implies a relation between B^n and the fundamental solutions:

$$B^n = B^n I = \begin{bmatrix} e_0(n+d-1) & e_1(n+d-1) & \cdots & e_{d-1}(n+d-1) \\ e_0(n+d-2) & e_1(n+d-2) & \cdots & e_{d-1}(n+d-2) \\ e_0(n+d-3) & e_1(n+d-3) & \cdots & e_{d-1}(n+d-3) \\ & \cdots & & \\ e_0(n) & e_1(n) & \cdots & e_{d-1}(n) \end{bmatrix}. \tag{5.2}$$

From the elementary theory of difference equations, every solution to (5.1)—including the fundamental ones—can be expressed as a linear combination of the sequences $e_0(n)$, $e_0(n+1)$, ..., $e_0(n+d-1)$. Therefore every entry in the right-hand side of (5.2) is actually a linear combination of shifts of $e_0(n)$. By taking determinants in (5.2) it follows that

$$P(e_0(n), e_0(n+1), \ldots, e_0(n+d-1)) = (\det B)^n$$

for some polynomial P. Laplace expansion implies $\det B = (-1)^d c_d$, so setting $c_d = (-1)^d$ makes the right-hand side 1.

The previous considerations lead to the following proposition.

Proposition 9. *For any integers $c_1, c_2, \ldots, c_{d-1}$, there is a nonzero polynomial $P(x_1, x_2, \ldots, x_d)$ such that the diophantine equation*

$$P(x_1, x_2, \ldots, x_d) = 1$$

has infinitely many solutions. In particular, the points $(a(n), a(n+1), \ldots, a(n+d-1))$ are solutions, where $a(n)$ satisfies

$$a(n) = \sum_{k=1}^{d-1} c_k a(n-k) + (-1)^d a(n-d)$$

and has initial conditions $0, 0, \ldots, 0, 1$.

Our goal is to show that the diophantine equations in Proposition 9 are sometimes solved by *only* the recurrence solutions. This goal is too lofty in general, but we have arguments which apply to an infinite family of recurrences, and one detailed case study concerning the Tribonacci numbers.

5.3 Tribonacci numbers

We begin with the Tribonacci numbers as a detailed example. The main idea is to show that all solutions to the associated diophantine equation are generated, in some sense, by increasing solutions, and then to construct all increasing solutions.

Definition 4. Define the numbers T_n by

$$T_0 = T_1 = 0$$

$$T_2 = 1$$

$$T_n = T_{n-1} + T_{n-2},$$

the polynomial P_T by

$$P_T(x, y, z) = x^3 + 2x^2y + x^2z + 2xy^2 - 2xyz - xz^2 + 2y^3 - 2yz^2 + z^3,$$

and the map R_T by

$$R_T(x,y,z) = (y,z,x+y+z).$$

Note that P_T is invariant under R_T, i.e., $P_T \circ R_T = P_T$.

Proposition 10. *If $P_T(x,y,z) = 1$ for integers (x,y,z), then (x,y,z) is the result of repeatedly applying R_T or its inverse to a nonnegative increasing solution. That is, there exist integers $0 \le a \le b \le c$ and a positive integer n such that $P(a,b,c) = 1$ and (x,y,z) is $R_T^n(a,b,c)$ or $R_T^{-n}(a,b,c)$.*

Proof. Repeatedly applying R_T to our initial point (x,y,z) produces a sequence $a(n)$ which satisfies

$$a(n) = a(n-1) + a(n-2) + a(n-3)$$

with initial conditions $(a(0),a(1),a(2)) = (x,y,z)$. Because P_T is invariant under R_T we have $P_T(a(n),a(n+1),a(n+2)) = 1$ for all n. The elementary theory of difference equations implies $a(n) \sim c \cdot \alpha^n$ where $\alpha = 1.8393$ is the unique real root of $X^3 - X^2 - X - 1$ and

$$c = \alpha \frac{(\alpha^2-\alpha-1)a(0) + (\alpha-1)a(1) + a(2)}{\alpha^2+2\alpha+3}.$$

Note that c is real. If $c < 0$, then we eventually obtain a strictly negative solution, which is impossible because $P_T(x,y,z) \le 0$ if $x,y,z \le 0$. If $c = 0$ then $(\alpha^2-\alpha-1)a(0) + (\alpha-1)a(1) + a(2) = 0$, and this is impossible because $\{1,\alpha,\alpha^2\}$ is linearly independent over the rationals. The remaining possibility is $c > 0$, which implies that we eventually have $0 < a(n) < a(n+1) < a(n+2)$, and we get back to (x,y,z) by applying the inverse map R_T^{-1}. ∎

Proposition 11. *If $P_T(x,y,z) = 1$ for integers $0 \le x \le y \le z$, then $(x,y,z) = (T_n, T_{n+1}, T_{n+2})$ for some integer $n \ge 0$.*

Proof. The map $R_T^{-1}(x,y,z) = (z-x-y,x,y)$ takes solutions to other solutions. Note that if $0 \le z-x-y \le x$, then the new solution is also nonegative and increasing, and in fact strictly smaller unless $x = y = 0$. (If $x = y = 0$ then $z = 1$ is the unique solution.) We will show that $0 \le z-x-y \le x$ for all increasing solutions with sufficiently large z.

If we divide both sides of the equation $P_T(x,y,z) = 1$ by z^3, and make the change of variables $(t,s) = (x/z, y/z)$, then we obtain

$$2s^3 + 2s^2t + 2st^2 + t^3 - 2st + t^2 - 2s - t + 1 = \frac{1}{z^3}. \tag{5.3}$$

Call the left-hand side of this equation $f(s,t)$ and note that it is a cubic defined on the unit square. It is a routine calculus exercise to show that the minimum of $f(s,t)$ on the region $1-t-s<0$ is

$$\frac{398-68\sqrt{34}}{27}.$$

Therefore we cannot have both (5.3) and $1-t-s<0$ for

$$z > \left(\frac{398-68\sqrt{34}}{27}\right)^{-1/3} = 2.6235.$$

It follows that $0 \leq z-x-y$ for all increasing solutions to $P_T(x,y,z)=1$ with $z \geq 3$. By an analogous argument on the region $1-t-s>t$, all increasing solutions to $P_T(x,y,z)=1$ with $z \geq 5$ satisfy $z-x-y \leq x$.

Repeatedly applying the "backwards" map R_T^{-1} produces smaller, nonnegative, increasing solutions as long as $z \geq 5$, and so this process terminates at a solution with $0 \leq x \leq y \leq z < 5$. It is simple to check that all such solutions return to the point $(0,0,1)$ under the map R_T^{-1}, and so *all* increasing nonnegative solutions come from applying the "forward" map R_T to $(0,0,1)$. This produces exactly the Tribonacci numbers. □

See Figure 5.1 for a visual representation of the maps and regions in Proposition 11.

Theorem 9. *If $P_T(x,y,z)=1$ for integers x,y,z, then $(x,y,z)=(T_n,T_{n+1},T_{n+2})$ for some integer n.*

Proof. By the previous two propositions, every solution comes from applying the maps $(x,y,z) \mapsto (y,z,x+y+z)$ and $(x,y,z) \mapsto (z-x-y,x,y)$ to the solution $(0,0,1)$, which produces exactly the Tribonacci numbers with positive and negative indices. □

5.4 Uniqueness in general

The arguments from the previous section carry over almost verbatim to the general third-order recurrence. The main difficulty is in establishing the minimum of the analogous cubic (5.3). For any specific recurrence it is completely routine to check whether the proof of Proposition 11 works, but Proposition 13 gives a weaker statement about an infinite family.

Definition 5. For any positive integers a and b, define the polynomial $P_{ab}(x,y,z)$ as

$$a^2y^2z+abxyz+aby^3+b^2xy^2+ax^2z+axy^2-2ayz^2+2bx^2y-bxz^2-by^2z+x^3-3xyz+y^3+z^3$$

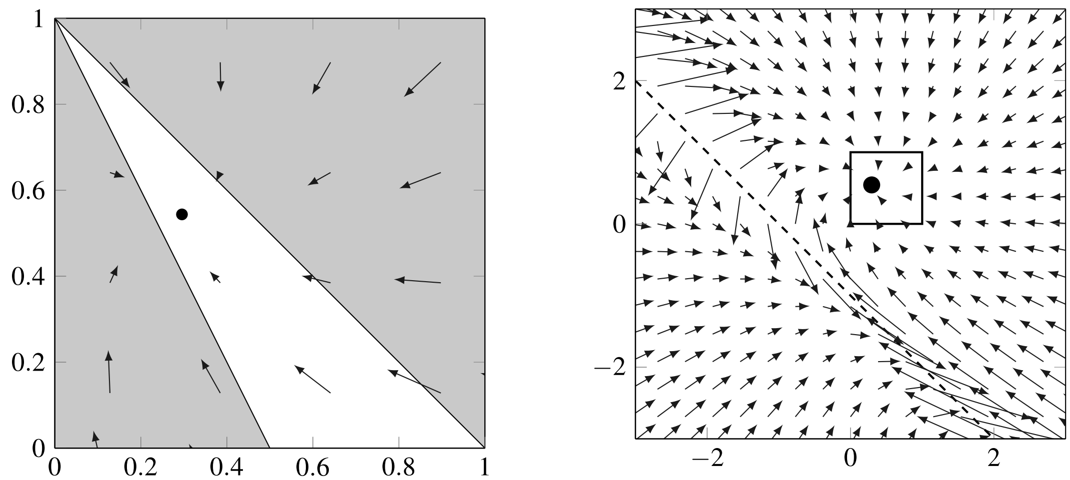


Figure 5.1: The map $(x,y,z) \mapsto (y,z,x+y+z)$ represented in the ts plane by its equivalent $(t,s) \mapsto (s/(s+t+1), 1/(s+t+1))$. Left: The map restricted to the unit square, with the region $\{s+t>1\} \cup \{s+2t<1\}$ shaded. The unique critical point in the first quadrant of the left-hand side of (5.3) is labeled by a black dot. Right: The map on a larger portion of the plane. The critical point is an attractor for $s+t+1>0$.

Proposition 12. *Let a and b be positive integers such that $X^3 - aX^2 - bX - 1$ is irreducible over $\mathbf{Q}$ and has a single largest root which is real and greater than 1. Then all integer solutions to $P_{ab}(x,y,z) = 1$ are generated by applying the map $(x,y,z) \mapsto (z-ay-bx,x,y)$ or its inverse to a nonnegative, increasing solution.*

Proof. The argument is the same as in Proposition 10. The irreducibility of $X^3 - aX^2 - bX - 1$, with largest root $\alpha > 1$, implies the linear independence of $\{1, \alpha, \alpha^2\}$ over the rationals and gives the correct asymptotics. □

Proposition 13. *Fix positive integers a and b and consider the recurrence*

$$u(n) = au(n-1) + bu(n-2) + u(n-3). \tag{5.4}$$

If a is sufficiently large relative to b, then all solutions $0 \le x \le y \le z$ to the diophantine equation $P_{ab}(x,y,z) = 1$ are generated by applying (5.4) to finitely many solutions.

It should be noted that while the following proof is non-constructive, the *method* is not. Carrying out the proof for any *specific* integers a and b will determine an exact bound under which the finitely many initial conditions can be found.

Proof. The polynomial $P_{ab}(x,y,z)$ is invariant under the map

$$(x,y,z) \mapsto (z-ay-bx,x,y), \tag{5.5}$$

so it takes solutions to solutions. In particular, the new solution is strictly closer to the origin unless $x = y = 0$ (which yields the unique solution $z = 1$). We will show that the new solution is also nonnegative and increasing for sufficiently large z.

If we divide both sides of $P_{ab}(x,y,z) = 1$ by z^3 and make the change of variables $(t,s) = (x/z, y/z)$, then we obtain $f_{ab}(t,s) = z^{-3}$ where f_{ab} is a cubic in t and s on the unit square. Because f_{ab} is a cubic, it is possible to exactly compute its critical points on the unit square, as well as the critical points of boundary functions such as $f_{ab}(0,s)$ and $f_{ab}(1,s)$. If we treat b as a constant and perform asymptotic expansions as $a \to \infty$ of these critical points, it turns out that the minimum of f_{ab} on the region $\{1 - as - bt < 0\} \cup \{1 - as - bt > t\}$ occurs on the line $1 - as - bt = 0$, and it equals

$$\frac{1}{a^6} - \frac{b^2}{4a^7} - \frac{9b}{2a^8} + O(a^{-9}).$$

So $f_{ab}(t,s) = z^{-3}$ fails if

$$z > a^2 + \frac{b^2}{12}a + \frac{3b}{2} + \frac{b^4}{72} + O(a^{-1}). \tag{5.6}$$

It follows that $0 < 1 - as - bt < t$, also known as $0 < z - ay - bx < x$ for any solution $0 \le x \le y \le z$ with sufficiently large z. We may therefore iterate (5.5) on such a solution until we reach one where z is below the bound implied by (5.6), and there are only finitely many of these. □

Theorem 10. *Let a and b be positive integers such that*

1. *$X^3 - aX^2 - bX - 1$ is irreducible over the rationals and has a single largest root which is real and greater than 1; and*

2. *a is sufficiently large relative to b (in the non-constructive sense of proposition 12).*

Then all integer solutions to $P_{ab}(x,y,z) = 1$ are generated by applying (5.4) forwards or backwards to finitely many initial solutions.

Note that the first condition is not very restrictive. The cubic $X^3 - aX^2 - bX - 1$ has a rational root only if $b = a + 2$.

5.5 Examples

A single family The characteristic equation

$$X^3 - 10X^2 - 3X - 1$$

leads to the diophantine equation

$$x^3+6x^2y+10x^2z+19xy^2+27xyz-3xz^2+31y^3+97y^2z-20yz^2+z^3=1.$$

Theorem 10 (along with explicit arguments from Proposition 13) shows that all solutions to this equation are generated by applying the maps $(x,y,z)\mapsto(y,z,10z+3y+x)$ and $(x,y,z)\mapsto(z-10y-3x,x,y)$ to the initial solution $(0,0,1)$.

Multiple families The characteristic equation

$$X^3-2X^2-3X-1$$

leads to the diophantine equation

$$x^3+6x^2y+2x^2z+11xy^2+3xyz-3xz^2+7y^3+y^2z-4yz^2+z^3=1.$$

Theorem 10 (along with explicit arguments from Proposition 13) shows that all solutions to this equation are generated by applying the maps $(x,y,z)\mapsto(y,z,3z+2y+x)$ and $(x,y,z)\mapsto(z-3y-2x,x,y)$ to the initial solutions

$$(0,0,1),(0,1,3),(0,2,7),(1,1,4).$$

A failure The characteristic equation

$$X^3-X^2-3X-1=(X-1)(X^2-2X-1)$$

corresponds to setting $a=1$ and $b=3$, which leads to the diophantine equation

$$x^3+6x^2y+x^2z+10xy^2-3xz^2+4y^3-2y^2z-2yz^2+z^3=1.$$

Our method fails here on two counts. First, the proof of Proposition 13 does not go through ($a=1$ is not big enough relative to $b=3$). Second, this recurrence has degenerate integer solutions like $(-1)^n$ which do not have the correct asymptotics.

Chapter 6

The Meta-C-Finite Ansatz

The Fibonacci numbers F_n satisfy the famous recurrence $F_n = F_{n-1} + F_{n-2}$. The sequence which takes every *other* Fibonacci number, F_{2n}, satisfies the similar recurrence $F_{2n} = 3F_{2(n-2)} - F_{2(n-2)}$. In fact, every sequence of the form F_{mn} satisfies such a recurrence. Here are the first few:

$$\begin{aligned}
F_n &= F_{n-1} + F_{n-2} \\
F_{2n} &= 3F_{2(n-1)} - F_{2(n-2)} \\
F_{3n} &= 4F_{3(n-1)} + F_{3(n-2)} \\
F_{4n} &= 7F_{4(n-1)} - F_{4(n-2)} \\
F_{5n} &= 11F_{5(n-1)} + F_{5(n-2)}.
\end{aligned} \tag{6.1}$$

If we look closely at the coefficients that appear—or plug them into the OEIS [OEI24]—there seems to be a general recurrence:

$$F_{mn} = L_m F_{m(n-1)} + (-1)^{m+1} F_{m(n-2)}. \tag{6.2}$$

This conjecture is right on the money, and we can prove it a dozen different ways—Binet's formula, induction, generatingfunctionology—but the *outline* is more interesting.

We began with a sequence which satisfied a nice recurrence (F_n), examined recurrences for a family of related sequences (F_{mn}), then noticed that the coefficients on the recurrences satisfied a *meta pattern* (equation (6.2)). This outline holds for any sequence which satisfies a linear recurrence relation with constant coefficients. Such sequences are called *C-finite* or *constant recursive* [Zei13; KP10]. Our goal is to prove that this outline holds for C-finite sequences and give some example applications.

The remainder of the paper is organized as follows. Section 6.1 gives a brief overview of C-finite sequences, Section 6.2 proves that an analogue of (6.2) holds for any C-finite sequence, Section 6.3 applies this to produce infinite families of summation identities, and Section 6.4 shows that a similar outline holds for products of C-finite sequences.

6.1 The C-finite ansatz

The theory of C-finite sequences is beautifully laid out in [KP10] and [Zei13]. What follows is a brief description of the principle results. For simplicity, assume that everything we do is over an algebraically closed field such as the complex numbers.

Given a sequence $a(n)$, let N be the shift operator defined by

$$Na(n) = a(n+1).$$

We say that $a(n)$ is *C-finite* if and only if there exists a polynomial $p(x)$ such that $p(N)a(n) = 0$ for all $n \geq 0$. We say that $p(x)$ *annihilates* $a(n)$. For example, $x^2 - x - 1$ annihilates the Fibonacci sequence $F(n)$ and $x-2$ annihilates the exponential sequence 2^n. The set of all polynomials which annihilate a fixed $a(n)$ is an ideal. The generator of this ideal is the *characteristic polynomial* of $a(n)$, and we call its degree the degree (or order) of $a(n)$.

Every C-finite sequence has a closed-form expression as a sum of polynomials times exponential sequences. More specifically,

$$a(n) = \sum_{k=1}^{m} f_k(n) r_k^n,$$

where $r_1, r_2, \ldots, r_m$ are the distinct roots of the characteristic equation of $a(n)$ and $f_k(n)$ is a polynomial in n with degree less than or equal to the multiplicity of the root r_k. We call these formulas *Binet-type formulas* after Binet's famous formula for the Fibonacci numbers. For example, $(x-2)^2$ is an annihilating polynomial of any sequence $a(n)$ which satisfies the recurrence $a(n+2) = 4a(n+1) - 4a(n)$, and this implies $a(n) = (\alpha + \beta n)2^n$ for some constants α and β.

We can go the other way and derive an annihilating polynomial from a closed form expression. A term of the form $n^d r^n$ is annihilated by $(x-r)^{d+1}$, so for each exponential r^n in the closed form, look for the highest power n^d which is multiplied by r^n and write down $(x-r)^{d+1}$. For example, the sequence $a(n) = n3^n - \frac{n^2}{2} + 5^n$ is annihilated by $(x-3)^2(x-1)^3(x-5)$.

Finally, if $a(n)$ and $b(n)$ are two C-finite sequences, then so are the following:

$$a(n)b(n) \qquad a(n) \pm b(n) \qquad \sum_{k=0}^{n} a(k)b(n-k).$$

C-finite sequences are a special subclass of *holonomic sequences*, sequences which satisfy a linear recurrence with *polynomial coefficients* [Kau13]. Holonomic sequences satisfy very similar properties, but do not have the readily computable closed forms which we need here.

6.2 Uniform recurrences

First up, we will prove the analogue of (6.2) for arbitrary C-finite sequences.

Proposition 14. *If $a(n)$ is a C-finite sequence of order d, then $n \mapsto a(nm)$ satisfies a recurrence of the form*

$$a(nm) = \sum_{k=1}^{d} c_k(m)a((n-k)m), \tag{6.3}$$

where $c_k(m)$ is C-finite with respect to m and has order at most $\binom{d}{k}$. The sequence $c_1(m)$ always satisfies the same recurrence as $a(n)$ itself, and $c_d(k) = \omega^k$, where ω is $(-1)^d$ times the constant coefficient of the characteristic polynomial of $a(n)$.

The following proof is constructive given the roots of the characteristic polynomial of $a(n)$, but [BGW15] gives formulas for $c_k(m)$ in terms of partial Bell polynomials without reference to the roots.

Proof. The Binet-type formula for $a(n)$ is a linear combination of terms of the form $n^i r^n$ where i is a nonnegative integer and r is a root of the characteristic polynomial of $a(n)$. Thus, the Binet-type formula for $a(nm)$ is a linear combination of terms of the form $(nm)^i r^{nm}$, which is equivalently a linear combination of terms of the form $n^i (r^m)^n$. The only thing that has changed is the exponential terms themselves, so if

$$\prod_{k=1}^{d}(x - r_k)$$

is the characteristic polynomial of $a(n)$ with possibly repeated roots $r_1, \ldots, r_d$, then

$$\prod_{k=1}^{d}(x - r_k^m). \tag{6.4}$$

annihilates $n \mapsto a(nm)$. From the elementary theory of polynomials, the coefficients of (6.4) are elementary symmetric functions of the roots r_k^m. C-finite sequences are closed under multiplication and addition, so the coefficients of the polynomial are C-finite with respect to m.

To obtain the degree bound, recall that the coefficient on x^{d-i} in (6.4) equals $(-1)^i e_i(r_1^m, \ldots, r_d^m)$, where $e_i(r_1^m, \ldots, r_d^m)$ is the sum of all products of i distinct r_k^m. Each of these products is of the form α^m for some constant α. The number of such terms is an upper bound on the degree of the sequence with respect to m, and there are exactly $\binom{d}{i}$ of them.

Finally, note that the coefficient on x^{d-1} is precisely the sum $\sum_k r_k^m$, which is annihilated by the characteristic polynomial of $a(n)$ itself, and the coefficient on x^{d-d} is precisely the product $(r_1 r_2 \ldots r_d)^m$. □

Example: Perrin numbers The Perrin numbers $P(n)$ are a third-order C-finite sequence defined by

$$P(0) = 0 \quad P(1) = 0 \quad P(2) = 2$$
$$P(n+3) = P(n+1) + P(n).$$

They are sometimes called the "skipponaci" numbers. They satisfy the interesting property that p divides $P(p)$ for every prime p. Tracing through the above proof reveals the meta-recurrence

$$P(mn) = P(m)P(m(n-1)) + c(m)P(m(n-2)) + P(m(n-3)), \tag{6.5}$$

where $c(m)$ is A078712 in the OEIS.

Example: General second-order Let $a(n)$ be annihilated by $(x-r_1)(x-r_2)$ for distinct reals r_1 and r_2. The proof of Proposition 14 shows that $n \mapsto a(mn)$ is annihilated by

$$(x - r_1^m)(x - r_2^m) = x^2 - (r_1^m + r_2^m)x + (r_1 r_2)^m.$$

In particular, if r_1 and r_2 are the golden ratio and its conjugate, respectively, then $r_1^m + r_2^m = L_m$ is the mth Lucas number, and $r_1 r_2 = -1$. This recovers (6.1).

Example: Square Fibonacci The square Fibonacci numbers F_n^2 are also C-finite. Going through the steps of the above proof and consulting the OEIS reveals the following general identity:

$$F_{mn}^2 = (5F_m^2 + 3(-1)^m)(F_{m(n-1)}^2 - (-1)^m F_{m(n-2)}^2) + (-1)^m F_{m(n-3)}^2. \tag{6.6}$$

Example: Tribonacci Consider the sequence T_n defined by

$$T_0 = 0 \quad T_1 = 0 \quad T_2 = 1$$
$$T_n = T_{n-1} + T_{n-2} + T_{n-3}.$$

The family of sequences $n \mapsto T_{nm}$ satisfy the following recurrences:

$$T_n = T_{n-1} + T_{n-2} + T_{n-3}$$
$$T_{2n} = 3T_{2(n-1)} + T_{2(n-2)} + T_{2(n-3)}$$
$$T_{3n} = 7T_{3(n-1)} - 5T_{3(n-2)} + T_{3(n-3)}$$
$$T_{4n} = 11T_{4(n-1)} + 5T_{4(n-2)} + T_{4(n-3)}$$
$$T_{5n} = 21T_{5(n-1)} + T_{5(n-2)} + T_{5(n-3)}$$
$$T_{6n} = 39T_{6(n-1)} - 11T_{6(n-2)} + T_{6(n-3)}.$$

In general,

$$T_{nm} = c_1(m)T_{(n-1)m} + c_2(m)T_{(n-1)m} + T_{(n-2)m},$$

where

$$c_1(1) = 1 \quad c_1(2) = 3 \quad c_1(3) = 7$$
$$c_1(m) = c_1(m-1) + c_1(m-2) + c_1(m-3)$$

and

$$c_2(1) = 1 \quad c_1(2) = 1 \quad c_1(3) = -5$$
$$c_2(m) = -c_2(m-1) - c_2(m-2) + c_2(m-3).$$

The sequences $c_k(m)$ were found via guessing. However, Proposition 14 establishes that these sequences *are* C-finite, and so proving our guess requires that we check only finitely many terms. In this case we must check no more than double the maximum degree, which is 6 terms. We have produced just enough examples above to constitute a proof.

6.3 Uniform sums

The Fibonacci numbers satisfy the famous summation identity

$$\sum_{k=0}^{n} F_k = F_{n+2} - 1. \tag{6.7}$$

There are as many ways to prove this identity as there are articles devoted to evaluating related Fibonacci sums [Lay77; Mel99; Fro18], but the most useful method at this juncture is the following method outlined in [KP10]. The annihilating polynomial of F_n can be written as

$$x^2 - x - 1 = (x-1)x - 1.$$

Applying this to F_n shows that $F_n = (x-1)F_{n+1} = F_{n+2} - F_{n+1}$. If we sum over n, then the right-hand side telescopes and we recover (6.7). In general, if $p(x)$ annihilates $a(n)$ and $p(1) \neq 0$, then we can write $p(x) = (x-1)q(x) + p(1)$ for some easily-computable polynomial $q(x)$. Applying this to $a(n)$ shows that $a(n) = (x-1)b(n)$ where $b(n) = -q(x)a(n)/p(1)$. Summing over n yields

$$\sum_{0 \leq k < n} a(k) = b(n) - b(0).$$

From this idea, the uniform recurrences we have derived for sequences of the form $n \mapsto a(mn)$ and $n \mapsto a(ni)a(nj)$ will help us discover uniform summation identities.

Here is one such identity for the Perrin numbers, using (6.5).

Proposition 15. *The Perrin numbers $P(n)$ satisfy*

$$\sum_{0 \leq k < n} P(mn) = \frac{(P(n)-3)(1-P(m)-c(m)) + P(n+1)(1-P(m)) + P(n+2) - 2}{P(m)+c(m)},$$

where $c(m)$ is A078712 in the OEIS.

Using (6.6), we can quickly rediscover the following infinite family of sums for the square of the Fibonacci numbers.

Proposition 16. *If m is odd, then*

$$\sum_{0 \leq k < n} F_{mk}^2 = \frac{F_{mn}F_{m(n-1)}}{L_m}.$$

Proof. Using (6.6), we obtain

$$\sum_{0\le k<n} F_{mk}^2 = \frac{F_{mn}^2(7-10F_m^2)+(F_{m(n+1)}^2-F_m^2)(4-5F_m^2)+F_{m(n+2)}^2-F_{2m}^2}{10F_m^2-8}.$$

This is far from the most economical representation. First, the numerator here contains $(5F_m^2-4)F_m^2-F_{2m}^2$. It is easy to check that

$$(5F_m^2-4)F_m^2-F_{2m}^2 = -8F_m^2\frac{(-1)^m+1}{2}, \tag{6.8}$$

so the expression on the left vanishes when m is odd. We are down to

$$\frac{F_{mn}^2(7-10F_m^2)+F_{m(n+1)}^2(4-5F_m^2)+F_{m(n+2)}^2}{10F_m^2-8}.$$

Applying the general recurrence (6.1) to $F_{m(n+2)}$ and simplifying the result brings us to

$$\frac{F_{mn}^2(8-10F_m^2)+F_{m(n+1)}^2((4-5F_m^2)+L_m^2)+2F_{mn}L_mF_{m(n+1)}}{10F_m^2-8}.$$

When m is odd, the identity $4-5F_m^2+L_m^2=0$ follows from dividing (6.8) by F_m^2 and recalling that $L_m = F_{2m}/F_m$. Using this and simplifying gives

$$\frac{F_{mn}(-L_mF_{mn}+F_{m(n+1)})}{L_m},$$

and applying the general recurrence (6.1) once more to $F_{m(n+1)}$ gives us the final answer $F_{mn}F_{m(n-1)}/L_m$. □

6.4 Uniform products

The proof of Proposition 14 relied on little more than the identity $r^{mn}=(r^m)^n$ and some structural facts about C-finite sequences. Unsurprisingly, these ideas apply to other settings. The below proposition shows how to apply the idea to prove that sequences of the form $n\mapsto a(ni)a(nj)$ also satisfy meta C-finite recurrences.

Proposition 17. *If $a(n)$ is C-finite of degree d whose characteristic polynomial has m distinct roots, then $P_{i,j}(n)=a(ni)a(nj)$ satisfies a recurrence of the form*

$$P_{i,j}(n)=\sum_{k=1}^{m(2d-m)} c_k(i,j)P_{i,j}(n-k),$$

where each $c_k(i,j)$ is C-finite with respect to i and j and $c_k(i,j)=c_k(j,i)$. The sequence $c_k(i,j)$ has order *(with respect to i or j) no more than* $\binom{d}{k}$.

Proof. Write the characteristic polynomial of $a(n)$ as $\prod_{k=1}^{m}(x-r_k)^{d_k+1}$ where the r_k are distinct and $d_1+d_2+\cdots+d_m = d-m$. Then,

$$a(n) = \sum_{k=1}^{m} p_k(n) r_k^n,$$

where p_k is a polynomial in n of degree d_k or less. Therefore

$$P_{i,j}(n) = \sum_{1\le k,v\le m} p_k(in)p_v(jn)(r_k^i r_v^j)^n.$$

Immediately, we see that $P_{i,j}(n)$ is annihilated by

$$\prod_{1\le k,v\le m}(x-r_k^i r_v^j)^{d_k+d_v+1}, \tag{6.9}$$

a polynomial of degree $\sum_{k,v}(d_k+d_v+1) = m(2d-m)$. The coefficients of this polynomial are elementary symmetric polynomials in the variables $\{r_k^i r_v^j\}_{1\le k,v\le d}$, and therefore C-finite with respect to i and j by the C-finite closure properties. The roots $r_k^i r_v^j$ are symmetric in i and j, so the coefficient sequences are as well.

The coefficient on x^{D-k} is essentially the sum of all products of k distinct elements from $\{r_k^i r_v^j\}_{1\le k,v\le d}$. As a sequence in i the r_v^j factors are irrelevant: The coefficient will be annihilated by the characteristic polynomial for the sum of all products of k distinct elements from $\{r_k^i\}_{1\le k\le d}$. Each term of this latter sum is of the form α^i for some constant α, and there are no more than $\binom{d}{k}$ distinct values of α. Therefore $c_k(i,j)$ has order no more than $\binom{d}{k}$ with respect to i (and also j). □

The previous proof can be slightly modified to produce a stronger statement. Namely, if we split the product (6.9) into diagonal and off-diagonal terms, we get the following corollary.

Corollary 3. *Let $a(n)$ be a C-finite sequence of degree d whose characteristic polynomial has m distinct roots. Then $n \mapsto a(ni)a(nj)$ is annihilated by a polynomial $C_{i,j}(x)$ which factors as*

$$C_{i,j}(x) = L_{i+j}(x)R_{i,j}(x), \tag{6.10}$$

where $\deg L_{i+j} = 2d-m$ and $\deg R_{i,j} = (m-1)(2d-m)$. The coefficients of $L_{i+j}(x)$ are C-finite sequences in $i+j$ and the coefficients of $R_{i,j}(x)$ are C-finite sequences which are symmetric in i and j.

There is one case of this corollary worth highlighting. Now that we know these annihilating polynomials with C-finite coefficients exist, we could find them by computing enough examples and guessing a pattern. However, if the degrees of $L_{i+j}(x)$ and $R_{i,j}(x)$ are the same, then it is not always clear which factor is L and which factor is R in a given example. This happens when $2d - m = (m-1)(2d-m)$. Since $m \leq d$, the interesting solution is $m = 2$. Thus sequences with exactly two roots in their characteristic polynomial should be handled "manually." We will show one example.

Example: Second-order annihilators Let $a(n)$ be a C-finite sequence annihilated by the quadratic $(x - r_1)(x - r_2)$ where $r_1 \neq r_2$. Then $n \mapsto a(ni)a(nj)$ is annihilated by

$$(x^2 - \mathscr{L}(i+j)x + (r_1 r_2)^{i+j})(x^2 - (r_1 r_2)^j \mathscr{L}(i-j)x + (r_1 r_2)^{i+j})$$

where $\mathscr{L}(n) = r_1^n + r_2^n$. If $a(n) = F(n)$ equals the nth Fibonacci number, then $\mathscr{L}(n) = L(n)$ is the nth Lucas number, $r_1 r_2 = -1$, and we obtain the annihilator

$$(x^2 - L(i+j)x + (-1)^{i+j})(x^2 - (-1)^j L(i-j)x + (-1)^{i+j}).$$

6.5 Computer demo

This article is joined by a corresponding Maple package `MetaCfinite`, obtainable from GitHub at https://github.com/rwbogl/MetaCfinite. With `MetaCfinite`, nearly all the propositions described in this article can be explored and checked empirically.

Guessing uniform recurrences Suppose that we want to discover (6.1) and the corresponding general pattern. The following Maple commands compute the five recurrences from (6.1):

```
Fib := [[0, 1], [1, 1]:
mSect(Fib, 1, 0); # [[0, 1], [1, 1]]
mSect(Fib, 2, 0); # [[0, 1], [3, -1]]
mSect(Fib, 3, 0); # [[0, 2], [4, 1]]
mSect(Fib, 4, 0); # [[0, 3], [7, -1]]
mSect(Fib, 5, 0); # [[0, 5], [11, 1]]
```

We are trying to guess the pattern followed by $1, 3, 4, 7, 11$, and $1, -1, 1, -1, 1$. The following command does this for us:

```
MetaMSect(Fib, 0); # [[[1, 3], [1, 1]], [[1], [-1]]]
```

This tells us that, for example, the coefficient on $F_{m(n-1)}$ is a sequence L_m which begins $L_1 = 1$, $L_2 = 3$, and satisfies $L_m = L_{m-1} + L_{m-2}$. These are the Lucas numbers.

Uniform summation identities The procedure `polysum(a, n, p, x` computes an expression for $\sum_{0 \leq k < n} a(k)$ where $a(n)$ is a C-finite sequence with characteristic polynomial $p(x)$. For example, the following command derives the famous identity (6.7):

```
polysum(F, n, x^2 - x - 1, x); # F(n + 1) - F(1).
```

This is most powerful when joined with uniform recurrences found by `MetaMSect`. For instance, the sequence $n \mapsto F(mn)$ has characteristic polynomial $p_m(x) = x^2 - L(m)x - (-1)^{m+1}$. The following commands derive a summation identity for $\sum_{0 \leq k < n} F(mk)$:

```
polysum(Fm, n, x^2 - L(m) * x - (-1)^(m + 1), x);
     (Fm(n) - Fm(0)) (1 - L(m)) + Fm(n + 1) - Fm(1)
   - ---------------------------------------------
                                m
                   1 - L(m) + (-1)
```

That is, we have automatically derived the famous identity

$$\sum_{0 \leq k < n} F(mk) = \frac{F(mn)(1 - L(m)) + F(m(n+1)) - F(m)}{L(m) - 1 - (-1)^m}.$$

6.6 Conclusion

We have used the theory of C-finite sequences to establish *meta-facts* about the recurrences C-finite sequences satisfy. Namely, we have shown that the recurrences satisfied by $n \mapsto a(nm)$ and $n \mapsto a(ni)a(nj)$ are uniform in a C-finite sense. This allowed us to state uniform families of summation identities for some C-finite sequences.

The summation identities our methods derive are automatic and uniform, but we do not claim that they are the "best possible." For instance, the first expression obtained for $\sum_{k=0}^{n-1} F_{mk}^2$ in Proposition 16 is quite cumbersome compared to the final answer:

$$\frac{F_{mn}^2(7 - 10F_m^2) + (F_{m(n+1)}^2 - F_m^2)(4 - 5F_m^2) + F_{m(n+2)}^2 - F_{2m}^2}{10F_m^2 - 8} = \frac{F_{mn}F_{m(n-1)}}{L_m}.$$

It still takes some (semi-automatic) sweat to discover this reduction. Can we automatically discover and prove such "complex = simple" identities? And might this apply to more complex sums, such as $\sum_{k=0}^{n-1} F_{mk}^5$? The answer is likely yes—and perhaps a C-finite simplification algorithm already exists—but we leave this as an open problem.

Chapter 7

Gosper's algorithm and Bell numbers

Evaluating the partial sums of sequences which are products and quotients of polynomials, binomial coefficients, factorials, and so on is a major theme in combinatorics, discrete probability, and computer science. The main tool in this area is Bill Gosper's marvelous hypergeometric summation algorithm [Gos78].

A sequence $f(k)$ is *hypergeometric* (or a *hypergeometric term*) provided that the consecutive quotient $f(k+1)/f(k)$ is a rational function in k. Gosper's algorithm completely solves the problem of hypergeometric summation in one variable. It constructively determines when $\sum_k f(k)$ itself is hypergeometric term. We call such hypergeometric terms "Gosper summable."

Unfortunately, many hypergeometric terms are *not* Gosper summable. For example, we cannot fill in the following blank with a hypergeometric term:

$$\sum_{k=0}^{n} \frac{1}{k!} = \underline{\hspace{6cm}}.$$

However, an upshot of Gosper's algorithm is that we can often *tweak* non-Gosper summable terms to make them summable. For example, while $1/k!$ is not Gosper summable, the term $(k-1)/k!$ *is*. In fact, lots of multiples of $1/k!$ are Gosper summable:

$$\begin{aligned}
\sum_{k=0}^{n} \frac{k-1}{k!} &= -\frac{1}{n!} \\
\sum_{k=0}^{n} \frac{k^2-2}{k!} &= -\frac{n+2}{n!} \\
\sum_{k=0}^{n} \frac{k^3-5}{k!} &= -\frac{n^2+3n+5}{n!} \\
\sum_{k=0}^{n} \frac{k^4-15}{k!} &= -\frac{n^3+4n^2+9n+15}{n!}.
\end{aligned}$$

This list suggests that there exists a sequence of integers $b(d)$—beginning $1,2,5,15$—such that $(k^d - b(d))/k!$ is Gosper summable. This turns out to be true. Even better, $b(d)$ turns out to be the dth Bell number, the number of partitions of d elements into any number of nonempty subsets.

Indeed, there is a similar statement for hypergeometric terms of the form $z^k a^{\overline{k}}$ and $z^k/a^{\overline{k}}$ for constant z, where $a^{\overline{k}} = a(a+1)\cdots(a+k-1)$ is the *rising factorial*. Specifically, there are explicit exponential generating functions $g_{a,z}(x)$ and $f_{a,z}(x)$ such that $(k^d - c(d))z^k a^{\overline{k}}$ is Gosper summable iff $c(d)$ is the coefficient on $x^d/d!$ in $g_{a,z}(x)$, and the analogous statement for $z^k/a^{\overline{k}}$ and $f_{a,z}(x)$. These generating functions happen to be related to the famous exponential generating function for the Bell numbers $B(x) = e^{e^x-1}$. Our goal is to explain and prove these facts.

The remainder of this paper is organized as follows. Section 7.1 gives a quick overview of Gosper's algorithm. Section 7.2 establishes the summability results and gives the explicit generating functions. Section 7.3 shows how to explicitly evaluate a special case of these sums in terms of well-known integer sequences. Section 7.4 explains how these generating functions are related to the Bell numbers.

7.1 Gosper's algorithm

This section provides a brief overview of Gosper's algorithm. For more details, see [Gos78] or [PWZ97].

A sequence $f(k)$ is *hypergeometric*, or a *hypergeometric term*, provided that $f(k+1)/f(k)$ is a rational function in k.

Every rational function $R(k)$ can be decomposed as

$$R(k) = \frac{a(k)}{b(k)} \frac{c(k+1)}{c(k)},$$

where a, b, and c are polynomials in k which satisfy $\gcd(a(k), b(k+i)) = 1$ for all nonnegative integers i. This is called the polynomial normal form of $R(k)$. If $f(k)$ is hypergeometric and $f(k+1)/f(k)$ has polynomial normal form

$$\frac{f(k+1)}{f(k)} = \frac{a(k)}{b(k)} \frac{c(k+1)}{c(k)},$$

then we call a/b the *kernel* of f, and c the *shell* of f. Note that

$$f(k) = zc(k) \prod_{j=0}^{k-1} (a(j)/b(j))$$

for some constant z. For this reason, we sometimes call $c(k)$ the "polynomial part" of $f(k)$ and the remaining product the "purely hypergeometric part."

Gosper's algorithm amounts to the following theorem.

Theorem. *The hypergeometric term $f(k)$ with polynomial normal form (a,b,c) is Gosper summable if and only if there is a polynomial solution $x(k)$ to*

$$x(k+1)a(k) - x(k)b(k-1) = c(k). \tag{7.1}$$

In that case,

$$\sum_k f(k) = \left(\frac{x(k)b(k-1)}{c(k)}\right) f(k).$$

For example, the term ratio of $f(k) = 1/k!$ has polynomial normal form

$$\frac{f(k+1)}{f(k)} = \frac{1}{k+1}$$

with $(a,b,c) = (1,k+1,1)$. Therefore $f(k)$ is summable if and only if $x(k+1) - x(k)k = 1$ has a polynomial solution $x(k)$, which it does not. On the other hand, the term ratio of $g(k) = (k-1)/k!$ has polynomial normal form

$$\frac{g(k+1)}{g(k)} = \frac{1}{k+1}\frac{k}{k-1}$$

with $(a,b,c) = (1,k+1,k-1)$. Therefore $g(k)$ is summable if and only if $x(k+1) - x(k)k = k-1$ has a polynomial solution $x(k)$, which it does, namely $x(k) = -1$. In addition,

$$\sum_k \frac{k-1}{k!} = -\frac{k}{k!} = -\frac{1}{(k-1)!}.$$

7.2 Summability

For pedagogical purposes, let us first prove the following proposition.

Proposition 18. *The term*

$$\frac{k^d - b(d)}{k!}$$

is Gosper summable if and only if $b(d)$ is the dth Bell number.

Proof. The consecutive term ratios of $1/k!$ are $1/(k+1)$, which has polynomial normal form $(1,k+1,1)$. Setting $x_d(k) = -k^d$ in (7.1) shows that

$$p_d(k) = k^{d+1} - (k+1)^d$$

is a sequence of polynomials such that $p_d(k)/k!$ is Gosper summable. Since the degree of $p_d(k)$ is exactly $d+1$, these polynomials are linearly independent, and therefore form a basis for the set of *all* polynomials $p(k)$ such that $p(k)/k!$ is Gosper summable. Our proposition amounts to the claim that $k^d + b(d)$ is a different basis for this space.

The degrees of the $p_d(k)$ start at 1 and increase by 1 every step, so by subtracting appropriate multiples of previous terms, we can cancel every power of k in $p_d(k)$ except the leading term and the constant. That is, there *is* a basis of the form $k + c(1), k^d + c(2), \ldots$, obtained by a linear operation on the $p_d(k)$. In particular, since

$$p_d(k) = k^{d+1} - \sum_j \binom{d}{j} k^j,$$

the correct multiples to subtract are as follows:

$$k^{d+1} + c(d+1) = p_d(k) + \sum_{j>0} \binom{d}{j} (k^j + c(j)). \tag{7.2}$$

If we set $c(0) = -1$ and look at the constant term of both sides, we obtain

$$c(d+1) = \sum_j \binom{d}{j} c(j).$$

This implies $c(d) = -b(d)$, where $b(d)$ is the dth Bell number, since the Bell numbers satisfy the same recurrence and begin with 1 rather than -1. $\square$

The above outline carries over nearly verbatim to other simple hypergeometric terms. A slight difference is that, most of the time, the sequence $c(d)$ is not well-known, and we have to settle for an explicit exponential generating function. The following propositions neatly summarize the results.

Proposition 19. *If $z \neq 0$ and a is not a nonpositive integer, then the hypergeometric term $(k^d - c(d))z^k/a^{\overline{k}}$ is Gosper summable if and only if*

$$c(d) = [x^d/d!]\exp(-z-(a-1)x+ze^x) = [x^d/d!]f_{a,z}(x).$$

Proof. The consecutive term ratios of $z^k/a^{\overline{k}}$ are $z/(a+k)$, so their polynomial normal form is $(z, a+k, 1)$. It follows that the sequence of polynomials

$$p_d(k) = k^d(a+k-1) - z(k+1)^d$$

for $d = 0, 1, 2, \ldots$ form a basis for the set of polynomials $p(k)$ such that $p(k)z^k/a^{\overline{k}}$ is Gosper summable. It suffices to transform this basis by iteratively eliminating all powers of k from $p_d(k)$ except its highest power and its constant term, then to show that the constant terms have the quoted exponential generating function.

Note that

$$p_d(k) = k^{d+1} - (z+1-a)k^d - z\sum_{j<d}\binom{d}{j}k^j.$$

Therefore,

$$k^{d+1} + c(d+1) = p_d(k) + (z+1-a)(k^d + c(d)) + z\sum_{0<j<d}\binom{d}{j}(k^j + c(j)).$$

Comparing constant terms, we see that

$$c(d+1) = -z + (z+1-a)c(d) + z\sum_{0<j<d}\binom{d}{j}c(j)$$

If we let $c(0) = -1$, then this becomes

$$c(d+1) = (1-a)c(d) + z\sum_{j}\binom{d}{j}c(j).$$

If $C(x) = \sum_{d\geq 0}\frac{c(d)}{d!}x^d$ is the exponential generating function of $c(d)$, then the previous equation implies

$$C'(x) = (1-a)C(x) + ze^x C(x).$$

Solving this linear differential equation yields

$$C(x) = -e^{-z-(a-1)x+ze^x}.$$

Therefore $(k^d - c(d))z^k/a^{\overline{k}}$ is Gosper summable if and only if $c(d)$ is the coefficient on $x^d/d!$ in $\exp(-z-(a-1)x+ze^x)$. □

Proposition 20. *If $z \neq 0$ and a is not a nonpositive integer, then the hypergeometric term $(k^d - c(d))z^k a^{\overline{k}}$ is Gosper summable with respect to k if and only if*

$$c(d) = [x^d/d!]\exp(z^{-1} - ax - z^{-1}e^{-x}) = [x^d/d!]g_{a,z}(x).$$

Proof. The consecutive term ratio of $z^k a^{\overline{k}}$ has polynomial normal form $(z(a+k), 1, 1)$. Therefore, as in the proof of Proposition 2, the sequence of polynomials

$$p_d(k) = z(k+1)^d(a+k) - k^d$$

form a basis for the set of all polynomials $p(k)$ such that $p(k)z^k a^{\overline{k}}$ is Gosper summable, and our job is to simplify it.

Note that

$$p_d(k) = zk^{d+1} + (z(a+d)-1)k^d + z\sum_{j<d}\left(a\binom{d}{j} + \binom{d}{j-1}\right)k^j.$$

Therefore, having constructed basis elements of the form $k+c(1)$, $k^2+c(2), \ldots, k^d+c(d)$, we have

$$z(k^{d+1}+c(d+1)) = p_d(k) - (z(a+d)-1)(k^d+c(d)) - z\sum_{0<j<d}\left(a\binom{d}{j} + \binom{d}{j-1}\right)(k^j+c(j))$$

Comparing constant coefficients yields

$$zc(d+1) = az - (z(a+d)-1)c(d) - z\sum_{0<j<d}\left(a\binom{d}{j} + \binom{d}{j-1}\right)c(j).$$

If we let $c(0) = -1$, then this becomes

$$zc(d+1) = c(d) - z\sum_{0\le j\le d}\left(a\binom{d}{j} + \binom{d}{j-1}\right)c(j).$$

If we move the sum to the left-hand side, the equation reads

$$c(d) = z\sum_{j}\left(a\binom{d}{j} + \binom{d}{j-1}\right)c(j).$$

If $C(x)$ is the exponential generating function of $c(d)$, then standard techniques give us

$$C(x) = z(e^x C(x) + e^x C'(x)),$$

whose unique solution with $C(0) = -1$ is $C(x) = -\exp(z^{-1} - ax - z^{-1}e^{-x})$. □

7.3 Explicit Formulas and Gould Numbers

In the previous section we proved that

$$\sum_k \frac{k^d - b(d)}{k!}$$

is Gosper summable when $b(d)$ is the dth Bell number. In this section we will explicitly evaluate this sum in terms of well-known integer sequences.

Equation (7.2) is essentially a change of basis equation. It tells us how to express the polynomials $p_d(k)$

in terms of the polynomials $k^d - b(d)$. The first basis, $p_d(k)$, has the benefit that

$$\sum_k \frac{p_d(k)}{k!} = -\frac{k^{d+1}}{k!}.$$

So, if we could invert (7.2) and express $k^d - b(d)$ in terms of $p_d(k)$, $p_{d-1}(k)$, and so on, we could apply linearity to evaluate $\sum_k \frac{k^d - b(d)}{k!}$.

Equation (7.2) amounts to the following matrix identity:

$$\begin{bmatrix} p_0(k) \\ p_1(k) \\ p_2(k) \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ -2 & -1 & 1 & 0 & \cdots \\ -3 & -3 & -1 & 1 & \cdots \\ \vdots \end{bmatrix} \begin{bmatrix} k - b(1) \\ k^2 - b(2) \\ k^3 - b(3) \\ \vdots \end{bmatrix}. \tag{7.3}$$

The coefficient matrix is invertible. The first few rows of A^{-1} are as follows:

$$A^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & 0 & \cdots \\ 3 & 1 & 1 & 0 & 0 & \cdots \\ 9 & 4 & 1 & 1 & 0 & \cdots \\ 31 & 14 & 5 & 1 & 1 & \cdots \\ \vdots \end{bmatrix}.$$

The OEIS [OEI24] suggests that the columns are the diagonals of A121207, which is a table of values T_{dj} defined by

$$T_{(d+1)j} = \sum_{i=0}^{d-j-1} \binom{r}{i} T_{(d-i)j}.$$

This table is a special case of a family of tables studied by Gould and Quaintance [GQ07]. The numbers T_{r1} are called the *Gould numbers* (see A040027). This suggestion turns out to be correct.

Proposition 21. *For any positive integer d,*

$$\sum_k \frac{k^d - b(d)}{k!} = -\frac{\sum_{j \geq 1} B_{dj} k^j}{k!},$$

where the matrix B_{dj} is defined by $B_{dd} = 1$ and

$$B_{(d+1)j} = \sum_{k=0}^{d-j} \binom{d}{k} B_{(d-k)j} \quad (d \geq j).$$

In particular, B_{dj} is the dth element of the $(j-1)$th diagonal of A121207.

Proof. The matrix in (7.3) is defined by

$$A_{dj} = [d = j] - \binom{d-1}{j}[d \neq j].$$

For B to be its matrix inverse, we must have

$$\sum_{k \geq 1} A_{(d+1)k} B_{kj} = [d+1 = j] \tag{7.4}$$

for all integers $d \geq 0$ and $j \geq 1$. If we expand $A_{(d+1)k}$, then this reads

$$B_{(d+1)j} = \sum_{k \geq 1} \binom{d}{k} B_{kj} + [d+1 = j]. \tag{7.5}$$

Note that this implies $B_{dj} = 0$ if $d < j$. Indeed, $B_{1j} = \binom{0}{1} B_{1k} = 0$, and if $d < j-1$, then we can apply induction to every term of the right-hand side of (7.5) to conclude that $B_{(d+1)j} = 0$. Hence we can define B_{dj} as follows:

$$\begin{aligned} B_{jj} &= 1 \\ B_{(d+1)j} &= \sum_{j \leq k \leq d} \binom{d}{k} B_{kj} \quad (d \geq j). \end{aligned}$$

This is A121207 shifted so that j begins at 1 rather than 0. $\square$

Written more concretely, this identity reads

$$\sum_{k=0}^{n-1} \frac{k^d - b(d)}{k!} = -\frac{\sum_{j \geq 1} B_{dj} n^j}{n!}.$$

If we multiply by $n!$ and rearrange things, we obtain the following equality for the bell numbers, valid for $n \geq 1$ and $d \geq 0$:

$$b(d) = \frac{\sum_{k=0}^{n-1} k^d n^{\underline{n-k}} + \sum_{j \geq 1} B_{dj} n^j}{\sum_{k=0}^{n-1} n^{\underline{n-k}}}. \tag{7.6}$$

It seems plausible that this has a combinatorial proof, but the author does not know one.

7.4 Connections with Bell numbers

The exponential generating functions from the previous section are

$$f_{a,z}(x) = \exp(-z-(a-1)x+ze^{x})$$
$$g_{a,z}(x) = \exp(z^{-1}-ax-z^{-1}e^{-x}).$$

These functions, and therefore the underlying sequences, are connected with the Bell numbers. In particular, if we let

$$B(x) = e^{e^{x}-1} = \sum_{j\geq 0} \frac{b(d)}{d!}x^{d}$$

be the exponential generating function for the Bell numbers, then for integral z we have the following identities:

$$f_{a,z}(x) = e^{(1-a)x}B(x)^{z} \tag{7.7}$$
$$g_{a,1/z}(x) = e^{-ax}B(-x)^{-z}. \tag{7.8}$$

If z is positive, the first equation says that the coefficients of $f_{1,z}(x)$ are the binomial convolution of $(1-a)^{k}$ with the convolution of the Bell numbers with themselves z times. If z is negative, the second equation says that the coefficients of $g_{1,1/z}(x)$ are the binomial convolution of $(-a)^{k}$ with the convolution of the alternating Bell numbers $(-1)^{d}b(d)$ with themselves z times.

Examples for $z^{k}/a^{\overline{k}}$ Setting $a=z=1$ in (7.7), we get $f_{1,1}(x)=B(x)$. If we translate this into the vocabulary of the previous section, this says that

$$\frac{k^{d}-b(d)}{k!}$$

is Gosper summable, and no other constants will work. Similarly, $f_{1,2}(x)=B(x)^{2}$, so

$$\frac{(k^{d}-c(d))2^{k}}{k!}$$

is Gosper summable only if $c(d) = \sum_j \binom{d}{j} b(d)b(d-j)$. This sequence begins $2, 6, 22, 94$, corresponding to the following identities:

$$\sum_{k=0}^{n} \frac{(k-2)2^k}{k!} = -\frac{2^{n+1}}{n!}$$
$$\sum_{k=0}^{n} \frac{(k^2-6)2^k}{k!} = -\frac{(n+3)2^{n+1}}{n!}$$
$$\sum_{k=0}^{n} \frac{(k^3-22)2^k}{k!} = -\frac{(n^2+4n+11)2^{n+1}}{n!}$$
$$\sum_{k=0}^{n} \frac{(k^4-94)2^k}{k!} = -\frac{(n^3+5n^2+17n+47)2^{n+1}}{n!}.$$

Setting $a = 1/2$ and $z = 1$ gives $g_{1/2,1}(x) = e^{x/2}B(x)$, which says that

$$\frac{k^d - c(d)}{(1/2)^{\overline{k}}} = (k^d - c(d))\frac{4^k k!}{(2k)!}$$

is Gosper summable only if $c(d) = \sum_j \binom{d}{j} b(d)/2^{d-j}$.

Examples for $z^k a^{\overline{k}}$ The connection for $g_{1,1/z}(x)$ is most convenient when z is a negative integer. Setting $z = -1$ in (7.8) gives $g_{1,-1}(x) = e^{-x}B(-x) = B'(-x)$, which says that

$$(k^d - (-1)^d b(d+1))(-1)^k k!$$

is Gosper summable, and no constant except $(-1)^d b(d+1)$ will work. Similarly, $g_{1,-1/2}(x) = e^{-x}B(-x)^2 = B'(-x)B(-x)$. Therefore,

$$(k^d - c(d))\frac{k!}{(-2)^k}$$

is Gosper summable only if $c(d) = (-1)^d \sum_j \binom{d}{j} b(j+1)b(d-j)$. For example,

$$\sum_{k=0}^{n} \frac{(k^2-11)k!}{(-2)^k} = \frac{(n-3)(n+1)!}{(-2)^n} - 8.$$

Setting $a = 1/2$ and $z = -1$ gives $g_{1/2,-1}(x) = e^{-x/2}B(-x)$, so

$$(k^d - c(d))(-1)^k (1/2)^{\overline{k}} = (k^d - c(d))(-1)^k \frac{(2k)!}{4^k k!}$$

is Gosper summable only if $c(d) = (-1)^d \sum_j \binom{d}{j} b(d)/2^{d-j}$.

7.5 Conclusion

We have given some explicit conditions for the Gosper summability of hypergeometric terms of the form

$$(k^d - c(d))z^k a^{\overline{k}} \quad \text{and} \quad (k^d - c'(d))\frac{z^k}{a^{\overline{k}}}.$$

Namely, $c(d)$ and $c'(d)$ must be the coefficients of explicit exponential generating functions which are related to the Bell numbers. In the special case of $1/k!$, we gave an explicit evaluation of these sums in terms of the inverse of a matrix involving binomial coefficients. The Bell numbers *probably* appear by accident. However, should some combinatorial connection be made, the author would like to hear about it.

We have made use of Gosper's algorithm for hypergeometric summation, but there is a continuous variant of Gosper's algorithm for hyperexponential integration [AZ90]. We may be able to make statements about when integrals of the form

$$\int (x^d - b(d))e^{-x^2}\, dx$$

are themselves hyperexponential. However, in contrast to the summation problem, we have a solid understanding of *all* elementary antiderivatives thanks to Liouville's theorem [GCL92, ch. 12], not just hyperexponential ones. Thus this could be a less satisfying problem.

Finally, we note that the results here work essentially because the space of polynomials $p(k)$ such that $p(k)z^k a^{\overline{k}}$ are Gosper summable contains polynomials of every degree greater than or equal to 1. More complicated hypergeometric terms will produce spaces with degrees only 2 or greater, or 3 or greater, and so on. In these cases, the basis could not be simplified down to leading powers and constants, so the results would be about terms of the form

$$(k^d - kc_1(d) - c_0(d))f(k),$$

or

$$(k^d - k^2c_2(d) - kc_1(d) - c_0(d))f(k),$$

and so on. The techniques here would certainly apply to such terms, though the results would be more difficult to state.

Chapter 8

Tweaking the Beukers Integrals In Search of More Miraculous Irrationality Proofs á La Apéry

In honor of our irrational guru Wadim Zudilin, on his $\lfloor 50\zeta(5) \rfloor$-th birthday

8.1 Hilbert's 0-th problem

Before David Hilbert [Hol20] stated his famous 23 problems, he mentioned two problems that he probably believed to be yet much harder, and indeed, are still wide open today. One of them was to prove that there are infinitely many prime numbers of the form $2^n + 1$, and the other one was to prove that the Euler-Mascheroni constant is irrational.

Two paragraphs later he stated his optimistic belief that "in mathematics there is no *ignorabimus*."

As we all know, he was proven wrong by Gödel and Turing *in general*, but even for such concrete problems, like the irrationality of a specific, natural, constant, like the Euler-Mascheroni constant (that may be defined in terms of the **definite** integral $-\int_0^\infty e^{-x} \log x \, dx$) , that is most probably **decidable** in the logical sense, (i.e. there *probably* **exists** a (rigorous) proof), we *lowly humans* did not yet find it, (and may never will!).

While the Euler-Mascheroni constant (and any other, natural, explicitly-defined, constant that is not obviously rational) is **surely** irrational, in the everyday sense of the word *sure* (like death and taxes), giving a *proof*, in the **mathematical** sense of 'proof' is a different matter. While e was proved irrational a long time ago (trivial exercise), and π was proved irrational by Lambert around 1750, we have no clue how to prove that $e+\pi$ is irrational. Ditto for $e\cdot\pi$. **Exercise**: Prove that *at least* one of them is irrational.

8.2 Apéry's Miracle

As Lindemann first proved in 1882, the number π is more than *just* irrational, it is *transcendental*, hence it follows that $\zeta(n)$ is irrational for all even arguments, since Euler proved that $\zeta(2n)$ is a multiple of π^{2n} by a rational number. But proving that $\zeta(3)$, $\zeta(5)$, $\dots$ are irrational remained wide open.

Since such problems are so hard, it was breaking news, back in 1978, when 64-years-old Roger Apéry announced and sketched a proof that $\zeta(3) := \sum_{n=1}^{\infty}\frac{1}{n^3}$ is irrational. This was beautifully narrated in a classic expository paper by Alf van der Poorten [Poo79], aided by details filled-in by Henri Cohen and Don Zagier. While beautiful in our eyes, most people found the proof ad-hoc and too complicated, and they did not like the heavy reliance on *recurrence relations*.

To those people, who found Apéry's original proof too *magical*, ad-hoc, and *computational*, another proof, by a 24-year-old PhD student by the name of Frits Beukers [Beu79] was a breath of fresh air. It was a marvelous *gem* in human-generated mathematics, and could be easily followed by a first-year student, using *partial fractions* and very easy estimates of a certain triple integral, namely

$$\int_0^1\int_0^1\int_0^1 \frac{(x(1-x)y(1-y)z(1-z))^n}{(1-z+xyz)^{n+1}}\,dx\,dy\,dz \quad .$$

The general approach of Apéry of finding concrete sequences of integers a_n, b_n such that

$$|\zeta(3)-\frac{a_n}{b_n}| < \frac{CONST}{b_n^{1+\delta}} \quad ,$$

(see below) for a **positive** δ was still followed, but the details were much more palatable and *elegant* to the average *mathematician in the street*.

As a warmup, Beukers, like Apéry before him, gave a new proof of the already proved fact that $\zeta(2)=\frac{\pi^2}{6}$ is irrational, using the double integral

$$\int_0^1\int_0^1 \frac{(x(1-x)y(1-y))^n}{(1-xy)^{n+1}}\,dx\,dy \quad .$$

Ironically, we will follow Beukers' lead, but heavily using **recurrence relations**, that will be the **engine** of our approach. Thus we will abandon the original *raison d'être* of Beukers' proof of **getting rid of recurrences**, and bring them back with a **vengeance**.

8.3 How Beukers' proofs could have been discovered

Once upon a time, there was a precocious teenager, who was also a *computer whiz*, let's call him/her/it/they *Alex*. Alex just got a new laptop that had `Maple`, as a birthday present.

Alex typed, *for no particular reason*,

```
int(int(1/(1-x*y),x=0..1),y=0..1);
```

and immediately got the answer: $\frac{\pi^2}{6}$. Then Alex wondered about the sequence

$$I(n) := \int_0^1 \int_0^1 \frac{(x(1-x)y(1-y))^n}{(1-xy)^{n+1}}\, dx\, dy \quad .$$

(why not, isn't it a natural thing to try out for a curious teenager?), and typed the following:

```
I1:=n->int(int(1/(1-x*y)*(x*(1-x)*y*(1-y)/(1-x*y))**n,x=0..1),y=0..1);
```

(`I` is reserved in Maple for $\sqrt{-1}$, so Alex needed to use `I1`) Alex looked at the first ten values by typing:

`L:=[seq(I1(i),i=1..10)];` ,

getting after a few seconds

$$[5-\frac{\pi^2}{2}, -\frac{125}{4}+\frac{19\,\pi^2}{6}, \frac{8705}{36}-\frac{49\,\pi^2}{2}, -\frac{32925}{16}+\frac{417\,\pi^2}{2},$$
$$\frac{13327519}{720}-\frac{3751\,\pi^2}{2}, -\frac{124308457}{720}+\frac{104959\,\pi^2}{6},$$
$$\frac{19427741063}{11760}-\frac{334769\,\pi^2}{2}, -\frac{2273486234953}{141120}+\frac{9793891\,\pi^2}{6},$$
$$\frac{202482451324891}{1270080}-\frac{32306251\,\pi^2}{2}, -\frac{2758128511985}{1728}+\frac{323445423\,\pi^2}{2}] \quad .$$

Alex immediately noticed that, at least for $n \leq 10$,

$$I(n) = a_n - b_n \frac{\pi^2}{6} \quad ,$$

for some **integers** b_n and some **rational numbers** a_n. By taking `evalf(L)`, Alex also noticed that $I(n)$ get smaller and smaller. Knowing that Maple could not be trusted with floating point calculations (unless you change the value of `Digits` from its default, to something higher, say, in this case `Digits:=30`), that they get smaller and smaller. Typing '`evalf(L,30);`', Alex got:

$$[0.0651977994553206905827545000 6, 0.00374727011630229297588816 63,$$

$$0.000247728866269394110526059, 0.0000176271312720269913734 7,$$

$$0.00000131246346593146768 53, 0.000000100776323486001254,$$

$$0.0000000079121296437194 6, 0.000000000631743771120 6,$$

$$5.1111100706 \times 10^{-11}, 4.17922459 \times 10^{-12}] \quad .$$

Alex realized that $I(n)$ seems to go to *zero* fairly fast, and since $I(10)/I(9)$ and $I(9)/I(8)$ were pretty close, Alex conjectured that the limit of $I(n)/I(n-1)$ tends to a certain constant. But ten data points do not suffice!

When Alex tried to find the first 2000 terms, Maple got slower and slower. Then Alex asked Alexa, the famous robot,

Alexa: how do I compute many terms of the sequence $I(n)$ given by that double-integral?

and Alexa replied:

Go to Doron Zeilberger's web-site and download the amazing program *https://sites.math.rutgers.edu/~zeilberg/tokhniot/MultiAlmkvistZeilberger.txt* that accompanied the article [AZ06].

Typing

`MAZ(1,1/(1-x*y),x*(1-x)*y*(1-y)/(1-x*y),[x,y],n,N, {})[1];`

immediately gave a recurrence satisfied by $I(n)$

$$I(n) = -\frac{\left(11\,n^2 - 11\,n + 3\right)}{n^2} \cdot I\left(n-1\right) + \frac{\left(n-1\right)^2}{n^2} \cdot I\left(n-2\right) \quad .$$

Using this recurrence, Alex easily computed the first 2000 terms, using the following Maple one-liner (calling the sequence defined by the recurrence `I2(n)`):

```
I2:=proc(n) option remember:  if n=0 then Pi**2/6 elif n=1 then 5-Pi**2/2 else -(11*n**2-11*n+
end:
```

and found out that indeed $I(n)/I(n-1)$ tends to a limit, about 0.09016994. Writing

$$I(n) = a_n - b_n \frac{\pi^2}{6}$$

and realizing that $I(n)$ is small, Alex found terrific **rational approximations** to $\frac{\pi^2}{6}$, a_n/b_n, that after **clearing denominators** can be written as a'_n/b'_n where *now* **both** numerator a'_n and denominator b'_n are **integers**.

$$\frac{\pi^2}{6} \approx \frac{a'_n}{b'_n} \quad .$$

Alex also noticed that for all n up to 2000, for some constant C,

$$|\frac{\pi^2}{6} - \frac{a'_n}{b'_n}| \leq \frac{C}{(b'_n)^{1+\delta}} \quad ,$$

where δ is roughly 0.09215925467. Then Alex concluded that this proves that $\frac{\pi^2}{6}$ is irrational, since if it were rational the left side would have been $\geq \frac{C_1}{b'_n}$, for some constant C_1. Of course, some details would still need to be filled-in, but that was not too hard.

8.4 The general strategy

Let's follow Alex's lead. (Of course our fictional Alex owes a lot to the real Beukers and also to Alladi and Robinson [AR80]).

Start with a constant, let's call it C, given by an explicit integral

$$\int_0^1 K(x)\,dx \quad ,$$

for some **integrand** $K(x)$, or, more generally, a d-dimensional integral

$$\int_0^1 \cdots \int_0^1 K(x_1,\ldots,x_k)\,dx_1 \ldots dx_k \quad .$$

Our goal in life is to prove that C is irrational. Of course C may turn out to be rational (that happens!), or more likely, an algebraic number, or expressible in terms of a logarithm of an algebraic number, for which, there already exist irrationality proofs (albeit not always *effective* ones). But who knows? Maybe this constant has never been proved irrational, and if it will happen to be famous (e.g. Catalan's constant, or $\zeta(5)$, or the Euler-Mascheroni constant mentioned above), we will be famous too. But even if it is a nameless constant, it is still extremely interesting, if it is the *first* irrationality proof, since these proofs are so hard, witness that, in spite of great efforts by experts like Wadim Zudilin, the proofs of these are still wide open.

In this article we will present numerous candidates. Our proofs of irrationality are modulo a 'divisibility lemma' (see below), that we are *sure* that someone like Wadim Zudilin, to whom this paper is dedicated, can fill-in. Our only doubts are whether these constants are not already proved to be irrational because they happen to be algebraic (probably not, since Maple was unable to *identify* them), or more complicated numbers (like logarithms of algebraic numbers). Recall that Maple's `identify` can't (yet) identify *everything* that God can.

Following Beukers and Alladi-Robinson, we introduce a sequence of integrals, parameterized by a nonnegative integer n

$$I(n) = \int_0^1 K(x)\,(x(1-x)K(x))^n\,dx \quad ,$$

and analogously for multiple integrals, or more generally

$$I(n) = \int_0^1 K(x)\,(x(1-x)S(x))^n\,dx \quad ,$$

for another function $S(x)$. Of course $I(0) = C$, our constant that we want to prove irrational.

It so happens that for a wide class of functions $K(x)$, $S(x)$, (for single or multivariable x) using the **Holonomic ansatz** [Zei90b], and implemented (for the single-variable case) in [AZ90], and for the multi-variable case in [AZ06], and *much more efficiently* in [Kou09], there exists a **linear recurrence equation with polynomial coefficients**, that can be actually computed (always in theory, but also often in practice, unless the dimension is high). In other words we can find a positive integer L, the **order** of the recurrence, and **polynomials** $p_0(n), p_1(n), \ldots, p_L(n)$, such that

$$p_0(n)I(n) + p_1(n)I(n+1) + \cdots + p_L(n)I(n+L) \,=\, 0 \quad .$$

If we are lucky (and all the cases in this paper fall into this case) the order L is 2. Furthermore, it would be evident in all the examples in this paper that $p_0(n), p_1(n), p_2(n)$ can be taken to have **integer coefficients**.

Another 'miracle' that happens in all the examples in this paper is that $I(0)$ and $I(1)$ are rationally-related, i.e. there exist integers c_0, c_1, c_2 such that

$$c_0 I(0) + c_1 I(1) = c_2 \quad ,$$

that our computers can easily find.

It then follows, by induction, that one can write

$$I(n) = b_n C - a_n \quad ,$$

for some sequences of **rational numbers** $\{a_n\}$ and $\{b_n\}$ that both satisfy the same recurrence as $I(n)$.

Either using trivial bounds on the integral, or using the so-called Poincaré lemma (see, e.g. [Poo79; ZZ21; ZZ20]) it turns out that

$$a_n \,=\, \Omega(\alpha^n) \quad , \quad b_n \,=\, \Omega(\alpha^n) \quad ,$$

for some constant $\alpha > 1$, and

$$|I(n)| = \Omega(\frac{1}{\beta^n}) \quad ,$$

for some constant $\beta > 1$.

[Please note that we use Ω in a looser-than-usual sense, for us $x(n) = \Omega(\alpha^n)$ means that $\lim_{n\to\infty} \frac{\log x(n)}{n} = \alpha$.]

In the tweaks of Beukers' integrals for $\zeta(2)$ and $\zeta(3)$ coming up later, α and β are equal, but in the tweaks of the Alladi-Robinson integrals, α is usually different than β.

It follows that

$$|C - \frac{a_n}{b_n}| = \Omega(\frac{1}{(\alpha\beta)^n}) \quad .$$

Note that a_n, and b_n are, usually, **not** integers, but rather *rational numbers* (In the original Beukers/Apéry cases, the b_n were integers, but the a_n were not, in the more general cases in this article, usually neither of them are integers).

It so happens, in all the cases that we discovered, that there exists another sequence of rational numbers $E(n)$ such that

$$a'_n := a_n E(n) \quad , \quad b'_n := b_n E(n) \quad ,$$

are always **integers**, and, of course $gcd(a'_n, b'_n) = 1$. We call $E(n)$ the **integer-ating factor**.

In some cases we were able to conjecture $E(n)$ exactly, in terms of products of primes satisfying certain conditions (see below), but in other cases we can only conjecture that such an explicitly-describable sequence exists.

In either case there exists a real number, that sometimes can be described exactly, and other times only estimated, let's call it ν, such that

$$\lim_{n\to\infty} \frac{\log E(n)}{n} = \nu \quad ,$$

or, in our notation, $E(n) = \Omega(e^{n\nu})$.

Since we have

$$|C - \frac{a'_n}{b'_n}| = \Omega(\frac{1}{(\alpha\beta)^n}) \quad ,$$

where $b'_n = \Omega(e^{\nu n}\alpha^n)$. We need a **positive** δ such that

$$(e^{\nu n}\alpha^n)^{1+\delta} = (\alpha\beta)^n \quad .$$

Taking log (and dividing by n) we have

$$(\nu+\log\alpha)(1+\delta)=\log\alpha+\log\beta \quad ,$$

giving

$$\delta=\frac{\log\beta-\nu}{\log\alpha+\nu} \quad .$$

If we are *lucky*, and $\log\beta>\nu$, then we have $\delta>0$, and an irrationality proof!, Yea! We also, at the same time, determined an *irrationality measure* (see [Poo79])

$$1+\frac{1}{\delta}=\frac{\log\alpha+\log\beta}{\log\beta-\nu} \quad .$$

If we are *unlucky*, and $\delta<0$, it is still an exponentially fast way to compute our constant C to any desired accuracy.

Summarizing: For each specific constant defined by a definite integral, we need to exhibit

• A second-oder recurrence equation for the numerator and denominator sequence a_n and b_n that feature in $I(n)=b_nC-a_n$.

• The *initial conditions* a_0,a_1, b_0,b_1 enabling a very fast computation of many terms of a_n,b_n.

• The constants α and β

• Exhibit a conjectured **integer-ating factor** $E(n)$, or else conjecture that one exists, and find, or estimate (respectively), $\nu:=\lim_{n\to\infty}\frac{\log E(n)}{n}$.

• Verify that $\beta>e^{\nu}$ and get (potentially) famous.

8.5 The three classical cases

$\log$ **2** ([AR80])

$$C=\int_0^1\frac{1}{1+x}\,dx=\log 2 \quad .$$

$$I(n)=\int_0^1\frac{(x(1-x))^n}{(1+x)^{n+1}}\,dx \quad .$$

Recurrence:

$$(n+1)X(n)+(-6n-9)X(n+1)+(n+2)X(n+2)=0 \quad .$$

$$\alpha = \beta = 3 + 2\sqrt{2} \quad .$$

Initial conditions

$$a_0 = 0\,, a_1 = 2 \quad ; \quad b_0 = 1\,, b_1 = 3 \quad .$$

Integer-ating factor $E(n) = lcm(1 \ldots n)$, $\nu = 1$.

$$\delta = \frac{\log\beta - \nu}{\log\alpha + \nu} = \frac{\log\beta - 1}{\log\alpha + 1} = \frac{\log(3+2\sqrt{2}) - 1}{\log(3+2\sqrt{2}) + 1} = 0.276082871862633587 \quad .$$

Implied irrationality measure: $1 + 1/\delta = 4.622100832454231334\ldots$.

$\zeta(\mathbf{2})$ ([Beu79])

$$C = \int_0^1 \int_0^1 \frac{1}{1-xy}\,dx\,dy = \zeta(2) \quad .$$

$$I(n) = \int_0^1 \int_0^1 \frac{(x(1-x)y(1-y))^n}{(1-xy)^{n+1}}\,dx\,dy \quad .$$

Recurrence:

$$-(1+n)^2 X(n) + \left(11\,n^2 + 33\,n + 25\right) X(n+1) + (2+n)^2 X(n+2) = 0 \quad .$$

$$\alpha = \beta = \frac{11}{2} + \frac{5\sqrt{5}}{2} \quad .$$

Initial conditions

$$a_0 = 0\,, a_1 = -5 \quad ; \quad b_0 = 1\,, b_1 = -3 \quad .$$

Integer-ating factor $E(n) = lcm(1 \ldots n)^2$, $\nu = 2$.

$$\delta = \frac{\log\beta - \nu}{\log\alpha + \nu} = \frac{\log\beta - 2}{\log\alpha + 2} = \frac{\log(11/2 + 5\sqrt{5}/2) - 2}{\log(11/2 + 5\sqrt{5}/2) + 2} = 0.09215925473323\ldots \quad .$$

Implied irrationality measure: $1 + 1/\delta = 11.85078219105234269595 28\ldots$.

$\zeta(\mathbf{3})$ ([Beu79])

$$C = \int_0^1 \int_0^1 \int_0^1 \frac{1}{1-z+xyz}\,dx\,dy\,dz = \zeta(3) \quad .$$

$$I(n) = \int_0^1 \int_0^1 \int_0^1 \frac{(x(1-x)y(1-y)z(1-z))^n}{(1-z+xyz)^{n+1}}\, dx\, dy\, dz \quad .$$

Recurrence:

$$(1+n)^3 X(n) - (2n+3)\left(17n^2+51n+39\right) X(n+1) + (n+2)^3 X(n+2) = 0 \quad .$$

$$\alpha = \beta = 17 + 12\sqrt{2} \quad .$$

Initial conditions

$$a_0 = 0\,, a_1 = 12 \quad ; \quad b_0 = 1\,, b_1 = 5 \quad .$$

Integer-ating factor $E(n) = lcm(1\ldots n)^3$, $\nu = 3$.

$$\delta = \frac{\log\beta - \nu}{\log\alpha + \nu} = \frac{\log\beta - 3}{\log\alpha + 3} = \frac{\log(17+12\sqrt{2}) - 3}{\log(17+12\sqrt{2}) + 3} = 0.08052943118906168518 6\ldots \quad .$$

Implied irrationality measure: $1 + 1/\delta = 13.41782023335376578458\ldots$.

8.6 Zudilin's Catalan constant integral

The *inspiration* for our *tweaks* came from Wadim Zudilin's brilliant discovery [Zud03] that the famous Catalan constant, that may be defined by the innocent-looking alternating series of the reciprocals of the odd perfect-squares

$$C := 1 - \frac{1}{3^2} + \frac{1}{5^2} - \frac{1}{7^2} + \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2},$$

can be written as the double integral

$$\frac{1}{8}\int_0^1 \int_0^1 \frac{x^{-\frac{1}{2}}(1-y)^{-\frac{1}{2}}}{1-xy}\, dx\, dy \quad .$$

This lead him to consider the sequence of Beukers-type double-integrals

$$I(n) = \int_0^1 \int_0^1 \frac{x^{-\frac{1}{2}}(1-y)^{-\frac{1}{2}}}{1-xy} \cdot \left(\frac{x(1-x)y(1-y)}{1-xy}\right)^n dx\, dy \quad .$$

Using the Zeilberger algorithm, Zudilin derived a three term recurrence for $I(n)$ leading to good diophantine approximations to the Catalan constant, alas not good enough to prove irrationality. This was elaborated and extended by Yu. V. Nesterenko [Nes16]. See also [Zud04].

Using the multivariable Almkvist-Zeilberger algorithm we can derive the recurrence much faster. Using Koutschan's package [Kou09], it is yet faster.

Inspired by Zudilin's Beukers-like integral for the Catalan constant, we decided to use our efficient tools for quickly manufacturing recurrences.

We systematically investigated the following families.

8.7 Generalizing the Alladi–Robinson integral for $\log 2$

Alladi and Robinson [AR80] gave a Beukers-style new proof of the irrationality of $\log 2$ using the elementary fact that

$$\log 2 = \int_0^1 \frac{1}{1+x}\,dx \quad ,$$

and more generally,

$$\frac{1}{c}\log(1+c) = \int_0^1 \frac{1}{1+cx}\,dx \quad .$$

They used the sequence of integrals

$$I(n) := \int_0^1 \frac{1}{1+cx}\left(\frac{x(1-x)}{1+cx}\right)^n dx \quad ,$$

and proved that for a wide range of choices of rational c, this leads to irrationality proofs and irrationality measures (see also [ZZ21]).

Our generalized version is the three-parameter family of constants

$$I_1(a,b,c) := \frac{1}{B(1+a,1+b)}\int_0^1 \frac{x^a(1-x)^b}{1+cx}\,dx$$

that is easily seen to equal ${}_2F_1(1,a+1;a+b+2;-c)$.

We use the sequence of integrals

$$I_1(a,b,c)(n) := \frac{1}{B(1+a,1+b)}\int_0^1 \frac{x^a(1-x)^b}{1+cx}\cdot\left(\frac{x(1-x)}{1+cx}\right)^n dx \quad .$$

Using the (original!) Almkvist-Zeilberger algorithm [AZ90], implemented in the Maple package `https://sites.math.rutgers.edu/~zeilberg/tokhniot/EKHAD.txt`, we immediately get a second-order recurrence that can be gotten by typing '`OpeL(a,b,c,n,N);`' in the Maple package `https://sites.math.`

rutgers.edu/~zeilberg/tokhniot/GenBeukersLog.txt

This enabled us to conduct a systematic search, and we found many cases of ${}_2F_1$ evaluations that lead to irrationality proofs, i.e. for which the δ mentioned above is **positive**. Many of them turned out to be (conjecturally) expressible in terms of algebraic numbers and/or logarithms of rational numbers, hence proving them irrational is not that exciting, but we have quite a few not-yet-identified (and inequivalent) cases. See the output file https://sites.math.rutgers.edu/~zeilberg/tokhniot/oGenBeukersLog1.txt for many examples. Whenever Maple was able to (conjecturally) identify the constants explicitly, it is mentioned. If nothing is mentioned then these are potentially explicit constants, expressible as a hypergeometric series ${}_2F_1$, for which this would be the **first** irrationality proof, once the details are filled-in.

We also considered the four-parameter family of constants

$$I_1'(a,b,c,d) := \frac{\int_0^1 \frac{x^a(1-x)^b}{(1+cx)^{d+1}}\,dx}{\int_0^1 \frac{x^a(1-x)^b}{(1+cx)^{d}}\,dx} \quad ,$$

and, using the more general recurrence, also obtained using the Almkvist-Zeilberger algorithm (to see it type '`OpeLg(a,b,c,d,n,Sn);`' in `GenBeukersLog.txt`), found many candidates for irrationality proofs that Maple was unable to identify. See the output file https://sites.math.rutgers.edu/~zeilberg/tokhniot/oGenBeukersLog2.txt.

8.8 Generalizing the Beukers Integral for $\zeta(2)$

Define

$$I_2(a_1,a_2,b_1,b_2)(n) = \frac{1}{B(1-a_1,1-a_2)B(1-b_1,1-b_2)} \cdot$$
$$\int_0^1\int_0^1 \frac{x^{-a_1}(1-x)^{-a_2}y^{-b_1}(1-y)^{-b_2}}{1-xy}\cdot\left(\frac{x(1-x)y(1-y)}{1-xy}\right)^n dx\,dy \quad ,$$

that happens to satisfy a linear-recurrence equation of second order, yielding Diophantine approximations to the constant $I_2(a_1,a_2,b_1,b_2)(0)$, let's call it $C_2(a_1,a_2,b_1,b_2)$

$$C_2(a_1,a_2,b_1,b_2) = \frac{1}{B(1-a_1,1-a_2)B(1-b_1,1-b_2)}\cdot\int_0^1\int_0^1 \frac{x^{-a_1}(1-x)^{-a_2}y^{-b_1}(1-y)^{-b2}}{1-xy}\,dx\,dy \quad .$$

It is readily seen that

$$C_2(a_1,a_2,b_1,b_2) = {}_3F_2\left(\begin{matrix}1\,,1-a_1\,,-b_1+1\\ 2-a_1-a_2\,,2-b_1-b_2\end{matrix};1\right) \quad .$$

Most choices of random a_1, a_2, b_1, b_2 yield disappointing, negative δ's, just like $C_2(\frac{1}{2}, 0, 0, \frac{1}{2})$ (alias 8 times the Catalan constant), but a systematic search yielded several hundred candidates that produce positive δ's and hence would produce irrationality proofs. Alas, many of them were conjecturally equivalent to each other via a fractional-linear transformation with integer coefficients, $C \to \frac{a+bC}{c+dC}$, with a, b, c, d integers, hence the facts that they are irrational are equivalent. Nevertheless we found quite a few that are (conjecturally) *not equivalent* to each other. Modulo filling-in some details, they lead to irrationality proofs. Amongst them some were (conjecturally) identified by Maple to be either algebraic, or logarithms of rational numbers, for which irrationality proofs exist for thousands of years (in case of $\sqrt{2}$ and $\sqrt{3}$ etc.), or a few hundred years (in case of $\log 2$, etc.).

But some of them Maple was **unable** to identify, so *potentially* our (sketches) of proofs would be the **first** irrationality proofs.

Denominator 2

We first searched for $C_2(a_1, a_2, b_1, b_2)$ where the parameters a_1, a_2, b_1, b_2 have denominator 2, there were quite a few of them, but they were all conjecturally equivalent to each other. Here is one of them:

- $C_2(0, 0, \frac{1}{2}, 0) = {}_3F_2(1, 1, 1/2; 2, 3/2; 1)$, alias $2\log 2$.

Denominator 3

There were also quite a few where the parameters a_1, a_2, b_1, b_2 have denominator 3, but again they were all equivalent to each other, featuring $\pi\sqrt{3}$. Here is one of them.

- $C_2(0, 0, \frac{1}{3}, -\frac{2}{3}) = {}_3F_2(1, 1, 2/3; 2, 7/3; 1)$, alias (conjecturally) $-6 + 4\pi\sqrt{3}/3$.

Denominator 4

There were also quite a few where the parameters a_1, a_2, b_1, b_2 have denominator 4, but again they were all equivalent to each other, featuring $\sqrt{2}$, yielding a new proof of the irrationality of $\sqrt{2}$ (for what it is worth). Here is one of them.

- $C_2(-\frac{3}{4}, -\frac{3}{4}, -\frac{1}{4}, -\frac{3}{4}) = {}_3F_2(1, 7/4, 5/4; 7/2, 3; 1)$, alias (conjecturally) $-240 + \frac{512}{3}\sqrt{2}$.

Denominator 5

There were also quite a few where the parameters a_1, a_2, b_1, b_2 have denominator 5, but again they were all equivalent to each other, featuring $\sqrt{5}$, yielding a new proof of the irrationality of $\sqrt{5}$ (for what it is worth). Here is one of them.

- $C_2(-\frac{4}{5}, -\frac{4}{5}, -\frac{2}{5}, -\frac{3}{5}) = {}_3F_2(1, 9/5, 7/5; 18/5, 3; 1)$, alias (conjecturally) $-\frac{845}{2} + \frac{2275}{12}\sqrt{5}$

Denominator 6 with identified constants

We found two equivalence classes where the parameters a_1, a_2, b_1, b_2 have denominator 6, for which the constants were identified. Here are one from each class.

- $C_2(-5/6, -5/6, -1/2, -1/2) = {}_3F_2(1, 11/6, 3/2; 11/3, 3; 1)$, alias (conjecturally) $-\frac{1344}{5} + \frac{16384\sqrt{3}}{105}$

• $C_2(-5/6,-5/6,-1/3,-2/3) = {}_3F_2(1,11/6,4/3;11/3,3;1)$, alias (conjecturally) $\frac{972\,2^{2/3}}{5} - \frac{1536}{5}$

denominator 7 with identified constants

We found two cases where the parameters a_1, a_2, b_1, b_2 have denominator 7, for which the constants were identified.

• $C_2(-6/7,-6/7,-4/7,-3/7) = {}_3F_2(1,13/7,11/7;26/7,3;1)$, alias (conjecturally) the positive root of $13824x^3 - 2757888x^2 - 10737789048x + 16108505539 = 0$.

• $C_2(-6/7,-1/7,4/7,2/7) = {}_3F_2(1,13/7,3/7;3,8/7;1)$, alias (conjecturally) the positive root of $2299968x^3 + 7074144x^2 - 11234916x - 12663217 = 0$

Maple was unable to identify the following constants, so we have *potentially* their first irrationality proofs.

Denominator 6 with not yet identified constants

We found two cases (up to equivalence):

• $C_2(0,-1/2,1/6,-1/2) = {}_3F_2(1,1,5/6;5/2,7/3;1)$

While Maple was unable to identify this constant, Mathematica came up with $-24 - \frac{81\sqrt{\pi}\Gamma(7/3)}{\Gamma(-1/6)}$.

• $C_2(-2/3,-1/2,1/2,-1/2) = {}_3F_2(1,5/3,1/2;19/6,2;1)$

While Maple was unable to identify this constant, Mathematica came up with $\frac{13}{2} - \frac{6\Gamma(19/6)}{\sqrt{\pi}\Gamma(8/3)}$.

Denominator 7 with not yet identified constants

We found six cases (up to equivalence):

• $C_2(-6/7,-6/7,-4/7,-5/7) = {}_3F_2(1,13/7,11/7;26/7,23/7;1)$

• $C_2(-6/7,-5/7,-3/7,-5/7) = {}_3F_2(1,13/7,10/7;25/7,22/7;1)$

• $C_2(-6/7,-5/7,-2/7,-1/7) = {}_3F_2(1,13/7,9/7;25/7,17/7;1)$

• $C_2(-6/7,-4/7,-1/7,-1/7) = {}_3F_2(1,13/7,8/7;24/7,16/7;1)$

• $C_2(-6/7,-3/7,-5/7,-3/7) = {}_3F_2(1,13/7,12/7;23/7,22/7;1)$

• $C_2(-5/7,-3/7,-4/7,-2/7) = {}_3F_2(1,12/7,11/7;22/7,20/7;1)$

For each of them, to get the corresponding theorem and proof, use procedure `TheoremZ2` in the Maple pacgage `GenBeukersZeta2.txt`.

To get a statement and full proof (modulo a divisibility lemma) type , in `GenBeukersZeta2.txt`

`TheoremZ2(a1,a2,b1,b2,K,0):`

with `K` at least 2000. For example, for the last constant in the above list ${}_3F_2(1,12/7,11/7;22/7,20/7;1)$, type

`TheoremZ2( -5/7, -3/7, -4/7, -2/7 ,3000,0):`

For more details (the recurrences, the estimated irrationality measures, the initial conditions) see the output file https://sites.math.rutgers.edu/~zeilberg/tokhniot/oGenBeukersZeta2g.txt.

8.9 Generalizing the Beukers Integral for $\zeta(3)$

The natural extension would be the **six**-parameter family (but now we make the exponents positive)

$$\frac{1}{B(1+a_1,1+a_2)B(1+b_1,1+b_2)B(1+c_1,1+c_2)} \cdot$$

$$\int_0^1 \int_0^1 \int_0^1 \frac{x^{a_1}(1-x)^{a_2}y^{b_1}(1-y)^{b_2}z^{c_1}(1-z)^{c_2}}{1-z+xyz} \cdot \left(\frac{x(1-x)y(1-y)z(1-z)}{1-z+xyz}\right)^n dx\,dy\,dz \quad .$$

However, for **arbitrary** a_1,a_2,b_1,b_2,c_1,c_2 the recurrence is **third order**. (Wadim Zudilin pointed out that this may be related to the work of Rhin and Viola in [RV01]).

Also, empirically, we did not find many promising cases. Instead, let's define

$$J_3(a_1,a_2,b_1,b_2,c_1,c_2;e)(n)$$

$$\int_0^1 \int_0^1 \int_0^1 \frac{x^{a_1}(1-x)^{a_2}y^{b_1}(1-y)^{b_2}z^{c_1}(1-z)^{c_2}}{(1-z+xyz)^e} \cdot \left(\frac{x(1-x)y(1-y)z(1-z)}{1-z+xyz}\right)^n dx\,dy\,dz \quad .$$

and

$$I_3(a_1,a_2,b_1,b_2,c_1,c_2;e)(n) := \frac{J_3(a_1,a_2,b_1,b_2,c_1,c_2;e+1)(n)}{J_3(a_1,a_2,b_1,b_2,c_1,c_2;e)(0)}$$

The family of constants that we hope to prove irrationality is the seven-parameter:

$$I_3(a_1,a_2,b_1,b_2,c_1,c_2;e)(0) \quad .$$

$$= \frac{\int_0^1 \int_0^1 \int_0^1 \frac{x^{a_1}(1-x)^{a_2}y^{b_1}(1-y)^{b_2}z^{c_1}(1-z)^{c_2}}{(1-z+xyz)^{e+1}}\,dx\,dy\,dz}{\int_0^1 \int_0^1 \int_0^1 \frac{x^{a_1}(1-x)^{a_2}y^{b_1}(1-y)^{b_2}z^{c_1}(1-z)^{c_2}}{(1-z+xyz)^{e}}\,dx\,dy\,dz} \quad .$$

Of course, for this more general, 7-parameter, family, there is no second-order recurrence, but rather a third-order one. But to our delight, we found a five-parameter family, let's call it

$$K(a,b,c,d,e)(n) := I_3(b,c,e,a,a,c,d)(n) \quad .$$

Spelled-out, our five-parameter family of constants is

$$K(a,b,c,d,e)(0) =$$

$$\frac{\int_0^1 \int_0^1 \int_0^1 \frac{x^b(1-x)^c y^e(1-y)^a z^a(1-z)^c}{(1-z+xyz)^{d+1}}\,dx\,dy\,dz}{\int_0^1 \int_0^1 \int_0^1 \frac{x^b(1-x)^c y^e(1-y)^a z^a(1-z)^c}{(1-z+xyz)^{d}}\,dx\,dy\,dz} \quad .$$

Now we found (see the section on finding recurrences below) a general second-order recurrence, that is too complicated to display here in full generality, but can be seen by typing

`OPEZ3(a,b,c,d,e,n,Sn);`

In the Maple package `GenBeukersZeta3.txt`. This enabled us, for each specific, numeric specialization of the parameters a,b,c,d,e to quickly find the relevant recurrence, and systematically search for those that give positive δ. Once again, many of them turned out to be (conjecturally) equivalent to each other.

Denominator 2:

We only found one class, up to equivalence, all related to $\log 2$. One of them is

$$K(0,0,0,1/2,1/2) = I_3(0,0,1/2,0,0,0,1/2) \quad ,$$

that is not that exciting since it is (conjecturally) equal to $-\frac{2-4\log(2)}{3-4\log(2)}$.

For details, type `TheoremZ3(0,0,0,1/2,1/2,3000,0);` in `GenBeukersZeta3.txt` .

Denominator 3:

We found three inequivalent classes, none of them Maple was able to identify.

$$K(0,0,0,1/3,2/3) = I_3(0,0,2/3,0,0,0,1/3) \quad ,$$

for details, type `TheoremZ3(0,0,0,1/3,2/3,3000,0);` in `GenBeukersZeta3.txt`.

$$K(0,0,0,2/3,1/3) = I_3(0,0,1/3,0,0,0,2/3) \quad ,$$

for details, type `TheoremZ3(0,0,0,2/3,1/3,3000,0);` in `GenBeukersZeta3.txt`.

$$K(0,1/3,2/3,1/3,2/3) = I_3(0,0,1/3,0,0,0,2/3) \quad ,$$

for details, type `TheoremZ3(0,1/3,2/3,1/3,2/3,3000,0);` in `GenBeukersZeta3.txt`,

These three constants are **candidates** for **'first-ever-irrationality proof'**.

Denominator 4: We only found one family, all expressible in terms of $\log 2$. Here is one of them.

For example

$$K(0,1/2,0,1/4,3/4) = I_3(1/2,0,3/4,0,0,0,1/4) \quad ,$$

that, conjecturally equals $-\frac{-30+45\log(2)}{-11+15\log(2)}$.

For details, type `TheoremZ3(0,1/2,0,1/4,3/4,3000,0);` in `GenBeukersZeta3.txt`.

Denominator 5: We only found one family, up to equivalence, but Maple was unable to identify the constant. So it is potentially the first irrationality proof of that constant

$$K(0,1/5,0,3/5,2/5) = I_3(1/5,0,2/5,0,0,0,3/5) \quad .$$

For details, type `TheoremZ3(0,1/5,0,3/5,2/5,3000,0);` in `GenBeukersZeta3.txt`.

Denominator 6: We found three families, up to equivalence, none of which Maple was able to identify. Once again, these are candidates for **first-ever irrationality proofs** for these constants.

$$K(0,1/2,1/2,1/3,1/6) = I_3(1/2,1/2,1/6,0,0,1/2,1/3) \quad .$$

For details, type `TheoremZ3(0,1/2,1/2,1/3,1/6,3000,0);` in `GenBeukersZeta3.txt`.

$$K(0,1/2,1/2,1/6,1/3) = I_3(1/2,1/2,1/3,0,0,1/2,1/6) \quad .$$

For details, type `TheoremZ3(0,1/2,1/2,1/6,1/3,3000,0);` in `GenBeukersZeta3.txt`.

$$K(1/3,0,2/3,1/2,5/6) = I_3(0,2/3,5/6,1/3,1/3,2/3,1/2) \quad .$$

For details, type `TheoremZ3(1/3,0,2/3,1/2,5/6,3000,0);` in `GenBeukersZeta3.txt`.

Denominator 7: We found five families, up to equivalence, none of which Maple was able to identify. Once again, these are candidates for first-ever irrationality proofs for these constants.

$$K(1/7,0,2/7,3/7,4/7) = I_3(0,2/7,4/7,1/7,1/7,2/7,3/7) \quad .$$

For details, type `TheoremZ3(1/7,0,2/7,3/7,4/7,3000,0);` in `GenBeukersZeta3.txt`.

$$K(1/7,0,2/7,5/7,3/7) = I_3(0,2/7,3/7,1/7,1/7,2/7,5/7) \quad .$$

For details, type `TheoremZ3(1/7,0,2/7,5/7,3/7,3000,0);` in `GenBeukersZeta3.txt`.

$$K(1/7,0,3/7,4/7,5/7) = I_3(0,3/7,5/7,1/7,1/7,3/7,4/7) \quad .$$

For details, type `TheoremZ3(1/7,0,3/7,4/7,5/7,3000,0);` in `GenBeukersZeta3.txt`.

$$K(1/7,0,4/7,2/7,5/7) = I_3(0,4/7,5/7,1/7,1/7,4/7,2/7) \quad .$$

For details, type `TheoremZ3(1/7,0,4/7,2/7,5/7,3000,0);` in `GenBeukersZeta3.txt`.

$$K(2/7,0,3/7,4/7,5/7) = I_3(0,3/7,5/7,2/7,2/7,3/7,4/7) \quad .$$

For details, type `TheoremZ3(2/7,0,3/7,4/7,5/7,3000,0);` in `GenBeukersZeta3.txt`.

If you don't have Maple, you can look at the output file `https://sites.math.rutgers.edu/~zeilberg/tokhniot/oGenBeukersZeta3All.txt` that gives detailed sketches of irrationality proofs of all the above constants, some with conjectured integer-ating factors.

Guessing an INTEGER-ating factor

In the original Beukers cases the integer-ating factor was easy to conjecture, and even to prove. For $\zeta(2)$ it was $lcm(1\dots n)^2$, and for $\zeta(3)$ it was $lcm(1\dots n)^3$. For the Alladi-Robinson case of $\log 2$ it was even simpler, $lcm(1\dots n)$.

But in other cases it is much more complicated. A natural 'atomic' object is, given a modulo M, a subset C of $\{0,...,M-1\}$, rational numbers e_1, e_2 between 0 and 1, rational numbers e_3, e_4, the following quantity, for positive integers n

$$Pp(e_1,e_2,e_3,e_4,C,M;n) := \prod_p p \quad ,$$

where p ranges over all primes such that (let $\{a\}$ be the fractional part of a, i.e. $a - \lfloor a \rfloor$)

- $e_1 < \{n/p\} < e_2$
- $e_3 < p/n < e_4$
- $p \; mod \; M \in C$

Using the prime number theorem, it follows (see e.g. [Zud04]) that

$$\lim_{n\to\infty} \frac{\log Pp(e_1,e_2,e_3,e_4,C,M;n)}{n} \quad ,$$

can be evaluated exactly, in terms of the function $\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ (see procedure `PpGlimit` in the Maple packages) thereby giving an exact value for the quantity δ whose positivity implies irrationality.

Of course, one still needs to rigorously prove that the conjectured integer-ating factor is indeed correct.

Looking under the hood: On Recurrence Equations

For 'secrets from the kitchen' on how we found the second-order, four-parameter recurrence operator `OPEZ2(a1,a2,b1,b2,n,N)` in the Maple package `GenBeukersZeta2.txt`, that was the engine driving the $\zeta(2)$ tweaks, and more impressively, the five-parameter second-order recurrence operator `OPEZ3(a,b,c,d,e,n,N)` in the Maple package `GenBeukersZeta3.txt`, that was the engine driving the $\zeta(3)$ tweaks, the reader is referred to the stand-alone appendix available from `https://sites.math.rutgers.edu/~zeilberg/`

mamarim/mamarimPDF/beukersAppendix.pdf.

Other Variations on Apéry's theme

Other attempts to use Apéry's brilliant insight are [Zei03; Zei24b; ZZ21]. Recently Marc Chamberland and Armin Straub [CS21] explored other fascinating aspects of the Apéry numbers, not related to irrationality.

Conclusion and Future Work

We believe that symbolic computational methods have great potential in irrationality proofs, in particular, and number theory in general. In this article we confined attention to approximating sequences that arise from *second-order recurrences*. The problem with higher order recurrences is that one gets linear combinations with rational coefficients of *several* constants, but if you can get two different such sequences coming from third-order recurrences, both featuring the same two constants, then the present method may be applicable. More generally if you have a k-th order recurrences, you need $k-1$ different integrals.

The general methodology of this article can be called *Combinatorial Number Theory*, but not in the usual sense, but rather as an analog of *Combinatorial Chemistry*, where one tries out many potential chemical compounds, most of them useless, but since computers are so fast, we can afford to generate lots of cases and pick the wheat from the chaff.

Encore: Hypergeometric challenges

As a *tangent*, we (or rather Maple) discovered many exact ${}_3F_2(1)$ evaluations. Recall that the Zeilberger algorithm can prove hypergemoetric identities only if there is at least one free parameter. For a *specific* ${}_3F_2(a_1\, a_2\, a_3\,; b_1\, b_2; 1)$, with *numeric* parameters, it is useless. Of course, it is sometimes possible to introduce such a parameter in order to conjecture a *general* identity, valid for 'infinitely' many n, and then specialize n to a specific value, but this remains an art rather than a science. The output file https://sites.math.rutgers.edu/~zeilberg/tokhniot/oGenBeukersZeta2f.txt contains many such conjectured evaluations, (very possibly many of them are equivalent via a hypergeometric transformation rule) and we challenge Wadim Zudilin, the *birthday boy*, or anyone else, to prove them.

Happy Ending

The *birthday boy* brilliantly met the challenges! See his brilliant note [Zud].

8.10 Accompanying Maple packages

This article is accompanied by three Maple packages, GenBeukersLog.txt, GenBeukersZeta2.txt, GenBeukersZeta3.txt all freely available from the *front* of this masterpiece https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/beukers.html, where one can find ample sample input and output files, that readers are welcome to extend.

Chapter 9

Enumerating restricted Dyck Paths with Context Free Grammars

9.1 Introduction

As Flajolet and Sedgewick demonstrate in their great text, *Analytic Combinatorics* [FS09], mathematicians have occasionally borrowed the study of formal languages from computer science and linguistics for combinatorial reasons. Many combinatorial classes can be reinterpreted as languages generated by certain grammars, and these grammars often make writing down generating functions, another favorite combinatorial tool, routine. Such grammars are sometimes called "combinatorial specifications."

For example, consider the well-known *Dyck paths*. A Dyck path is a finite list of $+1$'s and -1's whose partial sums are nonnegative, and whose sum is 0. We will write U (up) for $+1$ and D (down) for -1. Thus, the following are all Dyck paths:

$$UUDD$$
$$UDUD$$
$$UUUDUDDD$$

A Dyck path must have even length, and for this reason we often refer to Dyck paths of *semilength* n (length $2n$).

The number of Dyck paths of semilength n equals the nth *Catalan number*,

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

There are many proofs of this fact, but here is a *grammatical* proof.

Let $\mathscr{P}$ denote the set of all Dyck paths. Then, $\mathscr{P}$ is generated by the unambiguous, context-free grammar

$$\mathscr{P} = \varepsilon \cup U\mathscr{P}D\mathscr{P}, \tag{9.1}$$

where ε denotes the empty string. In words, a path is either empty or begins with a U, is followed by a Dyck path (shifted to height 1), a D, then another Dyck path.[1] This is a unique parsing of all Dyck paths.

Given a set of objects E each with a nonnegative integer size, let $GF(E) = \sum_{k\geq 0}|E(k)|z^k$ be a formal generating function, where $|E(k)|$ is the number of objects of size k in E. The main result about formal grammars is that, in an unambiguous context free grammar,

$$GF(A\cup B) = GF(A) + GF(B),$$

for disjoint clauses A and B, and

$$GF(AB) = GF(A)GF(B),$$

where $A\cup B$ is the union of the words of A and the words of B, and AB stands for "concatenation of words of A with words in B." The "sizes" of a grammar are the lengths of the words it generates.

In our case, if $P(z)$ is the generating function for the number of Dyck paths of semilength n, then the grammar (9.1) implies

$$\begin{aligned} P(z) &= GF(\varepsilon) + GF(U\mathscr{P}D\mathscr{P}) \\ &= 1 + zP(z)^2. \end{aligned}$$

The generating function $C(z)$ for the Catalan numbers *also* satisfies

$$C(z) = 1 + zC^2(z),$$

and since there are only two possible solutions, it is not hard to see that $P(z) = C(z)$.

The grammatical technique offers a unifying framework: Devise a grammar and you get an equation.

[1] Note that D denotes the first time the path returns to height 0.

Sometimes the equations turn out to be well-known or simple. Other times they are new and messy. The enumeration of all Dyck paths is one application of this framework, and here we want to demonstrate others. In particular, we will give grammatical proofs of several combinatorial facts about *restricted* Dyck paths, and also establish several infinite families of grammars in closed form.

First, let us define the restrictions we shall consider.

Definition 6. Given a Dyck path, the *height* of the path at position k is the partial sum of its first k terms. A *peak* of a Dyck path at height h (or simply "at h") is the bigram UD where the height of the path after the U is h. Similarly, a *valley* occurs at the bigram DU, and its height is analogously defined. The empty path has, by convention, a peak at 0 but no valley.

Definition 7. Given a sequence of steps L, define L^n to be the repetition of L n times. (For example, $U^2 = UU$ and $(UD)^3 = UDUDUD$.) A Dyck path has an *up-run of length n* provided that it contains at least one U^n that is not preceded nor followed by U. Similarly, it contains a *down-run of length n* provided that it contains at least one D^n that is neither preceded nor followed by D.

We will study Dyck paths whose peak heights, valley heights, up-run lengths, and down-run lengths avoid certain sets. We will, for example, discuss the set of all Dyck paths whose peak heights avoid $\{2,4,6,\dots\}$ and have no up-run of length greater than 2.

Definition 8. For arbitrary sets of positive integers A, B, C, and D, let $\mathscr{P}(A,B,C,D)$ be the set of Dyck paths whose peak heights avoid A, whose valley heights avoid B, whose up-run lengths avoid C, and whose down-run lengths avoid D. Let $P_{A,B,C,D}(z)$ be be the generating function for the number of Dyck paths of semilength n in $\mathscr{P}(A,B,C,D)$.

Some of these sets have been studied. In [PW01], Peart and Woan provide a continued-fraction recurrence for the generating functions $P_{\{k\},\emptyset,\emptyset,\emptyset}(z)$. In [ELY03], Eu, Liu, and Yeh take this idea further and express $P_{A,\emptyset,\emptyset,\emptyset}(z)$ as a finite continued fraction whenever A is finite or an arithmetic progression. In [HH17], Hein and Huang enumerate the number of Dyck paths which avoid up-runs of length k after a down step. In [EZ20], Zeilberger presents a rigorous experimental method to derive equations for $P_{A,B,C,D}(z)$ when the sets involved are finite or arithmetic progressions. Proving "by hand" some of Zeilberger's interesting discoveries *ex post facto* was a motivation for the present work. We generalize some of Zeilberger's results to infinite families which are likely out of reach for symbolic methods.

Our results include several explicit grammars (and therefore generating function equations) for infinite families of the sets A and B, and also grammatical proofs of several interesting special cases suggested in [EZ20]. Many of these—any grammars referencing restrictions on up- or down-runs—are not in [ELY03].

Some of our results are suggested in the OEIS [OEI24]; see, for example, A1006 (Motzkin numbers) and A004148 (generalized Catalan numbers).

The remainder of the paper is organized as follows. Section 9.2 presents some results discovered by experimentation with software from [EZ20] and proven with grammatical methods. Section 9.3 presents some infinite families of explicit grammars. Section 9.4 offers some concluding remarks about the limitations of grammars.

9.2 Combinatorial results

In this section we will present a number of results with grammatical proofs.

Proposition 22. *The number of Dyck paths of semilength n whose peak heights avoid $\{2r+3 \mid r \geq 0\}$ and whose up-runs are no longer than 2 is 1 when $n = 0$, and 2^{n-1} when $n \geq 1$.*

Proof. Let $\mathscr{P}$ be the set of all such Dyck paths, and $\mathscr{Q}$ the set of all Dyck paths which avoid peaks in $\{2r+2\}$ and up-runs longer than 2. Note that $\mathscr{P}$ and $\mathscr{Q}$ satisfy the following grammar:

$$\mathscr{P} = \varepsilon \cup UD\mathscr{P} \cup UUD\mathscr{Q}D\mathscr{P}$$
$$Q = \varepsilon \cup UDQ.$$

This implies the following system of equations:

$$P = 1 + zP + z^2QP$$
$$Q = 1 + zQ.$$

Thus, $Q(z) = (1-z)^{-1}$ (the only path in Q of semilength n is $(UD)^n$), and

$$P(z) = \frac{1-z}{1-2z}.$$

Therefore, $[z^0]P(z) = 1$, and $[z^n]P(z) = 2^{n-1}$. $\square$

The following proposition concerns *generalized Catalan numbers* (see A4148 in the OEIS and [SW79]).

These numbers are defined by the recurrence

$$G_0 = 1$$
$$G_1 = 1$$
$$G_{n+2} = G_{n+1} + \sum_{1 \le k < n+1} G_k G_{n-k}.$$

Proposition 23. *The number of Dyck paths of semilength n whose peak heights avoid* $\{2r+3 \mid r \ge 0\}$ *and whose up-runs are no longer than* 3 *equals the* $(n+1)$*th generalized Catalan number.*

Proof. Let $\mathscr{P}$, $\mathscr{O}$, and $\mathscr{E}$ be the set of all Dyck paths with up-runs no longer than 3, and whose peak heights avoid $\{2r+3 \mid r \ge 0\}$, $\{2r+2 \mid r \ge 0\}$, and $\{2r+1 \mid r \ge 0\}$, respectively. Observe that $\mathscr{P}$, $\mathscr{O}$, and $\mathscr{E}$ satisfy the following grammar:

$$\mathscr{P} = \varepsilon \cup UD\mathscr{P} \cup UUDOD\mathscr{P}$$
$$\mathscr{O} = \varepsilon \cup UD\mathscr{O} \cup UUUD\mathscr{O}D\mathscr{E}D\mathscr{O}$$
$$\mathscr{E} = \varepsilon \cup UUD\mathscr{O}D\mathscr{E}.$$

This grammar implies the following equations:

$$P = 1 + zP + z^2 OP$$
$$O = 1 + zO + z^3 EO^2$$
$$E = 1 + z^2 OE.$$

This system has two possible solutions for P, but only one is holomorphic near the origin, namely

$$P(z) = \frac{2}{1 - z - z^2 + (z^4 - 2z^3 - z^2 - 2z + 1)^{1/2}}.$$

The generating function $G(z)$ for the generalized Catalan numbers is (see A4148 in the OEIS)

$$G(z) = \frac{1 - z + z^2 - (1 - 2z - z^2 - 2z^3 + z^4)^{1/2}}{2z^2},$$

and it is routine to verify that $G(z) = zP(z) + 1$. Therefore $G_{n+1} = [z^n]P(z)$ for $n \ge 0$. $\square$

The following proposition is concerned with *Motzkin numbers* (see A1006 in the OEIS and [DS77]). A *Motzkin path* is like a Dyck path, but includes a “sideways” step S which does not change the height. The nth

Motzkin number M_n is the number of Motzkin paths of length n. The generating function $M = M(z)$ for M_n satisfies the quadratic equation

$$M = 1 + zM + z^2M^2.$$

There are numerous bijections between Motzkin paths and various restricted classes of Dyck paths. Such bijections are often variations of the "folding" map

$$\begin{aligned} UD &\mapsto S \\ DU &\mapsto S \\ UU &\mapsto U \\ DD &\mapsto D, \end{aligned}$$

which in general is not injective, but many restrictions on Dyck paths *make* it injective. For example, this idea shows that the Dyck paths of semilength n with no up-runs longer than 2 are in bijection with the Motzkin paths of length n. We offer a grammatical proof of this fact.

Proposition 24. *The number of Dyck paths of semilength n which avoid up-runs of length* 3 *or more equals the nth Motzkin number M_n.*

Proof. Let $\mathscr{P}$ be the set of such paths. A grammar for $\mathscr{P}$ is

$$\mathscr{P} = \varepsilon \cup UUD\mathscr{P}D\mathscr{P} \cup UD\mathscr{P}.$$

Our grammar implies that

$$P = 1 + zP + z^2P^2.$$

This is the same equation satisfied by the Motzkin generating function, and it is easy to check that $P(z) = M(z)$. $\square$

Proposition 25. *Consider the set of Dyck paths such that no peak or valley has positive, even height. The numbers of such paths of semilength* $2n$ *and* $2n+1$ *are* $\binom{2n-1}{n}$ *and* $\binom{2n}{n}$*, respectively.*

Proof. Let $\mathscr{P}$ denote the set of such paths, and let $\mathscr{O}$ denote the set of all Dyck paths whose peaks and valleys avoid odd heights. These sets satisfy the following grammars

$$\mathscr{P} = \varepsilon \cup U\mathscr{O}D\mathscr{P},$$
$$\mathscr{O} = \varepsilon \cup UU\mathscr{O}DD\mathscr{O}.$$

This grammar can be translated into the following equations:

$$P = 1 + zOP, \text{ and}$$
$$O = 1 + z^2O^2$$

Solving this system for O, we get two solutions for O, but only the following is holomorphic near the origin

$$O = \frac{1-\sqrt{1-4z^2}}{2z^2}.$$

Thus,

$$P = \frac{2z-1-\sqrt{1-4z^2}}{2(2z-1)},$$

and it is easy to check that

$$[z^{2n}]P(z) = \binom{2n-1}{n}, \text{ and}$$
$$[z^{2n+1}]P(z) = \binom{2n}{n}.$$

□

Now, let us define a mapping which allows us to translate restrictions on up-run (respectively, down-run) lengths into restrictions on down-run (respectively, up-run) lengths. Let $\mathscr{P}$ denote the set of all Dyck paths. Define the mapping

$$\phi : \mathscr{P} \to \mathscr{P}, \qquad P \mapsto Q, \tag{9.2}$$

where applying ϕ reverses the order and direction of the steps in P. For example,

$$\phi(UUUDUUDDDDUD) = UDUUUUDDUDDD.$$

It is obvious that $\phi(P)$ must be a Dyck path. Moreover, it is easy to check that ϕ is an involution. Note that the up-runs (respectively, down-runs) in P become down-runs (respectively, up-runs) in $\phi(P)$ of the same length.

Proposition 26. *Let A and B be arbitrary sets of positive integers. The number of Dyck paths of semi-length n which avoid up-runs and down-runs with lengths in A and B, respectively, equals the number of Dyck paths of semi-length n which avoid down-runs with lengths in A and up-runs with lengths in B.*

Proof. Let $\mathscr{P}(A,B)$ be the set of Dyck paths such that no up-run has length in A and no down-run has length in B, and $\mathscr{P}(B,A)$ be the set of Dyck paths such that no up-run has length in B and no down-run has length in A. Then ϕ – defined in equation 9.2 – gives a one-to-one correspondence between the Dyck paths of semi-length n in $\mathscr{P}(A,B)$ and the Dyck paths of semi-length n in $\mathscr{P}(B,A)$. $\square$

Note that ϕ also allows us to translate the grammar of $\mathscr{P}(A,B)$ into the grammar of $\mathscr{P}(B,A)$, as seen in the following section.

9.3 Grammatical families

In this section we provide some explicit grammars for infinite families of restricted Dyck paths. In many cases, such grammars are guaranteed to exist. The reasoning in [EZ20] shows that, for every set of Dyck paths whose peaks, valleys, and up- and down-runs avoid specific arithmetic progressions, we may construct a finite, context-free grammar which generates them. The method implied in [EZ20] to compute these grammars gives no hint as to their *form*, and this is what we try to provide here.

Our first two results are about Dyck paths whose up-run lengths avoid a fixed arithmetic progression $\{Ar+B \mid r \geq 0\}$; each of these is accompanied by a corollary on Dyck paths that avoid down-run lengths in $\{Ar+B \mid r \geq 0\}$. It turns out that when $B < A$, there is a simple context-free grammar for such paths. When $B \geq A$ the situation is more complicated, but we can derive a “grammatical equation” which again leads to a generating function.

Proposition 27. *Let $B < A$ be non-negative integers. The set $\mathscr{P}$ of Dyck paths whose up-run lengths avoid $\{Ar+B \mid r \geq 0\}$ has the unambiguous grammar*

$$\mathscr{P} = \left(\bigcup_{\substack{0 \leq k < A \\ k \neq B}} U^k (D\mathscr{P})^k \right) \cup U^A (\mathscr{P}D)^A \mathscr{P},$$

and therefore

$$P(z) = \left(\sum_{\substack{0 \le k < A \\ k \ne B}} z^k P^k(z) \right) + z^A P^{A+1}(z),$$

where $P(z)$ is the weight-enumerator of $\mathscr{P}$.

Proof. The grammar clearly uniquely parses the empty path, so suppose that $P \in \mathscr{P}$ has length $n > 0$. Then P starts with a up-run of length $k > 0$ for some $k \not\equiv B \mod A$. If $k < A$, then write $P = U^k DW$, where W is a walk from height $k-1$ to height 0 with the same restrictions on up-runs as P. For $0 \le i < k-1$, let D_i indicate the down-step in W which hits the height i for the first time. Then

$$W = P_{k-1} D_{k-2} P_{k-2} D_{k-3} ... P_1 D_0 P_0,$$

where P_i is a Dyck path shifted to height i with the same restrictions on up-runs as P. This uniquely parses P into the case $U^k (D\mathscr{P})^k$ in the grammar.

If the initial up-run has length $k \ge A$, then write $P = U^A W$, where W is a walk from height A to height 0 whose up-run lengths avoid $\{Ar + B \mid r \ge 0\}$. By argument analogous to the previous paragraph, we can decompose W as

$$W = P_A D_{A-1} P_{A-1} D_{A-2} ... P_1 D_0 P_0,$$

where $P_i \in \mathscr{P}$. Thus W is of the form $(\mathscr{P}D)^A \mathscr{P}$, and this uniquely parses P into the final case of the grammar.

We have shown that $\mathscr{P}$ is contained in the language generated by this grammar, and it is easy to see that the first k cases of the grammar are contained in $\mathscr{P}$. The final case, $U^A (\mathscr{P}D)^A \mathscr{P}$, is also contained in the grammar, because concatenating U^A to the beginning of a path does not change the length any of the up-runs modulo A. The different cases are clearly disjoint, so the grammar is also unambiguous. □

Corollary 4. *Let $A, B \in \mathbb{Z}_{\ge 0}$ such that $B < A$. The set $\mathscr{P}$ of Dyck paths avoiding down-run lengths in $\{Ar + B | r \in \mathbb{Z}_{\ge} 0\}$ has the unambiguous grammar*

$$\mathscr{P} = \left(\bigcup_{\substack{0 \le k < A \\ k \ne B}} (\mathscr{P}U)^k D^k \right) \cup \mathscr{P}(U\mathscr{P})^A D^A$$

and therefore

$$P(z) = \left(\sum_{\substack{0 \le k < A \\ k \ne B}} z^k P^k(z) \right) + z^A P^{A+1}(z),$$

where $P(z)$ is the weight-enumerator of $\mathscr{P}$.

Proof. Let ϕ be the involution defined in equation 9.2, and let $\mathscr{Q}$ be the set of Dyck paths avoiding up-run lengths in $\{Ar+B|r\in\mathbb{Z}_{\geq}0\}$. By proposition 27,

$$\mathscr{Q}=\bigcup_{\substack{0\leq k<A\\k\neq B}}U^k(D\mathscr{Q})^k\cup U^A(\mathscr{Q}D)^A\mathscr{Q}.$$

Since

$$\begin{aligned}\phi(\mathscr{Q})&=\mathscr{P},\\ \phi(U^k(D\mathscr{Q})^k)&=(\mathscr{P}U)^kD^k,\text{ for all }0\leq k<A,\text{ and}\\ \phi(U^A(\mathscr{Q}D)^A\mathscr{Q})&=\mathscr{P}(U\mathscr{P})^AU^A,\end{aligned}$$

ϕ translates the grammar of $\mathscr{Q}$ into the desired grammar for $\mathscr{P}$. □

Proposition 28. *Let $A\leq B$ be nonnegative integers. The set $\mathscr{P}$ of Dyck paths avoiding up-run lengths in $\{Ar+B\mid r\geq 0\}$ satisfies the "grammatical equation"*

$$\mathscr{P}\cup U^B(D\mathscr{P})^B=\left(\bigcup_{0\leq k<A}U^k(D\mathscr{P})^k\right)\cup\,U^A(\mathscr{P}D)^A\mathscr{P},$$

and therefore

$$P(z)+z^BP(z)^B=\left(\sum_{0\leq k<A}z^kP^k(z)\right)+z^AP^{A+1}(z),$$

where $P(z)$ is the weight-enumerator of $\mathscr{P}$.

Note that the right-hand side is nearly identical to proposition 6; the difference being that we can get paths in $U^B(D\mathscr{P})^B$, which we will show below.

Proof. If P is a path in $\mathscr{P}$, then we can uniquely parse P into a case of the right-hand side by the same argument given in the previous proposition. Note that

$$\begin{aligned}U^B(D\mathscr{P})^B&=U^AU^{B-A}(D\mathscr{P})^B\\&=U^A\{U^{B-A}(D\mathscr{P})^{B-A}\}(D\mathscr{P})^A\\&=U^A[\{U^{B-A}(D\mathscr{P})^{B-A}\}D(\mathscr{P}D)^{A-1}]\mathscr{P}.\end{aligned}$$

The expression in brackets, $U^{B-A}(D\mathscr{P})^{B-A}$, is in $\mathscr{P}$, which shows that $U^B(D\mathscr{P})^B$ is contained in $U^A(\mathscr{P}D)^A\mathscr{P}$.

Conversely, it remains to show that the left-hand side is *all* that the right-hand side can generate. $\bigcup_{0\leq k<A}U^k(D\mathscr{P})^k$

is contained in $\mathscr{P}$ as in the previous proposition. For $W \in U^A(\mathscr{P}D)^A\mathscr{P}$, write

$$W = U^A P_1 D \dots P_A D P_{A+1}.$$

Let ℓ be the length of the initial up-run in P_1. If $\ell \not\equiv B \pmod{A}$, then W contains no up-runs of lengths in $\{Ar+B \mid r \geq 0\}$ and is a path in $\mathscr{P}$. If $\ell \equiv B \pmod{A}$, then $\ell \leq B-A$. If $\ell < B-A$ then the initial run of W has length less than B. Thus, W contains no up-runs of lengths in $\{Ar+B \mid r \geq 0\}$. For $\ell = B-A$, let D_i denote the first time W steps down to height i for $A < i < B$ and write

$$\begin{aligned}
W &= U^A P_1 D \dots P_A D P_{A+1} \\
&= U^A (U^{B-A} D_{B-1} W_{B-1} \dots D_A W_A) D P_2 D \dots P_A D P_{A+1} \\
&= U^B D_{B-1} W_{B-1} \dots D_A W_A D P_2 D \dots P_A D P_{A+1}.
\end{aligned}$$

W_i is Dyck path shifted to height i by the definition of D_i. Hence, $W \in U^B(D\mathscr{P})^B$. □

Corollary 5. *Let $A, B \in \mathbb{Z}_{\geq 0}$ such that $B \geq A$. The set $\mathscr{P}$ of Dyck paths avoiding down-run lengths in $\{Ar+B | r \in \mathbb{Z}_{\geq}0\}$ satisfies the grammatical equation*

$$\mathscr{P} \cup (\mathscr{P}U)^B D^B = \left(\bigcup\nolimits_{0 \leq k < A} (\mathscr{P}U)^k D^k\right) \cup \mathscr{P}(U\mathscr{P})^A D^A.$$

and therefore

$$P(z) + z^B P^B(z) = \left(\sum_{0 \leq k < A} z^k P^k(z)\right) + z^A P^{A+1}(z).$$

where $P(z)$ is the weight-enumerator of $\mathscr{P}$.

Proof. Let ϕ be the involution defined in equation 9.2, and let $\mathscr{Q}$ be the set of Dyck paths avoiding up-run lengths in $\{Ar+B | r \in \mathbb{Z}_{\geq}0\}$. Applying ϕ to each clause of the grammar of $\mathscr{Q}$ given in proposition 28, we get

$$\mathscr{P} \cup (\mathscr{P}U)^B D^B = \left(\bigcup\nolimits_{0 \leq k < A} (\mathscr{P}U)^k D^k\right) \cup \mathscr{P}(U\mathscr{P})^A D^A,$$

as desired. □

Proposition 29. *Let $r \in \mathbb{Z}^+$. The set $\mathscr{P}$ of Dyck paths avoiding ascending and descending runs of lengths in $\{1, \dots, r\}$ satisfies the grammatical equation*

$$\mathscr{P} \cup UD\mathscr{P} = \varepsilon \cup U^{r+1} D^{r+1} \mathscr{P} \cup U\mathscr{P}D\mathscr{P}.$$

and therefore

$$P(z) + zP(z) = 1 + z^{r+1}P(z) + zP^2(z),$$

where $P(z)$ is the weight-enumerator of $\mathscr{P}$.

Proof. If $P \in \mathscr{P}$ is the empty path, then the grammar uniquely parses P. Otherwise, $P \in \mathscr{P}$ must begin with an ascending run of length $\ell > r$. If $\ell = r+1$, then clearly U^{r+1} must be immediately followed by the descending run D^{r+1}, and P is uniquely parsed into the case $U^{r+1}D^{r+1}\mathscr{P}$.

If $\ell > r+1$, then let D_0 denote the step where P returns to height 0 for the first time and write

$$P = UP_1D_0P_2.$$

It is obvious that $P_2 \in \mathscr{P}$ and P_1 is a Dyck path shifted to height 1. By restrictions on P, the final descending run in P_1 must have length $L \geq r$. If $L = r$ then the preceding ascending run ends at height $r+1$. But the ascending runs in P must have length of at least $r+1$, and hence P_1 hits height 0, contradicting the definition of D_0. From here, it is clear that P_1 has the same restrictions on ascending and descending runs as P. Thus, P is uniquely parsed into the case $U\mathscr{P}D\mathscr{P}$.

Since it is trivial that $UD\mathscr{P}$ is contained in $U\mathscr{P}D\mathscr{P}$, we have shown that the left-hand side of the given equation is generated by the right-hand side. It is also obvious that the cases defined on the right-hand side are disjoint and that $\varepsilon \cup U^{r+1}D^{r+1}\mathscr{P}$ is contained in $\mathscr{P}$. A path $UP_1DP_2 \in U\mathscr{P}D\mathscr{P}$ is contained in $UD\mathscr{P}$ if P_1 is the empty path and $\mathscr{P}$ otherwise. Thus, $\mathscr{P}$ satisfies the given grammatical equation.

□

Proposition 30. *Let $m,n \in \mathbb{Z}^+$. The set $\mathscr{P}$ of Dyck paths avoiding ascending runs of lengths in $\{1,...,m\}$ and descending runs of lengths in $\{1,...,n\}$ satisfies the grammatical equation*

$$\mathscr{P} \cup UD\mathscr{P} = \varepsilon \cup U\mathscr{P}D\mathscr{P} \cup U^{m+1}D^{n+1}(\mathscr{P}D)^{m-n}\mathscr{P}, \text{ if } m \geq n \tag{9.3}$$

$$\mathscr{P} \cup \mathscr{P}UD = \varepsilon \cup \mathscr{P}U\mathscr{P}D \cup \mathscr{P}(U\mathscr{P})^{n-m}U^{m+1}D^{n+1}, \text{ if } m \leq n. \tag{9.4}$$

Proof. We have already shown that this statement is true for $m = n$. Suppose $m > n$. If $P \in \mathscr{P}$ is the empty path, then the grammar uniquely parses P. Otherwise, P must begin with an ascending run of length $\ell > m$. If $\ell = m+1$ then U^{m+1} is followed by a descending chain of length of at least $n+1$. Let D_i denote the first time P returns to height i for $0 \leq i \leq m-n-1$, and write

$$P = U^{m+1}D^{n+1}P_{m-n}D_{m-n-1}...P_1D_0P_0.$$

It is obvious that P_i is a Dyck path, shifted to height i, that has the same restrictions on ascending runs and descending runs (with the exception of the final descending run) as P. Since P_i is a Dyck path, its final descending run must be at least as long as the ascending run preceding it. Thus, P_i is either the empty path or ends with a descending run of length $L > m > n$. Thus, P is uniquely parsed into the case $U^{m+1}D^{n+1}(\mathscr{P}D)^{m-n}\mathscr{P}$.

If $\ell > m+1$ then, letting D_0 denote the first time P returns to height 0, write

$$P = UP_1D_0P_0.$$

Clearly, $P_0 \in \mathscr{P}$, and P_1 is a Dyck path shifted to height 1 and has the same restrictions on ascending runs as P. Using the same argument as for P_i in the previous case, the descending runs in P_1 also have the same restrictions as P. This uniquely parses P into the case $U\mathscr{P}D\mathscr{P}$. Finally, it is obvious that $UD\mathscr{P}$ is contained in $U\mathscr{P}D\mathscr{P}$, so the left-hand side of (1) is generated by the right-hand side.

It is clear that the cases on the right-hand side are disjoint, and the empty path is an element of $\mathscr{P}$. Also, $UP_1DP_2 \in U\mathscr{P}D\mathscr{P}$ is contained in $\mathscr{P}$ if P_1 is not the empty path, and is contained in $UD\mathscr{P}$ otherwise. $U^{m+1}D^{n+1}(\mathscr{P}D)^{m-n}\mathscr{P}$ is contained in $\mathscr{P}$, since all ascending runs clearly avoid restrictions on $\mathscr{P}$ and the descending runs are formed by concatenating down-steps to descending runs of length of at least $n-1$. Thus, we have proved the grammar for the case $m \geq n$.

Now assume that $n \geq m$. Applying the involution ϕ from equation 9.2, we can directly translate the grammar 9.3 into the desired grammar 9.4. $\square$

Proposition 31. *Let $r, k \in \mathbb{Z}^+$ and let $\mathscr{P}$ be the set of Dyck paths avoiding ascending runs of length $\{1, ..., r\}$ and descending runs of length $\{k+1, ..., r\}$. Then the 'grammar' of $\mathscr{P}$ is*

$$\mathscr{P} \cup UD\mathscr{P} \cup U^{r+1}D^k(D\mathscr{P})^{r+1-k} = \varepsilon \cup U\mathscr{P}D\mathscr{P} \cup U^{r+1}D^{r+1}\mathscr{P} \cup U^{r+1}(DP)^{r+1}$$

Proof. If $P \in \mathscr{P}$ is the empty path, then the grammar uniquely parses P. Otherwise, P begins an ascending run of length $\ell > r$, and we can deduce that it also ends with a descending run of length $L > r$. If $\ell > r+1$, then let D_0 denote the first time that P returns to the $x-$axis and write

$$P = UP_1D_0P_0.$$

It is easy to see that P_0 is a path in $\mathscr{P}$ and P_1 is a Dyck path shifted to height 1. The initial ascending run in P_1 has length $\ell-1>r$. Thus, all ascending runs in P_1 have length of at least $r+1$ and, since P_1 is a shifted Dyck path, the final descending run in P_1 must also have length of at least $r+1$. From here, it is easy to see that P_1 has the same restrictions on ascending and descending runs as P. P is therefore uniquely parsed into the case $U\mathscr{P}D\mathscr{P}$.

Suppose $\ell=r+1$. Let D_i be the step where P returns to height i for the first time and write

$$P=U^{r+1}D_rP_r\ldots D_0P_0.$$

P_i is a Dyck path for all i and, if P_i is not the empty path, it must end with a descending run of length $r+1$ by restrictions on ascending runs. Thus P_i is a path in $\mathscr{P}$, and P is parsed into the case $U^{r+1}(D\mathscr{P})^{r+1}$.

It is trivial that $UD\mathscr{P}$ is contained in $U\mathscr{P}D\mathscr{P}$ and $U^{r+1}D^k(D\mathscr{P})^{r+1-k}$ is contained in $U^{r+1}(D\mathscr{P})^{r+1}$. Thus, the left-hand side is generated by the right-hand side. Note that, on the left-hand side,

$$UD\mathscr{P}\cap\mathscr{P}=UD\mathscr{P}\cap U^{r+1}D^k(D\mathscr{P})^{r+1-k}=\emptyset,$$

however

$$\mathscr{P}\cap U^{r+1}D^k(D\mathscr{P})^{r+1-k}=U^{r+1}D^{r+1}\mathscr{P}.$$

Looking at the right-hand side, it is clear that $\varepsilon, U\mathscr{P}D\mathscr{P}$, and $U^{r+1}(D\mathscr{P})^{r+1}$ are disjoint, and $U^{r+1}D^{r+1}\mathscr{P}$ is contained in $U^{r+1}(D\mathscr{P})^{r+1}$. Note that this resolves the issue of double counting paths in $U^{r+1}D^{r+1}\mathscr{P}$ on the left-hand side. Thus, all that remains to show is that all the paths generated by the right-hand side are contained in the left-hand side.

The path UP_1DP_0 in $U\mathscr{P}D\mathscr{P}$ is clearly in $\mathscr{P}$ if P_1 is not the empty path and in $UD\mathscr{P}$ otherwise. For W in $U^{r+1}(D\mathscr{P})^{r+1}$, write

$$W=U^{r+1}D_rP_r\ldots D_1P_1D_0P_0.$$

Choose the smallest i such that P_{r-i} is not the empty path or, if no such i exists, set $i=r$. Then the first descending run in W has length $i+1$. If $i\geq k$ then W is an element of $U^{r+1}D^k(D\mathscr{P})^{r+1-k}$. Otherwise, we claim that W is a path in $\mathscr{P}$. It is clear that W is a Dyck path and we have seen that nonempty $P_j\in\mathscr{P}$ must end in a descending run of length of at least $r+1$. Thus, we only need to show that the first descending run in W follows the restrictions in $\mathscr{P}$. This is clearly true since $i<k$. Hence $W\in\mathscr{P}$, and $\mathscr{P}$ satisfies the grammatical equation as desired. $\square$

9.4 Conclusion

We have given several grammatical proofs of various combinatorial results about restricted Dyck paths and established some infinite families of grammars. Our methods work because we are able to derive *context-free grammars* describing certain restricted classes Dyck paths, namely when our restrictions involved sets of arithmetic progressions.

It is natural to ask if context-free grammars exist for other types of restrictions. Parikh's theorem [Par66] states that the set of lengths of any context-free language is the union of finitely-many arithmetic progressions, so it seems likely that restrictions involving arithmetic progressions are *essentially* all that can be done. However, addressing this question in full is beyond our current scope.

Bibliography

[Abe14] Francine F. Abeles. "Chió's and Dodgson's determinantal identities". In: *Linear Algebra and its Applications* 454 (2014), pp. 130–137.

[Abl21] Jakob Ablinger. "Extensions of the AZ-Algorithm and the Package MultiIntegrate". In: *Anti-Differentiation and the Calculation of Feynman Amplitudes*. Ed. by Johannes Blümlein and Carsten Schneider. Cham: Springer International Publishing, 2021, pp. 35–61. URL: https://doi.org/10.1007/978-3-030-80219-6_2.

[Apé79] R. Apéry. "Irrationalité de $\zeta(2)$ et $\zeta(3)$". In: *Astérisque* 61 (1979), pp. 11–13.

[AR80] Krishna Alladi and Michael L. Robinson. "Legendre polynomials and irrationality". In: *J. Reine Angew. Math.* 318 (1980), pp. 137–155.

[AZ06] Moa Apagodu and Doron Zeilberger. "Multi-variable Zeilberger and Almkvist-Zeilberger algorithms and the sharpening of Wilf-Zeilberger Theory". In: *Adv. Appl. Math.* 37 (2006). https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/multiZ.html, pp. 139–152.

[AZ90] Gert Almkvist and Doron Zeilberger. "The method of differentiating under the integral sign". In: *J. Symbolic Computation* 10 (1990). https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/duis.html, pp. 571–591.

[BB87] J. Borwein and P. Borwein. *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*. Hoboken: Wiley-Interscience, 1987.

[BBP97] David Bailey, Peter Borwein, and Simon Plouffe. "On the rapid computation of various polylogarithmic constants". In: *Mathematics of Computation* 66.218 (1997), pp. 903–913.

[Beu79] Frits Beukers. "A note on the irrationality of $\zeta(2)$ and $\zeta(3)$". In: *Bull. Lond. Math. Soc.* 11.3 (1979), pp. 268–272.

[BGW15] D. Birmajer, J.B. Gil, and M.D. Weiner. *Linear recurrence sequences with indices in arithmetic progression and their sums*. 2015. URL: https://arxiv.org/abs/1505.06339.

[BM04] G. Boros and V. Moll. *Irresistible Integrals: Symbolics, Analysis and Experiments in the Evaluation of Integrals*. Cambridge: Cambridge University Press, 2004.

[Bos+17] Alin Bostan, Frédéric Chyzak, Mark van Hoeij, Manuel Kauers, and Lucien Pech. "Hypergeometric Expressions for Generating Functions of Walks with Small Steps in the Quarter Plane". In: *European Journal of Combinatorics* 61 (2017), pp. 242–275.

[CS21] Marc Chamberland and Armin Straub. "Apéry limits: experiments and proofs". In: *The American Mathematical Monthly* 128.9 (2021), pp. 811–824.

[DKZ22] Robert Dougherty-Bliss, Christoph Koutschan, and Doron Zeilberger. "Tweaking the Beukers integrals in search of more miraculous irrationality proofs á la Apéry". In: *The Ramanujan Journal* 58.3 (2022), pp. 973–994.

[Dou22] Robert Dougherty-Bliss. "Integral Recurrences from A to Z". In: *The American Mathematical Monthly* 129 (9 2022), pp. 805–815.

[DS77] R. Donaghey and L.W. Shapiro. "Motzkin numbers". In: *J. Combin. Theory Ser. A* 23.3 (1977), pp. 291–301.

[EHE] User E.H.E. *Math Stack Exchange Question 1084897*. https://math.stackexchange.com/questions/1084897. Accessed 2024-02-05.

[Elk88] Noam Elkies. "On $A^4 + B^4 + C^4 = D^4$". In: *Mathematics of Computation* 51 (1988), pp. 825–835.

[ELY03] S. Eu, S. Liu, and Y. Yeh. "Dyck paths with peaks avoiding or restricted to a given set". In: *Stud. Appl. Math.* 111.4 (2003), pp. 453–465.

[Eul55] Leonhard Euler. *Institutiones calculi differentialis*. 1755.

[EZ00] Anne E. Edlin and Doron Zeilberger. "The Goulden-Jackson Cluster method For cyclic Words". In: *Advances in Applied Mathematics* 25 (2000). https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/cgj.html, pp. 228–232.

[EZ20] S.B. Ekhad and Doron Zeilberger. *Automatic counting of restricted Dyck paths via (numeric and symbolic) dynamic programming*. 2020. URL: https://arxiv.org/abs/2006.01961.

[FBA99] Helaman Ferguson, David Bailey, and Steve Arno. "Analysis of PSLQ, an integer relation finding algorithm". In: *Mathematics of Computation* 68.225 (1999), pp. 351–369.

[Fro18] R. Frontczak. “Sums of powers of Fibonacci and Lucas numbers: A new bottom-up approach”. In: *Notes on Number Theory and Discrete Mathematics* 24.2 (2018), pp. 94–103.

[FS09] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[GCL92] K.O. Geddes, S.R. Czapor, and G. Labahn. *Algorithms for Computer Algebra*. Springer Science & Business Media, 1992.

[Gos78] R. Gosper. “Decision procedure for Indefinite Hypergeometric Summation”. In: *Proceedings of the National Academy of Sciences* 75.1 (1978), pp. 40–42.

[GQ07] H.W. Gould and J. Quaintance. “A Linear Binomial Recurrence and the Bell Numbers and Polynomials”. In: *Applicable Analysis and Discrete Mathematics* (2007), pp. 371–385.

[Gur90] Stanley Gurak. “Pseudoprimes for Higher-Order Linear Recurrence Sequences”. In: *Mathematics of Computation* 55 (1990), pp. 783–813.

[HH17] N. Hein and J. Huang. “Modular Catalan numbers”. In: *European J. Combin.* 61 (2017), pp. 197–218.

[Hol20] Stephan Holger. *Millions of Perrin pseudoprimes including a few giants*. https://arxiv.org/abs/2002.03756. 2020.

[Hul+22] Thomas C Hull, Manuel Morales, Sarah Nash, and Natalya Ter-Saakov. “Maximal origami flip graphs of flat-foldable vertices: properties and algorithms”. In: *Journal of Graph Algorithms and Applications* 26.4 (2022).

[Jac33] Carl Jacobi. “De binis quibuslibet functionibus homogeneis secundi ordinis per substitutiones lineares in alias binas transformandis, quae solis quadratis variabilium constant; una cum variis theorematis de transformationeet determinatione integralium multiplicium”. In: *Journal für die reine und angewandte Mathematik* 12 (1833), pp. 1–69.

[JM91] J. P. Jones and Y. V. Matiyasevich. “Proof of recursive unsolvability of Hilbert’s tenth problem”. In: *Amer. Math. Monthly* 98 (1991), pp. 689–709.

[Kau13] Manuel Kauers. “The Holonomic Toolkit”. In: *Computer Algebra in Quantum Field Theory*. Springer, Vienna, 2013, pp. 119–144.

[Kau23] Manuel Kauers. *D-finite Functions*. Springer, 2023.

[KK22] Manuel Kauers and Christoph Koutschan. “Guessing with Little Data”. In: *Proceedings of ISSAC’22*. 2022, pp. 83–90.

[KK23] Manuel Kauers and Christoph Koutschan. “Some D-finite and some possibly D-finite sequences in the OEIS”. In: *Journal of Integer Sequences* 23.4 (2023), p. 5.

[Kou09] Christoph Koutschan. “Advanced applications of the holonomic systems approach”. PhD thesis. Research Institute for Symbolic Computation (RISC), Johannes Kepler University, Linz, Austria, 2009. URL: http://www.koutschan.de/publ/Koutschan09/thesisKoutschan.pdf.

[Kou10] Christoph Koutschan. *HolonomicFunctions (User’s Guide)*. Tech. rep. 10-01. RISC Report Series, University of Linz, Austria, Jan. 2010. URL: http://www.risc.uni-linz.ac.at/research/combinat/software/HolonomicFunctions/.

[KP10] Manuel Kauers and Peter Paule. *The Concrete Tetrahedron: Symbolic sums, Recurrence Equations, Generating Functions, Asymptotic Estimates*. Springer, 2010.

[Kra15] Christian Krattenthaler. “Lattice Path Enumeration”. In: *Handbook of Combinatorics*. Ed. by Miklos Bona. Taylor & Francis, 2015, pp. 589–678.

[Kra99] Christian Krattenthaler. “Advanced Determinant Calculus”. In: *Seminaire Lotharingien Combinatoire* 42.B42q (1999).

[Lay77] J.W. Layman. “Certain general binomial-Fibonacci sums”. In: *Fibonacci Quart* 15.3 (1977), pp. 362–366.

[Mac95] Ian G. Macdonald. *Symmetric Functions and Hall Polynomials*. Second. Oxford: Clarendon Press, 1995.

[Mat93] Yuri V. Matiyasevich. *Hilbert’s Tenth Problem*. The MIT Press, 1993.

[Mel21] Stephen Melczer. *An Invitation to Analytic Combinatorics*. Springer, 2021.

[Mel99] R. Melham. “Sums involving Fibonacci and Pell numbers”. In: *Portugaliae Mathematica* 56.3 (1999), pp. 309–318.

[Mun] Robert Munafo. *RILYBOT Inverse Equation Solver*. https://www.mrob.com/pub/ries/index.html. Accessed 2024-02-05.

[Nem+97] I. Nemes, M. Petkovšek, H. Wilf, and D. Zeilberger. “How to do Monthly problems with your computer”. In: *Amer. Math. Monthly* 104.6 (1997), pp. 505–519.

[Nes16] Yu. V. Nesterenko. “On Catalan’s constant”. In: *Proceedings of the Steklov Institute of Mathematics* 292 (2016), pp. 153–170.

[OEI24] OEIS Foundation Inc. *The On-Line Encyclopedia of Integer Sequences*. 2024. URL: http://oeis.org.

[Par66] R.J. Parikh. "On context-free languages". In: *J. ACM* 13.4 (1966), pp. 570–581.

[Per99] Raoul Perrin. "Item 1484". In: *L'Intermédiare des math* 6 (1899), pp. 76–77.

[Poo79] Alf van der Poorten. "A proof that Euler missed... Apéry's proof of the irrationality of $\zeta(3)$". In: *Math. Intelligencer* 1 (1979), pp. 195–203.

[Pud20] Lara K Pudwell. "From Permutation Patterns to the Periodic Table". In: *Notices of the American Mathematical Society* 67.7 (2020), pp. 994–1001.

[PW01] P. Peart and W.J. Woan. "Dyck paths with no peaks at height k". In: *J. Integer Seq.* 4.1 (2001).

[PW13] Robin Pemantle and Mark C. Wilson. *Analytic Combinatorics in Several Variables*. Cambridge, 2013.

[PWZ97] Marko Petkovšek, Herb Wilf, and Doron Zeilberger. $A = B$. AK Peters, 1997.

[RIS] RISC. *ErgoSum Mathematica Package*. https://www3.risc.jku.at/research/combinat/software/ergosum/. Accessed 2024-02-07.

[RT07] Adrian Rice and Eve Torrence. ""Shutting up like a telescope": Lewis Carroll's "Curious" Condensation Method for Evaluating Determinants". In: *The College Mathematics Journal* 38.2 (2007), pp. 85–95.

[RV01] Georges Rhin and Carlo Viola. "The group structure of $\zeta(3)$". In: *Acta Arithmetica* 97 (2001), pp. 269–293.

[Ste96] Ian Stewart. "Tales of a Neglected Number". In: *Mathematical Recreations, Scientific American* 274.6 (1996), pp. 102–103.

[Sto23] David Stoutemyer. "How to hunt wild constants". In: *Maple Transactions* 3.1 (2023).

[SW79] P.R. Stein and M.S. Waterman. "On some new sequences generalizing the Catalan and Motzkin numbers". In: *Discrete Math.* 26.3 (1979), pp. 261–272.

[SZ94] Bruno Salvy and Paul Zimmermann. "Gfun: a Maple package for the manipulation of generating and holonomic functions in one variable". In: *ACM Transactions on Mathematical Software (TOMS)* 20.2 (1994), pp. 163–177.

[Tef04] A. Tefera. "What is... a Wilf–Zeilberger pair". In: *AMS Notices* 57.4 (2004), pp. 508–509.

[Vat22] Vince Vatter. "Social Distancing, Primes, and Perrin Numbers". In: *Math Horizons* 29.1 (2022). https://sites.math.rutgers.edu/~zeilberg/akherim/vatter23.pdf.

[Wik] Wikipedia. *Perrin Number*. URL: https://en.wikipedia.org/wiki/Perrin_number.

[Wil95] Andrew Wiles. "Modular Elliptic Curves and Fermat's Last Theorem". In: *Annals of Mathematics Second Series* 141.3 (May 1995), pp. 443–551.

[WZ85] Jet Wimp and Doron Zeilberger. "Resurrecting the asymptotics of linear recurrences". In: *J. Math. Anal. Appl.* 111.1 (1985), pp. 162–176.

[Zei03] Doron Zeilberger. "Computerized deconstruction". In: *Adv. Applied Math.* 30 (2003). https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/derrida.html, pp. 633–654.

[Zei07] Doron Zeilberger. "The Holonomic Ansatz II. Automatic Discovery(!) And Proof(!!) of Holonomic Determinant Evaluations". In: *Annals of Combinatorics* 11 (2007), pp. 241–247.

[Zei13] Doron Zeilberger. "The C-finite ansatz". In: *The Ramanujan Journal* 31.1 (2013), pp. 23–32. URL: https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/cfinite.html.

[Zei14] Doron Zeilberger. *Two Motivated Concrete Proofs (much better than the usual one) that the Square-Root of 2 is Irrational*. 2014. URL: https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/sqrt2.html.

[Zei24a] Doron Zeilberger. *Maple Packages and Programs*. https://sites.math.rutgers.edu/~zeilberg/programs.html. Accessed 2024-02-05. 2024.

[Zei24b] Doron Zeilberger. *Searching for Apéry-style miracles [using, inter-alia, the amazing Almkvist-Zeilberger algorithm]*. 2024. URL: https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/apery.html.

[Zei90a] Doron Zeilberger. "A Holonomic Systems Approach To Special Functions Identities". In: *Journal of Computational and Applied Mathematics* 32 (1990), pp. 321–368.

[Zei90b] Doron Zeilberger. "A Holonomic systems approach to special functions identities". In: *J. of Computational and Applied Math.* 32 (1990). https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/holonomic.html, pp. 321–368.

[Zei95] Doron Zeilberger. "Three recitations on holonomic systems and hypergeometric series". In: *J. Symb. Comput.* 20.5/6 (1995), pp. 699–724.

[Zud] Wadim Zudilin. *The birthday boy problem*. URL: https://arxiv.org/abs/2108.06586.

[Zud03] W Zudilin. "An Apéry-like Difference Equation for Catalan's Constant". In: *The Electronic Journal of Combinatorics* 10.1 (2003).

[Zud04] Wadim Zudilin. “Arithmetic of linear forms involving odd zeta values”. In: *J. Théorie Nombres Bordeaux* 16 (2004), pp. 251–291. URL: https://arxiv.org/abs/math/0206176.

[ZZ20] Doron Zeilberger and Wadim Zudilin. “The irrationality measure of π is at most 7.103205334137. . . ” In: *Mosc. J. Comb. Number Theory* 9.4 (2020), pp. 407–419. URL: https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/pimeas.html.

[ZZ21] Doron Zeilberger and Wadim Zudilin. “Automatic discovery of irrationality proofs and irrationality measures”. In: *International Journal of Number Theory* 17.03 (2021), pp. 815–825.