



École polytechnique

LAB RESEARCH PROJECT

On the integrality of some P-recursive sequences

Student:

Anastasia Matveeva, École polytechnique

Advisor:

Alin Bostan, Inria Saclay

February-May 2025

Acknowledgements

I am deeply grateful to Alin Bostan for his guidance, support, and for introducing me to computer algebra in the first place. The many insightful conversations we had were truly what made this project possible.

My warm thanks to Florian Fürnsinn and Sergey Yurkevich for illuminating discussions and valuable comments, and also for making my holidays in Vienna all the more enjoyable.

Finally, I would like to thank my friends and family for listening — with admirable patience — to every detail of the writing process.

Abstract

This project investigates the arithmetic nature of P-recursive sequences through the lens of their D-finite generating functions. Building on classical tools from differential algebra, we revisit the integrality criterion for Motzkin-type sequences due to Klazar and Luca, and propose a unified method for analysing global boundedness and algebraicity within a broader class of holonomic sequences. The central contribution is an algorithm that determines whether all, none, or a one-dimensional family of solutions to certain second-order recurrences are globally bounded. This approach generalizes earlier *ad hoc* methods and applies successfully to several well-known sequences from the [OEIS](#).

Contents

1	Introduction	4
2	Integrality analysis of Motzkin-type sequences	4
2.1	Approach of Klazar and Luca	5
2.2	Discussion	5
3	Arithmetic outlook	8
4	Bridging global boundedness and algebraicity	10
4.1	Hypergeometric case	10
4.2	Second-order recurrences	12
4.2.1	Deciding algebraicity of I_1, I_2	15
4.2.2	Transcendence in both components.	21
4.2.3	Examples	23
5	Integrality revisited	25
6	Integrality criteria for small Apéry numbers	26
6.1	Proof 1	27
6.2	Proof 2	28
7	Concluding remarks	29
	References	30

1 Introduction

Definition 1.1. A sequence $(u_n)_{n \in \mathbb{N}}$ of rational numbers is called *P-recursive* (or *P-finite*, or *holonomic*) if it satisfies a linear (homogeneous) recurrence relation with coefficients in $\mathbb{Q}[n]$.

Definition 1.2. A power series $f \in \mathbb{Q}[[x]]$ is called *D-finite* (“differentially finite”, or “holonomic”) if it satisfies a linear (homogeneous) differential equation with coefficients in $\mathbb{Q}[x]$.

Theorem 1.3. ([Sta80, Theorem 1.5]) *A power series is D-finite if and only if its coefficient sequence is P-recursive.*

Definition 1.4. A power series $f(x) = \sum_{n \geq 0} a_n x^n \in \mathbb{Q}[[x]]$ is called *almost integral* if there exists $C \in \mathbb{Z}^*$ such that $f(Cx) - f(0) \in \mathbb{Z}[[x]]$. Equivalently, if there exists $C \in \mathbb{Z}^*$ such that $C^n a_n \in \mathbb{Z}$ for all $n \geq 1$.

Definition 1.5. A power series $f \in \mathbb{Q}[[x]]$ is called *globally bounded* if it is almost integral and has a nonzero radius of convergence in $\mathbb{C}$.

P-recursive sequences appear in many combinatorics problems, as well as in number theory and the theory of special functions. Whether a given P-recursive sequence consists of integers only is a deceptively simple question, which is often approached with arguments tailored for particular sequences. Our starting point is the family of Motzkin-type sequences (defined properly in Section 2), that is, of sequences that satisfy

$$(n + 2)m_n = (2n + 1)m_{n-1} + (3n - 3)m_{n-2}.$$

with arbitrary initial values $m_0, m_1 \in \mathbb{Q}$.

In Section 2, we discuss the approach of Klazar and Luca to the integrality of Motzkin-type sequences and provide an algebraic argument for the integrality of Motzkin numbers. Section 3 generalizes this to a full characterisation of algebraicity and global boundedness for the entire Motzkin-type family. In Section 4, we present an algorithm that decides these properties for an even larger family of second-order P-recursive sequences with polynomial coefficients of degree 1, under certain assumptions on the recurrence coefficients. The algorithm is then applied to several sequences from the OEIS, including large Schröder numbers and central trinomial coefficients. The integrality question is explored in Section 5. Finally, in Section 6, we derive integrality criteria for small Apéry numbers, by adapting different approaches from the literature.

2 Integrality analysis of Motzkin-type sequences

A *Motzkin-type sequence* $M(\mu, \lambda) = (m_n(\mu, \lambda))_n$ is defined in [KL05] as a sequence of rational numbers that satisfies the recurrence:

$$(n + 2)m_n(\mu, \lambda) = (2n + 1)m_{n-1}(\mu, \lambda) + (3n - 3)m_{n-2}(\mu, \lambda). \quad (1.1)$$

with initial values $m_0(\mu, \lambda) = \mu$, $m_1(\mu, \lambda) = \lambda$. By abuse of notation, we sometimes write m_n instead of $m_n(\mu, \lambda)$ when it is clear which initial values are considered.

In particular, $M(1, 1)$ is the sequence of Motzkin numbers (A001006 in the OEIS). These numbers have a few interesting combinatorial interpretations, the most well-known being in terms of lattice paths. Motzkin numbers count the number of paths in the Cartesian plane from $(0, 0)$ to $(n, 0)$ that never dip below the horizontal axis and consist of up $(1, 1)$, down $(1, -1)$, and level $(1, 0)$ steps — known as *Motzkin paths*. They are also closely related to Catalan numbers, which count the special case where level steps are not allowed (*Dyck paths*).

In [KL05] Klazar and Luca prove that $M(\mu, \lambda)$ is integral, that is, consists of integers only, if and only if $\mu, \lambda \in \mathbb{Z}$ and $\mu = \lambda$. In this section we provide a summary of their approach and discuss its limitations.

2.1 Approach of Klazar and Luca

One starts by deducing the integrality of $M(\mu, \mu) = \mu M(1, 1)$ for $\mu \in \mathbb{Z}$. This is nothing more than a straightforward corollary of the integrality of $M(1, 1)$. If $\mu \notin \mathbb{Z}$, then $M(\mu, \mu) \notin \mathbb{Z}[[x]]$ as its first coefficient μ is already not an integer. Regarding the case $M(\mu, \lambda)$ with $\mu \neq \lambda$, the following result holds, where $P(k)$ denotes the largest prime factor of $k \in \mathbb{Z}$ with the convention that $P(0) = P(\pm 1) = 1$.

Theorem 2.1. ([KL05], Theorem 1). *Let $M(\mu, \lambda) = (m_n(\mu, \lambda))_{n \geq 0} = (m_n)_{n \geq 0}$ be any Motzkin-type sequence of rational numbers with $m_0 = \mu \neq \lambda = m_1$. For any n write $m_n = \frac{a_n}{b_n}$, where a_n and b_n are coprime integers with $b_n \geq 1$. Then $\limsup_{n \rightarrow \infty} P(b_n) = \infty$.*

In other words, if $\mu \neq \lambda$, then for infinitely many primes p there exists $n \in \mathbb{N}$ such that $m_n(\mu, \lambda)$ has a denominator divisible by p . Moreover, it can be shown that for p large enough it is in fact $m_{p-2}(\mu, \lambda)$ that has a denominator divisible by p .

Below is a concise summary of the proof of Theorem 2.1. We refer to [KL05] for more details.

Proof strategy for Theorem 2.1. It suffices to exhibit a single pair (α, β) with $\alpha \neq \beta$ such that $M(\alpha, \beta)$ contains terms whose denominators are divisible by arbitrarily large primes, since the set of solutions to (1.1) is two-dimensional. Indeed, every sequence $M(a, b)$ can be written as a linear combination of $M(1, 1)$ and $M(\alpha, \beta)$. Since $M(1, 1)$ is integral, $M(a, b)$ is almost integral for $a \neq b$ if and only if $M(\alpha, \beta)$ is.

To construct an easy-to-work-with example, consider the shifted generating power series $M(x) = \sum_{n \geq 0} m_n x^{n+2}$ and complete it to a power series $S(x) = a + bx + M(x)$ that satisfies an inhomogeneous linear differential equation of order 1. With a clever choice of a, b, m_0, m_1 , this differential equation rewrites as

$$gS' - \frac{1}{2}g'S = g, \quad \text{where } g(x) = 1 - 2x - 3x^2. \quad (1.2)$$

This equation admits the solution

$$S(x) = \sqrt{g(x)} \int \frac{1}{\sqrt{g(x)}} dx,$$

whose coefficients m_k satisfy

$$m_{k-2} = \sum_{n=0}^{k-1} \frac{c_n d_{k-n-1}}{k-n} = \frac{d_{k-1}}{k} + \sum_{n=1}^{k-1} \frac{c_n d_{k-n-1}}{k-n}, \quad \text{for all } k \geq 2,$$

where c_n and d_n are the coefficients of the power series expansions of $\sqrt{g(x)} = 1 + \sum_{n \geq 1} c_n x^n$ and $1/\sqrt{g(x)} = 1 + \sum_{n \geq 1} d_n x^n$, respectively. Both $(c_n)_n$ and $(d_n)_n$ are integer sequences.

The crucial point is that d_{p-1} is not divisible by p for any prime $p > 3$. This is established by analysing the explicit binomial sum representation of $1 + \sum_{n \geq 1} d_n x^n$. As a result, m_{p-2} has a denominator divisible by p for large p , and since p is arbitrary, Theorem 2.1 follows. $\square$

2.2 Discussion

A natural question to ask is:

To what extent can the approach described in [KL05] be generalized and applied to other P-recursive sequences?

Comment 1. The first claim made by Klazar and Luca is the integrality of Motzkin-type sequences of the form $M(\mu, \mu)$, $\mu \in \mathbb{Z}$. Although the argument appears to be direct, it relies

heavily on the combinatorial properties of $M(1, 1)$. Hence, the first limitation: in the general case the underlying combinatorial interpretation (if any) of a given P-recursive sequence is not known beforehand. This already arises two questions:

1.1 How to “guess” that the choice $m_0 = m_1 = 1$ results in an integral solution of (1.1)?

1.2 How to prove it?

Question 1.1 can be approached algorithmically with the approach presented in Section 4.2.

One way to answer question 1.2 is to consider the generating function $y(x) = \sum_{n \geq 0} m_n x^n$. Its coefficient sequence is P-recursive thanks to (1.1), so we know that $y(x)$ is D-finite. Let us carry out some computations in Maple¹

```
> rec := {(n + 2)m(n) = (2n + 1)m(n - 1) + (3n - 3)m(n - 2), m(0) = 1, m(1) = 1} :
> with(gfun):
> rectodiffeq(rec, m(n), y(x));
```

$$(-3x^2 - 3x + 2)y(x) + (-3x^3 - 2x^2 + x) \left(\frac{d}{dx} y(x) \right) - 2.$$

```
> dsolve({%, y(0) = 1}, y(x));
```

$$y(x) = \frac{I\sqrt{3x-1}\sqrt{x+1} - x + 1}{2x^2}.$$

In the procedure shown above we computed the differential equation satisfied by the generating function of Motzkin numbers with the *rectodiffeq* command, and solved it using *dsolve*. The initial condition $y(0) = 1$ comes from the fact that $y(0) = m_0 = 1$. Equivalently:

$$y(x) = \frac{1 - x - \sqrt{1 - 2x - 3x^2}}{2x^2}.$$

and one recognizes the closed form for the generating function of Motzkin numbers.

It is now easy to see why all $m_n \in \mathbb{Z}$: let $\sum_{n \geq 0} a_n x^n \in \mathbb{Q}[[x]]$ be the power series expansion of $\sqrt{1 - 2x - 3x^2}$. Then

$$\left(\sum_{n \geq 0} a_n x^n \right)^2 = \sum_{n \geq 0} x^n \sum_{k=0}^n a_k a_{n-k} = 1 - 2x - 3x^2.$$

For all $k \in \mathbb{N}$ and $P(x) \in \mathbb{Q}[[x]]$ let $[x^k]P(x)$ denote the coefficient of x^k in the power series expansion of $P(x)$. Matching the first few coefficients gives

$$\begin{aligned} [x^0] \sum_{n \geq 0} x^n \sum_{k=0}^n a_k a_{n-k} &= a_0 = 1, \\ [x^1] \sum_{n \geq 0} x^n \sum_{k=0}^n a_k a_{n-k} &= a_0 a_1 + a_1 a_0 = -2 \implies a_1 = -1, \\ [x^2] \sum_{n \geq 0} x^n \sum_{k=0}^n a_k a_{n-k} &= a_0 a_2 + a_1 a_1 + a_2 a_0 = -3 \implies a_2 = -2. \end{aligned}$$

¹Using the *gfun* package, version 4.10.

Hence

$$\frac{1 - x - \sqrt{1 - 2x - 3x^2}}{2x^2} = \frac{1 - x - 1 + x - \sum_{n \geq 2} a_n x^n}{2x^2} = \frac{1}{2} \sum_{n \geq 2} a_n x^{n-2}.$$

It remains to prove that a_n is an even integer for $n \geq 2$.

Proof. We proceed by induction. The statement holds for $n = 2$ as $a_2 = -2$. Assume $a_k \in 2\mathbb{Z}$ for all $k \leq N$ for some $N \geq 2$. Then

$$0 = [x^{N+1}] \sum_{n \geq 0} x^n \sum_{k=0}^n a_k a_{n-k} = 2a_0 a_{N+1} + 2a_1 a_N + \sum_{k=2}^{N-1} a_k a_{N+1-k} \implies a_{N+1} = a_N - \frac{1}{2} \sum_{k=2}^{N-1} a_k a_{N+1-k}.$$

By the inductive assumption, a_k is even for $2 \leq k \leq N$, so $2|a_N$ and $4|a_k a_{N-k}$ for $2 \leq k \leq N-1$. Hence,

$$2 \text{ divides } a_N - \frac{1}{2} \sum_{k=2}^{N-1} a_k a_{N+1-k} = a_{N+1}.$$

□

Remark 2.2. The above reasoning could have been made shorter by appealing to known identities for Motzkin numbers. For example, one could argue that

$$m(x) = \frac{1 - x - \sqrt{1 - 2x - 3x^2}}{2x^2}$$

is a standard way of expressing terms of $M(1, 1)$ and therefore the computation by hand (or within a computer algebra system) is redundant. However, most of such results utilize implicitly the connections with combinatorial objects and this is precisely what we are trying to avoid for the sake of generality.

Comment 2. It is noteworthy that the proof of Theorem 2.1 depends significantly on the structure of the differential equation (1.2). As a result, even slight changes to the recurrence parameters require adjustments to the method.

In view of Definition 1.5, and using the fact that $M(\mu, \mu)$ is globally bounded (because it belongs to $\mu\mathbb{Z}[[x]]$) for any $\mu \in \mathbb{Q}$, we can reformulate Theorem 2.1 as follows:

Theorem 2.3. *A Motzkin-type sequence $M(\mu, \lambda)$ with $\mu, \lambda \in \mathbb{Q}$ is globally bounded if and only if $\mu = \lambda$.*

Let us now invoke an important result, first stated by Eisenstein in 1852 ([Eis52]) and subsequently proved by Heine ([Hei53], [Hei54]):

Theorem 2.4. *(Eisenstein's theorem). If $f(x) \in \mathbb{Q}[[x]]$ is algebraic, then it is globally bounded.*

Using the fact that the generating function of Motzkin numbers (and, consequently, of any $M(\mu, \mu)$ by linearity) is algebraic, we obtain the following corollary:

Corollary 2.5. *(of Theorem 2.3 and Theorem 2.4) The generating function of a Motzkin-type sequence $M(\mu, \lambda)$ is algebraic if and only if $\mu = \lambda$.*

3 Arithmetic outlook

We recall that Motzkin-type sequences satisfy the recurrence

$$(n+2)m_n = (2n+1)m_{n-1} + (3n-3)m_{n-2}, \quad n \geq 2 \quad (2.1)$$

and are uniquely defined by a pair of initial values m_0, m_1 . We will also write $m_n(\lambda, \mu)$ for the n th term of a Motzkin-type sequence with $m_0 = \mu, m_1 = \lambda$ when it is necessary to specify the initial values.

In what follows, we will also make use of the notion of diagonals of multivariate power series:

Definition 3.1. Let

$$R(x_1, \dots, x_k) = \sum_{n_1, \dots, n_k \geq 0} c(n_1, \dots, n_k) x_1^{n_1} \dots x_k^{n_k} \in \mathbb{Q}[[x_1, \dots, x_k]].$$

The *diagonal* of R is a univariate power series defined by

$$\text{Diag}(R) = \sum_{n \geq 0} c(n, \dots, n) t^n \in \mathbb{Q}[[t]].$$

Proposition 3.2. (Furstenberg, [Fur67]). *Any algebraic power series is the diagonal of a rational function in two variables.*

The converse of Proposition 3.2, known as Pólya's theorem [P622], is also true.

Proposition 3.3. ([Chr90, Proposition 5, p. 49]) *Diagonals of rational functions are globally bounded.*

Set $M(x) = \sum_{n \geq 0} m_n x^n$. The next result describes several equivalent conditions involving the arithmetic nature of $M(x)$, and the parameters defining the sequence $(m_n)_n$.

Theorem 3.4. *The following statements are equivalent:*

1. $M(x)$ is algebraic.
2. $M(x)$ is globally bounded.
3. There exists $b \in \mathbb{Z}^*$ such that $m_n \in \frac{1}{b}\mathbb{Z}$ for all $n \geq 0$.
4. $m_0 = m_1$.
5. $M(x)$ is a diagonal of some bivariate rational function.

Proof.

1 $\implies$ 2. By Eisenstein's theorem.

2 $\implies$ 4. By the contrapositive of Theorem 2.3.

1 $\iff$ 4. By Corollary 2.5.

4 $\implies$ 3. By the fact that the sequence of Motzkin numbers, i.e. $(m_n(1, 1))_n$, is integral and $m_n(\frac{a}{b}, \frac{a}{b}) = \frac{a}{b} m_n(1, 1)$ for all $n \geq 0$ and all $a, b \in \mathbb{Z}$.

3 $\implies$ 2. Trivial.

1 $\implies$ 5. By Proposition 3.2.

5 $\implies$ 2. By Proposition 3.3.

□

The figure below visualizes the implications in Theorem 3.4:

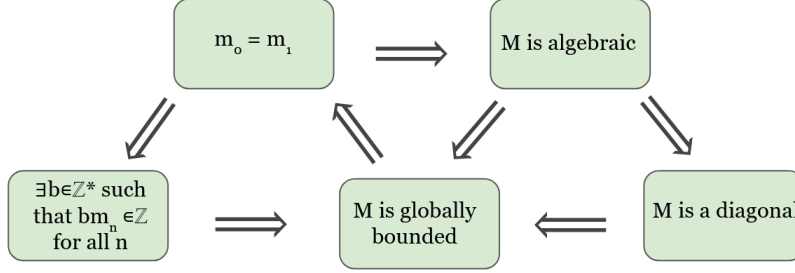


FIGURE 1

Such a result is surprising as it establishes, in the case of this specific recurrence, the converse of Eisenstein's theorem (the implication 1. $\implies$ 2.). Of course, this is no longer true in the general setting, and there are many examples of transcendental functions with almost integral, or even integral power series expansions. For instance, the generating function of the squared central binomial coefficients lives in $\mathbb{Z}[[x]]$, but at the same time is transcendental ([Fur67, p. 271–272]). It writes as

$${}_2F_1\left[\begin{matrix} \frac{1}{2} & \frac{1}{2} \\ 1 \end{matrix}; 16x\right] = \sum_{n \geq 0} \binom{2n}{n}^2 x^n,$$

where ${}_2F_1$ denotes the Gaussian hypergeometric function defined as

$${}_2F_1\left[\begin{matrix} a & b \\ c \end{matrix}; x\right] = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{x^n}{n!},$$

and $(q)_n$ is the rising factorial (also known as *Pochhammer symbol*):

$$(q)_n = \begin{cases} 1 & \text{if } n = 0, \\ q(q+1) \dots (q+n-1) & \text{if } n > 0. \end{cases}$$

There is no contradiction with the general falsity of the converse. At a closer look, the equivalence of 1. and 2. relies on an intermediate step: namely the condition $m_0 = m_1$. The phenomenon is captured by the chain of implications 1 $\implies$ 2 $\implies$ 4.

Another curious observation is that the implication 2 $\implies$ 5 entails Christol's conjecture ([Chr90], Conjecture 4, p. 55) for P-recursive sequences that are solutions of (2.1).

Conjecture 3.5. ([Chr90]) *Every D-finite globally bounded function is the diagonal of a rational function.*

It is worth mentioning that Theorem 3.4 is not easily generalizable to other linear recurrences. Even in the very specific case of second-order recurrences, the implications 2 $\implies$ 1 and 3 $\implies$ 1 do not hold if we drop the linearity of polynomial coefficients. As a counterexample one can take the recurrence for big Apéry numbers (OEIS A005259):

$$n^3 A_n = (34n^3 - 51n^2 + 27n - 5)A_{n-1} - (n-1)^3 A_{n-2}. \quad (2.2)$$

The sequence solving the recurrence above with $A_0 = 1, A_1 = 5$ appears in Apéry's famous proof of the irrationality of $\zeta(3)$ ([Ap679]), and is known to be integral. Integrality can be deduced, for example, from the binomial sums representation of $(A_n)_n$:

$$A_n = \sum_{k=0}^n \binom{n}{k}^2 \binom{n+k}{k}^2, \quad n \in \mathbb{N}.$$

There exist several proofs of the fact that $A_0 = 1, A_1 = 5$ is the only pair of initial values (up to multiplication by an integer) that produces an integer sequence ([Cha84], [AJ91]). However, the generating function $\sum_{n \geq 0} A_n x^n$ of any sequence that solves (2.2) is transcendental (one reason for this is asymptotics incompatible with algebraicity, the so-called *Flajolet criterion*).

A similar argument applies to small Apéry numbers (OEIS A005258): the generating function of any sequence produced by the recurrence for small Apéry numbers is transcendental, while there do exist almost integral and even integral solutions of this recurrence. A detailed analysis of such cases is provided in Section 6.

4 Bridging global boundedness and algebraicity

One possible generalization of Theorem 3.4 is the following one, which we state as a conjecture:

Conjecture 4.1. *Let $(s_n)_n$ be a sequence of rational numbers that satisfies a linear homogeneous recurrence relation with polynomial coefficients of degree 1:*

$$\sum_{k=0}^d (a_k n + b_k) s_{n-k} = 0, \quad a_i \neq 0 \text{ for } i = 0, \dots, d \quad d \in \mathbb{N}.$$

Let $S(x) = \sum_{n \geq 0} s_n x^n$ be its generating function. The following statements are equivalent:

1. $S(x)$ is algebraic.
2. $S(x)$ is a diagonal.
3. $S(x)$ is globally bounded.

As before,

- 1 $\implies$ 2 by Proposition 3.2,
- 2 $\implies$ 3 by Proposition 3.3.

The interesting direction is 3 $\implies$ 1. We prove it for first-order recurrences in Section 4.1 and for a specific subclass of second-order recurrences in Section 4.2. The general case of Conjecture 4.1 is currently unsolved.

4.1 Hypergeometric case

The most basic setting of Conjecture 4.1 is the case $d = 1$. It defines a certain subclass of hypergeometric functions, that is, of power series whose coefficient sequences satisfy first-order linear recurrences with polynomial coefficients. Specifically, we consider $(s_n)_n$ such that

$$(a_0 n + b_0) s_n + (a_1 n + b_1) s_{n-1} = 0, \quad a_0, a_1 \in \mathbb{Z}^*, \quad b_0, b_1 \in \mathbb{Z}, \quad \frac{b_0}{a_0} \notin -\mathbb{N}^*.$$

Or, equivalently

$$(n + b_0) s_n + (a_1 n + b_1) s_{n-1} = 0, \quad a_1 \in \mathbb{Q}^*, \quad b_0 \in \mathbb{Q} \setminus -\mathbb{N}^*, \quad b_1 \in \mathbb{Q}, \quad (3.1)$$

where the condition $\frac{b_0}{a_0} \notin -\mathbb{N}^*$ comes from the fact that if $a_0 n + b_0 = 0$ for some $n \in \mathbb{N}^*$, then s_n is undefined unless $s_k = 0$ for all $k \in \mathbb{N}$.

Let $S(x) = \sum_{n \geq 0} s_n x^n$ denote the generating function of $(s_n)_n$.

The usual scheme of transforming (3.1) into a differential equation and then solving it yields:

$$S(x) = x^{-b_0} (a_1 x + 1)^{-1+b_0-\frac{b_1}{a_1}} \left(s_0 b_0 \int x^{-1+b_0} (a_1 x + 1)^{\frac{-a_1 b_0 + b_1}{a_1}} dx + c_1 \right). \quad (3.2)$$

We can clearly see that $S(x)$ is a combination of algebraic functions and a primitive of an algebraic function

$$x^{-1+b_0} (a_1 x + 1)^{\frac{-a_1 b_0 + b_1}{a_1}}.$$

Fortunately, the problem of integrating algebraic functions has been extensively studied and goes back to the work of Liouville [Lüt12, Chapter IX]. In our analysis we make use of the following (rather involved) theorem due to André:

Theorem 4.2. (André, [And89, p. 149]). *Let y be a power series. Then y is algebraic if and only if y is globally bounded and $\frac{dy}{dx}$ is algebraic.*

Corollary 4.3. (of Theorem 4.2 and Theorem 2.4). *The primitive of an algebraic power series is globally bounded if and only if it is algebraic.*

Therefore, $S(x)$ is algebraic if and only if it is globally bounded, which immediately implies Conjecture 4.1 for this case $d = 1$.

The next theorem characterises the conditions under which $S(x)$ is algebraic and globally bounded, noting that these properties are equivalent by Conjecture 4.1.

Theorem 4.4. *Let $(s_n)_n$ be a sequence of rational numbers that satisfies (3.1). Then the generating function $S(x) = \sum_{n \geq 0} s_n x^n$ is globally bounded (and algebraic) if and only if one of the following conditions holds:*

1. $b_0 \in \mathbb{N}$, $\frac{b_1}{a_1} \notin \mathbb{Z}$, or
2. $b_0 \notin \mathbb{Z}$, $b_0 - \frac{b_1}{a_1} \in -\mathbb{N}$, or
3. $b_0, b_0 - \frac{b_1}{a_1} \notin \mathbb{Z}$, $\frac{b_1}{a_1} + 1 \in -\mathbb{N}$, or
4. $b_0, \frac{b_1}{a_1} \in \mathbb{Z}$, $\frac{b_1}{a_1} < 0 < b_0$, $\frac{b_1}{a_1} \leq -1$, or
5. $b_0, \frac{b_1}{a_1} \in \mathbb{Z}$, $0 < b_0 \leq \frac{b_1}{a_1}$.

Proof. $S(x)$ can be expressed from (3.2) as

$$S(x) = x^{-b_0} (a_1 x + 1)^{-1+b_0-\frac{b_1}{a_1}} \left(s_0 x^{b_0} {}_2F_1 \left[\begin{matrix} b_0 & b_0 - \frac{b_1}{a_1} \\ 1 + b_0 \end{matrix} ; -a_1 x \right] + c_1 \right). \quad (3.3)$$

Therefore, $S(x)$ is algebraic precisely when ${}_2F_1 \left[\begin{matrix} b_0 & b_0 - \frac{b_1}{a_1} \\ 1 + b_0 \end{matrix} ; -a_1 x \right]$ is.

An algorithm to determine the algebraicity of hypergeometric functions with arbitrary parameters is described by Fürsinn and Yurkevich in [FY24]. In the same paper, a useful criterion for deciding the algebraicity of Gaussian hypergeometric functions is presented:

Corollary 4.5. ([FY24], Corollary 3.7)

The Gaussian hypergeometric function ${}_2F_1([\alpha, \beta], [\alpha + k]; x)$ for $k \in \mathbb{Z}$, is algebraic if and only if either $k \leq 0$ or:

1. $\alpha \in \mathbb{Z}$, $\beta \notin \mathbb{Z}$, or
2. $\alpha \notin \mathbb{Z}$, $\beta \in -\mathbb{N}$, or
3. $\alpha, \beta \notin \mathbb{Z}$, $\beta - \alpha - k \in \mathbb{N}$, or
4. $\alpha, \beta \in \mathbb{Z}$, $\alpha < \beta \leq 0$, $\beta - \alpha \geq k$, or
5. $\alpha, \beta \in \mathbb{Z}$, $0 < \alpha < \beta$, $\beta - \alpha \geq k$, or
6. $\alpha, \beta \in \mathbb{Z}$, $\beta \leq 0 < \alpha$.

In our case of ${}_2F_1 \left[\begin{smallmatrix} b_0 & b_0 - \frac{b_1}{a_1} \\ 1 + b_0 \end{smallmatrix} ; -a_1 x \right]$, the values are $\alpha = b_0, \beta = b_0 - \frac{b_1}{a_1}, k = 1$. By substituting them into Corollary 4.5, we obtain the necessary and sufficient conditions for the algebraicity of $S(x)$, captured in Theorem 4.4.

Observe that case 4. in Corollary 4.5 is not possible for the choice $\alpha = b_0, \beta = b_0 - \frac{b_1}{a_1}, k = 1$. Indeed, the three requirements in that line become

$$b_0, \frac{b_1}{a_1} \in \mathbb{Z}, \quad b_0 < b_0 - \frac{b_1}{a_1} \leq 0, \quad -\frac{b_1}{a_1} \geq 1.$$

which simplify to

$$b_0, \frac{b_1}{a_1} \in \mathbb{Z}, \quad b_0 \leq \frac{b_1}{a_1} \leq -1.$$

This forces b_0 to be a nonpositive integer, contradicting our assumption $b_0 \in \mathbb{Q} \setminus -\mathbb{N}^*$. Hence case 4. is ruled out for the parameters that occur in Theorem 4.4. $\square$

4.2 Second-order recurrences

Let $(s_n)_n$ be a P-recursive sequence of rational numbers that satisfies a second-order recurrence relation:

$$(n + b_0)s_n + (a_1n + b_1)s_{n-1} + (a_2n + b_2)s_{n-2} = 0 \quad (4.1)$$

with $a_1, a_2 \in \mathbb{Q}^*$, $b_0, b_1, b_2 \in \mathbb{Q}$, $b_0 \notin -\mathbb{N} \setminus \{0, -1\}$.

Let $S(x) = \sum_{n \geq 0} s_n x^n$ be the generating function of $(s_n)_n$. In the general case, $S(x)$ takes the following form:

```
> rec := {(n + b_0)*s(n) + (a_1*n + b_1)*s(n-1) + (a_2*n + b_2)*s(n-2),
          s(0) = s_0, s(1) = s_1}:
> dsolve(rectodiffeq(rec, s(n), S(x)), S(x));
```

$$S(x) = (a_2x^2 + a_1x + 1)^{-\frac{b_2}{2a_2} - 1 + \frac{b_0}{2}} \left(\int (a_1s_0x + b_0s_1x + b_1s_0x + b_0s_0 + s_1x) (a_2x^2 + a_1x + 1)^{-\frac{a_2b_0 - b_2}{2a_2}} x^{b_0-1} \right. \\ \left. e^{\frac{\operatorname{artanh}\left(\frac{2a_2x+a_1}{\sqrt{a_1^2-4a_2}}\right)(a_1a_2b_0+a_1b_2-2a_2b_1)}{\sqrt{a_1^2-4a_2a_2}}} dx + c_1 \right) x^{-b_0} e^{-\frac{\operatorname{artanh}\left(\frac{2a_2x+a_1}{\sqrt{a_1^2-4a_2}}\right)(a_1a_2b_0+a_1b_2-2a_2b_1)}{\sqrt{a_1^2-4a_2a_2}}},$$

where c_1 is some constant that comes from solving a differential equation with unspecified initial conditions. Supposedly, if e does not vanish in the equation above, $S(x)$ is not globally bounded. However, this question remains open for now.

From now on we shall restrict ourselves to the case $b_2 = \frac{2a_2b_1 - a_1a_2b_0}{a_1}$, setting the exponent of e to zero.

```
> simplify(subs(b_2 = (2*a_2*b_1 - a_1*a_2*b_0)/a_1, %));
```

$$S(x) = (a_2x^2 + a_1x + 1)^{-\frac{(b_0-1)a_1-b_1}{a_1}} \left(\int x^{b_0} (a_2x^2 + a_1x + 1)^{-b_0 + \frac{b_1}{a_1}} \left(a_1s_0 + b_0s_1 + b_1s_0 + \frac{b_0s_0}{x} + s_1 \right) dx + c_1 \right) x^{-b_0}.$$

Since the integrand in the code snippet above is algebraic, André's theorem ensures that global boundedness and algebraicity of $S(x)$ are equivalent in this context. Hence, Conjecture 4.1 is proved for the case $b_2 = \frac{2a_2b_1 - a_1a_2b_0}{a_1}$.

In light of this result, we treat algebraicity and global boundedness of $S(x)$ as interchangeable throughout this section.

Given a recurrence of type (4.1), a sequence $(s_n)_n$ is uniquely defined by a pair (s_0, s_1) . The algebraic nature of $S(x)$ falls therefore into one of three cases:

- (C1) $S(x)$ is algebraic for all $(s_0, s_1) \in \mathbb{Q}^2$.
- (C2) $S(x)$ is transcendental for all $(s_0, s_1) \in \mathbb{Q}^2 \setminus \{(0, 0)\}$.
- (C3) The set of pairs (s_0, s_1) such that $S(x)$ is algebraic forms a one-dimensional subspace of $\mathbb{Q}^2$.

We now present an algorithm that determines which of the three cases holds for a given recurrence of type (4.1) with

$$b_2 = \frac{2a_2b_1 - a_1a_2b_0}{a_1}, \quad b_0 \in \mathbb{N}.$$

If $S(x)$ is algebraic for all initial conditions or for none (C1 and C2), the algorithm will detect this directly. In C3, where algebraicity occurs only along a one-dimensional subspace of $\mathbb{Q}^2$, the algorithm computes a nonzero rational pair (s_0, s_1) that generates all such algebraic solutions, up to scalar multiplication. Our approach involves a case-by-case analysis based on the recurrence structure.

Firstly, it might happen that $a_1^2 - 4a_2 = 0$. That is, $1 + a_1x + a_2x^2$ has a double root $a = -\frac{a_1}{2a_2}$. This case is simple². One computes:

$$S(x) = x^{-b_0} (1 + ax)^{2(b_0-1) - \frac{b_1}{a}} \left(c_1 + x^{b_0} \left(s_0 (1 + ax)^{1-2b_0 + \frac{b_1}{a}} + (as_0 + s_1) {}_2F_1 \left(\left[1 + b_0, 2b_0 - \frac{b_1}{a} \right], [2 + b_0]; -ax \right) \right) \right).$$

It is immediate that $S(x)$ has a nontrivial algebraic solution when $as_0 + s_1 = 0$ ($\iff -\frac{a_1}{2a_2}s_0 + s_1 = 0 \iff a_1s_0 = 2a_2s_1$). For example, one can take $(s_0, s_1) = (2a_2, a_1)$. Moreover, by Corollary 4.5, $S(x)$ is algebraic for all $(s_0, s_1) \in \mathbb{Q}^2$ if and only if:

1. $b_0 \in \mathbb{Z}$, $\frac{b_1}{a} \notin \mathbb{Z}$, or
2. $b_0 \notin \mathbb{Z}$, $2b_0 - \frac{b_1}{a} \in -\mathbb{N}$, or
3. $b_0, 2b_0 - \frac{b_1}{a} \notin \mathbb{Z}$, $b_0 - \frac{b_1}{a} - 2 \in \mathbb{N}$, or
4. $b_0, \frac{b_1}{a} \in \mathbb{Z}$, $b_0 > -1$, $b_0 - \frac{b_1}{a} \geq 2$, or
5. $b_0, \frac{b_1}{a} \in \mathbb{Z}$, $b_0 > -1$, $2b_0 - \frac{b_1}{a} \leq 0$.

If we insist that $b_0 \in \mathbb{N}$ and substitute back $a = -\frac{a_1}{2a_2}$, the list above becomes

²Assuming your institution provides a Mathematica license.

1. $\frac{2a_2b_1}{a_1} \notin \mathbb{Z}$, or
2. $\frac{2a_2b_1}{a_1} \in \mathbb{Z}$, $b_0 + \frac{2a_2b_1}{a_1} \geq 2$, or
3. $\frac{2a_2b_1}{a_1} \in \mathbb{Z}$, $b_0 + \frac{2a_2b_1}{a_1} \leq 0$.

We now turn to the case $a_1^2 - 4a_2 \neq 0$ and study the integral

$$\int \left(x^{b_0-1} b_0 s_0 + x^{b_0} (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) \right) (1 + a_1 x + a_2 x^2)^{-b_0 + \frac{b_1}{a_1}} dx, \quad (4.2)$$

and ask when it is algebraic.

From now on we also assume $b_0 \in \mathbb{N}^*$, leaving the general case for future work.

The key idea is to represent (4.2) as a linear combination of two integrals of the type

$$\int x^n (1 + a_1 x + a_2 x^2)^q dx, \quad a_1, a_2 \in \mathbb{Q}^*, \quad n \in \mathbb{N}, \quad q \in \mathbb{Q}. \quad (4.3)$$

Indeed, let

$$I_1 = \int x^{b_0-1} (1 + a_1 x + a_2 x^2)^{-b_0 + \frac{b_1}{a_1}} dx, \quad I_2 = \int x^{b_0} (1 + a_1 x + a_2 x^2)^{-b_0 + \frac{b_1}{a_1}} dx.$$

Then

$$\int \left(x^{b_0} (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) + x^{b_0-1} b_0 s_0 \right) (1 + a_1 x + a_2 x^2)^{-b_0 + \frac{b_1}{a_1}} dx = b_0 s_0 I_1 + (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) I_2.$$

Observe that

$$\begin{aligned} b_0 s_0 = 0 &\iff s_0 = 0 \quad \text{since } b_0 \geq 1 \\ a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1 = 0 &\iff (b_0 + 1) s_1 = -s_0 (a_1 + b_1). \end{aligned}$$

The algorithm proceeds as follows:

1. Decide algebraicity of I_1 and I_2 .
2. Proceed according to the table below:

I_1	I_2	Conclusion
Algebraic	Algebraic	C1 .
Algebraic	Transcendental	C3 , return $(b_0 + 1, -a_1 - b_1)$.
Transcendental	Algebraic	C3 , return $(0, 1)$.
Transcendental	Transcendental	Check whether the combination $b_0 s_0 I_1 + (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) I_2$ is algebraic for some $(s_0, s_1) \in \mathbb{Q}^2 \setminus \{(0, 0)\}$. If so, then C3 , return (s_0, s_1) . Otherwise, C2 .

4.2.1 Deciding algebraicity of I_1, I_2 .

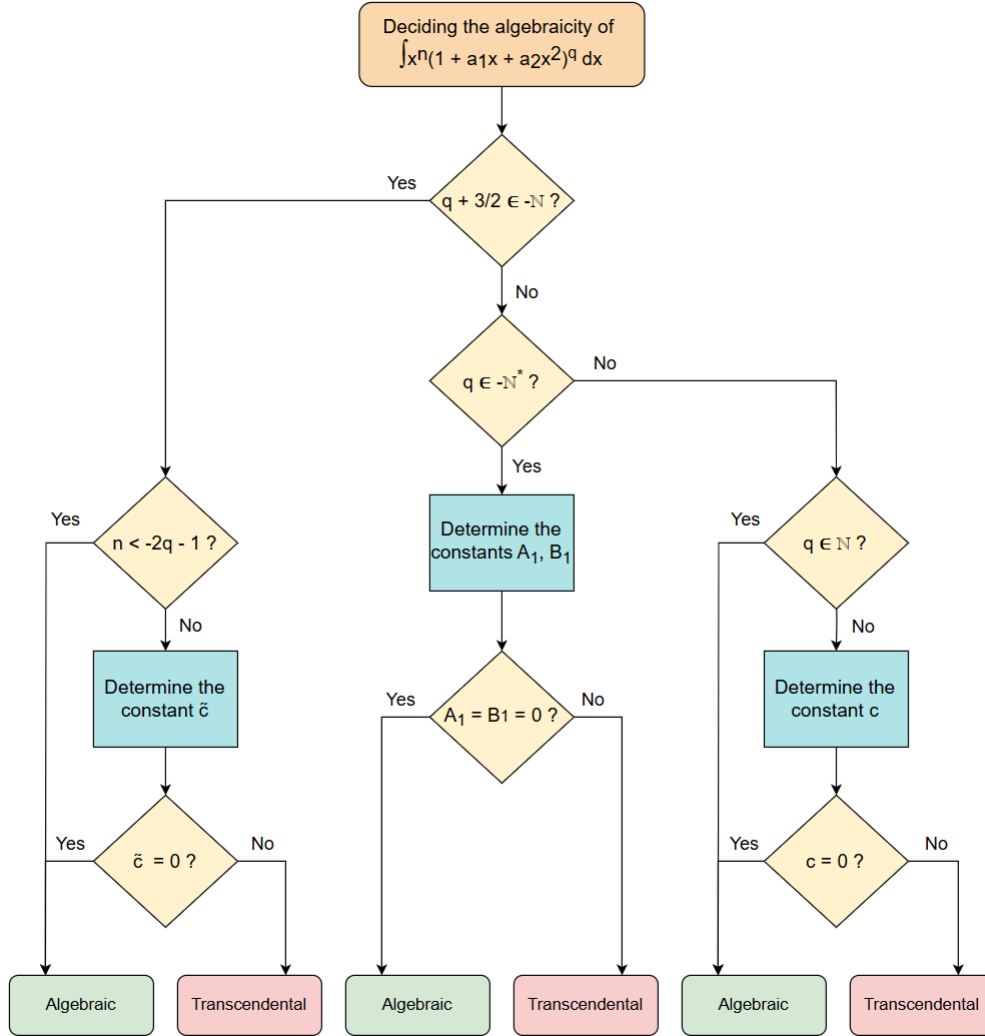


FIGURE 2: Deciding whether $\int x^n(1 + a_1x + a_2x^2)^q dx$, such that $a_1, a_2 \in \mathbb{Q}^*, n \in \mathbb{N}, q \in \mathbb{Q}$, is algebraic or transcendental. The constants $c, \tilde{c}, (A_1, B_1)$ come from (4.4), (4.7), (4.8), respectively.

Consider an integral of the type (4.3). There are several cases:

1. $\mathbf{n = 0}$.

In order to study the integral

$$\int (1 + a_1x + a_2x^2)^q dx.$$

we distinguish 2 subcases:

1.2 Assume $\mathbf{a_1^2 - 4a_2 > 0}$.

If $q = -1$,

$$\int (1 + a_1x + a_2x^2)^{-1} dx = -\frac{2 \operatorname{artanh}\left(\frac{2a_2x+a_1}{\sqrt{a_1^2-4a_2}}\right)}{\sqrt{a_1^2-4a_2}},$$

which is transcendental. Otherwise,

$$\int (1 + a_1x + a_2x^2)^q dx = \frac{1}{a_2(q+1)} 2^{q-1} (2a_2x + a_1 - \sqrt{a_1^2 - 4a_2}) \left(\frac{2a_2x + a_1 + \sqrt{a_1^2 - 4a_2}}{\sqrt{a_1^2 - 4a_2}} \right)^{-q} \\ (a_2x^2 + a_1x + 1)^q {}_2F_1 \left([-q, q+1], [q+2]; \frac{-a_1 - 2a_2x + \sqrt{a_1^2 - 4a_2}}{2\sqrt{a_1^2 - 4a_2}} \right) + C.$$

In other words, we get a particular ${}_2F_1$ multiplied by an algebraic function. Therefore, the answer to “Is $\int (1 + ax + bx^2)^q dx$ algebraic?” depends solely on this ${}_2F_1$.

Corollary 4.5 allows to establish all cases when ${}_2F_1 \left([-q, q+1], [q+2]; \frac{-a_1 - 2a_2x + \sqrt{a_1^2 - 4a_2}}{2\sqrt{a_1^2 - 4a_2}} \right)$ is algebraic:

(a) $q \in \mathbb{N}$, or

(b) $q + \frac{3}{2} \in -\mathbb{N}$.

1.2 Assume $a_1^2 - 4a_2 < 0$.

If $q = -1$,

$$\int (1 + a_1x + a_2x^2)^{-1} dx = \frac{2 \operatorname{artanh} \left(\frac{2a_2x + a_1}{\sqrt{-a_1^2 + 4a_2}} \right)}{\sqrt{-a_1^2 + 4a_2}}.$$

which is transcendental. Otherwise, the integral admits a solution in terms of algebraic functions and a Gaussian ${}_2F_1$:

$$\int (1 + a_1x + a_2x^2)^q dx = \left(x + \frac{a_1}{2a_2} \right) \left(\frac{4a_2 - a_1^2}{4a_2} \right)^q {}_2F_1 \left(\left[\frac{1}{2}, -q \right], \left[\frac{3}{2} \right]; \frac{(a_1 + 4a_2x)^2}{a_1^2 - 4a_2} \right) + C.$$

which is algebraic if and only if:

(a) $q \in \mathbb{N}$, or

(b) $q + \frac{3}{2} \in -\mathbb{N}$.

Observe that algebraicity conditions turn out to be the same for either of the subcases.

2. $n \geq 1, q \notin -\mathbb{N}^*, q + \frac{3}{2} \notin -\mathbb{N}$.

Lemma 4.6. For all $n \in \mathbb{N}, q \in \mathbb{Q} \setminus \{-1, -\frac{3}{2}, -2, \dots, -\frac{n+1}{2}\}$ there exist $c \in \mathbb{Q}$ and $C(x) \in \mathbb{Q}[x]_{n-1}$ such that

$$\int x^n (1 + a_1x + a_2x^2)^q dx = c \int (1 + a_1x + a_2x^2)^q dx + C(x)(1 + a_1x + a_2x^2)^{q+1}. \quad (4.4)$$

Proof. Differentiating both sides of (4.4) yields

$$x^n (1 + a_1x + a_2x^2)^q = c(1 + a_1x + a_2x^2)^q + C'(x)(1 + a_1x + a_2x^2)^{1+q} + C(x)(1+q)(a_1 + 2a_2x)(1 + a_1x + a_2x^2)^q.$$

Divide both sides of the equation above by $(1 + a_1x + a_2x^2)^q$:

$$x^n = c + C'(x)(1 + a_1x + a_2x^2) + C(x)(a_1 + 2a_2x)(1 + q). \quad (4.5)$$

We write $C(x) = \sum_{i=0}^{n-1} c_i x^i$ and show that it is always possible to find $c, c_0, c_1, \dots, c_{n-1} \in \mathbb{Q}$

such that equation (4.5) holds. Each side of (4.5) is simply a polynomial of degree n .

If $n = 1$, then one solves

$$x = c + c_0(a_1 + 2a_2x)(1 + q).$$

for c and c_0 and gets

$$c_0 = \frac{1}{2a_2(1+q)},$$

$$c = -\frac{a_1}{2a_2}.$$

Otherwise, write

$$x^n = c + (1 + a_1x + a_2x^2) \sum_{i=0}^{n-2} (i+1)c_{i+1}x^i + (1+q)(a_1 + 2a_2x) \sum_{i=0}^{n-1} c_i x^i.$$

Rearranging the terms gives:

$$\begin{aligned} x^n &= c + c_1 + a_1c_0(q+1) \\ &+ \sum_{i=1}^{n-2} (c_{i+1}(i+1) + c_i(a_1i + a_1(q+1)) + c_{i-1}(a_2(i-1) + 2a_2(q+1)))x^i \\ &+ (c_{n-1}(a_1(n-1) + a_1(q+1)) + c_{n-2}(a_2(n-2) + 2a_2(q+1)))x^{n-1} \\ &+ (a_2(n-1)c_{n-1} + 2(q+1)a_2c_{n-1})x^n. \end{aligned} \quad (4.6)$$

One can now derive explicit expressions for $c, c_0, \dots, c_{n-1}$:

$$\begin{aligned} c_{n-1} &= \frac{1}{a_2(n+1+2q)}, \\ c_{n-2} &= -\frac{a_1(n+q)}{a_2(n+2q)}c_{n-1}, \\ c_i &= -\frac{c_{i+2}(i+2) + c_{i+1}a_1(i+2+q)}{a_2(i+2+2q)}, \quad i = 0, \dots, n-3 \\ c &= -c_1 - a_1c_0(q+1). \end{aligned}$$

Starting from c_{n-1} , one can iteratively compute all c_i for $i = 0, \dots, n-1$, as well as c . This procedure only fails if a zero appears in a denominator during computation. Since $a_2 \neq 0$ by assumption, we need to check that $i+2+2q \neq 0$ for $i = 0, \dots, n-1$. Equivalently, that

$$q \neq -1 - \frac{i}{2} \text{ for } i = 0, \dots, n-1.$$

This is precisely where the condition $q \in \mathbb{Q} \setminus \{-1, -\frac{3}{2}, -2, \dots, -\frac{n+1}{2}\}$ comes from: it ensures that $c, c_0, \dots, c_{n-1}$ are well defined. $\square$

Hence, for $q \in \mathbb{Z} \setminus \{-1, -\frac{3}{2}, \dots, -\frac{n+1}{2}\}$ $\int x^n(1 + a_1x + a_2x^2)^q dx$ is algebraic if and only if $\int (1 + ax + bx^2)^q dx$ is algebraic, or $c = 0$. Using our classification for the case $n = 0$, we deduce the following:

Corollary 4.7. *If $n \in \mathbb{N}, q \notin -\mathbb{N}^*, q + \frac{3}{2} \notin -\mathbb{N}$, then $\int x^n(1 + a_1x + a_2x^2)^q dx$ is algebraic if and only if*

- (a) $q \in \mathbb{N}$, or
- (b) $c = 0$, where c is the constant from (4.4).

Proof. By (4.4),

$$\int x^n(1 + a_1x + a_2x^2)^q dx = c \int (1 + a_1x + a_2x^2)^q dx + \underbrace{C(x)(1 + a_1x + a_2x^2)^{q+1}}_{\text{algebraic}}.$$

The corollary follows immediately from the fact that the set of algebraic functions is closed under addition. $\square$

3. $\mathbf{n} \geq \mathbf{1}, \mathbf{q} + \frac{3}{2} \in -\mathbb{N}$. There are three possible scenarios depending on n :

3.1 $\mathbf{n} < -2\mathbf{q} - \mathbf{1}$, i.e. $q < -\frac{n+1}{2}$. Then we simply apply Corollary 4.7 to deduce that $\int x^n(1 + ax + bx^2)^q dx$ is algebraic if and only if $\int (1 + ax + bx^2)^q dx$ is algebraic. Note that $\int (1 + ax + bx^2)^q dx$ is always algebraic by the case analysis performed in 2.

3.2 $\mathbf{n} = -2\mathbf{q} - \mathbf{1}$. Equivalently, $q = -\frac{n+1}{2}$. Since $q + \frac{3}{2} \in -\mathbb{N}$, n must be even.

Lemma 4.8.

$$\int x^n(1 + a_1x + a_2x^2)^{-\frac{n+1}{2}} dx, \quad a_1, a_2 \in \mathbb{Q}^*, \quad n \in 2\mathbb{N}$$

is not globally bounded.

Our proof technique mirrors the computation from [KL05, p. 71].

Proof. Let $(1 + a_1x + a_2x^2)^{-\frac{n+1}{2}} = \sum_{k \geq 0} d_k x^k$. Then

$$\begin{aligned} \int x^n(1 + a_1x + a_2x^2)^{-\frac{n+1}{2}} dx &= \int \sum_{k \geq 0} d_k x^{n+k} dx = \sum_{k \geq n+1} \frac{d_{k-n-1}}{k} x^k. \\ (1 + a_1x + a_2x^2)^{-\frac{n+1}{2}} &= \sum_{k \geq 0} \binom{-(n+1)/2}{k} (a_1x + a_2x^2)^k \\ &= \sum_{k \geq 0} \binom{-(n+1)/2}{k} x^k \sum_{i=0}^k \binom{k}{i} x^i a_1^{k-i} a_2^i, \\ d_m = [x^m] (1 + a_1x + a_2x^2)^{-\frac{n+1}{2}} &= \sum_{k=0}^m \binom{-(n+1)/2}{k} \binom{k}{m-k} a_1^{2k-m} a_2^{m-k}. \end{aligned}$$

In addition,

$$\begin{aligned} \binom{-(n+1)/2}{k} &= \frac{-\frac{n+1}{2}(-\frac{n+1}{2}-1) \dots (-\frac{n+1}{2}-k+1)}{k!} \\ &= \frac{(-n-1)(-n-3) \dots (-n-2k+1)}{2^k k!} \\ &= \frac{(-1)^k (n+1)(n+3) \dots (n+2k-1)}{2^k k!}. \end{aligned}$$

Let $p > 3$ be a prime number, pairwise coprime with a_1 and a_2 . Then, for k between 0 and p :

$$\begin{aligned} \binom{-(n+1)/2}{k} &\not\equiv 0 \pmod{p} \iff n+2k-1 < p \iff 2k < p-n+1, \\ \binom{k}{p-n-1-k} &= 0 \text{ for } k < p-n-1-k \iff 2k < p-n-1. \end{aligned}$$

Consequently,

$$\binom{-(n+1)/2}{k} \binom{n}{p-n-1-k} \not\equiv 0 \pmod{p}$$

if and only if

$$p-n-1 \leq 2k < p-n+1.$$

Observe that since p is odd and n is even, the condition above is equivalent to $k = \frac{p-n-1}{2}$.

Hence, if $m = p-n-1$, the only product of the form

$$\binom{-(n+1)/2}{k} \binom{k}{p-n-1-k}$$

that appears in d_{p-n-1} and is not divisible by p is the one corresponding to $k = \frac{p-n-1}{2}$. Therefore, d_{p-n-1} is not divisible by p , which proves that the denominators of $\frac{d_{k-n-1}}{k}$ become divisible by arbitrarily large primes. So $\sum_{k \geq n+1} \frac{d_{k-n-1}}{k} x^k$ is not almost integral. $\square$

Hence, $\int x^{-2q-1}(1+ax+bx^2)^q dx$ is not globally bounded and not algebraic.

3.3 **$n > -2q - 1$.** Essentially, this case is not very different from the one where $q \in \mathbb{Q} \setminus \{-1, -\frac{3}{2}, -2, \dots, -\frac{n+1}{2}\}$, as the following identity holds:

$$\begin{aligned} \int x^n(1+a_1x+a_2x^2)^q dx &= c \underbrace{\int (1+a_1x+a_2x^2)^q dx}_{\text{algebraic}} \\ &+ \tilde{c} \underbrace{\int x^{-2q-1}(1+a_1x+a_2x^2)^q dx}_{\text{transcendental}} \\ &+ \underbrace{C(x)(1+a_1x+a_2x^2)^{q+1}}_{\text{algebraic}}. \end{aligned} \quad (4.7)$$

Where $c, \tilde{c} \in \mathbb{Q}$ and $C(x) \in \mathbb{Q}[x]_{n-1}$. In particular, this implies $\int x^n(1+a_1x+a_2x^2)^q dx$ is not algebraic unless $\tilde{c} = 0$.

The proof is analogous to that of (4.4) except for the computation of c_{-2q-2} . Similarly, write

$$x^n = c + \tilde{c}x^{-2q-1} + (1+a_1x+a_2x^2) \sum_{i=0}^{n-2} (i+1)c_{i+1}x^i + (1+q)(a_1+2a_2x) \sum_{i=0}^{n-1} c_i x^i.$$

We apply the same coefficient matching but with additional “patching” at x^{-2q-1} : If $n = -2q$,

$$\begin{aligned} x^{-2q} &= c + c_1 + a_1c_0(q+1) \\ &+ \sum_{1 \leq i \leq -2q-2} ((i+1)c_{i+1} + a_1(i+q+1)c_i + a_2(i+2q+1)c_{i-1})x^i \\ &+ (\tilde{c} - a_1qc_{-2q-1})x^{-2q-1} \\ &+ a_2c_{-2q-1}x^{-2q}. \end{aligned}$$

Otherwise,

$$\begin{aligned}
 x^n &= c + c_1 + a_1 c_0(q+1) \\
 &+ \sum_{\substack{1 \leq i \leq n-2 \\ i \neq -2q-1}} ((i+1)c_{i+1} + a_1(i+q+1)c_i + a_2(i+2q+1)c_{i-1})x^i \\
 &+ (\tilde{c} - 2qc_{-2q} - a_1qc_{-2q-1})x^{-2q-1} \\
 &+ (a_1(n+q)c_{n-1} + a_2(n+2q)c_{n-2})x^{n-1} \\
 &+ (a_2(n+2q+1)c_{n-1})x^n.
 \end{aligned}$$

Observe that previously c_{-2q-2} was computed from the equation

$$(i+1)c_{i+1} + a_1(i+q+1)c_i + a_2(i+2q+1)c_{i-1} = 0, \quad i = -2q-1.$$

Or, in the case $n = -2q$, from

$$a_1(n+q)c_{n-1} + a_2(n+2q)c_{n-2} = 0, \quad n = -2q.$$

However, when $i = -2q-2$, the coefficient of c_{-2q-2} in the expressions above becomes zero. This is when $\tilde{c}$ comes into play: it allows to cancel out the coefficient of x^{-2q-1} . Hence, if $n \neq -2q$:

$$\begin{aligned}
 c_{n-1} &= \frac{1}{a_2(n+1+2q)}, \\
 c_{n-2} &= -\frac{a_1(n+q)}{a_2(n+2q)}c_{n-1}, \\
 c_i &= -\frac{c_{i+2}(i+2) + c_{i+1}a_1(i+2+q)}{a_2(i+2+2q)}, \quad i = 0, \dots, n-3, \quad i \neq -2q-2, \\
 c_{-2q-2} &= 0 \quad (\text{could be any value}), \\
 c &= -c_1 - a_1c_0(q+1), \\
 \tilde{c} &= 2qc_{-2q} + a_1qc_{-2q-1}.
 \end{aligned}$$

If $n = -2q$:

$$\begin{aligned}
 c_{-2q-1} &= \frac{1}{a_2}, \\
 c_{-2q-2} &= 0 \quad (\text{could be any value}), \\
 c_i &= -\frac{c_{i+2}(i+2) + c_{i+1}a_1(i+2+q)}{a_2(i+2+2q)}, \quad i = 0, \dots, -2q-3, \\
 c &= -c_1 - a_1c_0(q+1), \\
 \tilde{c} &= a_1qc_{-2q-1}.
 \end{aligned}$$

4. $\mathbf{n} \geq \mathbf{1}, \mathbf{q} \in -\mathbb{N}^*$. Since the double root case was treated earlier, we may assume

$$1 + a_1x + a_2x^2 = (1 + \alpha x)(1 + \beta x), \quad \alpha \neq \beta.$$

Moreover, $\alpha, \beta \neq 0$. Apply a partial fraction decomposition to the integrand:

$$\frac{x^n}{(1 + a_1x + a_2x^2)^{-q}} = \frac{x^n}{(1 + \alpha x)^{-q}(1 + \beta x)^{-q}} = \sum_{m=1}^{-q} \left(\frac{A_m}{(1 + \alpha x)^m} + \frac{B_m}{(1 + \beta x)^m} \right). \quad (4.8)$$

Where A_m and B_m are coefficients to be determined. Let us integrate the sum in (4.8):

$$\begin{aligned} \int \frac{x^n}{(1 + a_1x + a_2x^2)^{-q}} dx &= \int \sum_{m=1}^{-q} \left(\frac{A_m}{(1 + \alpha x)^m} + \frac{B_m}{(1 + \beta x)^m} \right) dx \\ &= A_1 \frac{\ln(1 + \alpha x)}{\alpha} + B_1 \frac{\ln(1 + \beta x)}{\beta} + \sum_{m=2}^{-q} \left(\frac{A_m}{\alpha(1 - m)(1 + \alpha x)^{m-1}} + \frac{B_m}{\beta(1 - m)(1 + \beta x)^{m-1}} \right). \end{aligned} \quad (4.9)$$

In other words, the result consists of a sum of rational functions and logarithmic terms. The only scenario in which (4.9) can be algebraic is if $A_1 = B_1 = 0$. The coefficients A_1 and B_1 can be computed either via the generalized residue formula:

$$\begin{aligned} A_1 &= \frac{1}{(-q-1)!} \frac{d^{-q-1}}{dx^{-q-1}} \left[\frac{x^n}{(1 + \beta x)^{-q}} \right] \Big|_{x=-\frac{1}{\alpha}} \neq 0, \\ B_1 &= \frac{1}{(-q-1)!} \frac{d^{-q-1}}{dx^{-q-1}} \left[\frac{x^n}{(1 + \alpha x)^{-q}} \right] \Big|_{x=-\frac{1}{\beta}} \neq 0. \end{aligned}$$

or symbolically, as they naturally arise when integrating rational functions, for instance using Hermite reduction [gG13, Chapter 22]. This is a standard and largely automated procedure in computer algebra systems.

Remark 4.9. Although the case $b_0 = 0$ falls outside the class we are considering, it is still easy to treat. $S(x)$ writes as

$$S(x) = (a_2x^2 + a_1x + 1)^{-1-\frac{b_1}{a_1}} \left(((a_1 + b_1)s_0 + s_1) \int (a_2x^2 + a_1x + 1)^{\frac{b_1}{a_1}} + c_1 \right).$$

Hence, if $\frac{b_1}{a_1} \in \mathbb{N}$ or $\frac{b_1}{a_1} + \frac{3}{2} \in -\mathbb{N}$, then C1 by the analysis performed for $n = 0$. Otherwise, C3 with $(1, -a_1 - b_1)$.

4.2.2 Transcendence in both components.

“Though this be madness, yet there is method in’t.”
— W. Shakespeare, *Hamlet*, Act II, Scene II.

Even if both I_1 and I_2 are transcendental, the linear combination

$$b_0s_0I_1 + (a_1s_0 + b_0s_1 + b_1s_0 + s_1)I_2$$

might still be algebraic. In this section, we provide an algorithm to detect such cases.

1. $\mathbf{q} \notin -\mathbb{N}^*$, $\mathbf{q} + \frac{3}{2} \notin -\mathbb{N}$. Then, by Lemma 4.6:

$$\begin{aligned} I_1 &= c_I \int (1 + a_1x + a_2x^2)^q dx + C_1(x)(1 + a_1x + a_2x^2)^{q+1}, \\ I_2 &= c_{II} \int (1 + a_1x + a_2x^2)^q dx + C_2(x)(1 + a_1x + a_2x^2)^{q+1}. \end{aligned}$$

The transcendence of I_1 and I_2 means two things: $\int (1 + a_1x + a_2x^2)^q dx$ is transcendental, and $c_I, c_{II} \neq 0$. So

$$\begin{aligned} b_0s_0I_1 + (a_1s_0 + b_0s_1 + b_1s_0 + s_1)I_2 &= \\ &= (b_0s_0c_I + (a_1s_0 + b_0s_1 + b_1s_0 + s_1)c_{II}) \underbrace{\int (1 + a_1x + a_2x^2)^q dx}_{\text{transcendental}} + \\ &+ \underbrace{(b_0s_0C_1(x) + (a_1s_0 + b_0s_1 + b_1s_0 + s_1)C_2(x))(1 + a_1x + a_2x^2)^{q+1}}_{\text{algebraic}}. \end{aligned}$$

can only be algebraic if

$$b_0 s_0 c_I + (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) c_{II} = 0,$$

which rewrites as

$$s_1 = \frac{s_0(b_0 c_I + a_1 c_{II} + b_1 c_{II})}{(b_0 + 1) c_{II}}.$$

Hence, we are in **C3** and the algorithm returns $(c_{II}(b_0 + 1), b_0 c_I + a_1 c_{II} + b_1 c_{II})$.

2. $\mathbf{q} + \frac{3}{2} \in -\mathbb{N}$.

Each of I_1, I_2 can be written in the form

$$c \underbrace{\int (1 + a_1 x + a_2 x^2)^q dx}_{\text{algebraic}} + \tilde{c} \underbrace{\int x^{-2q-1} (1 + a_1 x + a_2 x^2)^q dx}_{\text{transcendental}} + \underbrace{C(x)(1 + a_1 x + a_2 x^2)^{q+1}}_{\text{algebraic}},$$

where the constant $\tilde{c}$ will be denoted as $\tilde{c}_1$ (respectively, $\tilde{c}_2$) for I_1 (respectively, I_2). Hence,

$$\begin{aligned} b_0 s_0 I_1 + (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) I_2 \\ = (b_0 s_0 \tilde{c}_1 + (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) \tilde{c}_2) \int x^{-2q-1} (1 + a_1 x + a_2 x^2)^q dx + \mathcal{A}(x). \end{aligned}$$

where $\mathcal{A}(x)$ is some algebraic function. Similarly to the previous case, we are in **C3** and the algorithm returns $(\tilde{c}_2(b_0 + 1), b_0 \tilde{c}_1 + a_1 \tilde{c}_2 + b_1 \tilde{c}_2)$.

3. $\mathbf{q} \in -\mathbb{N}^*$. Then

$$\begin{aligned} I_1 &= A_1^{(1)} \frac{\ln(1 + \alpha x)}{\alpha} + B_1^{(1)} \frac{\ln(1 + \beta x)}{\beta} + \mathcal{A}_1(x), \\ I_2 &= A_1^{(2)} \frac{\ln(1 + \alpha x)}{\alpha} + B_1^{(2)} \frac{\ln(1 + \beta x)}{\beta} + \mathcal{A}_2(x). \end{aligned}$$

where $\mathcal{A}_1(x), \mathcal{A}_2(x)$ are some algebraic functions. So

$$\begin{aligned} b_0 s_0 I_1 + (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) I_2 &= (b_0 s_0 A_1^{(1)} + (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) A_1^{(2)}) \frac{\ln(1 + \alpha x)}{\alpha} \\ &\quad + (b_0 s_0 B_1^{(1)} + (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) B_1^{(2)}) \frac{\ln(1 + \beta x)}{\beta} \\ &\quad + \mathcal{A}(x). \end{aligned}$$

Equivalently,

$$\begin{aligned} b_0 s_0 I_1 + (a_1 s_0 + b_0 s_1 + b_1 s_0 + s_1) I_2 &= (s_0(b_0 A_1^{(1)} + a_1 A_1^{(2)} + b_1 A_1^{(2)}) + s_1 A_1^{(2)}(b_0 + 1)) \frac{\ln(1 + \alpha x)}{\alpha} \\ &\quad + (s_0(b_0 B_1^{(1)} + a_1 B_1^{(2)} + b_1 B_1^{(2)}) + s_1 B_1^{(2)}(b_0 + 1)) \frac{\ln(1 + \beta x)}{\beta} \\ &\quad + \mathcal{A}(x). \end{aligned}$$

which is algebraic if and only if the coefficients multiplying the logarithmic terms are simultaneously equal to zero (we recall that $\alpha \neq \beta$). In other words, one needs

$$\begin{pmatrix} b_0 A_1^{(1)} + (a_1 + b_1) A_1^{(2)} & (b_0 + 1) A_1^{(2)} \\ b_0 B_1^{(1)} + (a_1 + b_1) B_1^{(2)} & (b_0 + 1) B_1^{(2)} \end{pmatrix} \begin{pmatrix} s_0 \\ s_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Hence, we are in **C3** if the matrix

$$\mathcal{M} = \begin{pmatrix} b_0 A_1^{(1)} + (a_1 + b_1) A_1^{(2)} & (b_0 + 1) A_1^{(2)} \\ b_0 B_1^{(1)} + (a_1 + b_1) B_1^{(2)} & (b_0 + 1) B_1^{(2)} \end{pmatrix}$$

is not invertible. Then the algorithm returns a nonzero element of its kernel. Otherwise, the conclusion is **C2**.

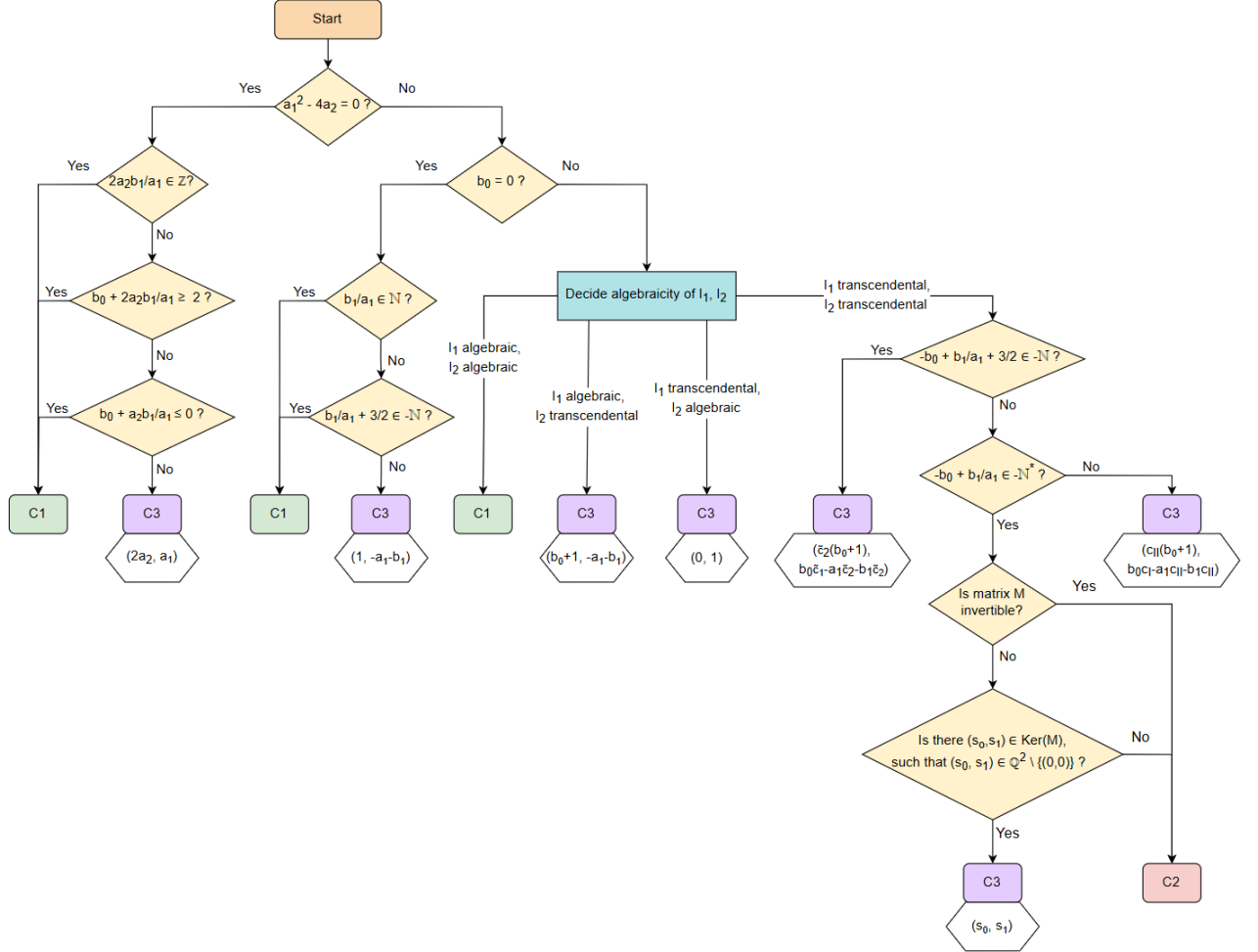


FIGURE 3: Deciding which of the cases **C1**, **C2**, **C3** holds for a recurrence relation of type (4.1) with $b_2 = \frac{2a_2b_1 - a_1a_2b_0}{a_1}$. If the answer is **C3**, a pair $(s_0, s_1) \in \mathbb{Q}^2 \setminus \{(0,0)\}$ such that $S(x)$ is algebraic is provided. The step “Decide algebraicity of I_1, I_2 ” is described in Fig. 2

4.2.3 Examples

1. Motzkin numbers (OEIS A001006).

The recurrence

$$(n+2)m_n = (2n+1)m_{n-1} + (3n-3)m_{n-2}$$

corresponds to

$$b_0 = 2, a_1 = -2, b_1 = -1, a_2 = -3, b_2 = 3.$$

We have $b_2 = \frac{2a_2b_1 - a_1a_2b_0}{a_1}$, $a_1^2 - 4a_2 \neq 0$, $b_0 \in \mathbb{N}^*$. Moreover, $q = -b_0 + \frac{b_1}{a_1} = -\frac{3}{2} \implies q + \frac{3}{2} = 0 \in -\mathbb{N}$.

$$I_1 = \int x(1 - 2x - 3x^2)^{-\frac{3}{2}} dx, \quad I_2 = \int x^2(1 - 2x - 3x^2)^{-\frac{3}{2}} dx.$$

$b_0 - 1 = 1 < -2q - 1 = 2 \implies I_1$ is algebraic.

$b_0 = 2 = -2q - 1 \implies I_2$ is transcendental.

The algorithm returns **C3** with $(b_0 + 1, -(a_1 + b_1)) = (3, 3)$, proving that Motzkin-type sequences are globally bounded if and only if the two first terms coincide.

Alternatively, one can verify the transcendence of any solution with $m_0 \neq m_1$ by initializing the recurrence with, for instance, $m_0 = 0, m_1 = 1$, converting it into a homogeneous differential equation, and then using the *istranscendental* function from Maple's [gfun](#) package (available since version 4.10).

2. Central trinomial coefficients ([OEIS A002426](#)).

The recurrence

$$ns_n = (2n - 1)s_{n-1} + (3n - 3)s_{n-2}$$

corresponds to

$$b_0 = 0, a_1 = -2, b_1 = 1, a_2 = -3, b_2 = 3.$$

We have $b_2 = \frac{2a_2b_1 - a_1a_2b_0}{a_1}$, $a_1^2 - 4a_2 \neq 0$. $b_0 = 0$, so we look at $\frac{b_1}{a_1} = -\frac{1}{2}$. Since $\frac{b_1}{a_1} = -\frac{1}{2} \notin \mathbb{N}$ and $\frac{b_1}{a_1} + \frac{3}{2} = 1 \notin -\mathbb{N}$, the algorithm returns **C3** with $(1, -(a_1 + b_1)) = (1, 1)$.

This is consistent with the fact that $s_0 = s_1 = 1$ yields not just an almost integral, but an integer sequence of the Central trinomial coefficients. In particular, we deduce that the only integer sequences satisfying the recurrence for the central trinomial coefficients are those with $s_0 = s_1$, mirroring the case of Motzkin numbers.

3. Large Schröder numbers ([OEIS A006318](#)) : fully algebraic example.

The recurrence

$$(n + 1)s_n = (6n - 3)s_{n-1} - (n - 2)s_{n-2}$$

corresponds to

$$b_0 = 1, a_1 = -6, b_1 = 3, a_2 = 1, b_2 = -2.$$

We have $b_2 = \frac{2a_2b_1 - a_1a_2b_0}{a_1}$, $a_1^2 - 4a_2 \neq 0$, $b_0 \in \mathbb{N}^*$. Moreover, $q = -b_0 + \frac{b_1}{a_1} = -\frac{3}{2} \implies q + \frac{3}{2} \in \mathbb{N}$.

$$I_1 = \int (1 - 6x + x^2)^{-\frac{3}{2}} dx, \quad I_2 = \int x(1 - 6x + x^2)^{-\frac{3}{2}} dx.$$

$b_0 - 1 = 0 < -2q - 1 = 2 \implies I_1$ is algebraic.

$b_0 = 1 = -2q - 1 \implies I_2$ is algebraic.

The algorithm returns **C1**.

4. Fully transcendental example.

Consider the following innocent-looking recurrence:

$$(n + 3)s_n = (n + 1)s_{n-1} + (2n - 2)s_{n-2}.$$

It corresponds to

$$b_0 = 3, a_1 = -1, b_1 = -1, a_2 = -2, b_2 = 2.$$

We have $b_2 = \frac{2a_2b_1 - a_1a_2b_0}{a_1}$, $a_1^2 - 4a_2 \neq 0$, $b_0 \in \mathbb{N}^*$ and $q = -b_0 + \frac{b_1}{a_1} = -2 \in -\mathbb{N}^*$.

$$I_1 = \int \frac{x^2}{(1-x-2x^2)^2} dx = \frac{2}{27} \ln(2x-1) - \frac{2}{27} \ln(x+1) - \frac{1}{18(2x-1)} - \frac{1}{9(x+1)},$$

$$I_2 = \int \frac{x^3}{(1-x-2x^2)^2} dx = \frac{7}{108} \ln(2x-1) + \frac{5}{27} \ln(x+1) - \frac{1}{36(2x-1)} + \frac{1}{9(x+1)}.$$

So,

$$A_1^{(1)} = \frac{2}{27}, \quad B_1^{(1)} = -\frac{2}{27}, \quad A_1^{(2)} = \frac{7}{108}, \quad B_1^{(2)} = \frac{5}{27}.$$

None of these numbers are equal to zero, in particular this implies that both I_1 and I_2 are transcendental. Moreover, the matrix

$$\begin{pmatrix} b_0A_1^{(1)} + (a_1 + b_1)A_1^{(2)} & (b_0 + 1)A_1^{(2)} \\ b_0B_1^{(1)} + (a_1 + b_1)B_1^{(2)} & (b_0 + 1)B_1^{(2)} \end{pmatrix} = \begin{pmatrix} 5/54 & 7/27 \\ -16/27 & 20/27 \end{pmatrix}.$$

is invertible, as its determinant equals $\frac{2}{9} \neq 0$. Hence, no choice of s_0, s_1 can make $S(x) = \sum_{n \geq 0} s_n x^n$ algebraic. The algorithm returns **C2**.

5 Integrality revisited

“Coming back to where you started is not the same as never leaving.”
— T. Pratchett, *A Hat Full of Sky*.

Although proving algebraicity does not directly address the problem of finding integer solutions to a recurrence, it often greatly simplifies the analysis. In the specific setting of second-order recurrences of the form

$$(n + b_0)s_n + (a_1n + b_1)s_{n-1} + (a_2n + b_2)s_{n-2} = 0$$

with $b_2 = \frac{2a_2b_1 - a_1a_2b_0}{a_1}$, integrality implies global boundedness, which in turn is equivalent to algebraicity. Thus, the problem of deciding whether a solution is integral reduces to detecting algebraicity — something that can be done using the algorithm developed in Section 4 (see Fig. 3). Analysing integrality within this algebraic framework is generally easier.

For instance, if an algebraic solution has a closed-form expression involving rational powers of polynomials, its integrality can be established using the criteria derived by Pomerat and Straub in [PS24] (Theorem 1.1 and multiple examples applied to specific cases).

Example. Consider a recurrence relation:

$$ns_n + (2n + 3)s_{n-1} + 9(n + 3)s_{n-2} = 0. \quad (5)$$

Applying our algorithm (note that $b_0 = 0$), we find that $S(x) = \sum_{n \geq 0} s_n x^n$ is globally bounded if and only if $s_1 = -5s_0$. In that case, one has

$$S(x) = s_0(1 + 2x + 9x^2)^{-\frac{5}{2}}.$$

Clearly, all integral solutions are among the globally bounded ones. Thus, the problem reduces to identifying the values of s_0 for which $s_0(1 + 2x + 9x^2)^{-\frac{5}{2}}$ lies in $\mathbb{Z}[[x]]$.

Proposition 5.1. *If $s_0 \in \mathbb{Z}$, then*

$$s_0(1 + 2x + 9x^2)^{-\frac{5}{2}} \in \mathbb{Z}[[x]].$$

Proof. The claim is a direct consequence of the following result:

Lemma 5.2. ([PS24, Example 1.8]) *Let $a, b \in \mathbb{Z}$, $\lambda \in \mathbb{Q}$. Let k be the denominator of λ brought to the lowest terms. Then $(1 + ax + bx^2)^\lambda \in \mathbb{Z}[[x]]$ if and only if*

- $a, b \in k \operatorname{rad}(k)\mathbb{Z}$, or
- $k = 2\kappa$ and $a, b \in \kappa \operatorname{rad}(\kappa)\mathbb{Z}$ as well as $(a, b) \equiv (2, 1) \pmod{4}$.

Here $\operatorname{rad}(k)$ denotes the largest squarefree integer dividing k (for example, $\operatorname{rad}(24) = 6$).

In our setting $a = 2, b = 9, \lambda = -\frac{5}{2}, k = 2, \kappa = 1$. Since $2, 9 \in \mathbb{Z}$ and $(2, 9) \equiv (2, 1) \pmod{4}$ we get $(1 + 2x + 9x^2)^{-\frac{5}{2}} \in \mathbb{Z}[[x]]$. Hence, for any $s_0 \in \mathbb{Z}$, $s_0(1 + 2x + 9x^2)^{-\frac{5}{2}} \in \mathbb{Z}[[x]]$. $\square$

Since taking $s_0 \notin \mathbb{Z}$ would already violate the condition $s_n \in \mathbb{Z}$ for all $n \in \mathbb{N}$, we conclude that $S(x) \in \mathbb{Z}[[x]]$ if and only if $s_1 = -5s_0$ and $s_0 \in \mathbb{Z}$.

The next figure summarizes the classification of solutions according to their arithmetic properties. Two solutions are called *independent* if one is not a rational multiple of the other, which in our case translates to the condition $s_0\tilde{s}_1 \neq \tilde{s}_0s_1$. Remarkably, all six areas in the figure below are nonempty: such recurrence relations occur in practice.

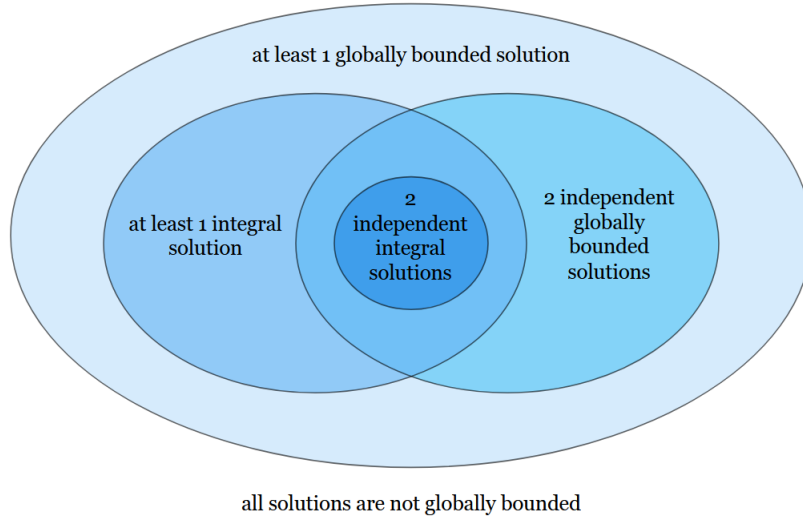


FIGURE 4: Classification of the solution space of second-order linear recurrences with polynomial coefficients, according to integrality and global boundedness properties.

6 Integrality criteria for small Apéry numbers

In this section, we investigate the sequences $(a_n)_n$ satisfying the recurrence relation

$$n^2 a_n = (11n^2 - 11n + 3)a_{n-1} + (n-1)^2 a_{n-2}, \quad n \geq 2. \quad (5.1)$$

In the standard setting, this recurrence is initialized with $a_0 = 1, a_1 = 3$, yielding the sequence of small Apéry numbers (OEIS A005258). These numbers also admit a binomial sum representation:

$$a_n = \sum_{k=0}^n \binom{n}{k}^2 \binom{n+k}{k}, \quad (5.2)$$

which makes it clear that they are all integers. However, the situation becomes more complex when considering arbitrary initial conditions. Let $(a(\mu, \lambda)_n)_n$ denote the solution of (5.1) with $a_0 = \mu, a_1 = \lambda$. The behaviour of these generalized Apéry-type sequences is described in the following theorem, which, to the best of our knowledge, is new, though similar to the case of big Apéry numbers (OEIS A005259).

Theorem 6.1. *A sequence of rational numbers $(a_n)_n$ satisfying (5.1) is integral if and only if $(a_0, a_1) = (\alpha, 3\alpha)$ for $\alpha \in \mathbb{Z}$.*

Proof. Let us first show that $a(1, 3)_n \in \mathbb{Z}$ for all $n \geq 0$. Obviously, this would imply that $a(\alpha, 3\alpha)_n \in \mathbb{Z}$ for all $n \geq 0, \alpha \in \mathbb{Z}$.

The sequence $(\tilde{a}_n)_n$ defined by $\tilde{a}_n = \sum_{k=0}^n \binom{n}{k}^2 \binom{n+k}{k}$ is a solution to (5.1) (a result due to Apéry).

While coming up with such a relation is nontrivial, its verification is straightforward with creative telescoping techniques [Zei91]. In Maple this can be done by calling *Zeilberger* function from the *SumTools/Hypergeometric* package, which applies Zeilberger's algorithm to find a recurrence satisfied by a given binomial sum.

```
> A := binomial(n, k)^2 * binomial(n + k, k):
> simplify(SumTools[Hypergeometric][Zeilberger](A, n, k, a_n)[1]);
```

$$(n+2)^2 a_n^2 + (-11n^2 - 33n - 25)a_n - (n+1)^2,$$

where the powers of a_n correspond to the coefficient shift, i.e. the expression above is equivalent to

$$(n+2)^2 a_{n+2} - (11(n+2)^2 - 11(n+2) + 3)a_{n+1} - (n+1)^2 a_n = 0, \quad n \geq 0$$

which is simply a shift of (5.1).

In addition, every $\tilde{a}_n$ is a sum of binomials, hence an integer. We check the first two values:

$$\begin{aligned} \tilde{a}_0 &= \binom{0}{0}^2 \binom{0}{0} = 1, \\ \tilde{a}_1 &= \binom{1}{0}^2 \binom{1}{0} + \binom{1}{1}^2 \binom{2}{1} = 1 + 2 = 3. \end{aligned}$$

Since any solution of (5.1) is uniquely defined by a pair of initial values (a_0, a_1) , we deduce that $(a(1, 3)_n)_n = (\tilde{a}_n)_n$. Thus, $(a(1, 3)_n)_n$ is indeed integral. $\square$

We now present two proofs that establish the uniqueness of this integral solution.

6.1 Proof 1

For every $n \geq 2$

$$a_n = \frac{(11n^2 - 11n + 3)}{n^2} a_{n-1} + \frac{(n-1)^2}{n^2} a_{n-2}. \quad (5.3)$$

Consider two sequences $(u_n)_n$ and $(v_n)_n$ defined by

$$u_n = \frac{(11n^2 - 11n + 3)}{n^2}, \quad v_n = \frac{(n-1)^2}{n^2}$$

for all $n \geq 0$. Then (5.3) becomes

$$a_n = u_n a_{n-1} + v_n a_{n-2}. \quad (5.4)$$

The following result by Chang [Cha84] proves to be remarkably useful in our setting.

Theorem 6.2. ([Cha84]): Suppose that

- a) For all integers $n > 2$, $v_n \neq 0$.
- b) The limit $\lim_{n \rightarrow \infty} \prod_{k=2}^n |v_k|$ exists.
- c) The recurrence relation (5.4) has two linearly independent integer solutions.

Then $|v_n| = 1$ for all large n .

Theorem 6.2 is naturally applicable to certain 2-order P-recursive sequences. In particular, it can be used to prove the uniqueness of an integer solution should one exist. The following reasoning mimics the one for big Apéry numbers in [Cha84], presented as an application of Theorem 6.1 in [Cha84].

Let us prove the $\Rightarrow$ direction of Theorem 6.1 by contradiction. Assume there exist $\mu, \lambda \in \mathbb{Z}$ with $\lambda \neq 3\mu$ such that $a(\mu, \lambda)_n \in \mathbb{Z}$ for all $n \in \mathbb{N}$. Then all the conditions of Theorem 6.2 are satisfied. Indeed, for all $k > 2$

$$v_k = \frac{(k-1)^2}{k^2} \neq 0$$

and

$$\lim_{n \rightarrow \infty} \prod_{k=2}^n |v_k| = \lim_{n \rightarrow \infty} \prod_{k=2}^n \frac{(k-1)^2}{k^2} = \lim_{n \rightarrow \infty} \frac{1}{n^2} = 0.$$

By Theorem 6.2, $|v_n|$ must be equal to 1 for all $n > N$ for some $N \in \mathbb{N}$. This is clearly not the case since $v_n = \frac{(n-1)^2}{n^2}$, hence a contradiction. Therefore, (5.1) has at most 1 integral solution, and we have already seen that it is given by $(a(1, 3)_n)_n$.

6.2 Proof 2

This proof is largely inspired by the work of André-Jeannin [AJ91] on the integrality of big Apéry numbers.

Again, we argue by contradiction. Assume there exist $\mu, \lambda \in \mathbb{Z}$ with $\lambda \neq 3\mu$ such that $a(\mu, \lambda)_n \in \mathbb{Z}$ for all $n \in \mathbb{N}$. By the linearity of (5.1) this implies the integrality of the sequence

$$(a(0, \lambda - 3\mu)_n)_n = (a(\mu, \lambda)_n)_n - \mu(a(1, 3)_n)_n.$$

In other words, there exists an integer $c \neq 0$ such that for all $n \in \mathbb{N}$ $a(0, c)_n \in \mathbb{Z}$. To avoid notational ambiguity, we introduce a new sequence $(b_n)_n = (a(0, c)_n)_n$, and write $(a_n)_n = (a(1, 3)_n)_n$ for the sequence of small Apéry numbers.

Recurrence (5.3) for $(b_n)_n$ can be rewritten as

$$b_n - b_{n-1} = \frac{10n^2 - 11n + 3}{n^2} b_{n-1} + \frac{(n-1)^2}{n^2} b_{n-2} = \frac{10(n - \frac{3}{5})(n - \frac{1}{2})}{n^2} b_{n-1} + \frac{(n-1)^2}{n^2} b_{n-2}. \quad (5.5)$$

Now,

$$\frac{10(n - \frac{3}{5})(n - \frac{1}{2})}{n^2} > 0, \quad \frac{(n-1)^2}{n^2} > 0 \quad \text{for } n \geq 2 \text{ and } b_1 > b_0 = 0,$$

so that the RHS of (5.5) is strictly positive for all $n \geq 2$. Consequently, the sequence $(b_n)_n$ is increasing, and so is $(a_n)_n$ by the exact same reasoning. Moreover, $a_n > 0$ and $b_n > 0$ for $n \geq 1$.

We compute $b_2 = \frac{25}{4} b_1$. Therefore, $b_1 = \frac{b_1}{a_1} < \frac{b_2}{a_2} = \frac{25}{12} b_1$. The next step is to prove

$$\frac{b_{n-1}}{a_{n-1}} < \frac{b_n}{a_n}. \quad (5.6)$$

for all $n > 2$.

Equivalently to (5.1),

$$(11n^2 - 11n + 3)a_{n-1} = n^2a_n - (n-1)^2a_{n-2}. \quad (5.7)$$

Let $\lambda_i = \frac{b_i}{a_i}$, $i \geq 1$. If $\lambda_{n-2} < \lambda_n$, then from (5.7) we have

$$(11n^2 - 11n + 3)\lambda_{n-1}a_{n-1} = n^2\lambda_na_n - (n-1)^2\lambda_{n-2}a_{n-2} < \lambda_n(n^2a_n - (n-1)^2a_{n-2}),$$

and

$$(11n^2 - 11n + 3)\lambda_{n-1}a_{n-1} > \lambda_{n-2}(n^3a_n - (n-1)^3a_{n-2}).$$

Hence,

$$\lambda_{n-2} < \lambda_n \implies \lambda_{n-2} < \lambda_{n-1} < \lambda_n.$$

Similarly,

$$\lambda_{n-2} \geq \lambda_n \implies \lambda_{n-2} \geq \lambda_{n-1} \geq \lambda_n.$$

Therefore, $\lambda_{n-2} < \lambda_{n-1} \implies \lambda_{n-1} < \lambda_n$. Since (5.6) holds for $n = 2$, it also holds for all $n \geq 2$.

Substituting $n+1$ in (5.7) and dividing by the same equation written for $(b_n)_n$ yields

$$\frac{b_n}{a_n} = \frac{(n+1)^2b_{n+1} - n^2b_{n-1}}{(n+1)^2a_{n+1} - n^2a_{n-1}}.$$

which up to rearrangement of terms is equivalent to

$$(2n+1) \left(\frac{b_{n+1}}{a_{n+1}} - \frac{b_n}{a_n} \right) a_n a_{n+1} = n^2((a_n b_{n-1} - b_n a_{n-1}) - (a_n b_{n+1} - b_n a_{n+1})). \quad (5.8)$$

The LHS of (5.8) is positive by (5.6). Hence, $a_n b_{n+1} - b_n a_{n+1} < a_n b_{n-1} - b_n a_{n-1}$ for all $n \geq 2$. We deduce from (5.6) that $a_n b_{n-1} - b_n a_{n-1} < 0$ for all $n \geq 2$. Thus,

$$a_n b_{n+1} - b_n a_{n+1} < 0 \quad (5.9)$$

for all $n \geq 2$. But (5.6) directly implies

$$a_n b_{n+1} - b_n a_{n+1} > 0 \quad (5.10)$$

for all $n \geq 2$. The proof follows by observing that (5.9) and (5.10) are incompatible.

7 Concluding remarks

The story, however, is far from complete, as several questions remain open. The case of $b_2 \neq \frac{2a_2b_1 - a_1a_2b_0}{a_1}$ is still poorly understood, although we believe it leads to not globally bounded solutions. Moreover, our algorithm is currently not applicable to the recurrences with $b_0 \notin \mathbb{N}$. Beyond that, higher-order recurrences or those with nonlinear polynomial coefficients present further challenges. Their solution spaces are more complex, and the connection between global boundedness and algebraicity becomes harder to characterise.

Although our main goal was to decide algebraicity and global boundedness, the original question of integrality motivated much of this work. By revealing a structural connection between these notions for second-order recurrences, we were able to approach parts of the integrality problem via concepts from differential algebra. This perspective not only simplifies the analysis, but also suggests a path toward a more systematic, algorithmic treatment of arithmetic properties in holonomic sequences.

References

- [AJ91] R. Andre-Jeannin. Sequences of integers satisfying recurrence relations. *The Fibonacci Quarterly*, 29(3):205–208, 1991.
- [And89] Y. André. *G-Functions and Geometry*, volume E13 of *Aspects of Mathematics*. Friedr. Vieweg & Sohn, Braunschweig, 1989.
- [Apé79] R. Apéry. Irrationalité de $\zeta(2)$ et $\zeta(3)$. *Astérisque*, (61):11–13, 1979.
- [Cha84] D. Chang. A note on Apéry numbers. *The Fibonacci Quarterly*, 22(2):178–180, 1984.
- [Chr90] G. Christol. Globally bounded solutions of differential equations. In K. Nagasaka and E. Fouvry, editors, *Analytic Number Theory*, volume 1434 of *Lecture Notes in Mathematics*, pages 45–64. Springer, 1990.
- [Eis52] G. Eisenstein. Über eine allgemeine Eigenschaft der Reihenentwicklungen aller algebraischen Funktionen. *Berichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, pages 441–443, 1852.
- [Fur67] H. Furstenberg. Algebraic functions over finite fields. *Journal of Algebra*, 7(2):271–277, 1967.
- [FY24] F. Fürnsinn and S. Yurkevich. Algebraicity of hypergeometric functions with arbitrary parameters. *Bull. Lond. Math. Soc.*, 56(9):2824–2846, 2024.
- [gG13] J. von zur Gathen and J. Gerhard. *Modern computer algebra*. Cambridge University Press, Cambridge, third edition, 2013.
- [Hei53] E. Heine. Der Eisensteinsche Satz über Reihen-Entwicklung algebraischer Functionen. *Journal für die Reine und Angewandte Mathematik*, 45:285–302, 1853.
- [Hei54] E. Heine. Über die Entwicklung von Wurzeln algebraischer Gleichungen in Potenzreihen. *Journal für die Reine und Angewandte Mathematik*, 48:267–275, 1854.
- [KL05] M. Klazar and F. Luca. On integrality and periodicity of the Motzkin numbers. *Aequationes Math.*, 69(1-2):68–75, 2005.
- [Lüt12] J. Lützen. *Joseph Liouville 1809–1882: Master of pure and applied mathematics*, volume 15. Springer Science & Business Media, 2012.
- [PS24] J. Pomerat and A. Straub. Criteria for the integrality of n th roots of power series. *Acta Arith.*, 215(1):1–10, 2024.
- [Pó22] G. Pólya. Sur les séries entières, dont la somme est une fonction algébrique. *Enseign. Math.*, 22:38–47, 1921/1922.
- [Sta80] R. P. Stanley. Differentiably finite power series. *European J. Combin.*, 1(2):175–188, 1980.
- [Zei91] D. Zeilberger. The method of creative telescoping. *J. Symbolic Comput.*, 11(3):195–204, 1991.