Riv. Mat. Univ. Parma, Vol. **13** (2022), 1-18

Roberto Dvornicich, Francesco Veneziano
and Umberto Zannier

On the integral values of a curious recurrence

Dedicated to Roberto Dvornicich
on the occasion of his seventieth birthday

Abstract. We discuss an elementary problem, initially proposed for the Romanian Mathematical Olympiad, which leads to interesting remarks of various nature. We relate the problem to the theory of linear recurrence sequences with non-constant coefficients and their p-adic behaviour. Our considerations can be applied to a larger set of similarly-defined recurrences.

Keywords. Mathematical Olympiad, Recurrence relation, Integral values.

Mathematics Subject Classification: 11B83, 11B99, 11D99.

This note arises from an investigation carried out by the three authors several years ago, out of a shared interest for elementary mathematical problems and their non-elementary interpretations. It has been a pleasure for the last two authors to revisit this topic on this special occasion and to dedicate this expanded version to Roberto.

1 - Introduction

The purpose of this note is to comment on a problem shortlisted for the Romanian Mathematical Olympiad 2010 (see [**RMS**, Shortlisted Problems for the 61st NMO, No. 16]).

Received: February 24, 2021; accepted in revised form: April 7, 2021.

Problem. *Let $x_0, x_1, x_2, \dots$ be the sequence defined by*

$$(1)\qquad \begin{aligned} x_0 &= 1 \\ x_{n+1} &= 1 + \frac{n}{x_n}, \qquad \forall n \geq 0. \end{aligned}$$

What are the values of n for which x_n is an integer?

We list here the first few values of the sequence:

$$(2)\qquad 1, 1, 2, 2, \frac{5}{2}, \frac{13}{5}, \frac{38}{13}, \frac{58}{19}, \frac{191}{58}, \frac{655}{191}, \dots$$

The author of this problem is Gheorghe Iurea, but his solution does not appear in the booklet mentioned above.

We have found this problem of interest, not only in itself, but also because *a posteriori* it may be approached in different ways, each of which involves mathematical arguments of various nature. We will give two proofs of the following:

Theorem 1. *$x_n \in \mathbb{Z}$ if and only if $n = 0, 1, 2, 3$.*

One proof, in Section 2, relates the sequence (x_n) to a second-order linear recurrence of combinatorial significance; the theorem follows from some arithmetical properties of this sequence.

A second proof, in Section 4, takes the point of view of real dynamics, with arithmetics only playing a role at the very end of the argument.

In fact, the first proof gives a lower bound for the reduced denominator of the sequence (x_n), which we refine in Section 3.2 through a careful computation of its 2-adic valuation, showing the following theorem:

Theorem 1′. *For every $n \geq 1$ we have*

$$D_n \geq \frac{\sqrt{(n-1)!}}{2^{\frac{n+1}{4}}},$$

where D_n is the reduced denominator of x_n.

The proof of this bound is entirely elementary; compare it with the actual order of growth, which can be shown (see Section 3.2) to be

$$D_n \asymp \frac{e^{\sqrt{n}}}{n^{1/4}} \frac{\sqrt{(n-1)!}}{2^{n/4}}.$$

As a natural generalization of the original problem, we consider also sequences arising from the same recursion and different initial values.

Let us fix a real value $X > 0$, and consider the sequence defined by the recurrence

$$\begin{aligned} X_1 &= X \\ X_{n+1} &= 1 + \frac{n}{X_n}, \qquad \forall n \geq 1. \end{aligned} \tag{3}$$

With a variant of the second proof and a representation of X_n as a linear fractional function of X we are able to prove the following theorem in Section 5.1.

Theorem 2. *For every $X > 0$ and every $n \geq 14$, the value X_n is not an integer.*

For every $X \neq \frac{4}{51}, \frac{1}{3}, \frac{2}{3}, 1$ and every $n \geq 3$, the value X_n is not an integer.

Finally, we also considered the question of the growth of the reduced denominator of the more general sequence (X_n).

When trying to generalize Theorem $1'$ to the sequence (X_n) for any rational value of X we studied a new recurrence sequence, seeking to replicate the arithmetical arguments of Section 2. This leads, however, to much more complicated problems about the p-adic values of certain converging series. Section 3.3 contains these observations and some open questions on the matter.

Nevertheless we were able to modify the second proof to show in Section 5.2 the following theorem:

Theorem 3. *Let X be a positive rational number. Then for every positive integer k there exists an effective constant $C_{k,X}$ such that for every $n \geq 1$ the reduced denominator of X_n is at least $n^k - C_{k,X}$.*

The proof of this theorem hinges on the existence of a certain asymptotic expansion, which we derive by a general theory of linear recurrences due to Birkhoff and Trjitzinsky.

Some numerical experiments and considerations on the matrices appearing in Section 5.1 would suggest that lower bounds of the same order as those for D_n should hold for any initial value, but in order to reach this precision we would need to overcome the arithmetic difficulties outlined in Section 3.3.

2 - First solution

To study the sequence (x_n) from an arithmetic point of view we define two integer sequences $(a_n), (b_n)$ by the recurrences

$$\begin{aligned} a_0 &= 1 \\ b_0 &= 1 \\ a_{n+1} &= a_n + nb_n, && \forall n \geq 0 \qquad (4)\\ b_{n+1} &= a_n, && \forall n \geq 0. \qquad (5)\end{aligned}$$

Comparing (4) and (5) with (1) we see immediately that they satisfy $x_n = \dfrac{a_n}{b_n}$, so a_n, b_n are the numerator and denominator in some fractional representation of x_n; however a_n, b_n *a priori* need not be coprime, so the said fraction can be possibly simplified.

We also see that b_n may be eliminated from the recurrence to get

$$\begin{aligned} a_0 &= 1 \\ a_1 &= 1 \\ a_{n+2} &= a_{n+1} + (n+1)a_n, && \forall n \geq 0, \qquad (6)\end{aligned}$$

with $x_n = \dfrac{a_n}{a_{n-1}}$.

Let us define $d_n = \gcd(a_n, a_{n-1})$; d_n measures how much the reduced denominator of x_n differs from a_{n-1}. In order to show that $x_n \notin \mathbb{Z}$ we seek a lower bound for a_{n-1} and an upper bound for d_n.

Remark 4. By the recurrence (6) we see that $d_{n+1} | a_{n+2}$, and so $d_{n+1} | d_{n+2}$; this will be helpful in establishing an upper bound for d_n.

A lower bound for a_n is easily obtained in the following lemma.

Lemma 5. *For every $n \geq 0$ we have $a_n \geq \sqrt{n!}$.*

Proof. We argue by induction on $n \geq 0$. We check that $a_0 = 1 = \sqrt{0!}$, $a_1 = 1 = \sqrt{1!}$, and assuming the bound for a_n and a_{n+1} we get

$$\begin{aligned} a_{n+2} &= a_{n+1} + (n+1)a_n \geq \sqrt{(n+1)!} + (n+1)\sqrt{n!} = \\ &= \sqrt{(n+2)!}\frac{1+\sqrt{n+1}}{\sqrt{n+2}} = \sqrt{(n+2)!}\sqrt{1 + \frac{2\sqrt{n+1}}{n+2}} \geq \sqrt{(n+2)!}. \quad \square \end{aligned}$$

To get an upper bound for d_n we introduce the exponential generating function of the sequence $(a_n)_{n\in\mathbb{N}}$, namely

$$F(x) = \sum_{n=0}^{\infty} \frac{a_n}{n!} x^n.$$

We consider $F(x)$ merely as a formal power series, although one could prove that it converges for every complex value of x.

From the recurrence on (a_n) we can obtain a differential equation for F; in fact, we can multiply (6) by $\frac{x^n}{n!}$ and sum it for $n \geq 0$; since clearly $F'(x) = \sum_{n=0}^{\infty} \frac{a_{n+1}}{n!} x^n$ and $F''(x) = \sum_{n=0}^{\infty} \frac{a_{n+2}}{n!} x^n$, we obtain that F satisfies the conditions

$$\begin{cases} F(0) & = 1 \\ F'(0) & = 1 \\ F''(x) & = (x+1)F'(x) + F(x). \end{cases} \tag{7}$$

The Cauchy problem (7) may be solved (in the ring of formal power series) to get

$$F(x) = e^{x+\frac{x^2}{2}},$$

and we can use this explicit form to get a formula for a_n. In fact

$$\sum_{n=0}^{\infty} \frac{a_n}{n!} x^n = e^{x+\frac{x^2}{2}} = e^x e^{\frac{x^2}{2}} = \sum_{m=0}^{\infty} \frac{x^m}{m!} \sum_{s=0}^{\infty} \frac{x^{2s}}{2^s s!}$$

$$\frac{a_n}{n!} = \sum_{2s+m=n} \frac{1}{2^s s!} \frac{1}{m!}$$

$$a_n = \sum_{2s\leq n} \frac{n!}{2^s s!(n-2s)!} = \sum_{2s\leq n} \binom{n}{2s} (2s-1)!!, \tag{8}$$

where the *semifactorial* $(2s-1)!!$ denotes as usual the product $(2s-1)\cdot(2s-3)\cdots 3\cdot 1$ and is defined to be 1 for $s = 0$. Here and in the following we always assume the standard convention that $\binom{n}{k} = 0$ if $k < 0$ or $k > n$.

From formula (8) we can derive arithmetic informations about the sequence a_n.

Lemma 6. *Let p be and odd prime. Then the congruence*

$$a_{m+p} \equiv a_m \pmod{p}$$

holds for all $m \geq 0$.

In particular, if $p|n$ *then* $a_n \equiv 1 \pmod{p}$.

Proof. By formula (8) we have

$$a_m = \sum_s \binom{m}{2s}(2s-1)!! \qquad a_{m+p} = \sum_s \binom{m+p}{2s}(2s-1)!!$$

If $2s < p$ then $\binom{m+p}{2s} \equiv \binom{m}{2s} \pmod{p}$, and the corresponding summands in both sums have the same remainder modulo p.

If $p < 2s$, then $p|(2s-1)!!$, as p itself appears as one of the factors in the definition of $(2s-1)!!$. Then the corresponding summands in both sums are equal to zero modulo p.

This shows the first assertion. If we now set $m = 0$ we obtain that

$$a_p \equiv a_0 = 1 \pmod{p},$$

and so on by induction for any multiple of p. □

As a corollary we learn about the prime factors of d_n.

Corollary 7. *For every* $n \geq 1$, d_n *is a power of* 2.

Proof. If an odd prime p divides d_m for some $m \geq 1$, then, by Remark 4, p divides d_n (and hence a_n) for all $n \geq m$, so also for $n = pm$. But this is not possible because $a_{pm} \equiv 1 \pmod{p}$ by Lemma 6. □

We are now ready to prove an upper bound for d_n, which we obtain using again the exponential generating function $F(x)$.

Proposition 8. *For every* $n \geq 1$ *we have that* $d_n \leq 2^{n-1}$.

Proof. We have the following identities concerning the above generating function $F(x)$:

$$\left(\sum_{m=0}^{\infty} a_m \frac{x^m}{m!}\right)\left(\sum_{r=0}^{\infty}(-1)^r a_r \frac{x^r}{r!}\right) = F(x)F(-x) = e^{x^2} = \sum_{n=0}^{\infty} \frac{x^{2n}}{n!}.$$

Comparing the coefficients of x^{2n} for any $n \geq 1$, we obtain

$$\sum_{m+r=2n} (-1)^r \binom{2n}{m} a_m a_r = \frac{(2n)!}{n!} = 2^n \cdot (2n-1)!!. \tag{9}$$

Now, as observed in Remark 4 above, d_{n+1} divides any a_m with $m \geq n$, so it divides the left-hand side of (9), and we know from the Corollary 7 that it is a power of 2, therefore $d_{n+1} \leq 2^n$ for $n \geq 1$; of course $d_1 = 1 = 2^0$. □

To conclude the proof of Theorem 1, denote by D_n the *reduced* denominator of x_n. We have:

$$D_n = \frac{a_{n-1}}{d_n} \geq \frac{\sqrt{(n-1)!}}{2^{n-1}} \qquad \forall n \geq 1. \tag{10}$$

It is easily seen that

$$\sqrt{(n-1)!} > 2^{n-1} \qquad \forall n \geq 10,$$

so we are left to inspect the values of x_n with $0 \leq n \leq 9$, which are exactly the values listed in (2).

3 - Further observations to the first solution

The sequence a_n and its exponential generating function $F(x)$ are widely known in the literature for their combinatorial significance: the sequence a_n counts the number of involutions in the symmetric group S_n (see for instance [**Wil06**, Thm 3.16] and sequence A000085 in [**OEIS**]). With analytic tools — for instance, Cauchy integrals — it can be shown (see [**MW56**]) from the generating function $F(x)$ that

$$a_n \sim 2^{-\frac{1}{2}} e^{-\frac{1}{4}} e^{\sqrt{n}} \left(\frac{n}{e}\right)^{\frac{n}{2}} \sim (8\pi e n)^{-\frac{1}{4}} e^{\sqrt{n}} \sqrt{n!}. \tag{11}$$

3.1 - *A combinatorial proof of Corollary 7*

The combinatorial interpretation of a_n as the cardinality of

$$A_n = \left\{\sigma \in S_n \mid \sigma^2 = e\right\}$$

suggests a different way to prove Corollary 7. If $n = pm$, consider the subgroup $G < S_n$ generated by the cycles $(1, \ldots, p), (p+1, \ldots, 2p), \ldots, ((m-1)p+1, \ldots, n)$. G has cardinality p^m, and it acts on A_n by conjugation.

It is not difficult to see that the identity is the only fixed point of this action. Indeed take $\sigma \in A_n$ not the identity, and let (ij) with $i < j$ be one of the transpositions appearing in the cycle decomposition of σ. If $i \leq kp < j$ for some integer k, then σ is not fixed by $((k-1)p + 1, \ldots, kp)$. Similarly, if $(k-1)p + 1 \leq i < j \leq kp$, then conjugating σ by $((k-1)p + 1, \ldots, kp)$ and its powers cannot yield disjoint transpositions because p is odd, and again σ cannot be a fixed point.

This shows that no element of A_n other than the identity is fixed by G, and therefore $a_n \equiv 1 \pmod p$, because when a p-group acts on a finite set, the cardinality of the set is congruent modulo p to the number of fixed points of the action.

3.2 - *The 2-adic valuation of a_n*

We will now compute the exact power of 2 which divides a_n.

Proposition 9. *For every $n \geq 0$, 2^{e_n} is the highest power of 2 dividing a_n, where*

$$e_n = \begin{cases} k & \text{if } n = 4k \\ k & \text{if } n = 4k+1 \\ k+1 & \text{if } n = 4k+2 \\ k+2 & \text{if } n = 4k+3. \end{cases}$$

Proof. Setting $q_n := a_n/2^{e_n}$, we need to prove that q_n is an odd integer for $n \geq 0$.

We begin with the following lemma.

Lemma 10. *For every $n \geq 2$ we have*

$$a_{n+6} = 2(n^2 + 9n + 19)a_{n+2} - n(n-1)(n+2)(n+5)a_{n-2}. \tag{12}$$

Proof. Indeed, since the sequence (a_n) verifies a linear recurrence of the *second* order, any *three* sequences of the shape $(a_n), (a_{n+r}), (a_{n+s})$ are linearly related by an equation with coefficients which are polynomials in n; they may be found by easy elimination. Presently, we are interested in the case $r = 4, s = 8$, where this elimination is hidden in the following explicit calculations:

$$\begin{aligned} a_{n+6} &= a_{n+5} + (n+5)a_{n+4} = (n+6)a_{n+4} + (n+4)a_{n+3} \\ &= (2n+10)a_{n+3} + (n+6)(n+3)a_{n+2} \\ &= (n^2+11n+28)a_{n+2} + 2(n+5)(n+2)a_{n+1} \\ &= 2(n^2+9n+19)a_{n+2} - (n+2)(n+5)a_{n+2} + 2(n+2)(n+5)a_{n+1} \\ &= 2(n^2+9n+19)a_{n+2} + (n+2)(n+5)a_{n+1} - (n+2)(n+5)(n+1)a_n \\ &= 2(n^2+9n+19)a_{n+2} - n(n+2)(n+5)a_n + n(n+2)(n+5)a_{n-1} \\ &= 2(n^2+9n+19)a_{n+2} - n(n-1)(n+2)(n+5)a_{n-2}. \end{aligned}$$ □

Remark 11. It is worth noticing that the relation with *polynomial* coefficients that we have obtained is “monic”, in the sense that the coefficient of

a_{n+6} is 1. This feature, which is for us important, is not *a priori* guaranteed for a linear recurrence with polynomial coefficients, and appears to us as a piece of good luck.

Now divide the recurrence (12) by $2^{e_{n+6}}$, obtaining the recurrence

$$q_{n+6} = (n^2 + 9n + 19)q_{n+2} - \frac{n(n-1)(n+2)(n+5)}{4} q_{n-2}.$$

Observe that $n^2 + 9n + 19$ is odd for every n, while $\frac{n(n-1)(n+2)(n+5)}{4}$ is an even integer for every n: indeed, for n even (resp. n odd), the product $n(n + 2)$ (resp. $(n - 1)(n + 5)$) is divisible by 8.

Thus, after noticing that

$$q_0 = 1, q_1 = 1, q_2 = 1, q_3 = 1, q_4 = 5, q_5 = 13, q_6 = 19, q_7 = 29$$

are odd integers, we obtain by induction that every q_n is an odd integer. □

Proposition 9 and Corollary 7 yield immediately the following corollary:

Corollary 12. *The following formulae hold for* $n \geq 1$*:*

$$d_n = 2^{\lfloor \frac{n+1}{4} \rfloor}; \qquad v_2(x_n) = \begin{cases} -1 & \text{if } n \equiv 0 \pmod 4, \\ 0 & \text{if } n \equiv 1 \pmod 4, \\ 1 & \text{if } n \equiv 2 \pmod 4, \\ 1 & \text{if } n \equiv 3 \pmod 4. \end{cases}$$

The proof of Theorem 1′ now follows from (10) and the exact value $d_n = 2^{\lfloor \frac{n+1}{4} \rfloor}$ given in the previous corollary. Furthermore, combining (10), (11) and Corollary 12, one gets the correct order of growth of D_n, to which we alluded in the introduction.

3.3 - *The second (linearly-independent) solution to the recurrence* (6)

Disregarding now the initial conditions, the set of solutions of the differential equation (7) is a vector space of dimension two. A second solution, linearly independent from $F(x)$, can be found with standard methods, so let us define

$$G(x) = F(x) \int_0^x F(t)^{-1} \, dt,$$

which satisfies (7) together with $G(0) = 0, G'(0) = 1$. The function $G(x)$ is the exponential generating function of an integer sequence $(g_n)_{n \in \mathbb{N}}$ (sequence

A000932 in [**OEIS**]) which satisfies the same recurrence (6) of the sequence $(a_n)_{n\in\mathbb{N}}$ with the initial values $g_0 = 0, g_1 = 1$.

It is possible to derive some formulae for (g_n) similar to the ones in Section 2, but the situation in this case is more complicated, especially when considering the greatest common divisor of g_n and g_{n+1}.

Writing $e^{-x-\frac{x^2}{2}} = \sum_{n=0}^{\infty} \frac{c_n}{n!}x^n$ and arguing as in Section 2 gives

$$c_n = \sum_{2s\le n} (-1)^{n+s}\binom{n}{2s}(2s-1)!!,$$

from which we obtain with some manipulations

$$\begin{aligned} g_n &= \sum_{r\ge 1}\binom{n}{r}a_{n-r}c_{r-1} \\ &= \sum_{r\ge 1}\binom{n}{r}\cdot\left(\sum_h \binom{n-r}{2h}(2h-1)!!\right)\cdot\left(\sum_k (-1)^{r+k-1}\binom{r-1}{2k}(2k-1)!!\right) \\ &= \sum_{h,k}(2h-1)!!(2k-1)!!\cdot\left(\sum_{r=2k+1}^{n-2h}(-1)^{r+k-1}\binom{n}{r}\binom{n-r}{2h}\binom{r-1}{2k}\right) \end{aligned}$$

where, for a fixed n, the summation range of the inner sum is nonempty only for finitely many h and k.

This expression may be further simplified. Rearranging $\binom{n}{r}\binom{n-r}{2h}$ into $\binom{n}{2h}\binom{n-2h}{r}$ and applying the more standard formula

$$\sum_{r=b+1}^{a}(-1)^r\binom{a}{r}\binom{r-1}{b} = (-1)^{b-1} \quad \text{for } a\ge b+1,$$

we see that, if n, h, k are three non-negative integers such that $2k+1\le n-2h$, the equality

$$\sum_{r=2k+1}^{n-2h}(-1)^{r+k-1}\binom{n}{r}\binom{n-r}{2h}\binom{r-1}{2k} = (-1)^k\binom{n}{2h}$$

holds.

Therefore we can write

$$g_n = \sum_h \binom{n}{2h}(2h-1)!! \sum_{2k\le n-2h-1}(-1)^k(2k-1)!!. \tag{13}$$

In order to understand the arithmetic relationship between a_n and g_n, fix an odd prime ℓ and define σ_ℓ to be the limit, in the ℓ-adic topology, of the converging series

$$\sigma_\ell = \sum_{k=0}^{\infty} (-1)^k (2k-1)!! \ \in \mathbb{Z}_\ell.$$

The product $a_n\sigma_\ell$ can be written, in $\mathbb{Z}_\ell$, as

$$\begin{aligned} a_n\sigma_\ell &= \left(\sum_h \binom{n}{2h}(2h-1)!!\right)\left(\sum_k (-1)^k(2k-1)!!\right) \\ &= \sum_{h,k} (2h-1)!!(2k-1)!!(-1)^k \binom{n}{2h}. \end{aligned} \tag{14}$$

By comparing the expansions (13) and (14) we see that, as n grows to infinity, more and more terms of the two series agree, thus proving that

$$g_n - a_n\sigma_\ell \xrightarrow{n\to\infty} 0 \quad \text{in } \mathbb{Z}_\ell.$$

In fact the convergence can be quantified, and we obtain that

$$g_n \equiv a_n\sigma_\ell \pmod{\ell^e} \quad \forall n \geq 2\ell(e+1)-1.$$

Denote now by v_ℓ the ℓ-adic valuation, and define $d'_n = \gcd(g_n, g_{n-1})$. From these congruences we see that

$$v_\ell(d'_n) = v_\ell(d_n) + v_\ell(\sigma_\ell) = v_\ell(\sigma_\ell)$$

holds for all n sufficiently large (if $\sigma_\ell = 0$ the statement should be interpreted as the fact that $\lim_{n\to\infty} v_\ell(d'_n) = \infty$).

The matter of understanding the valuations $v_\ell(\sigma_\ell)$ seems very complicated. There are primes for which this valuation is greater than zero — we found 3, 17, 1789 and 12583 — but we do not know if this happens for infinitely many primes; we do not even know if there is a prime ℓ such that $\sigma_\ell = 0$.

A theorem of Chudnovsky (see [**DGS94**, Chapter VIII]) implies the rather weak conclusion that there exist infinitely many odd primes ℓ such that $\sigma_\ell \neq 0$. Notice that other behaviours are indeed possible for similarly-defined series: it is an elementary exercise to show that $\sum_{n=0}^{N} n \cdot n! = (N+1)! - 1$, therefore the series $\sum_{n\geq 0} n \cdot n!$ converges ℓ-adically to -1 for every prime ℓ.

We conclude with the computation of the 2-adic valuation of g_n.

Proposition 13.

$$v_2(g_n) = v_2(a_n) = e_n \quad \forall n \geq 8,$$

where e_n is the exponent defined in Proposition 9.

Proof. We argue as in Proposition 9. It is enough to notice that the sequence g_n satisfies the same recurrence as the sequence a_n and so it also satisfies the recurrence (12). By direct computation we check that $g_n/2^{e_n}$ is an odd integer for $8 \leq n \leq 15$, and the same proof applies. □

4 - Second solution

Now we present a solution which turns out to be essentially the same of Gheorghe's, which he kindly sent us.

4.1 - *Another elementary solution*

If we define $f_n(x) = 1 + \frac{n}{x}$, we see that $x_{n+1} = f_n(x_n)$, so that one is led to study the dynamics of the sequence of functions f_n; we note that in this solution the *arithmetic* comes into play only at the end, whereas one starts just by studying the dynamics from the real variable viewpoint.

Let us call $y_n = \frac{1+\sqrt{4n+1}}{2}$ the (positive) fixed point of f_n. Plainly we have that if $x < y_n$ then $f_n(x) > y_n$ and vice versa.

We can prove by induction that

Lemma 14. *For every $n \geq 4$ we have*

$$y_{n-1} = \frac{1+\sqrt{4n-3}}{2} < x_n < \frac{1+\sqrt{4n+1}}{2} = y_n. \tag{15}$$

Proof. By a direct computation, we have $\frac{1+\sqrt{13}}{2} < \frac{5}{2} < \frac{1+\sqrt{17}}{2}$, which establishes the basis of the induction. Assuming (15) holds for n, by the previous remark we have that $y_n < x_{n+1}$, so we need only to prove that

$$1 + \frac{n}{x_n} < y_{n+1}.$$

By the inductive hypothesis, it is enough to show that

$$1 + \frac{n}{y_{n-1}} < y_{n+1}, \qquad i.e.,$$

$$1 + \frac{2n}{1+\sqrt{4n-3}} < \frac{1+\sqrt{4n+5}}{2},$$

which is an elementary, though tedious, computation. □

Remark 15. For the values of n smaller than 4, we have $x_3 = y_2 = x_2 = 2$ and $x_1 = y_0 = x_0 = 1$.

Lemma 16. *For every $n \geq 1$ the interval (y_{n-1}, y_n) does not contain any integer.*

Proof. Let us now assume that for some $n \geq 1$ there is an integer $k \in (y_{n-1}, y_n)$. Then we have

$$\frac{1+\sqrt{4n-3}}{2} < k < \frac{1+\sqrt{4n+1}}{2},$$
$$\sqrt{4n-3} < 2k-1 < \sqrt{4n+1},$$
$$4n-3 < (2k-1)^2 < 4n+1.$$

However the last line gives a contradiction modulo 4. □

The two lemmas together immediately imply that the only integral values of the sequence are $x_0, \ldots, x_3$.

4.2 - *A different interpretation of the same argument*

Essentially the same solution may be reached by a slightly different approach.

The same conclusion as before can be reached if we show that $n-1 < x_n^2 - x_n < n$ for $n \geq 4$.

We argue by induction. The inequalities are verified by direct inspection for $n = 4$, since $3 < \frac{25}{4} - \frac{5}{2} < 4$.

Now let $t_n = x_n^2 - x_n$, and assume that the inequalities hold up to n. We may write t_{n+1} as

$$t_{n+1} = x_{n+1}^2 - x_{n+1} = x_{n+1}(x_{n+1}-1) = \left(1+\frac{n}{x_n}\right)\frac{n}{x_n} = \frac{n(x_n+n)}{x_n^2}.$$

By the induction hypothesis we have:

$$x_n^2 < x_n + n \Longrightarrow t_{n+1} > n\frac{(x_n+n)}{x_n+n} = n$$

and

$$x_n^2 > x_n + n - 1 \Longrightarrow t_{n+1} < n\frac{x_n+n}{x_n+n-1} < n+1$$

since $x_n > 1$.

5 - Further observations to the second solution

5.1 - *A linear fractional representation for the sequence* (X_n)

We can iterate the recursive formula (1) to obtain $X_{n+2} = 1 + \frac{n+1}{1+\frac{n}{X_n}} = \frac{(n+2)X_n + n}{X_n + n}$; in general, X_{n+k} may be expressed in two different ways: first, by a kind of continued fraction involving X_n and the integers in $\{n, \ldots, n + k - 1\}$, second, by a linear fractional transformation in X_n, namely $X_{n+k} = M_{n,k}(X_n) = \frac{\alpha_{n,k}X_n + \beta_{n,k}}{\gamma_{n,k}X_n + \delta_{n,k}}$, for suitable integer coefficients depending on n, k; here $M_{n,k}$ denotes the matrix $\begin{pmatrix} \alpha_{n,k} & \beta_{n,k} \\ \gamma_{n,k} & \delta_{n,k} \end{pmatrix}$.

These matrices satisfy the recurrence

$$M_{n,k+1} = \begin{pmatrix} 1 & n+k \\ 1 & 0 \end{pmatrix} M_{n,k}.$$

In particular, if we consider the expression of X_n in terms of X, we can express the matrices involved in terms of the sequences (a_n) and (g_n) introduced previously, and it is easily shown by induction that

$$X_n = M_{1,n-1}(X) = \frac{g_n X + a_n - g_n}{g_{n-1}X + a_{n-1} - g_{n-1}} \qquad \forall n \geq 1.$$

Using this representation, it is possible to extend the arguments in Section 4.1 to prove the full Theorem 2.

P r o o f o f T h e o r e m 2. The first statement follows as in Section 4.1 if we show that

$$y_{n-1} < X_n < y_n \qquad \forall n \geq 14.$$

The inductive step has been already shown in the proof of Lemma 14, so we only need to show that the inequality holds for $n = 14$.

By expanding X_{14} as $M_{1,13}(X_1)$ as explained above, we obtain that

$$X_{14} = \frac{195330X + 103480}{46779X + 24284},$$

and therefore

$$y_{13} = \frac{1+\sqrt{53}}{2} < \frac{195330}{46779} < X_{14} < \frac{103480}{24284} < \frac{1+\sqrt{57}}{2} = y_{14};$$

this proves the first assertion.

For $n = 3, \dots, 13$ one can examine the analogous expressions of X_n as a rational function of X and check when the image of $(0, +\infty)$ under the map $M_{1,n-1}(X)$ contains an integer number. This only happens when X is one of the four values listed in the statement, which generate the following sequences, the last of which is the sequence (x_n) of the original problem.

$$\frac{4}{51}, \frac{55}{4}, \frac{63}{55}, \frac{76}{21}, \frac{40}{19}, \frac{27}{8}, \frac{25}{9}, \frac{88}{25}, \frac{36}{11}, \frac{15}{4}, \frac{11}{3}, 4, 4, \frac{17}{4}, \dots$$
$$\frac{1}{3}, 4, \frac{3}{2}, 3, \frac{7}{3}, \frac{22}{7}, \frac{32}{11}, \frac{109}{32}, \frac{365}{109}, \frac{1346}{365}, \frac{2498}{673}, \frac{9901}{2498}, \frac{39877}{9901}, \frac{168590}{39877}, \dots$$
$$\frac{2}{3}, \frac{5}{2}, \frac{9}{5}, \frac{8}{3}, \frac{5}{2}, 3, 3, \frac{10}{3}, \frac{17}{5}, \frac{62}{17}, \frac{116}{31}, \frac{457}{116}, \frac{1849}{457}, \frac{7790}{1849}, \dots$$
$$1, 2, 2, \frac{5}{2}, \frac{13}{5}, \frac{38}{13}, \frac{58}{19}, \frac{191}{58}, \frac{655}{191}, \frac{2374}{655}, \frac{4462}{1187}, \frac{17519}{4462}, \frac{71063}{17519}, \frac{298810}{71063}, \dots \qquad \square$$

5.2 - *The denominators tend to infinity*

A variation of the argument in Section 4.2 can be used to show that the reduced denominator of x_n grows to infinity.

One can show that

$$x_n = \sqrt{n} + \frac{1}{2} - \frac{1}{8\sqrt{n}} - \frac{1}{8n} + o\left(\frac{1}{n}\right) \quad \text{for } n \to \infty, \tag{16}$$

which implies that

$$x_n^2 - x_n - n + \frac{1}{2} = O\left(\frac{1}{\sqrt{n}}\right) \quad \text{for } n \to \infty. \tag{17}$$

From this we can deduce that $D_n \gg n^{1/4}$, where D_n is the reduced denominator of x_n. Indeed if it were not so, equation (17) would imply that, for infinitely many n, the equality

$$x_n^2 - x_n - n + \frac{1}{2} = 0$$

holds. For those values of n we would have that

$$x_n = \frac{1 + \sqrt{4n - 1}}{2} = \sqrt{n} + \frac{1}{2} - \frac{1}{8\sqrt{n}} + O\left(\frac{1}{n^{3/2}}\right),$$

which is in contradiction with (16).

Appealing to a general theory of asymptotic expansions of recurrence sequences, a refinement of this argument gives us the proof of Theorem 3.

Proof of Theorem 3. It can be shown (see [**BT33**] and [**WZ85**, Example 2.1]) that for any natural number r there exist coefficients c_i such that an asymptotic expansion of the form

$$X_n = \sum_{i=-1}^{r} c_i n^{-i/2} + O(n^{-(r+1)/2}) \quad \text{for } n \to \infty \tag{18}$$

holds; these coefficients must be rational and can be worked out by imposing that this expansion satisfies the recurrence (3).

Therefore, for any positive integer k we can find polynomials $a, b, c \in \mathbb{Q}[n]$ of degree at most k, not all zero and such that equation (17) can be improved to

$$a(n)X_n^2 - b(n)X_n - c(n) = O(n^{-k/2}).$$

Indeed finding the coefficients of such polynomials amounts to the resolution of a linear system with $3k+3$ unknowns and $3k+2$ equations whose coefficients are polynomial expressions of the c_i with integer coefficents; this guarantees the existence of a non-zero solution. Furthermore, the fact that $X_n \sim \sqrt{n}$ assures us that $a(n)$ is not identically zero.

Then either the reduced denominators of X_n grow at least as fast as $n^{k/4}$, or

$$a(n)X_n^2 - b(n)X_n - c(n) = 0 \tag{19}$$

holds for infinitely many n.

If this is the case, then the expansion at infinity of the algebraic function

$$\varphi(t) = \frac{b(t) + \sqrt{b(t)^2 + 4a(t)c(t)}}{2a(t)}$$

must coincide with (18), and $\varphi(t)$ satisfies the recurrence

$$\varphi(t+1) = 1 + \frac{t}{\varphi(t)}. \tag{20}$$

Now we show that this is impossible.

Let $\delta(t) \in \mathbb{Q}[t]$ be the squarefree part of $b(t)^2 + 4a(t)c(t)$, so that $\varphi(t) \in \mathbb{Q}(t, \sqrt{\delta(t)})$. The set of the roots of δ is the ramification locus of φ, if we think φ as a map between the algebraic curve defined by $b(t)^2 + 4a(t)c(t) = u^2$ and $\mathbb{P}_1$.

The functional equation (20) implies that $\sqrt{\delta(t+1)} \in \mathbb{Q}(t, \sqrt{\delta(t)})$, which implies that the set of roots of δ is invariant under translation by 1. But this is a finite set, so it must be empty, which means that φ is a rational function. This is impossible because the expansion (18) begins with $\sqrt{n}$.

This shows that the denominators of X_n grow faster than $n^{k/4}$. The same argument holds for all k and all initial data X and therefore a bound as in the statement of the theorem may be obtained by changing the exponent and taking $C_{k,X}$ sufficiently large.

To check that the result is effective notice that the coefficients c_i can be worked out algorithmically from the recurrence; for every fixed r, expansion (18) may be converted in an explicit inequality which may be proved by induction; the coefficients of the polynomials a, b, c are determined by a linear system involving only the c_i. □

Remark 17. We sketch an alternative argument that can be used to conclude the proof. Arguing as above, we see that $X_n = \varphi(n)$ holds for all n large enough, because both sequences satisfy the same recurrence.

Notice that, using the fact that $X_n = \sqrt{n} + \frac{1}{2} + o(1)$, we have $X_n^2 = n + \sqrt{n} + o(\sqrt{n})$, and we see that in (19) the leading term of a must cancel out the leading term of c in the term of degree $k+1$, while the leading term of a must cancel out the leading term of b in the term of degree $k + \frac{1}{2}$. This shows that $\deg c = 1 + \deg a$ and $\deg b = \deg a$, and that we can assume the polynomials a, b, c to be all three monic; we can also assume them to be coprime.

From (19), using (3) to express X_n in terms of X_{n+1}, clearing out denominators, and shifting indices by one, we obtain a second quadratic relation for X_n with coefficients in $\mathbb{Q}[n]$. This relation must be proportional to (19), otherwise it would be possible to eliminate X_n^2 from the two relations and express X_n as a rational function of n, against the fact that $X_n \sim \sqrt{n}$.

Writing out what it means for these two relations to be proportional we obtain the system

$$\text{(21)} \qquad \begin{cases} a(n)\left((n-1)b(n-1) - 2c(n-1)\right) = -b(n)c(n-1) \\ a(n)\left((n-1)^2 a(n-1) + (n-1)b(n-1) - c(n-1)\right) = c(n)c(n-1). \end{cases}$$

Because a, b, c are coprime we see that $a(n) | c(n-1)$, so that $c(n-1)$ is equal to $a(n)$ times a linear factor. This linear factor must be equal to $n-1$ because otherwise it would divide $b(n-1)$ and $a(n-1)$ as well, against the coprimality.

Therefore we can write $c(n-1) = (n-1)a(n)$, obtain from the first equation of (21) that $a(n) = (b(n) + b(n-1))/2$, and use it to eliminate both a and c from the second one. In this way we reach

$$\text{(22)} \qquad nb(n+1) + (n+1)b(n) - nb(n-1) - (n-1)b(n-2) = 0,$$

which gives a contradiction when comparing the coefficients of the terms of degree $\deg b$.

Acknowledgments. We thank the anonymous referee for his/her useful comments and interesting new questions.

References

[**BT33**] G. D. Birkhoff and W. J. Trjitzinsky, *Analytic theory of singular difference equations*, Acta Math. **60** (1933), 1–89.

[**DGS94**] B. Dwork, G. Gerotto and F. J. Sullivan, *An introduction to G-functions*, Ann. of Math. Stud., **133**, Princeton University Press, Princeton, NJ, 1994.

[**MW56**] L. Moser and M. Wyman, *Asymptotic expansions*, Canadian J. Math. **8** (1956), 225–233.

[**OEIS**] The OEIS Foundation Inc., The On-Line Encyclopedia of Integer Sequences, 2020, https://oeis.org.

[**RMS**] Romanian Mathematical Society, *RMC 2010*, Paralela 45 Publishing House, Bucharest, 2010.

[**Wil06**] H. S. Wilf, *Generatingfunctionology*, 3rd ed., A K Peters, Wellesley, MA, 2006.

[**WZ85**] J. Wimp and D. Zeilberger, *Resurrecting the asymptotics of linear recurrences*, J. Math. Anal. Appl. **111** (1985), 162–176.

Roberto Dvornicich
Dipartimento di Matematica
Università di Pisa
Largo Pontecorvo 5
56127 Pisa, Italia
e-mail: dvornic@dm.unipi.it

Francesco Veneziano
Department of Mathematics
University of Genova
Via Dodecaneso 35
16146 Genova, Italy
e-mail: veneziano@dima.unige.it

Umberto Zannier
Scuola Normale Superiore
Piazza dei Cavalieri 7
56126 Pisa, Italia
e-mail: u.zannier@sns.it