Decision problems for second-order holonomic recurrences

Eike Neumann
Max Planck Institute for Software Systems, Saarland Informatics Campus, Germany
eike@mpi-sws.org

Joël Ouaknine
Max Planck Institute for Software Systems, Saarland Informatics Campus, Germany
joel@mpi-sws.org

James Worrell
Department of Computer Science, Oxford University, United Kingdom
jbw@cs.ox.ac.uk

Abstract

We study decision problems for sequences which obey a second-order holonomic recurrence of the form $f(n+2) = P(n)f(n+1) + Q(n)f(n)$ with rational polynomial coefficients, where P is non-constant, Q is non-zero, and the degree of Q is smaller than or equal to that of P. We show that existence of infinitely many zeroes is decidable. We give partial algorithms for deciding the existence of a zero, positivity of all sequence terms, and positivity of all but finitely many sequence terms. If Q does not have a positive integer zero then our algorithms halt on almost all initial values $(f(1), f(2))$ for the recurrence. We identify a class of recurrences for which our algorithms halt for all initial values. We further identify a class of recurrences for which our algorithms can be extended to total ones.

2012 ACM Subject Classification Mathematics of Computing → Discrete Mathematics

Keywords and phrases holonomic sequences, Positivity Problem, Skolem Problem

Digital Object Identifier 10.4230/LIPIcs...

1 Introduction

A sequence $(f(n))_{n\geq 1}$ of real numbers is called *holonomic* or *P-finite* if its terms satisfy an algebraic equation of the form

$$P_r(n)f(n+r) + P_{r-1}(n)f(n+r-1) + \cdots + P_0(n)f(n) = 0,$$

where $P_0, \ldots, P_r \in \mathbb{R}[X]$ are polynomials, not all zero. The number r is called the *order* of the recurrence. When all polynomials $P_r, \ldots, P_0$ are constant we recover the familiar example of ordinary linear recurrence sequences. Alternatively, holonomic sequences are characterised as the coefficients of formal power series which satisfy a non-trivial homogeneous linear ordinary differential equation with polynomial coefficients [10]. Strikingly, holonomic sequences with rational polynomial coefficients can be tested for equality automatically [11]. This allows for automatic proving of highly non-trivial special function identities with numerous applications in mathematics and the sciences [7].

It is natural to ask if holonomic sequences can be automatically tested for inequality as well. This reduces to the problem of deciding whether all terms of a given holonomic sequence $(f(n))_n$ are positive. In full generality this question seems to be completely out of reach, even in the case where the polynomials $P_r, \ldots, P_0$ are all constant. While this problem, often called the Positivity Problem, is widely believed to be decidable in the constant coefficient case, a feasible decision method for linear recurrences of order six or higher would entail major breakthroughs in Diophantine approximation [6]. Nonetheless, decision methods are known for constant coefficient linear recurrence sequences of order up to five.

There is hence some hope that one can obtain decidability results on the Positivity Problem for low-order holonomic sequences as well. In this paper we investigate the problem for sequences satisfying a second-order holonomic recurrence of the form

$$f(n+2) = P(n)f(n+1) + Q(n)f(n) \tag{1}$$

with P non-constant and $0 \leq \deg Q \leq \deg P$. This constitutes arguably the simplest class of non-trivial instances of the Positivity Problem. By a straightforward reduction, our results extend to the larger class of holonomic recurrences of the form

$$R(n)f(n+2) = P(n)f(n+1) + Q(n)f(n)$$

with P non constant, R without positive integer zeroes, and $0 \leq \deg R + \deg Q \leq \deg P$.

We study the possible behaviours that a sequence satisfying a recurrence of the form (1) may exhibit as $n \to \infty$. Up to shifting the recurrence by finitely many terms we may assume that Q does not have any positive integer zeroes. We then show that the plane of initial values $(f(1), f(2)) \in \mathbb{R}^2$ decomposes into five disjoint pieces — the origin O, two rays L^+ and L^-, and two open half-planes H^+ and H^- — such that the behaviour of the sequence $(f(n))_{n\geq 1}$ for large n depends only on the piece that contains the initial values and the signs of the leading coefficients of P and Q. Depending on this data, unless the sequence is identically zero, it will be eventually strictly positive, strictly negative, or alternating between strictly positive and strictly negative. Moreover, we can compute on each of the five pieces a number N such that the sequence $(f_n)_n$ has the described behaviour for all $n \geq N$.

Up to potentially shifting the recurrence by finitely many terms, the line $L = L^+ \cup L^- \cup O$ has a well-defined slope, given by the (necessarily convergent) continued fraction $-\mathrm{K}_{n=1}^{\infty} \frac{Q(n)}{P(n)}$. This allows us to approximate the line L numerically to any given finite precision. We can hence determine, by means of a potentially non-terminating algorithm, if a given pair of rational initial values $(f(1), f(2)) \in \mathbb{Q}^2$ is outside the line and in that case determine the behaviour of the sequence for large n. This yields a partial algorithm for deciding the Positivity Problem and related problems.

While we do not obtain a total algorithm for all second-order holonomic recurrences of the form (1), we identify a class of recurrences for which the slope of the line L is an effectively computable rational number. In this case we can extend our algorithm to a total one. One can effectively check if a given recurrence belongs to this class.

Our algorithm is also total for the class of all recurrences such that the slope of the line L is irrational. We establish non-trivial effective criteria that assert this.

Related Work. Decidability of the Positivity Problem for second-order holonomic sequences is investigated in [4]. It is shown that for sequences with linear polynomial coefficients, the Positivity Problem reduces to the problem of deciding the equality of certain effectively given quantities, closely related to periods [5] whose equality is conjectured to be decidable [5, Conjecture 1].

Gerhold and Kauers [2] give a partial algorithm for deciding Positivity for general holonomic sequences based on symbolic methods from real algebraic geometry. To the best of our knowledge its precise termination behaviour is not known, even for low-order sequences. Further practical partial algorithms in the spirit of [2] are introduced in [3, 8, 9]. In those papers sufficient termination criteria are given for recurrences of the form

$$P_r(n)f(n+r) + \cdots + P_0(n)(f) = 0$$

with $\deg P_0 = \deg P_r$ and $\deg P_j \leq \deg P_0$ for all $j \leq r$. This situation is disjoint from the one we investigate. Similarly to our main result, one of the algorithms in [3] is shown to terminate for all second-order recurrences of the above form on almost all initial values. All further termination criteria established in the aforementioned papers put restrictions on the eigenvalues of the holonomic recurrence but no restrictions on the initial values. The algorithms may fail to converge for all initial values of a recurrence that does not meet the restrictions on the eigenvalues.

Key contributions. There remains a dearth of algorithmic results on positivity and inequality problems for second-order (and higher) holonomic sequences. The present paper makes two substantial contributions to these outstanding open problems: (i) we identify a large class of second-order holonomic recurrences for which we can precisely characterise all the possible asymptotic behaviours (Theorem 1); and (ii) building upon this, we identify a substantial subclass of holonomic sequences for which we exhibit total algorithms for the Positivity and Skolem problems.

2 Results

Let us first introduce the decision problems we seek to investigate. The Skolem Problem is the problem of deciding for a given recurrence of the form (1) and for given initial values $f(1), f(2)$ if the induced recurrence sequence $(f(n))_n$ has a zero. The Infinite Zero Problem asks if the sequence $(f(n))_n$ thus given has infinitely many zeroes. The Positivity Problem asks if all terms of the sequence $(f(n))_n$ are positive. The Ultimate Positivity Problem is the problem of deciding if there exists an index N such that all terms $f(n)$ with $n \geq N$ are positive.

Our results are best stated for holonomic recurrences that are normalised in the following sense. A second-order holonomic recurrence $f(n+2) = P(n)f(n+1) + Q(n)f(n)$ with P non-constant and $0 \leq \deg Q \leq \deg P$ is said to be *in normal form* if $P, Q \in \mathbb{Z}[X]$ are integer polynomials such that P has a positive leading coefficient, P and Q have no positive real zeroes, $P(n)^2 + 4Q(n) > 0$ for all positive integers n, and there is no prime number p such that p divides all coefficients of P and p^2 divides all coefficients of Q.

For the most part, the above assumptions on P and Q do not present essential restrictions: Given any second-order holonomic recurrence of the form (1) we can effectively compute integers c and N such that the holonomic recurrence $g(n+2) = cP(N+n)g(n+1) + c^2Q(N+n)g(n)$ is in normal form. For any given pair of initial values $f(1), f(2) \in \mathbb{Q}$, we can effectively compute $g(1) = c^{N+1}f(N+1)$ and $g(2) = c^{N+2}f(N+2)$. The sequence $(g(n))_n$ is then equal to $(c^n f(n))_{n>N}$. Thus, the behaviour of $(f(n))_n$ is easily deduced from that of $(g(n))_n$ and the finite sequence $f(1), \ldots, f(N)$. In particular, the above mentioned decision problems reduce in this way to their specialisation to recurrences in normal form. However, since in general there may exist initial values for which our algorithm is not guaranteed to terminate, we also need to understand which initial values for the original recurrence get mapped to such ones. We will discuss this below, after we have stated our main results.

It is worth pointing out that our results extend to holonomic recurrences of the form $R(n)f(n+2) = P(n)f(n+1) + Q(n)f(n)$ with P non constant, R without positive integer zeroes, and $0 \leq \deg R + \deg Q \leq \deg P$. Indeed, if $(f(n))_n$ satisfies a recurrence of this form then the sequence $g(n) = R(1) \cdot \dots \cdot R(n)f(n)$ satisfies the recurrence $g(n+2) = R(n+2)P(n)f(n+2) + R(n+2)R(n+1)Q(n)$, which falls within the class we investigate. Note that up to shifting the recurrence appropriately we may assume that $R(n)$ has constant (non-zero) sign, so that the behaviour of $(f(n))_n$ is easily deduced from that of $(g(n))_n$.

Our first result is a complete classification of the possible behaviours that a holonomic recurrence of the form (1) may exhibit for large n. We say that a sequence $(x_n)_n$ of real numbers is *eventually positive* if there exists an $N \in \mathbb{N}$ such that $x_n > 0$ for all $n \geq N$. We say that it is *eventually negative* if there exists an $N \in \mathbb{N}$ such that $x_n < 0$ for all $n \geq N$. We say that it is *eventually alternating* if there exists an $N \in \mathbb{N}$ such that $x_N \neq 0$ and $\operatorname{sgn}(x_{n+1}) = -\operatorname{sgn}(x_n)$ for all $n \geq N$. We say that it is *eventually zero* if there exists an $N \in \mathbb{N}$ such that $x_N = 0$ for all $n \geq N$. In each of these cases we call any admissible choice for N a *witness* for the respective behaviour.

▶ **Theorem 1.** Let $f(n+2) = P(n)f(n+1) + Q(n)f(n)$ be a holonomic recurrence in normal form. Then there exists a partition of $\mathbb{R}^2$ into five pieces, the origin O, two rays L^+ and L^-, and two open half-planes H^+ and H^-, such that for all pairs of initial values $(f(1), f(2)) \in \mathbb{R}^2$ we have:

1. If $(f(1), f(2)) \in O$ then the sequence is constant equal to zero.
2. If $(f(1), f(2)) \in H^+$ then the sequence is eventually positive.
3. If $(f(1), f(2)) \in H^-$ then the sequence is eventually negative.
4. If $(f(1), f(2)) \in L^+$ then the sequence is eventually positive if the leading coefficient Q is negative, and eventually alternating if the coefficient is positive.
5. If $(f(1), f(2)) \in L^-$ then the sequence is eventually negative if the leading coefficient Q is negative, and eventually alternating if the coefficient is positive.

We will call the line $L = O \cup L^+ \cup L^-$ the *critical line* of the holonomic recurrence. We can compute its slope to any given finite precision thanks to the following continued fraction representation:

▶ **Proposition 2.** Let $f(n+2) = P(n)f(n+1) + Q(n)f(n)$ be a holonomic recurrence in normal form. Then we can compute a number N such that for the shifted recurrence $g(n+2) = P(n+N)g(n+1) + Q(n+N)g(n)$, in the notation of Theorem 1, the line $L = O \cup L^+ \cup L^-$ has slope

$$-\mathop{\mathrm{K}}_{n=N}^{\infty} \frac{Q(n)}{P(n)} = -\cfrac{Q(N)}{P(N) + \cfrac{Q(N+1)}{P(N+1) + \dots}}.$$

Theorem 1 suggests the following computational problem: given a holonomic recurrence in normal form and initial values $(f(1), f(2)) \in \mathbb{Q}^2$, report whether the sequence $(f(n))_n$ thus defined is eventually positive, eventually negative, eventually alternating, or eventually zero and output a witness N for this. Let us call this the *Ultimate Sign Problem*. By Theorem 1 this problem is well-defined. It is clear that the Skolem Problem, the Positivity Problem, the Ultimate Positivity Problem, and the Infinite Zero Problem reduce to this.

Unfortunately we only obtain a partial computability result:

▶ **Theorem 3.** There exists an algorithm which takes as input a holonomic recurrence $f(n+2) = P(n)f(n+1) + Q(n)f(n)$ in normal form, together with a pair $(f(1), f(2)) \in \mathbb{Q}^2$ of rational initial values and halts if and only if $(f(1), f(2)) \notin L^+ \cup L^-$. Upon halting the algorithm reports if $(f(1), f(2))$ is zero, belongs to H^+, or belongs to H^-, and in the latter two cases returns a number N such that the sequence $(f(n))_n$ has constant sign for all $n \geq N$.

Theorem 3 yields a total algorithm for deciding the Infinite Zero Problem and partial algorithms for deciding the Skolem Problem, the Positivity Problem, and the Ultimate

Positivity Problem. The set of problem instances where the algorithm does not halt is “small” in the sense that it is contained in a set of codimension one.

While we do not obtain a total algorithm in general, there are special instances for which we do. To describe these instances we need to introduce further concepts. The *companion matrix* of the holonomic recurrence (1) is the matrix

$$M(n) = \begin{pmatrix} 0 & 1 \\ Q(n) & P(n) \end{pmatrix}.$$

Note that we have

$$\begin{pmatrix} f(n) \\ f(n+1) \end{pmatrix} = \prod_{j=1}^{n-1} M(j) \begin{pmatrix} f(1) \\ f(2) \end{pmatrix}.$$

Its n^{th} *characteristic polynomial* is given by

$$z^2 - P(n)z - Q(n).$$

For holonomic recurrences in normal form, the discriminant of this polynomial is by definition strictly positive for all $n \in \mathbb{N}$. Hence the characteristic polynomial has two distinct real roots

$$\lambda_1(n) = \frac{1}{2}\left(P(n) + \left(P(n)^2 + 4Q(n)\right)^{1/2}\right)$$

and

$$\lambda_2(n) = \frac{1}{2}\left(P(n) - \left(P(n)^2 + 4Q(n)\right)^{1/2}\right).$$

In the case where $\lambda_2(n)$ is a constant function of n we can compute the slope of the line L, yielding a total algorithm.

▶ **Proposition 4.** Let $f(n+2) = P(n)f(n+1) + Q(n)f(n)$ be a holonomic recurrence in normal form. If $\lambda_2(n) = \lambda_2$ is a constant function of n then, in the notation of Theorem 1,

$$L^+ \cup L^- \cup O = \left\{(x, y) \in \mathbb{R}^2 \mid y = \lambda_2 x\right\}.$$

▶ **Corollary 5.** The Ultimate Sign Problem is computable for the class of all holonomic recurrences on normal form which have the additional property that $\lambda_2(n)$ is a constant function of n.

The following criterion allows us to check whether λ_2 is constant, increasing, or decreasing:

▶ **Proposition 6.** Write $P(n) = a_d n^d + \cdots + a_0$, $Q(n) = b_d n^d + \cdots + b_0$ with $a_d > 0$. Let

$$\chi_j = \det\begin{pmatrix} b_d & b_j \\ a_d & a_j \end{pmatrix}$$

for $j = 1, \ldots, d-1$. Let

$$\chi_0 = \det\begin{pmatrix} b_d & b_0 \\ a_d & a_0 \end{pmatrix} + b_d^2/a_d$$

The function λ_2 is either constant, strictly monotonically increasing for sufficiently large n, or strictly monotonically decreasing for large n. It is constant if and only if $\chi_0 = \cdots = \chi_d = 0$. It is decreasing if and only if there exists a j_0 such that $\chi_{j_0} > 0$ and $\chi_j = 0$ for $j > j_0$. It is increasing if and only if there exists a j_0 such that $\chi_{j_0} < 0$ and $\chi_j = 0$ for $j > j_0$.

We also obtain a total algorithm for the Ultimate Sign Problem in the case where the critical line contains no rational points. We collect some sufficient conditions that guarantee this.

▶ **Theorem 7.** Let $f(n+2) = P(n)f(n+1) + Q(n)f(n)$ be a holonomic recurrence in normal form with integer polynomial coefficients $P(n) = a_d n^d + \cdots + a_1 n + a_0$ and $Q(n) = b_d n^d + \cdots + b_1 n + b_0$. Then the critical line L contains no non-trivial rational points if any of the following sufficient conditions is met:

1. $b_d = 0$.
2. $|\operatorname{lcof}(Q)/\operatorname{lcof}(P)| < 1$.
3. $|\operatorname{lcof}(Q)/\operatorname{lcof}(P)| = 1$ and $\lambda_2(n)$ is non-constant, positive and increasing for large n.
4. $|\operatorname{lcof}(Q)/\operatorname{lcof}(P)| = 1$ and $\lambda_2(n)$ is non-constant, negative and decreasing for large n.
5. $|\operatorname{lcof}(Q)/\operatorname{lcof}(P)| = 1$ and $\lambda_2(n)$ is non-constant, positive, decreasing for large n, and

$$\begin{cases} |a_0 + b_0 - 1| < 3a_1 & \text{if } d = 1, \\ |a_{d-1} + b_{d-1}| < (d+2)a_d & \text{otherwise.} \end{cases}$$

6. $|\operatorname{lcof}(Q)/\operatorname{lcof}(P)| = 1$ and $\lambda_2(n)$ is non-constant, negative, increasing for large n, and

$$\begin{cases} |a_0 - b_0 + 1| < 3a_1 & \text{if } d = 1, \\ |a_{d-1} - b_{d-1}| < (d+2)a_d & \text{otherwise.} \end{cases}$$

Finally, let us discuss how the reduction to normal form affects the termination behaviour of our algorithm. Let $f(n+2) = P(n)f(n+1) + Q(n)f(n)$ be a holonomic recurrence of the form (1), and let $g(n+2) = cP(n+N)g(n+1) + c^2Q(n+N)g(n)$ be a recurrence in normal form as above. The map which sends initial values $(f(1), f(2))$ for the original recurrence to the initial values $(c^{N+1}f(N+1), c^{N+2}f(N+2))$ for the recurrence in normal form is a linear map $A\colon \mathbb{R}^2 \to \mathbb{R}^2$. The map A is bijective if and only if Q does not have any positive integer zeroes. If Q does have positive integer zeroes then the kernel of A is one-dimensional. In the cases where we have a total algorithm for the Ultimate Sign Problem for the recurrence in normal form the reduction of course yields a total algorithm. Assume that we only have a partial algorithm that halts outside the union of the two rays L^+ and L^-. If Q does not have any positive integer zeroes, then applying the partial algorithm for the Ultimate Sign Problem after the reduction yields an algorithm that halts outside the union of the two rays $A^{-1}(L^+)$ and $A^{-1}(L^-)$. Thus, the behaviour of the algorithm is unchanged. If Q has positive integer zeroes then either A sends all initial values outside its kernel into the union of the rays L^+ and L^-, or it sends all such initial values into the union of H^+ and H^-. In the latter case we obtain a total algorithm, but in the former case we obtain an algorithm that only halts on the one-dimensional kernel of A. Thus, in the former scenario the dimension of the set of inputs which lead to termination decreases by one.

Let us illustrate some of our results with the help of a simple example.

▶ **Example 8.** Consider the holonomic recurrence

$$f(n+2) = (n-1)f(n+1) + nf(n).$$

We have $\lambda_1(n) = n$ and $\lambda_2(n) = -1$ for all $n \in \mathbb{N}$. Proposition 4 allows us to easily compute the critical line:

$$L = \left\{(x,y) \in \mathbb{R}^2 \mid x = -y\right\}.$$

Using elementary linear algebra we can explicitly compute the n^{th} term of the sequence:

$$f(n) = (-1)^n \frac{f(1) - f(2)}{2} + \left(\frac{n!}{n+1} + \sum_{k=1}^{n-1} \frac{(-1)^{n-k} k!}{(k+1)(k+2)} \right) \frac{f(1) + f(2)}{2}.$$

We hence have $H^+ = \left\{(x, y) \in \mathbb{R}^2 \mid x > -y\right\}$ and $H^- = \left\{(x, y) \in \mathbb{R}^2 \mid x < -y\right\}$.

Thus, if we fix $f(1) > 0$ and let $f_t(2) = -f(1) + t$ with $t \in [0, 1]$ the sequence with initial values $(f(1), f_t(2))$ is eventually positive for all $t > 0$. For sufficiently small $t > 0$ the sequence will alternate between positive and negative values a finite number $N(t)$ of times before attaining only positive values. We have $N(t) \to \infty$ as $t \to 0$. For $t = 0$ the sequence alternates between positive and negative values forever, a witness for this being given by $N(0) = 1$.

If we let the sequence start at the index $n = 0$ then the matrix product has the following closed form:

$$\prod_{j=0}^{n} M(j) = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}.$$

Thus, every initial value gets mapped onto the critical line for the same recurrence with starting index $n = 1$. The sequence is alternating for all initial values. Since λ_2 is a constant function our algorithm is total and hence able to detect this.

3 Proof of the Results

3.1 Preliminaries

Consider a second-order holonomic recurrence $f(n+2) = P(n)f(n+1) + Q(n)f(n)$ in normal form. Let $M(n)$ be its companion matrix. Using that $M(n)$ has two distinct real eigenvalues for all n, write $M(n) = S(n)D(n)S(n)^{-1}$, where

$$D(n) = \begin{pmatrix} \lambda_2(n) & 0 \\ 0 & \lambda_1(n) \end{pmatrix}, \qquad S(n) = \begin{pmatrix} 1 & 1 \\ \lambda_2(n) & \lambda_1(n) \end{pmatrix},$$

$$S(n)^{-1} = \frac{1}{\lambda_1(n) - \lambda_2(n)} \begin{pmatrix} \lambda_1(n) & -1 \\ -\lambda_2(n) & 1 \end{pmatrix}.$$

Then we have:

$$M(n)M(n-1) \cdot \dots \cdot M(k) = S(n)D(n)S(n)^{-1}S(n-1)D(n-1)S(n-1)^{-1} \cdot \dots \cdot S(k)D(k)S(k)^{-1}.$$

Intuitively, the products $S(n+1)^{-1}S(n)$ are very close to the identity matrix for large n, but we need to study the error terms precisely. Thus, define real-valued functions $\varepsilon_{i,j}(n)$ by:

$$S(n+1)^{-1}S(n) = \begin{pmatrix} 1 + \varepsilon_{1,1}(n) & \varepsilon_{1,2}(n) \\ \varepsilon_{2,1}(n) & 1 + \varepsilon_{2,2}(n) \end{pmatrix}.$$

More explicitly:

$$\varepsilon_{1,1}(n) = \frac{\lambda_2(n+1) - \lambda_2(n)}{\lambda_1(n+1) - \lambda_2(n+1)} \qquad \varepsilon_{1,2}(n) = \frac{\lambda_1(n+1) - \lambda_1(n)}{\lambda_1(n+1) - \lambda_2(n+1)}$$

$$\varepsilon_{2,1}(n) = \frac{\lambda_2(n) - \lambda_2(n+1)}{\lambda_1(n+1) - \lambda_2(n+1)} \qquad \varepsilon_{2,2}(n) = \frac{\lambda_1(n) - \lambda_1(n+1)}{\lambda_1(n+1) - \lambda_2(n+1)}$$

We want to study the product $M(n)M(n-1)\cdot\dots\cdot M(k)$. To this end, define functions $a(k,n), b(k,n), c(k,n)$, and $d(k,n)$ via:

$$\prod_{j=k}^{n} M(j) = S(n)\begin{pmatrix} a(k,n) & b(k,n) \\ c(k,n) & d(k,n)\end{pmatrix} S(k)^{-1}.$$

Define functions `stay-small`, `switch-big`, `switch-small`, and `stay-big` as follows:

$$\begin{aligned}\texttt{stay-small}(n) &= \lambda_2(n+1)(1+\varepsilon_{1,1}(n)) & \texttt{switch-big}(n) &= \lambda_2(n+1)\varepsilon_{1,2}(n)\\ \texttt{switch-small}(n) &= \lambda_1(n+1)\varepsilon_{2,1}(n) & \texttt{stay-big}(n) &= \lambda_1(n+1)(1+\varepsilon_{2,2}(n)).\end{aligned}$$

A straightforward calculation then shows that we have recursive equations:

$$\begin{aligned}a(k,n+1) &= \texttt{stay-small}(n)a(k,n) + \texttt{switch-big}(n)c(k,n)\\ b(k,n+1) &= \texttt{stay-small}(n)b(k,n) + \texttt{switch-big}(n)d(k,n)\\ c(k,n+1) &= \texttt{stay-big}(n)c(k,n) + \texttt{switch-small}(n)a(k,n)\\ d(k,n+1) &= \texttt{stay-big}(n)d(k,n) + \texttt{switch-small}(n)b(k,n).\end{aligned} \tag{2}$$

By definition we have the following initial values:

$$a(k,k) = \lambda_2(k) \qquad b(k,k) = 0 \qquad c(k,k) = 0 \qquad d(k,k) = \lambda_1(k).$$

The next three lemmas constitute the key steps in the proof of Theorem 1. We defer their proof to the appendix.

▶ **Lemma 9.**

1. The function $\lambda_1(n)$ is positive, strictly monotonically increasing, and satisfies $\lambda_1(n) = \Theta(P(n))$ as $n \to \infty$.
2. The function $\lambda_2(n)$ is either positive for all n or negative for all n. It is either constant, strictly monotonically decreasing for large n, or strictly monotonically increasing for large n. It satisfies $\lambda_2(n) = \Theta(n^{\deg Q - \deg P})$ as $n \to \infty$.
3. We have $\texttt{stay-big}(n) = \Theta(P(n))$ as $n \to \infty$.
4. We have $\texttt{switch-big}(n) = \Theta(n^{\deg Q - \deg P - 1})$ as $n \to \infty$.
5. We have $\texttt{stay-small}(n) = \Theta(n^{\deg Q - \deg P})$ as $n \to \infty$.
6. If λ_2 is constant then $\texttt{switch-small} = 0$ for all n. Otherwise $\texttt{switch-small} = O(n^{-2})$ and $\texttt{switch-small} = \Omega(n^{1-3\deg P})$ as $n \to \infty$.

▶ **Lemma 10.** We can compute a number $K \in \mathbb{N}$ such that for all $k \geq K$ there exists $N \in \mathbb{N}$ such that $d(k,n) > 0$ for all $n \geq N$.

▶ **Lemma 11.** Let K be as in Lemma 10. Let $k \geq K$ be fixed and $n \geq k+5$ such that $d(k,n') > 0$ for all $n' \geq n$. Then $|a(k,n)/d(k,n)| \in O(1/nP(n))$ and $|b(k,n)/d(k,n)| \in O(1/nP(n))$.

▶ **Lemma 12.** Let K be as in Lemma 10. Let $k \geq K$ be fixed and $n \geq k+3$ such that $d(k,n') > 0$ for all $n' \geq n$. Then the sequence $c(k,n)/d(k,n)$ converges to a limit $L(k)$ as $n \to \infty$. We have $L(k) = O(1/k^2P(k)^2)$ as $k \to \infty$. The number $L(k)$ is equal to zero if λ_2 is constant. If λ_2 is decreasing and positive or increasing and negative then the number $L(k)$ is positive. If λ_2 is increasing and positive or decreasing and negative then the number $L(k)$ is negative. Moreover, for any given $p \in \mathbb{N}$ we can compute a rational number $\tilde{L} \in Q$ with $|\tilde{L} - L(k)| < 2^{-p}$.

3.2 Proof of Theorem 1.

With the asymptotic behaviour of the matrix entries being established, we can study the asymptotic behaviour of the sequence $(f(n))_n$. Using Lemmas 10 and 12 we can compute a number K such that for all $k \geq K$, the number $L(k)$ is defined and $1 - L(k) > 0$. By definition of $M(n)$ we have for all $k \geq K$:

$$\begin{aligned}\begin{pmatrix} f(n) \\ f(n+1) \end{pmatrix} &= \prod_{j=k}^{n-1} M(j) \begin{pmatrix} f(k) \\ f(k+1) \end{pmatrix} \\ &= S(n) \begin{pmatrix} a(k,n-1) & b(k,n-1) \\ c(k,n-1) & d(k,n-1) \end{pmatrix} S(k)^{-1} \begin{pmatrix} f(k) \\ f(k+1) \end{pmatrix}.\end{aligned}$$

By calculating the right hand side explicitly we obtain:

$$\begin{aligned}&(\lambda_1(k) - \lambda_2(k))f(n) = \\ &\quad \big(a(k,n-1)\,(\lambda_1(k)f(k) - f(k+1)) \qquad +b(k,n-1)\,(f(k+1) - \lambda_2(k)f(k)) \\ &\quad +c(k,n-1)\,(\lambda_1(k)f(k) - f(k+1)) \qquad +d(k,n-1)\,(f(k+1) - \lambda_2(k)f(k))\,\big).\end{aligned}$$

Using Lemma 11 we obtain:

$$(\lambda_1(k) - \lambda_2(k))\tfrac{f(n)}{d(k,n-1)} = \tag{3}$$

$$\tfrac{c(k,n-1)}{d(k,n-1)}\,(\lambda_1(k)f(k) - f(k+1)) + (f(k+1) - \lambda_2(k)f(k)) + O(1/nP(n)) \tag{4}$$

Passing to the limit as $n \to \infty$ we obtain, using Lemma 12:

$$\lim_{n\to\infty} (\lambda_1(k) - \lambda_2(k))\,\tfrac{f(n)}{d(k,n-1)} = (1 - L(k))f(k+1) + (\lambda_1(k)L(k) - \lambda_2(k))f(k). \tag{5}$$

Let $\ell(k) = (1 - L(k))f(k+1) + (\lambda_1(k)L(k) - \lambda_2(k))f(k)$. Note that this defines a straight line for all k, since $1 - L(k) \neq 0$ by assumption.

Now there are two cases:

1. There exists a $k \geq K$ such that $\ell(k) \neq 0$. In this case the sign $f(n)$ is eventually constant and the same as that of $\ell(k)$. It follows from Lemma 12 that $\ell(k)$ is computable. This together with the estimate (4) with an effective constant for the $O(1/nP(n))$ term allows us to compute an index N such that the sign of $f(n)$ is equal to that of $\ell(k)$ for all $n \geq N$.
2. We have $\ell(k) = 0$ for all $k \geq K$. Then the sequence satisfies the first-order recurrence relation

$$f(k+1) = \frac{\lambda_2(k) - \lambda_1(k)L(k)}{1 - L(k)} f(k)$$

for all $k \geq K$. In particular, if λ_2 is negative for all k then the sequence $(f(n))_{n\geq K}$ is zero or alternating, and if λ_2 is positive then the sign of every sequence element $f(n)$ with $n \geq K$ is equal to that of $f(K)$.

Now, since $Q(n)$ is assumed to have no integer zeroes, the matrices $M(n)$ are non-singular for all n. It follows with the above that if $\ell(n) \neq 0$ then the behaviour of the sequence as $n \to \infty$ is robust under small perturbations, while if $\ell(n) \neq 0$ then the behaviour changes under arbitrarily small perturbations of the initial values. It follows that for all $n \geq K$, the matrix $M(n)$ sends the line $L_n = \left\{(x,y) \in \mathbb{R}^2 \mid (1 - L(n))y + (\lambda_1(n)L(n) - \lambda_2(n))x\right\}$ to the line L_{n+1}. Hence $\ell(k) = 0$ for some $k \geq K$ if and only if $\ell(k) = 0$ for all $k \geq K$. Thus, the first case in the above case alternative occurs if and only if $\ell(K) \neq 0$. Also note that since the matrices $M(j)$ are all invertible, if the sequence $(f(n))_n$ is eventually zero then it is everywhere zero. It follows that in the second case alternative above the sequence is either eventually alternating or eventually has constant sign.

Let $h = M(K-1) \cdot \dots \cdot M(1)$. Let

$$H^+ = h^{-1}\left(\left\{(x,y) \in \mathbb{R}^2 \mid (1-L(K))y + (\lambda_1(K)L(K) - \lambda_2(K))x > 0\right\}\right)$$
$$H^- = h^{-1}\left(\left\{(x,y) \in \mathbb{R}^2 \mid (1-L(K))y + (\lambda_1(K)L(K) - \lambda_2(K))x < 0\right\}\right)$$
$$L^+ = h^{-1}\left(\left\{(x,y) \in \mathbb{R}^2 \mid (1-L(K))y + (\lambda_1(K)L(K) - \lambda_2(K))x = 0, x > 0\right\}\right)$$
$$L^- = h^{-1}\left(\left\{(x,y) \in \mathbb{R}^2 \mid (1-L(K))y + (\lambda_1(K)L(K) - \lambda_2(K))x = 0, x < 0\right\}\right).$$

Theorems 1 and 3 follow.

Proposition 2 is now proved as follows: For $n \geq K$, let $S(n)$ denote the slope of the line $L_n = \left\{(x,y) \in \mathbb{R}^2 \mid (1-L(n))y + (\lambda_1(n)L(n) - \lambda_2(n))x\right\}$. Using that $M(n)$ maps L_n onto L_{n+1} we obtain the equation $S(n) = -\frac{Q(n)}{P(n)-S(n+1)}$. This yields $S(K) = -\mathrm{K}_{m=K}^{\infty} \frac{Q(m)}{P(m)}$, and this is the slope of the critical line of the recurrence shifted by K.

If λ_2 is a constant function of n then $L_n = 0$ for all n, as is readily seen from the recursive equation (2) for $c(n,k)$. Thus, $\ell(K) = y - \lambda_2 x$. Since the vector $(1, \lambda_2)$ is an eigenvector for all matrices $M(1), \dots, M(K-1)$ we have $h^{-1}(\{(x,y) \mid y - \lambda_2 x = 0\}) = \{(x,y) \mid y - \lambda_2 x = 0\}$. It follows that we have $L^+ \cup L^- \cup O = \{(x,y) \mid y = \lambda_2 x\}$. This establishes Proposition 4. Corollary 5 follows together with the above discussion.

Let us now prove Proposition 6. Write $P(n) = a_d n^d + \dots + a_0$ and $Q(n) = b_d n^d + \dots + b_0$ with $a_d > 0$. Then the limit of $\lambda_2(n)$ as $n \to \infty$ is equal to $-b_d/a_d$. This follows for instance from the series representation (10) in the appendix. We have seen in Lemma 9 that λ_2 is either constant or strictly monotone. It follows that λ_2 is decreasing or constant if and only if $\lambda_2(n) \geq -b_d/a_d$ for all sufficiently large n, with λ_2 being decreasing if and only if the inequality is strict. By writing out the definition of $\lambda_2(n)$ and applying basic algebra we obtain that this is equivalent to:

$$P(n) + 2b_d/a_d \geq \left(P(n)^2 + 4Q(n)\right)^{\frac{1}{2}}.$$

For sufficiently large n the expressions on both sides are positive, so the inequality is equivalent to the same inequality with both sides squared:

$$P(n)^2 + 4P(n)b_d/a_d + 4b_d^2/a_d^2 \geq P(n)^2 + 4Q(n).$$

This is further equivalent to the inequality:

$$P(n)b_d - a_d Q(n) + b_d^2/a_d \geq 0.$$

Proposition 6 follows.

3.3 Proof of Theorem 7

By the proof of Theorem 1 the critical line is, up to potentially shifting the recurrence, given by the equation

$$(1 - L(1))f(2) + (\lambda_1(1)L(1) - \lambda_2(1))f(2).$$

If the equation has a non-zero rational solution then it has a non-zero integer solution. Thus, assume that the equation has an integer solution $(f(1), f(2))$ with $f(1)$ and $f(2)$ not both zero. Then the recurrence sequence $(f(n))_n$ satisfies

$$f(n+1) = \frac{\lambda_2(n) - L(n)\lambda_1(n)}{1 - L(n)} f(n). \tag{6}$$

If $\deg Q < \deg P$ or $\deg Q = \deg P$ and $|\operatorname{lcof}(Q)| < |\operatorname{lcof}(P)|$ then it follows from Lemma 9 that $\lambda_2(n) \in o(1)$ as $n \to \infty$. It follows from (6) that $f(n) \in o(1)$. But since $(f(n))_n$ is an integer sequence it follows that $f(n) = 0$ for all large n. But then $f(n) = 0$ for all n by Theorem 1. Hence, the only integer solution is $(0, 0)$.

It remains to consider the cases where $\deg Q = \deg P$ and $|\operatorname{lcof}(Q)| = |\operatorname{lcof}(P)|$.

We claim that in the case where λ_2 is negative and decreasing or positive and increasing the sequence $(f(n))_n$ is bounded. In this case $|\lambda_2(n)| < 1$ for all n. It follows from (6) and Lemma 12 that there exists a constant c such that $|f(n)| \leq \prod_{k=1}^{n}(1 + \frac{c}{k^2})|f(1)|$. Boundedness of the sequence follows by taking logarithms on both sides and noting that the sequence $\sum_{n=1}^{\infty} n^{-2}$ converges.

We claim that if λ_2 is positive and decreasing and $d \geq 2$ then $|f(n)| \leq Cn^{\left|\frac{a_{d-1}+b_{d-1}}{a_d}\right|}$. By assumption on λ_2 and Lemma 12 the number $L(n)$ is positive, so that we have for all sufficiently large n:

$$\left|\frac{\lambda_2(n) - L(n)\lambda_1(n)}{1 - L(n)}\right| \leq \lambda_2(n)$$

There exist constants c and c' such that for all sufficiently large n we have:

$$|\lambda_2(n)| \leq 1 + \left|\frac{b_{d-1} + a_{d-1}}{a_d}\right| \frac{1}{n - c'} + \frac{c}{n^2}.$$

It follows that, for sufficiently large M and all $N \geq M$ we have:

$$f(N) \leq |f(M)| \prod_{n=M}^{N} \left(1 + \left|\frac{b_{d-1} + a_{d-1}}{a_d}\right| \frac{1}{n - c'} + \frac{c}{n^2}\right).$$

Taking logarithms on both sides we obtain:

$$\log f(N) \leq \log |f(M)| + \sum_{n=M}^{N} \log \left(1 + \left|\frac{b_{d-1} + a_{d-1}}{a_d}\right| \frac{1}{n - c'} + \frac{c}{n^2}\right)$$

Using $\log(1 + x) = x + O(x^2)$ we obtain:

$$\begin{aligned} \log f(N) &\leq \log |f(M)| + \sum_{n=M}^{N} \left(\left|\frac{b_{d-1} + a_{d-1}}{a_d}\right| \frac{1}{n - c'} + \frac{c}{n^2} + O(1/n^2)\right) && (7) \\ &\leq \log |f(M)| + \left|\frac{b_{d-1} + a_{d-1}}{a_d}\right| \log(N) + C', && (8) \end{aligned}$$

where C' is a constant.

We have used in (7) that $\sum_{n=1}^{N} \frac{1}{n} \leq \log(N) + \gamma + 1$ for all large N, where $\gamma \approx 0.5572\ldots$ is the Euler-Mascheroni constant. Since $\sum_{n=1}^{N} \frac{1}{n}$ diverges, it follows that $\sum_{n=M}^{N} \frac{1}{n} \leq \log(N)$ for large M. Now apply the exponential function to both sides of (8):

$$f(N) \leq |f(M)| N^{\left|\frac{b_{d-1}+a_{d-1}}{a_d}\right|} e^{C'}.$$

This proves the claim. An analogous argument shows that $f(N) \in O\left(N^{\left|\frac{b_{d-1}+a_{d-1}}{a_d}\right|}\right)$ if λ_2 is negative and increasing and $d \geq 2$. Analogous claims hold for the case that $d = 1$. Now, by assumption we have

$$L(n) = \frac{\lambda_2(n)f(n) - f(n+1)}{\lambda_1(n)f(n) - f(n+1)}. \tag{9}$$

By our previous considerations, the denominator of this expression is in $O(\frac{1}{P(n)})$ if $\lambda_2(n)$ is positive and increasing or if $\lambda_2(n)$ is negative and decreasing. If $\lambda_2(n)$ is, say, positive and decreasing and $d \geq 2$ then the expression is in $O\left(n^{\left|\frac{b_{d-1}+a_{d-1}}{a_d}\right|}P(n)\right)$. Similarly for the other cases we consider. Thus, if $\lambda_2(n)$ is positive and increasing or negative and decreasing, or for instance if $d \geq 2$ and it is positive and decreasing and $\left|\frac{b_{d-1}+a_{d-1}}{a_d}\right| < \deg P + 2$, then the numerator of (9) has to be in $o(1)$.

In other words, we have a sequence of pairs of non-zero integers (x_n, y_n) with $|x_n|, |y_n| \leq n^p$ for some positive integer p such that $|\lambda_2(n)x_n - y_n| \to 0$ as $n \to \infty$. We may assume that $p \geq 2$. Now use the series representation (10) of $\lambda_2(n)$ computed in Appendix A:

$$\begin{aligned}
&|\lambda_2(n)x_n - y_n| \\
&= \left| -x_n \sum_{m=1}^{\infty} \frac{2^{m-1}Q(n)^m}{m!} \prod_{j=1}^{m-1} (1-2j)\, P(n)^{1-2m} - y_n \right| \\
&= \left| x_n \sum_{m=1}^{2p} \frac{2^{m-1}Q(n)^m}{m!} \prod_{j=1}^{m-1} (1-2j)\, P(n)^{1-2m} + y_n + O\left(n^p Q(n)^{2p}/P(n)^{4p-1}\right) \right| \\
&= \left| P(n)^{1-4p} \left(x_n \left(\sum_{m=1}^{2p} \frac{2^{m-1}Q(n)^m}{m!} \prod_{j=1}^{m-1} (1-2j)\, P(n)^{4p-2m} \right) + y_n P(n)^{4p-1} \right) \right| \\
&\qquad + O\left(n^p Q(n)^{2p}/P(n)^{4p-1}\right).
\end{aligned}$$

In order for this to converge to zero we must have

$$x_n \left(\sum_{m=1}^{2p} \frac{2^{m-1}Q(n)^m}{m!} \prod_{j=1}^{m-1} (1-2j)\, P(n)^{4p-2m} \right) + y_n P(n)^{4p-1} \in o(P(n)^{4p-1}).$$

But the left hand side is an integer linear combination of two polynomials of degree $(4p-1)\deg P$. By assumption, their leading coefficients are either equal or additive inverses of each other - depending on the sign of $\lambda_2(n)$. It follows that $x_n = y_n$ or $x_n = -y_n$ for all sufficiently large n. Let us without loss of generality assume that the former holds true.

Then, for all sufficiently large n the lines

$$L_n = \{(x, y) \mid (\lambda_1(n)L(n) - \lambda_2(n))x + (1 - L(n)) = 0\}$$

are all equal to the diagonal $x = y$. Since the matrix $M(n)$ sends the line L_n line onto the line L_{n+1}, it follows that the vector $(1, 1)$ is an eigenvector for every $M(n)$. The corresponding eigenvalue must be $\lambda_2(n)$. But since the vector gets mapped to itself it follows that $\lambda_2(n) = 1$ for all n, contradicting our initial assumption.

References

1. Jason P. Bell, Stanley N. Burris, and Karen Yeats. On the set of zero coefficients of a function satisfying a linear differential equation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 153(2):235–247, 2012.
2. Stefan Gerhold and Manuel Kauers. A procedure for proving special function inequalities involving a discrete parameter. In Manuel Kauers, editor, *Symbolic and Algebraic Computation, International Symposium ISSAC 2005, Beijing, China, July 24-27, 2005, Proceedings*, pages 156–162. ACM, 2005.

3 Manuel Kauers and Veronika Pillwein. When can we detect that a P-finite sequence is positive? In Wolfram Koepf, editor, *Symbolic and Algebraic Computation, International Symposium, ISSAC 2010, Munich, Germany, July 25-28, 2010, Proceedings*, pages 195–201. ACM, 2010.

4 George Kenison, Oleksiy Klurman, Engel Lefaucheux, Florian Luca, Pieter Moree, Joël Ouaknine, Markus A. Whiteland, and James Worrell. On positivity and minimality for second-order holonomic sequences. *CoRR*, abs/2007.12282, 2020. URL: `https://arxiv.org/abs/2007.12282`.

5 Maxim Kontsevich and Don Zagier. *Periods*, pages 771–808. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

6 Joël Ouaknine and James Worrell. Positivity problems for low-order linear recurrence sequences. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 366–379. SIAM, 2014.

7 Marko Petkovšek, Herbert S. Wilf, and Doron Zeilberger. *A = B*. A.K. Peters, 1997.

8 Veronika Pillwein. Termination conditions for positivity proving procedures. In *Proceedings of the 38th International Symposium on Symbolic and Algebraic Computation*, ISSAC '13, page 315–322, New York, NY, USA, 2013. Association for Computing Machinery.

9 Veronika Pillwein and Miriam Schussler. An efficient procedure deciding positivity for a class of holonomic sequences. *ACM Communications in Computer Algebra*, 49(3):90–93, 2015. Extended abstract of the poster presentation at ISSAC 2015.

10 R.P. Stanley. Differentiably finite power series. *European Journal of Combinatorics*, 1(2):175–188, 1980.

11 Doron Zeilberger. A holonomic systems approach to special functions identities. *Journal of Computational and Applied Mathematics*, 32(3):321–368, 1990.

A Proof of Lemmas 11 and 12

We will expand the terms $a(k,n)$, $b(k,n)$, $c(k,n)$, and $d(k,n)$ into large sums based on the above recursive equations. It will be convenient to describe these sums with the help of a finite automaton. Consider the finite automaton $\mathcal{A}$ over the alphabet $\Sigma = \{\texttt{stay-big}, \texttt{switch-small}, \texttt{stay-small}, \texttt{switch-big}\}$ defined in Figure 1.

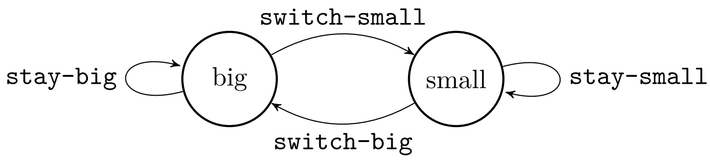

Figure 1 The automaton $\mathcal{A}$

Let [small → small] $\subseteq \Sigma^*$ denote the set of all words that are accepted by $\mathcal{A}$ with initial state "small" and accepting state "small". Define the sets [small → big], [big → small], and [big → big] analogously.

For each symbol $s \in \Sigma$, let $[\![s]\!] \colon \mathbb{N} \to \mathbb{R}$ be the obvious function associated with it. For a word $w = w_1 \cdot \dots \cdot w_s$ over the alphabet Σ and $n \geq s$, let $[\![w]\!](n) = [\![w_1]\!](n) \cdot \dots \cdot [\![w_s]\!](n-s+1)$.

For $k \leq n$, let

$$A(k,n) = [\text{small} \to \text{small}] \cap \Sigma^{n-k} \qquad B(k,n) = [\text{small} \to \text{big}] \cap \Sigma^{n-k}$$
$$C(k,n) = [\text{big} \to \text{small}] \cap \Sigma^{n-k} \qquad D(k,n) = [\text{big} \to \text{big}] \cap \Sigma^{n-k}$$

From the above recursive equations and initial values we obtain for all $n > k$:

$$a(k,n) = \lambda_2(k) \sum_{w \in A(k,n)} [\![w]\!](n-1) \qquad b(k,n) = \lambda_1(k) \sum_{w \in B(k,n)} [\![w]\!](n-1)$$
$$c(k,n) = \lambda_2(k) \sum_{w \in C(k,n)} [\![w]\!](n-1) \qquad d(k,n) = \lambda_1(k) \sum_{w \in D(k,n)} [\![w]\!](n-1).$$

We want to study the asymptotic behaviour of the quotients $a(k,n)/d(k,n)$, $b(k,n)/d(k,n)$, and $c(k,n)/d(k,n)$. In order to do so, we first need to study the asymptotic behaviour of the functions `stay-big`, `switch-small`, `stay-small`, `switch-big`. In the sequel we will denote these functions by `stb`, `sws`, `sts`, `swb` for short.

Recall that we have

$$\lambda_1(n) = \tfrac{1}{2}\left(P(n) + \left(P(n)^2 + 4Q(n)\right)^{1/2}\right)$$

and

$$\lambda_2(n) = \tfrac{1}{2}\left(P(n) - \left(P(n)^2 + 4Q(n)\right)^{1/2}\right)$$

The Taylor series expansion of $h(x) = x^{1/2}$ about $x = P(n)^2$ is

$$x^{1/2} = \sum_{m=0}^{\infty} \frac{(x - P(n)^2)^m}{2^m m!} \left(\prod_{j=1}^{m-1} (1-2j)\right) P(n)^{1-2m}.$$

Hence:

$$\left(P(n)^2 + 4Q(n)\right)^{1/2} = \sum_{m=0}^{\infty} \frac{2^m Q(n)^m}{m!} \left(\prod_{j=1}^{m-1} (1-2j)\right) P(n)^{1-2m}. \tag{10}$$

Let

$$\rho(n) = \sum_{m=1}^{\infty} \frac{2^{m-1} Q(n)^m}{m!} \prod_{j=1}^{m-1} (1-2j)\, P(n)^{1-2m}.$$

Then we have

$$\lambda_1(n) = P(n) + \rho(n),$$

and

$$\lambda_2(n) = -\rho(n).$$

Note that the series

$$\sum_{m=1}^{\infty} \frac{2^{m-1} Q(n)^m}{m!} \prod_{j=1}^{m-1} (1-2j)\, P(n)^{1-2m}$$

is majorised by the geometric series $P(n) \sum_{m=1}^{\infty} \left(\frac{4Q(n)}{P(n)^2}\right)^m$. In particular we have

$$\sum_{m=k}^{\infty} \frac{2^{m-1} Q(n)^m}{m!} \prod_{j=1}^{m-1} (1-2j)\, P(n)^{1-2m} = O\left(\frac{Q(n)^k}{P(n)^{2k-1}}\right).$$

▶ **Lemma 13** (Lemma 9).

1. The function $\lambda_1(n)$ is positive, strictly monotonically increasing, and satisfies $\lambda_1(n) = \Theta(P(n))$ as $n \to \infty$.
2. The function $\lambda_2(n)$ is either positive for all n or negative for all n. It is either constant, strictly monotonically decreasing for large n, or strictly monotonically increasing for large n. It satisfies $\lambda_2(n) = \Theta(n^{\deg Q - \deg P})$ as $n \to \infty$.
3. We have $\texttt{stb}(n) = \Theta(P(n))$ as $n \to \infty$.
4. We have $\texttt{swb}(n) = \Theta(n^{\deg Q - \deg P - 1})$ as $n \to \infty$.
5. We have $\texttt{sts}(n) = \Theta(n^{\deg Q - \deg P})$ as $n \to \infty$.
6. If λ_2 is constant then $\texttt{sws} = 0$ for all n. Otherwise $\texttt{sws} = O(n^{-2})$ and $\texttt{sws} = \Omega(n^{1-3\deg P})$ as $n \to \infty$.

Proof. It is clear that $\lambda_1(n)$ is positive and monotonically increasing. It follows easily from the series representation (10) that $\lambda_1(n)$ has the claimed asymptotic behaviour.

If $Q(n)$ is negative for all n then $\left(P(n)^2 + 4Q(n)\right)^{1/2} < P(n)$ and $\lambda_2(n)$ is positive for all n. If $Q(n)$ is positive for all n then $\left(P(n)^2 + 4Q(n)\right)^{1/2} > P(n)$ and $\lambda_2(n)$ is negative for all n. It follows easily from the series representation (10) that $\lambda_2(n)$ has the claimed asymptotic behaviour.

The claimed asymptotic behaviour of the functions $\texttt{stb}$, $\texttt{swb}$, and $\texttt{sts}$ is easily verified.

It remains to prove that λ_2 is either constant or monotone and to study the asymptotic behaviour of $\texttt{sws}$. To this end we compute the derivative of $\lambda_2(z)$ for $z \in \mathbb{R}$:

$$\lambda_2'(z) = \frac{P'(z)(P(z)^2 + 4Q(z))^{1/2} - P(z)P'(z) - 2Q'(z)}{2(P(z)^2 + 4Q(z))^{-1/2}}.$$

Let $A(z) = P'(z)(P(z)^2 + 4Q(z))^{1/2}$ and $B(z) = P(z)P'(z) - 2Q'(z)$. Then $A(z)^2$ and $B(z)^2$ are polynomials. Hence, if the functions $A(z)$ and $B(z)$ are not equal everywhere, then there exists a positive constant c such that $|A(z)^2 - B(z)^2| \geq c$ for all sufficiently large z. Thus,

$$|A(z) - B(z)| = |A(z)^2 - B(z)^2| / |A(z) + B(z)| \geq c / |A(z) + B(z)|.$$

This already establishes that λ_2 is either constant or monotone.

We have $\texttt{sws} \in \Theta(\lambda_2(n) - \lambda_2(n+1))$. If λ_2 is constant then clearly $\texttt{sws} = 0$. Assume now that λ_2 is not constant. The upper bound $\texttt{sws} \in O(n^{-2})$ can be deduced from the easily established fact that if the degree of P and Q is bounded by d, then the degree of $Q(n)P(n+1) - Q(n+1)P(n)$ is bounded by $2d - 2$.

From the series representation (10) we obtain $|A(z) + B(z)| \in O(z^{2\deg P - 1})$ and $2(P(z)^2 - 4Q(z))^{-1/2} \in O(z^{\deg P})$. It follows that $|\lambda_2'(z)| \in \Omega(z^{1-3\deg P})$. Then, by the mean value theorem, $|\lambda_2(n) - \lambda_2(n+1)| \in \Omega(n^{1-3\deg P})$. ◀

The signs of the coefficients $\texttt{stb}, \texttt{sws}, \texttt{swb}, \texttt{sts}$ can be easily deduced from Lemma 9. They depend on the behaviour of the function $\lambda_2(n)$. Recall that this function is either constant, positive, or negative and either increasing or decreasing. This leads to four possible sign configurations, indicated in Figures 2 - 6.

We will treat the case where λ_2 is constant and the case where λ_2 is non-constant separately. Let us focus on the latter case for now. Let $D(k,n)^+$ denote the set of words in $D(k,n)$ which contain each of the symbols $\texttt{sws}$ and $\texttt{sts}$ an even number of times. Let $D(k,n)^-$ denote its complement. Clearly, for every word $w \in D(k,n)^+$, the number $[\![w]\!](n)$ is positive.

The following proposition is trivial but useful when comparing sums over large index sets:

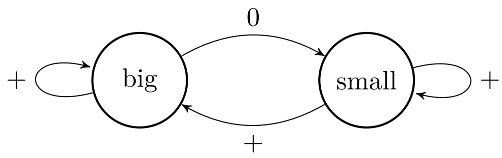

Figure 2 λ_2 constant

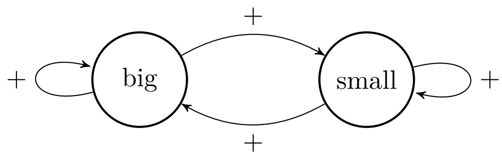

Figure 3 λ_2 positive and decreasing

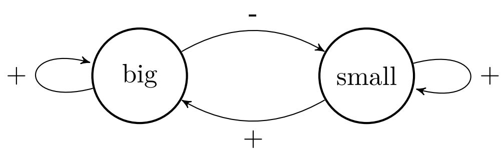

Figure 4 λ_2 positive and increasing

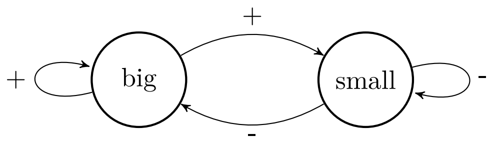

Figure 5 λ_2 negative and increasing

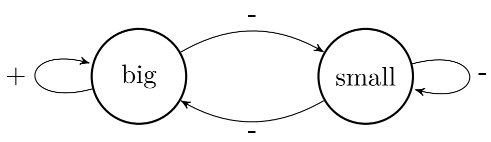

Figure 6 λ_2 negative and decreasing

▶ **Proposition 14.** Let A and B be finite sets of positive real numbers. Let $\mu\colon A \to B$ be a function. Assume that $a/\mu(a) < \varepsilon$ for some $\varepsilon > 0$ and $\mu^{-1}(b)$ contains at most c elements. Then

$$\sum_{a\in A} a / \sum_{b\in B} b < c\varepsilon.$$

▶ **Lemma 15.** Assume that λ_2 is non-constant. For all sufficiently large k and $n \geq k$ we have

$$\sum_{w\in D(k,n)^+} [\![w]\!](n-1) > 2 \sum_{w\in D(k,n)^-} |[\![w]\!](n-1)|$$

Proof. Let $D(k,n)_1$ denote the set of words with an odd number of $\mathtt{sws}$. Let $D(k,n)_2$ denote the set of words with an even number of $\mathtt{sws}$. Consider the map $\mu\colon D(k,n)_1 \to D(k,n)_2$ defined as follows: For a word $w \in D(k,n)_1$ there exist unique words p, r such that $w = p\cdot\mathtt{sws}\cdot r$ with $\mathtt{sws}$ not occurring in r. Let $\mu(w) = p\cdot\mathtt{stb}^{|r|+1}$.

For $w = p \cdot \texttt{sws} \cdot r$ we have

$$\mu^{-1}(\mu(w)) = \left\{ p \cdot \texttt{sws} \cdot \texttt{sts}^j \cdot \texttt{swb} \cdot \texttt{stb}^{|r|-j-1} \mid j \in \{0, \dots, |r|-1\} \right\}.$$

We have

$$\begin{aligned}
&\sum_{v \in \mu^{-1}(\mu(w))} |[\![v]\!](n-1)| / |[\![\mu(w)]\!](n-1)| \\
&= \sum_{j=0}^{n-k-|p|-2} \left| \frac{\texttt{sws}(n-|p|-1) \cdot \prod_{l=1}^{j} \texttt{sts}(n-|p|-l-1) \cdot \texttt{swb}(n-|p|-j-2) \cdot \prod_{l=k}^{n-|p|-j-3} \texttt{stb}(l)}{\texttt{stb}(n-|p|-1) \cdot \dots \cdot \texttt{stb}(k)} \right| \\
&\leq \sum_{j=0}^{n-k-|p|-2} \left| \frac{c/(n-|p|-1)^2 \cdot c^j \cdot c/(n-|p|-j-2)}{P(n-|p|-1) \cdot \dots \cdot P(n-|p|-j-2)} \right|
\end{aligned}$$

for some constant c. Now, for large k we have $P(k) > c$ and we can estimate:

$$\begin{aligned}
&\sum_{j=0}^{n-k-|p|-2} \left| \frac{c/(n-|p|-1)^2 \cdot c^j \cdot c/(n-|p|-j-2)}{P(n-|p|-1) \cdot \dots \cdot P(n-|p|-j-2)} \right| \\
&\leq \frac{c^2}{k(n-|p|-1)^2} \sum_{j=0}^{n-k-|p|-2} \left(\frac{c}{P(k)} \right)^j \\
&\leq \frac{c^2}{k^3} \frac{P(k)}{P(k)-c}.
\end{aligned}$$

It follows that

$$\sum_{w \in D(k,n)_1} |[\![w]\!](n-1)| / \sum_{w \in D(k,n)_2} |[\![w]\!](n-1)| \leq \frac{c^2}{k^3} \frac{P(k)}{P(k)-c}.$$

Define a map $\sigma \colon D(k,n)_2 \to D(k,n)^+$ as follows: For a word $w \in D(k,n)_2$, if w contains the symbol $\texttt{sts}$ an even number of times, let $\sigma(w) = w$. If w contains the symbol $\texttt{sts}$ an odd number of times then there exists a unique integer $e \geq 1$ and unique words p, q such that

$$w = p \cdot \texttt{sws} \cdot \texttt{sts}^e \cdot q$$

and q does not contain the symbol $\texttt{sts}$. Now, let

$$\sigma(w) = p \cdot \texttt{stb} \cdot \texttt{sws} \cdot \texttt{sts}^{e-1} \cdot q.$$

Then σ is a well-defined map of type $D(k,n)_2 \to D(k,n)^+$. Every word in $D(k,n)^+$ has at most two preimages under σ.

Let $w = p \cdot \texttt{sws} \cdot \texttt{sts}^e \cdot q \in D(k,n)_2$ be a word which contains the symbol $\texttt{sts}$ an odd number of times. Then

$$|[\![w]\!](n-1)| / [\![\sigma(w)]\!](n-1) = \frac{|\texttt{sws}(n-|p|-1) \cdot \texttt{sts}(n-|p|-2)|}{|\texttt{stb}(n-|p|-1) \cdot \texttt{sws}(n-|p|-2)|} \leq \frac{c}{P(k)}$$

with the same constant $c > 0$ as above. It follows that

$$\sum_{w \in D(k,n)_2} |[\![w]\!](n-1)| / \sum_{w \in D(k,n)^+} [\![w]\!](n-1) \leq 2.$$

Further, letting $D(k,n)_2^-$ denote the set of words in $D(k,n)_2$ which contain the symbol $\texttt{sts}$ an odd number of times, the same estimate shows that

$$\sum_{w\in D(k,n)_2^-} |[\![w]\!](n-1)| / \sum_{w\in D(k,n)^+} [\![w]\!](n-1) \leq \frac{c}{P(k)}.$$

Thus, in total, we have:

$$\begin{aligned}
&\frac{\sum_{w\in D(k,n)^-} |[\![w]\!](n-1)|}{\sum_{w\in D(k,n)^+} [\![w]\!](n-1)} \\
&\leq \frac{\sum_{w\in D(k,n)_2^-} |[\![w]\!](n-1)| + \sum_{w\in D(k,n)_1} |[\![w]\!](n-1)|}{\sum_{w\in D(k,n)^+} [\![w]\!](n-1)} \\
&= \frac{\sum_{w\in D(k,n)_2^-} |[\![w]\!](n-1)|}{\sum_{w\in D(k,n)^+} [\![w]\!](n-1)} + \frac{\sum_{w\in D(k,n)_1} |[\![w]\!](n-1)|}{\sum_{w\in D(k,n)_2} |[\![w]\!](n-1)|} \cdot \frac{\sum_{w\in D(k,n)_2} |[\![w]\!](n-1)|}{\sum_{w\in D(k,n)^+} [\![w]\!](n-1)} \\
&\leq \frac{c}{P(k)} + 2\frac{c^2}{k^3}\frac{P(k)}{P(k)-c}.
\end{aligned}$$

The right hand side converges to zero as $k \to \infty$. In particular it is smaller than $1/2$ for all sufficiently large k, which yields the claim. ◀

Lemma 15 immediately implies Lemma 10. The computability of the constant K is obtained by observing that we can effectively find all constants that appear in the estimates in the proof. We are now ready to prove Lemmas 11 and 12.

▶ **Lemma 16** (Lemma 11). Let K be as in Lemma 10. Let $k \geq K$ be fixed and $n \geq k+5$ such that $d(k,n') > 0$ for all $n' \geq n$. Then $|a(k,n)/d(k,n)| \in O(1/nP(n))$ and $|b(k,n)/d(k,n)| \in O(1/nP(n))$.

Proof. We only prove the claim for $b(k,n)$. The claim for $a(k,n)$ is proved analogously.

Let us first assume that λ_2 is non-constant. Then by Lemma 15 it suffices to show that

$$\sum_{w\in B(k,n)} |[\![w]\!](n-1)| / \sum_{w\in D(k,n)} |[\![w]\!](n-1)| = O(1/nP(n)).$$

By Lemma 9 we have $\texttt{sws} \in \Omega(n^{1-3\deg P})$. Define a map $\mu\colon B(k,n) \to D(k,n)$ as follows: for a word $w = p \cdot q$ in $B(k,n)$ with $|p| = 5$, let $\mu(w) = \texttt{stb}^4 \cdot s \cdot q$, where $s = \texttt{sws}$ if $q \in [\text{small} \to \text{big}]$ and $s = \texttt{stb}$ if $q \in [\text{big} \to \text{big}]$. Then the set $\mu^{-1}(\mu(w))$ contains at most 16 elements. By Proposition 14 it suffices to show that $|[\![w]\!](n-1)|/|[\![\mu(w)]\!](n-1)| \in O(1/nP(n))$.

Consider two cases. The first case is that $w = p \cdot q$ with $p \in [\text{small} \to \text{small}]$, $|p| = 5$. Note that $|[\![p]\!](n-1)|$ is smaller than $[\![\texttt{sts}^5]\!](n-1)$ or $[\![\texttt{swb} \cdot \texttt{stb}^3 \cdot \texttt{sws}]\!](n-1)$. Now, $\mu(\texttt{sts}^5 \cdot q) = \texttt{stb}^4 \cdot \texttt{sws} \cdot q$, with

$$\begin{aligned}
\frac{[\![\texttt{sts}^5 \cdot q]\!](n-1)}{[\![\texttt{stb}^4 \cdot \texttt{sws} \cdot q]\!](n-1)} &= \frac{\texttt{sts}(n-1) \cdot \cdots \cdot \texttt{sts}(n-5)}{\texttt{stb}(n-1) \cdot \cdots \cdot \texttt{stb}(n-4) \cdot \texttt{sws}(n-5)} \\
&= \frac{\Theta(n^{5(\deg P - \deg Q)})}{\Theta(n^{4\deg P} n^{1-3\deg P})} \\
&= O\left(\frac{1}{nP(n)}\right).
\end{aligned}$$

It remains to check the other possibility. $\mu(\texttt{swb} \cdot \texttt{stb}^3 \cdot \texttt{sws} \cdot q) = \texttt{stb}^4 \cdot \texttt{sws} \cdot q$, with

$$\frac{[\![\texttt{swb} \cdot \texttt{stb}^3 \cdot \texttt{sws} \cdot q]\!](n-1)}{[\![\texttt{stb}^4 \cdot \texttt{sws} \cdot q]\!](n-1)} = \frac{\texttt{swb}(n-1)}{\texttt{stb}(n-1)} = \frac{O(1/n)}{\Theta(P(n))} = O\left(\frac{1}{nP(n)}\right).$$

The second case is that $w = p \cdot q$ with $p \in [\text{small} \to \text{big}]$, $|p| = 5$. Again, $|[\![p]\!](n-1)|$ is smaller than $[\![\texttt{swb} \cdot \texttt{stb}^4]\!](n-1)$ or $[\![\texttt{sts}^4 \cdot \texttt{swb}]\!](n-1)$. We have $\mu(\texttt{swb} \cdot \texttt{stb}^4 \cdot q) = \texttt{stb}^5 \cdot q$ with

$$\frac{[\![\texttt{swb} \cdot \texttt{stb}^4 \cdot q]\!](n-1)}{[\![\texttt{stb}^5 \cdot q]\!](n-1)} = \frac{\texttt{swb}(n-1)}{\texttt{stb}(n-1)} = O\left(\frac{1}{nP(n)}\right)$$

and $\mu(\texttt{sts}^4 \cdot \texttt{swb} \cdot q) = \texttt{stb}^5 \cdot q$ with

$$\frac{[\![\texttt{sts}^4 \cdot \texttt{swb} \cdot q]\!](n-1)}{[\![\texttt{stb}^5 \cdot q]\!](n-1)} = \frac{\texttt{sts}(n-1) \cdot \dots \cdot \texttt{sts}(n-4) \cdot \texttt{swb}(n-5)}{\texttt{stb}(n-1) \cdot \dots \cdot \texttt{stb}(n-5)} = O\left(\frac{1}{nP(n)}\right).$$

This proves the claim.

It remains to examine the case where λ_2 is constant. In this case, $\texttt{sws} = 0$, so that

$$d(k,n) = \texttt{stb}(n-1) \cdot \dots \cdot \texttt{stb}(k)\lambda_1(k).$$

We have

$$B(k,n) = \left\{\texttt{sts}^e \cdot \texttt{swb} \cdot \texttt{stb}^{n-k-e} \mid e \in \{0, \dots, n-k\}\right\},$$

so that

$$b(k,n) = \lambda_2(k) \sum_{e=0}^{n-k} \left(\prod_{j=1}^{e} \texttt{sts}(n-j)\right) \texttt{swb}(n-e-1) \left(\prod_{j=e+1}^{n-k} \texttt{stb}(n-j-1)\right).$$

It follows that

$$b(k,n)/d(k,n) = \tfrac{\lambda_2(k)}{\lambda_1(k)} \sum_{e=0}^{n-k} \frac{\left(\prod_{j=1}^{e} \texttt{sts}(n-j)\right) \texttt{swb}(n-e-1)}{\prod_{j=0}^{e} \texttt{stb}(n-j-1)}.$$

Now, by Lemma 9 there exists a positive constant c such that $|\texttt{sts}(n-j)| \leq c$, $|\texttt{swb}(n-e-1)| \leq c/(n-e)$, and $|\texttt{stb}(n-j-1)| \geq P(n)/c$. Thus:

$$\begin{aligned}
|b(k,n)/d(k,n)| &\leq \sum_{e=0}^{n-k} \frac{c^{2e+1}}{(n-e)P(n) \cdot P(n-1) \cdot \dots \cdot P(n-e)} \\
&\leq \frac{c}{nP(n-k)} \sum_{e=0}^{n-k} \frac{(c^2)^e}{(n-1) \cdot \dots \cdot (n-e)} \\
&\leq \frac{c}{nP(n-k)} \sum_{e=0}^{n-k} \frac{(c^2)^e}{e!} \\
&\leq \frac{c\exp(c^2)}{nP(n-k)} \\
&= O(1/nP(n)).
\end{aligned}$$

◀

▶ **Lemma 17** (Lemma 12)**.** Let K be as in Lemma 10. Let $k \geq K$ be fixed and $n \geq k+3$ such that $d(k,n') > 0$ for all $n' \geq n$. Then the sequence $c(k,n)/d(k,n)$ converges to a limit $L(k)$ as $n \to \infty$. We have $L(k) = O(1/k^2P(k)^2)$ as $k \to \infty$. The number $L(k)$ is equal to zero if λ_2 is constant. If λ_2 is decreasing and positive or increasing and negative then the number $L(k)$ is positive. If λ_2 is increasing and positive or decreasing and negative then the number $L(k)$ is negative. Moreover, for any given $p \in \mathbb{N}$ we can compute a rational number $\tilde{L} \in Q$ with $|\tilde{L} - L(k)| < 2^{-p}$.

Proof. If $\texttt{sws} = 0$ then $c(k,n) = 0$ for all k and n, so that the claim is trivial.

Let us hence assume that $\texttt{sws} \neq 0$. Define a map $\mu\colon C(k,n) \to D(k,n)$ as follows: Let $w = p \cdot q \in C(k,n)$ with $|q| = 3$. If $p \in [\text{big} \to \text{big}]$ then let $\mu(w) = p \cdot \texttt{stb}^3$. If $p \in [\text{big} \to \text{small}]$ then let $\mu(w) = p \cdot \texttt{swb} \cdot \texttt{stb}^2$. One easily verifies that $|[\![w]\!](n-1)/[\![\mu(w)]\!](n-1)| \in O(1/k^2 P(k))$ with a constant that does not depend on n. It follows from Lemma 15, Lemma 9, and Proposition 14 that

$$|c(k,n)/d(k,n)| = O(1/k^2 P(k)^2) \tag{11}$$

with a constant that does not depend on n. In particular, for all fixed k the sequence $(c(k,n)/d(k,n))_n$ is bounded.

Now, by (2) we have:

$$\begin{aligned} \frac{c(k,n+1)}{d(k,n+1)} &= \frac{\texttt{stb}(n)c(k,n) + \texttt{sws}(n)a(k,n)}{\texttt{stb}(n)d(k,n) + \texttt{sws}(n)b(k,n)} \\ &= \frac{\texttt{stb}(n)c(k,n)}{\texttt{stb}(n)d(k,n) + \texttt{sws}(n)b(k,n)} + \frac{\texttt{sws}(n)a(k,n)}{\texttt{stb}(n)d(k,n) + \texttt{sws}(n)b(k,n)}. \end{aligned}$$

Thus,

$$\begin{aligned} &\left|\frac{c(k,n+1)}{d(k,n+1)} - \frac{c(k,n)}{d(k,n)}\right| \\ &\leq \frac{\texttt{swb}(n)c(k,n)}{\texttt{swb}(n)d(k,n)} \frac{\texttt{sws}(n)b(k,n)}{\texttt{stb}(n)d(k,n) + \texttt{sws}(n)b(k,n)} + \frac{\texttt{sws}(n)a(k,n)}{\texttt{stb}(n)d(k,n)} \\ &= O(1)O\left(\tfrac{1}{nP(n)^2}\right) + O\left(\tfrac{1}{nP(n)^2}\right) \\ &= O\left(\tfrac{1}{nP(n)^2}\right). \end{aligned}$$

It follows that the distance $\left|\frac{c(k,n+m)}{d(k,n+m)} - \frac{c(k,n)}{d(k,n)}\right|$ is majorised by the tail of a convergent series. The convergence of $\left(\frac{c(k,n)}{d(k,n)}\right)_n$ follows.

The asymptotics of $L(k)$ follow from (11). The computability of $L(k)$ to any given finite precision is obtained by observing that we can make all implicit constants in the above estimates into explicit ones, yielding explicit error estimates.

It remains to compute the sign of $L(k)$. The case where λ_2 is constant is trivial. For the other cases, note that essentially the same argument as in Lemma 15 shows that the sum $\sum_{w \in C(k,n)} [\![w]\!](n-1)$ is dominated by the terms with an even number of $\texttt{sts}$ and an odd number of $\texttt{sws}$. The number of $\texttt{swb}$ is even in all such terms, so that the sign of the sum $\sum_{w \in C(k,n)} [\![w]\!](n-1)$ for sufficiently large n is the sign of $\texttt{sws}(k)$. Since $c(k,n) = \lambda_2(k) \sum_{w \in C(k,n)} [\![w]\!](n-1)$, the sign of $c(k,n)$ is the sign of $\lambda_2(k) \cdot \texttt{sws}(k)$. Since $d(k,n)$ is positive for large n it follows that the sign of $L(k)$ is the sign of $\lambda_2(k) \cdot \texttt{sws}(k)$. ◀