Periodic Behavior of the Minimal Colijn-Plazzotta Rank for Trees with a Fixed Number of Leaves

Michael R. Doboli [1]
Department of Mathematics, Stanford University, CA, USA

Hsien-Kuei Hwang
Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

Noah A. Rosenberg
Department of Biology, Stanford University, CA, USA

Abstract

The Colijn-Plazzotta ranking is a certain bijection between the unlabeled binary rooted trees and the positive integers, such that the integer associated with a tree is determined from the integers associated with the two immediate subtrees of its root. Letting a_n denote the minimal Colijn-Plazzotta rank among all trees with a specified number of leaves n, the sequence $\{a_n\}$ begins 1, 2, 3, 4, 6, 7, 10, 11, 20, 22, 28, 29, 53, 56, 66, 67 (OEIS A354970). Here we show that $a_n \sim 2[2^{P(\log_2 n)}]^n$, where P varies as a periodic function dependent on $\{\log_2 n\}$ and satisfies $1.24602 < 2^{P(\log_2 n)} < 1.33429$.

2012 ACM Subject Classification Mathematics of computing → Combinatoric problems

Keywords and phrases Colijn–Plazzotta ranking, recurrences, unlabeled trees

Digital Object Identifier 10.4230/LIPIcs.AofA.2024.18

Funding *Noah A. Rosenberg*: National Institutes of Health grant R01 HG005855.

Acknowledgements We are grateful to Michael Fuchs and Daniel Krenn for discussions of this work, and to Stephan Wagner for discussions and for comments on a draft of the manuscript.

1 Introduction

Consider an unlabeled binary rooted tree t with $m(t)$ leaves. Colijn & Plazzotta [2] introduced a ranking for the unlabeled binary rooted trees, according to which the rank $f(t)$ of t is determined from the ranks $\ell(t)$ of its left subtree and $r(t)$ of its right subtree: $f(t) = 1$ for $m(t) = 1$, and $f(t) = f\big(\ell(t)\big)\big[f\big(\ell(t)\big) - 1\big]/2 + 1 + f\big(r(t)\big)$ for $m(t) \geq 2$. To compute the Colijn-Plazzotta rank, or CP rank, of a tree t, the "left" and "right" subtrees of t are arranged in a canonical order, such that $f\big(\ell(t)\big) \geq f\big(r(t)\big)$.

The ranking f bijectively associates positive integers to unlabeled binary rooted trees – which number 1, 1, 2, 3, 6, 11, 23, 46, 98 for trees of $n = 1$ to 10 leaves (the Wedderburn-Etherington numbers, OEIS A001190). Among trees with n leaves, CP ranks vary greatly; for example, the 8-leaf symmetric tree has rank 11 and the 8-leaf caterpillar has rank 2,598,062.

The CP rank has been proposed for various uses in the mathematical study of evolutionary trees [2, 3, 9]. It provides a tree encoding with the property that similar shapes often have nearby ranks, even if they possess different numbers of leaves. As a result, it gives a basis for computing a distance between unlabeled trees of differing size – a useful metric for the evolutionary trees that might be produced from genetic sequences in pathogens and other organisms. Because highly balanced shapes have the smallest rank among trees with a fixed

[1] Corresponding author.

35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024).
Editors: Cécile Mailler and Sebastian Wild; Article No. 18; pp. 18:1–18:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

number of leaves and highly unbalanced shapes have the largest rank, the CP rank can serve as a measure of the balance of an unlabeled tree – for example, in studies that seek to compare the balance of observed trees to that predicted by models of evolutionary processes.

The minimal and maximal CP ranks across all trees with a fixed number of leaves n can assist in assessing the CP ranks of specific trees, for example by normalizing the CP rank as a measure of tree balance. Rosenberg [9] studied the minimal and maximal CP ranks across trees with n leaves, identifying the trees that give rise to those ranks. The maximal rank, denoted b_n, recursively follows $b_n = b_{n-1}(b_{n-1} - 1)/2 + 2$ for $n \geq 2$, with $b_1 = 1$ [9, Theorem 9]. As a quadratic recursion of a form studied by Aho & Sloane [1], b_n has asymptotic growth $b_n \sim 2\beta^{2^n}$ for a constant $\beta \approx 1.05653$ [9, Corollary 14].

The minimal CP rank, denoted a_n, recursively follows [9, Theorem 6]

$$a_n = \begin{cases} 1, & n = 1 \\ a_{\lceil n/2 \rceil}(a_{\lceil n/2 \rceil} - 1)/2 + 1 + a_{\lfloor n/2 \rfloor}, & n \geq 2. \end{cases} \tag{1}$$

For n equal to a power of 2, Rosenberg [9] showed that the recursion for a_n is related to that for b_n, producing $a_n \sim 2\alpha^n$ for a constant $\alpha = \beta^4$, $\alpha \approx 1.24602$ [9, Theorem 13]. For general n, however, Rosenberg [9] gave only an upper bound, $a_n < (\frac{3}{2})^n$ [9, Proposition 15].

Here we obtain the asymptotic growth of a_n. Informally, our main result, obtained in Theorems 6 and 9 and summarized in Corollary 10, states that the minimal Colijn-Plazzotta rank a_n across trees with n leaves is approximately equal to $2[2^{P(\log_2 n)}]^n$, where P is a bounded periodic function of period 1. Moreover, the minimum and supremum of 2^P are given by constants $c_1 \approx 1.24602$ and $c_2 \approx 1.33429$.

Extremal properties of the non-differentiable periodic functions arising from recursions such as eq. 1 that involve $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ are often difficult to characterize; many examples therefore rely on case-dependent approaches. The computation here uses an inductive method for studying the extrema.

2 An elementary improvement to the bounds on a_n

We begin by providing a refined exponential upper and lower bound on a_n for $n \geq 66$ that improves upon the $(\frac{3}{2})^n$ exponential upper bound in [9].

▶ **Proposition 1.** *For all integers $n \geq 66$, $3(1.2)^n < a_n < (1.34)^n$.*

Proof. We proceed via induction on n. For the base case, we verify computationally from eq. 1 that $3(1.2)^n < a_n < (1.34)^n$ for all integers $66 \leq n \leq 132$. For the inductive hypothesis, assume that $3(1.2)^k < a_k < (1.34)^k$ for all k, $66 \leq k < n$. Because we have already considered $66 \leq n \leq 132$, suppose $n > 132$. Writing $a_n = \frac{1}{2}a_{\lceil n/2 \rceil}(a_{\lceil n/2 \rceil} - 1) + a_{\lfloor n/2 \rfloor} + 1$, by the inductive hypothesis, we have for the lower bound

$$\begin{aligned} a_n &> \frac{3(1.2)^{\lceil n/2 \rceil}[3(1.2)^{\lceil n/2 \rceil} - 1]}{2} + 3(1.2)^{\lfloor n/2 \rfloor} + 1 \\ &= \frac{9}{2}(1.2)^{2\lceil n/2 \rceil} - \frac{3}{2}(1.2)^{\lceil n/2 \rceil} + 3(1.2)^{\lfloor n/2 \rfloor} + 1 \\ &\geq \frac{9}{2}(1.2)^{2\lceil n/2 \rceil} + \left[-\frac{3}{2}(1.2) + 3 \right](1.2)^{\lfloor n/2 \rfloor} + 1 > \frac{9}{2}(1.2)^n + 1.2(1.2)^{\lfloor n/2 \rfloor} + 1 > 3(1.2)^n. \end{aligned}$$

For the upper bound, by the recursive formula for a_n and the inductive hypothesis,

$$\begin{aligned}
a_n &< \frac{(1.34)^{\lceil n/2\rceil}[(1.34)^{\lceil n/2\rceil}-1]}{2} + (1.34)^{\lfloor n/2\rfloor}+1\\
&= \frac{1}{2}(1.34)^{2\lceil n/2\rceil} - \frac{1}{2}(1.34)^{\lceil n/2\rceil} + (1.34)^{\lfloor n/2\rfloor}+1\\
&\le \frac{1}{2}(1.34)^{2\lceil n/2\rceil} + \Big(-\frac{1}{2}+1\Big)(1.34)^{\lfloor n/2\rfloor}+1\\
&= \frac{1}{2}(1.34)^{2\lceil n/2\rceil} + \frac{1}{2}(1.34)^{\lfloor n/2\rfloor}+1 \le \frac{1}{2}(1.34)^{n+1} + \frac{1}{2}(1.34)^{n/2}+1\\
&= (1.34)^n + \big[-0.33(1.34)^n + 0.5(1.34)^{n/2}+1\big] < (1.34)^n,
\end{aligned}$$

where the last step follows by noting that $-0.33(1.34)^n + 0.5(1.34)^{n/2}+1<0$ for $1.34^{n/2} > (0.5+\sqrt{1.57})/0.66$, or $n > 2\log[(0.5+\sqrt{1.57})/0.66]/\log 1.34 \approx 6.6754$. ◀

We continue now with a more precise analysis via the methods of [4].

3 Obtaining the periodically varying exponential order

Hwang et al. [4] studied recurrences with the nth term written in terms of $\lfloor \frac{n}{2}\rfloor$ and $\lceil \frac{n}{2}\rceil$. Such recurrences can arise in tree problems, in which a quantity associated with the root is written in terms of corresponding quantities for subtrees (see also e.g. [6, 8]). The floor and ceiling function give rise to periodicity in the exponential orders of the associated sequences.

Theorem 5 of [4], which considers recurrences that involve $\lceil \frac{n}{2}\rceil$, enables asymptotic evaluation of a_n from eq. 1. Denote $\{t\} = t - \lfloor t\rfloor$, writing $\{t^-\}$ as the left-continuous version of $\{t\}$: $\{t^-\}=1$ for integer t, and $\{t^-\}=\{t\}$ otherwise. In other words, $\{t^-\} = 1-\{-t\}$.

▶ **Theorem 2** ([4]). *Suppose $f(n) = 2f(\lceil \frac{n}{2}\rceil) + g(n)$ for $n \ge 2$, where $f(1)$ is given and $g(1)=0$. Suppose further that the function $G_m(t) = \sum_{k=0}^{m} 2^{-k} g(\lceil 2^k t\rceil)$ converges uniformly to $G(t) = \sum_{k=0}^{\infty} 2^{-k} g(\lceil 2^k t\rceil)$ for $t\in[1,2]$.*

Then for $n\ge 1$, we have $f(n) = nP(\log_2 n) - Q(n)$, with P and Q defined by

$$P(t) = 2^{-\{t^-\}}\Big[G\big(2^{\{t^-\}}\big)+2f(1)\Big]$$

$$Q(n) = G(n)-g(n) = \sum_{k=1}^{\infty} 2^{-k} g(2^k n).$$

The theorem states that for a class of recurrences in which $f(n)$ is expressed in terms of $f(\lceil \frac{n}{2}\rceil)$, $f(n)$ can be written in terms of a periodic function P that varies with the fractional part of $\log_2 n$. We rewrite a_n from eq. 1 in a form suited to the theorem.

Expanding eq. 1, for $n\ge 2$, we have $a_n = \frac{1}{2}a^2_{\lceil n/2\rceil} - \frac{1}{2}a_{\lceil n/2\rceil} + a_{\lfloor n/2\rfloor}+1$, with $a_1=1$. We augment the definition by writing $a_0=0$. Writing $a_n = 2g_n - \frac{1}{2}$, we have $g_0=\frac{1}{4}$, $g_1=\frac{3}{4}$, and for $n\ge 2$, $g_n = g^2_{\lceil n/2\rceil}+h_n$, where $h_n = g_{\lfloor n/2\rfloor} - g_{\lceil n/2\rceil} + \frac{11}{16}$, with $h_0=\frac{11}{16}$ and $h_1 = \frac{3}{16}$.

Let $\lambda_n = \log_2 g_n$. Then $\lambda_0=-2$, $\lambda_1 = \log_2\frac{3}{4}$, and for $n\ge 2$, $\lambda_n = 2\lambda_{\lceil n/2\rceil}+\mu_n$, where $\mu_n = \log_2(1+h_n/g^2_{\lceil n/2\rceil})$ for $n\ge 2$. We set $\mu_1 = 0$; a value for μ_0 is not needed.

▶ **Proposition 3.** *For $n\ge 2$, the sequence λ_n can be written $\lambda_n = nP(\log_2 n) - Q(n)$, where*

$$P(t) = 2^{-\{t^-\}}\left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+\{t^-\}}\rceil}\right) \tag{2}$$

$$Q(n) = \sum_{k=1}^{\infty} 2^{-k}\mu_{2^k n}. \tag{3}$$

Proof. First, λ_n has the correct recursive form for the theorem: $\lambda_n = 2\lambda_{\lceil n/2\rceil} + \mu_n$ for $n \geq 2$, with λ and μ in the roles of f and g. λ_1 is given, equaling $\log_2 \frac{3}{4}$, and $\mu_1 = 0$ by definition.

Note that the μ_n depend on the λ_n, which is not the case for $g(n)$ in Theorem 2 in relation to $f(n)$, so that Theorem 2 does not immediately apply. However, because $f(n)$ here is solved in closed form without error, we can check the conditions of the theorem – which amounts to showing the convergence of an infinite series – and still apply the resulting solution to λ_n.

If we can show uniform convergence of $G_m(t) = \sum_{k=0}^{m} 2^{-k}\mu_{\lceil 2^k t\rceil}$ to $G(t) = \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^k t\rceil}$ on $t \in [1,2]$, then the proposition will follow by Theorem 2, with f replaced by λ and g by μ. To prove this uniform convergence result, we first note that

$$\mu_n = \log_2\left(1 + \frac{g_{\lfloor n/2\rfloor} - g_{\lceil n/2\rceil} + \frac{11}{16}}{g^2_{\lceil n/2\rceil}}\right) = \log_2\left(1 + \frac{g_{\lfloor n/2\rfloor}}{g^2_{\lceil n/2\rceil}} - \frac{1}{g_{\lceil n/2\rceil}} + \frac{\frac{11}{16}}{g^2_{\lceil n/2\rceil}}\right)$$
$$\leq \log_2\left(1 + \frac{1}{g_{\lceil n/2\rceil}} - \frac{1}{g_{\lceil n/2\rceil}} + \frac{\frac{11}{16}}{g^2_{\lceil n/2\rceil}}\right) = \log_2\left(1 + \frac{\frac{11}{16}}{g^2_{\lceil n/2\rceil}}\right).$$

The inequality follows from $g_{\lfloor n/2\rfloor} \leq g_{\lceil n/2\rceil}$, which holds because $g_n = \frac{1}{2}(a_n + \frac{1}{2})$ and $\{a_n\}_{n=1}^{\infty}$ is strictly increasing [9, Lemma 5]. Then $\{g_n\}_{n=1}^{\infty}$ is also strictly increasing. We conclude that there exists a constant upper bound on $\log_2\left[1 + (\frac{11}{16})/g^2_{\lceil n/2\rceil}\right]$ that is applicable for all $n \geq 1$. Next, notice that

$$\mu_n = \log_2\left(1 + \frac{g_{\lfloor n/2\rfloor}}{g^2_{\lceil n/2\rceil}} - \frac{1}{g_{\lceil n/2\rceil}} + \frac{\frac{11}{16}}{g^2_{\lceil n/2\rceil}}\right) \geq \log_2\left(1 - \frac{1}{g_{\lceil n/2\rceil}}\right).$$

Because $\{g_n\}_{n=1}^{\infty}$ is strictly increasing and $g_2 = \frac{5}{4}$, we can conclude that for all $n \geq 4$, $\mu_n \geq \log_2\left(1 - 1/g_{\lceil n/2\rceil}\right) \geq \log_2\left(1 - 1/g_2\right) = \log_2\left(\frac{1}{5}\right)$. Hence, $\mu_n \geq \min\left(\mu_1, \mu_2, \mu_3, \log_2(\frac{1}{5})\right)$ for all $n \geq 1$, showing that μ_n is also bounded below by a constant applicable for all $n \geq 1$.

Because μ_n is bounded below and above by constants applicable for all n, there exists a constant M such that $|\mu_n| < M$ for all $n \geq 1$. We use this constant to show that $G_m(t)$ converges uniformly to $G(t) = \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^k t\rceil}$ for $t \in [1,2]$. Indeed, if we let $g_k(t) = 2^{-k}\mu_{\lceil 2^k t\rceil}$, we then have that $G_m(t) = \sum_{k=0}^{m} g_k(t)$. Because $|g_k(t)| = |2^{-k}\mu_{\lceil 2^k t\rceil}| \leq 2^{-k}M$ for all $t \in [1,2]$ and $k \geq 0$ and $\sum_{k=0}^{\infty} 2^{-k}M = 2M < \infty$, it follows by the Weierstrass M-test that $G_m(t)$ converges uniformly to $G(t)$ on $t \in [1,2]$, as desired.

By Theorem 2, we deduce that $\lambda_n = nP(\log_2 n) - Q(n)$, where P and Q are defined by

$$P(t) := 2^{-\{t^-\}}\left[G(2^{\{t^-\}}) + 2\lambda_1\right] = 2^{-\{t^-\}}\left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+\{t^-\}}\rceil}\right) \tag{4}$$

$$Q(n) := G(n) - \mu_n = \sum_{k=1}^{\infty} 2^{-k}\mu_{2^k n}. \tag{5}$$

◀

Examples of $a_n, g_n, h_n, \lambda_n, \mu_n, P(\log_2 n)$ and $Q(n)$ for small values of n appear in Table 1. Values for $P(\log_2 n)$ and $Q(n)$ are numerical approximations, and values for λ_n and μ_n are rounded to four decimal places. To find the asymptotic growth of a_n, we use Proposition 3:

$$a_n = 2g_n - \frac{1}{2} = 2(2^{\lambda_n}) - \frac{1}{2} = 2[2^{nP(\log_2 n) - Q(n)}] - \frac{1}{2} = 2[2^{-Q(n)}]\,[2^{P(\log_2 n)}]^n - \frac{1}{2}.$$

Table 1 Examples of $a_n, g_n, h_n, \lambda_n, \mu_n, P(\log_2 n)$ and $Q(n)$ for $0 \le n \le 5$. a_n is calculated recursively using eq. 1, and g_n is evaluated from $a_n = 2g_n - \frac{1}{2}$. h_n is evaluated as $h_n = g_{\lfloor n/2\rfloor} - g_{\lceil n/2\rceil} + \frac{11}{16}$, and λ_n as $\lambda_n = \log_2 g_n$. μ_n is defined as $\log_2(1 + h_n/g^2_{\lceil n/2\rceil})$ for $n \ge 2$ with $\mu_1 = 0$. The values of $P(\log_2 n)$ and $Q(n)$ are approximated via eqs. 2 and 3, using the values of μ_n and λ_1.

n	a_n	g_n	h_n	λ_n	μ_n	$P(\log_2 n)$	$Q(n)$
0	0	0.25	0.6875	-2	-	-	-
1	1	0.75	0.1875	-0.4150	0	0.3173	0.7324
2	2	1.25	0.6875	0.3219	1.1520	0.3173	0.3127
3	3	1.75	0.1875	0.8074	0.1635	0.3237	0.1639
4	4	2.25	0.6875	1.1699	0.5261	0.3173	0.0994
5	6	3.25	0.1875	1.7004	0.0857	0.3496	0.0474

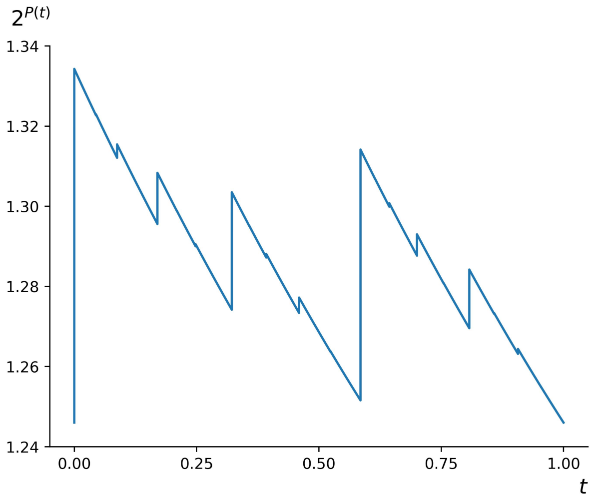

Figure 1 $2^{P(t)}$ as a function of t. $2^{P(t)}$ is a periodic function with period 1. The plot computes all μ_n for $1 \le n \le 2056$ using $\mu_n = \log_2(1 + h_n/g^2_{\lceil n/2\rceil})$ for $n \ge 2$, where $h_n = g_{\lfloor n/2\rfloor} - g_{\lceil n/2\rceil} + \frac{11}{16}$ and $g_n = \frac{a_n}{2} + \frac{1}{4}$ for $n \ge 2$ with $g_1 = \frac{3}{4}$ and $h_1 = \frac{3}{16}$; a_n is defined by eq. 1. We use the μ_n to approximate $P(t)$ as in eq. 2, evaluating $P(t)$ for all $t = k/100000$ for integers $0 \le k < 100000$.

We use a lemma from A to show in B that $Q(n) \to 0$ as $n \to \infty$. It follows that $a_n \sim 2[2^{P(\log_2 n)}]^n$. The asymptotic exponential growth of a_n thus depends only on the value of $\{\log_2 n\}$. Because P is a periodic function with period 1, we have that $P[\log_2(2n)] = P(1 + \log_2 n) = P(\log_2 n)$. The base of the exponent of a_{2n} is the same as that of a_n for any n. A plot of $2^{P(t)}$ as a function of $t \in [0, 1]$ appears in Figure 1.

The function in Figure 1 appears to have many discontinuities, the most visually apparent of which lies at $t = 0$. In the next section, we show that $2^{P(t)}$ has its supremum as $P(t)$ approaches 0 from the right and its minimum at $t = 0$.

4 The upper bound on the exponential order

From Section 3, $a_n \sim 2[2^{P(\log_2 n)}]^n$. Hence, to find upper and lower bounds on the exponential order of a_n, we must find the extreme values of $2^{P(\log_2 n)}$. Because P is a 1-periodic function, it suffices to find the extrema of $2^{P(t)}$ on $t \in [0, 1)$.

We obtain the upper bound in Theorem 6 and the lower bound in Theorem 9. The proof of Theorem 6 requires an inequality that concerns a certain sum involving the μ_n. To prove the inequality, Lemma 4 obtains a term-wise result for terms in the sum that have a sufficiently high index. The term-wise result does not hold for terms with a small index, and Lemma 5 addresses their sum all at once. The lemmas are proven in C.

▶ **Lemma 4.** $\mu_{2^k+1} - 2^{-t}\mu_{\lceil 2^{k+t}\rceil} - \lambda_1(2^{-t}-1) > -[\mu_2 - 2^{-t}\mu_2 - \lambda_1(2^{-t}-1)]$ *for all integers* $k \geq 11$ *and all* $t \in (0,1)$.

▶ **Lemma 5.** *For all* $t \in (0,1)$,

$$\sum_{k=1}^{10}\left[2^{-k}\mu_{2^k+1} - 2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k}\lambda_1(2^{-t}-1)\right] > -\sum_{k=1}^{10}2^{-k}\left[\mu_2 - 2^{-t}\mu_2 - \lambda_1(2^{-t}-1)\right].$$

▶ **Theorem 6.** $\sup_{t\in(0,1)} 2^{P(t)} = \lim_{t\to 0^+} 2^{P(t)}$.

Proof. Because 2^x is a strictly increasing function with respect to x, finding the supremum of $2^{P(t)}$ on $(0,1)$ is equivalent to finding the supremum of $P(t)$. For $t \in (0,1)$, $\{t^-\} = t$. Hence, applying the definition of $P(t)$ from eq. 4,

$$\lim_{t\to 0^+} P(t) = \lim_{t\to 0^+} 2^{-\{t^-\}}\left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+\{t^-\}}\rceil}\right) = 2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{2^k+1}.$$

Proving that $P(t) < \lim_{t\to 0^+} P(t)$ for $t \in (0,1)$ is equivalent to proving

$$2^{-t}\left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+t}\rceil}\right) < 2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{2^k+1}. \tag{6}$$

Rearranging eq. 6 and noting $2 = \sum_{k=0}^{\infty} 2^{-k}$, we must prove $\sum_{k=0}^{\infty}[2^{-k}\mu_{2^k+1} - 2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k}\lambda_1(2^{-t}-1)] > 0$, or equivalently, extracting the $k=0$ term,

$$\mu_2 - 2^{-t}\mu_2 - \lambda_1(2^{-t}-1) + \sum_{k=1}^{\infty}\left[2^{-k}\mu_{2^k+1} - 2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k}\lambda_1(2^{-t}-1)\right] > 0. \tag{7}$$

By Lemmas 4 and 5, we have the following:

$$\begin{aligned}&[\mu_2 - 2^{-t}\mu_2 - \lambda_1(2^{-t}-1)] + \sum_{k=1}^{10}\left[2^{-k}\mu_{2^k+1} - 2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k}\lambda_1(2^{-t}-1)\right]\\ &\quad + \sum_{k=11}^{\infty}\left[2^{-k}[\mu_{2^k+1} - 2^{-t}\mu_{\lceil 2^{k+t}\rceil} - \lambda_1(2^{-t}-1)]\right]\\ &\quad > [\mu_2 - 2^{-t}\mu_2 - \lambda_1(2^{-t}-1)]\left[1 - \sum_{k=1}^{10}2^{-k} - \sum_{k=11}^{\infty}2^{-k}\right] = 0.\end{aligned}$$

The chain of inequalities verifies eq. 7, proving the theorem. ◀

5 The lower bound on the exponential order

We can use techniques similar to those of Section 4 to find the minimum of $2^{P(t)}$ for $t \in [0,1)$. Again, we need two lemmas, one for terms with a sufficiently large index, and another for terms with small values for the index. The lemmas are proven in D.

▶ **Lemma 7.** *For integers* $k \geq 11$ *and all* $t \in (0,1)$:

$$2^{-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-1}\mu_{2^{k+1}} - \lambda_1(2^{-1} - 2^{-t}) > -[2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1} - 2^{-t})].$$

▶ **Lemma 8.** *For all $t \in (0,1)$,*

$$\sum_{k=1}^{10} \left[2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k-1}\mu_{2^{k+1}} - 2^{-k}\lambda_1(2^{-1}-2^{-t})\right]$$
$$> -\sum_{k=1}^{10} 2^{-k}\left[2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1}-2^{-t})\right].$$

▶ **Theorem 9.** $\min_{t\in[0,1)} 2^{P(t)} = 2^{P(0)}$.

Proof. As before, 2^x is an increasing function in x, so that finding the minimum of the 1-periodic $2^{P(t)}$ is equivalent to finding the minimum of $P(t)$ over $[0,1)$. We must show that

$$P(0) = \frac{1}{2}\left[2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{2^{k+1}}\right] < 2^{-t}\left[2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+t}\rceil}\right] = P(t),$$

for all $t \in (0,1)$. Equivalently, replacing 2 by $\sum_{k=0}^{\infty} 2^{-k}$, we must show $\sum_{k=0}^{\infty}[2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k-1}\mu_{2^{k+1}} - 2^{-k}\lambda_1(2^{-1}-2^{-t})] > 0$. Using Lemmas 7 and 8,

$$\begin{aligned}
&\sum_{k=0}^{\infty}[2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k-1}\mu_{2^{k+1}} - 2^{-k}\lambda_1(2^{-1}-2^{-t})] \\
&\quad > 2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1}-2^{-t}) - \sum_{k=1}^{10} 2^{-k}\left[2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1}-2^{-t})\right] \\
&\qquad - \sum_{k=11}^{\infty} 2^{-k}[2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1}-2^{-t})] \\
&\quad = [2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1}-2^{-t})]\left[1 - \sum_{k=1}^{10} 2^{-k} - \sum_{k=11}^{\infty} 2^{-k}\right] = 0.
\end{aligned}$$

◀

6 Summary of exponential bounds

Theorems 6 and 9 produce the following corollary. We define two constants, c_1 and c_2:

$$c_1 = 2^{P(0)} = 2^{\frac{1}{2}(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{2^{k+1}})} \approx 1.2460208329836624$$
$$c_2 = \lim_{\log_2 n \to 0^+} 2^{P(\log_2 n)} = 2^{(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{2^{k+1}})} \approx 1.3342827071604892.$$

▶ **Corollary 10.** $\limsup_{n\to\infty}[a_n/(2c_2^n)] = 1$, *and* $\liminf_{n\to\infty}[a_n/(2c_1^n)] = 1$.

Proof. $a_n \sim 2[2^{P(\log_2 n)}]^n$ by Proposition 3, or $\lim_{n\to\infty}\left[a_n/\left(2[2^{P(\log_2 n)}]^n\right)\right] = 1$. By Theorem 6, the supremum of $2^{P(\log_2 n)}$ on $[0,1)$ is attained as $\{\log_2 n\} \to 0^+$. By definition of c_2, the supremum is $\lim_{\log_2 n\to 0^+} 2^{P(\log_2 n)} = c_2$. Hence, $\limsup_{n\to\infty}[a_n/(2c_2^n)] = 1$.

Similarly, by Theorem 9, the minimum of $2^{P(\log_2 n)}$ is attained at $\{\log_2 n\} = 0$. The minimum is thus $2^{P(0)} = c_1$. We conclude that $\liminf_{n\to\infty}[a_n/(2c_1^n)] = 1$. ◀

Note that the constant c_1 is equal to the value of α in [9, Theorem 13], which finds $a_{2^n} \sim 2\alpha^{(2^n)}$. We can also improve on the upper bound $a_n < (\frac{3}{2})^n$ from [9, Proposition 15], producing a corollary that gives the strictest exponential upper bound possible for a_n.

▶ **Corollary 11.** $a_n < 2c_2^n$ *for all* $n \geq 1$.

Proof. By Proposition 3, $a_n = 2\,[2^{-Q(n)}]\,[2^{P(\log_2 n)}]^n - \frac{1}{2}$ for all $n \geq 1$. In B, we prove $Q(n) > 0$ for every integer $n \geq 1$. By Theorem 6, we have $a_n = 2\left[2^{-Q(n)}\right]\left[2^{P(\log_2 n)}\right]^n - \frac{1}{2} < 2\left[2^{-Q(n)}\right]\left[2^{P(\log_2 n)}\right]^n < 2\left[2^{P(\log_2 n)}\right]^n < 2c_2^n$. ◀

A result of [4] (see also [5]) enables a computation of $\mathbb{E}[P(t)]$, producing an approximation for the mean of the exponential order $2^{P(t)}$ over the unit interval for t: Theorem 3 of [4] obtains the analogous quantity to $\mathbb{E}[P(t)]$ for recursions $f(n) = f(\lfloor n/2 \rfloor) + f(\lceil n/2 \rceil) + g(n)$. To obtain our next result, we follow its reasoning for a recursion of the form $f(n) = 2f(\lceil n/2 \rceil) + g(n)$.

We wish to compute the mean $\int_0^1 P(t)\,dt$. In the proof of Proposition 3, we showed that $\sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^k t \rceil}$ is bounded above by a constant applicable for all t, so that the 1-periodic $P(t)$ is also bounded on unit intervals for t, say $t \in [0,1]$. To show that the bounded $P(t)$ is integrable on $[0,1]$, it remains to show that $P(t)$ is continuous almost everywhere.

We show $P(t) = 2^{-\{t^-\}}(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+\{t^-\}} \rceil})$ is continuous outside the countable set $S = \bigcup_{k=1}^{\infty} S_k$, where $S_k = \{t : t \in [0,1] \text{ and } 2^{k+t} \in \mathbb{N}\}$. It suffices to show that $2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+\{t^-\}} \rceil} = 2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+t} \rceil}$ is continuous for $t \in (0,1)\backslash S$.

For a positive integer N, recall from the proof of Proposition 3 the uniform convergence of $G_m(t) = \sum_{k=0}^{m} 2^{-k}\mu_{\lceil 2^k t \rceil}$ on $t = [1,2]$ (and uniform boundedness of $|\mu_n|$ by M for all n). Choose $\epsilon_N > 0$ such that $\sum_{k=N+1}^{\infty} |2^{-k}\mu_{\lceil 2^{k+t} \rceil}| < \epsilon_N/2$ for all $t \in (0,1)$. For all $t \in (0,1)\backslash S$, given N, there exists $\delta_N > 0$ such that for all $x \in (t-\delta_N, t+\delta_N)$, $2\lambda_1 + \sum_{k=0}^{N} 2^{-k}\mu_{\lceil 2^{k+x} \rceil} = 2\lambda_1 + \sum_{k=0}^{N} 2^{-k}\mu_{\lceil 2^{k+t} \rceil}$. Therefore, for all $x \in (t-\delta_N, t+\delta_N)$,

$$\left|\left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+t} \rceil}\right) - \left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+x} \rceil}\right)\right| = \left|\sum_{k=N+1}^{\infty} 2^{-k}(\mu_{\lceil 2^{k+t} \rceil} - \mu_{\lceil 2^{k+x} \rceil})\right|.$$

Then $\left|\sum_{k=N+1}^{\infty} 2^{-k}(\mu_{\lceil 2^{k+t} \rceil} - \mu_{\lceil 2^{k+x} \rceil})\right| < \sum_{k=N+1}^{\infty} |2^{-k}\mu_{\lceil 2^{k+t} \rceil}| + \sum_{k=N+1}^{\infty} |2^{-k}\mu_{\lceil 2^{k+x} \rceil}| < \epsilon_N$. We let N grow large, concluding that $P(t)$ is continuous almost everywhere and integrable.

For the integral, with $G(t) = \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^k t \rceil}$, we have that $\int_0^1 P(t)\,dt = \int_0^1 2^{1-t}\lambda_1\,dt + \int_0^1 2^{-t}G(2^t)\,dt = \lambda_1/\log 2 + \int_0^1 2^{-t}G(2^t)\,dt$. Define $\mu(v) = \mu_{\lceil v \rceil}$. It remains to compute

$$\begin{aligned}
\int_0^1 2^{-t}G(2^t)\,dt &= \frac{1}{\log 2}\int_1^2 v^{-2}G(v)\,dv = \frac{1}{\log 2}\int_1^2 v^{-2}\sum_{k=0}^{\infty} 2^{-k}\mu(2^k v)\,dv \\
&\overset{DCT}{=} \frac{1}{\log 2}\lim_{m\to\infty}\int_1^2 \sum_{k=0}^{m} v^{-2}2^{-k}\mu(2^k v)\,dv = \frac{1}{\log 2}\lim_{m\to\infty}\sum_{k=0}^{m}\int_1^2 v^{-2}2^{-k}\mu(2^k v)\,dv \\
&= \frac{1}{\log 2}\lim_{m\to\infty}\sum_{k=0}^{m}\int_{2^k}^{2^{k+1}} y^{-2}\mu(y)\,dy = \frac{1}{\log 2}\lim_{m\to\infty}\int_1^{2^{m+1}} y^{-2}\mu(y)\,dy.
\end{aligned}$$

To justify use of the dominated convergence theorem (DCT), we note that for $v \in [1,2]$, $|\sum_{k=0}^{m} v^{-2}2^{-k}\mu(2^k v)| \leq \sum_{k=0}^{m} |v^{-2}2^{-k}\mu(2^k v)| \leq \sum_{k=0}^{\infty} |v^{-2}2^{-k}\mu(2^k v)| \leq \sum_{k=0}^{\infty} |2^{-k}\mu(2^k v)|$, a quantity that is uniformly bounded for all v. Next, notice that

$$\begin{aligned}
\int_1^{2^{m+1}} y^{-2}\mu(y)\,dy &= \sum_{n=2}^{2^{m+1}}\int_{n-1}^{n} y^{-2}\mu(y)\,dy = \sum_{n=2}^{2^{m+1}}\int_{n-1}^{n} y^{-2}\mu(\lceil y \rceil)\,dy \\
&= \sum_{n=2}^{2^{m+1}} \mu_n \int_{n-1}^{n} y^{-2}\,dy = \sum_{n=2}^{2^{m+1}} \frac{\mu_n}{n(n-1)}.
\end{aligned}$$

Therefore,

$$\frac{1}{\log 2}\lim_{m\to\infty}\int_1^{2^{m+1}} y^{-2}\mu(y)\,dy = \frac{1}{\log 2}\lim_{m\to\infty}\sum_{n=2}^{2^{m+1}}\frac{\mu_n}{n(n-1)} = \frac{1}{\log 2}\sum_{n=2}^{\infty}\frac{\mu_n}{n(n-1)}.$$

We have proven the following proposition.

▶ **Proposition 12.** *The mean value of $P(t)$ on the unit interval $[0,1]$ is*

$$\int_0^1 P(t)\,dt = \frac{\lambda_1}{\log 2} + \frac{1}{\log 2}\sum_{n=2}^{\infty}\frac{\mu_n}{n(n-1)}.$$

By Jensen's inequality on the convex function $\varphi(x) = 2^x$, we obtain a numerical lower bound on the mean of the exponential order $2^{P(t)}$,

$$\int_0^1 2^{P(t)} dt = \mathbb{E}[2^{P(t)}] \geq 2^{\mathbb{E}[P(t)]} = 2^{\int_0^1 P(t)dt} \approx 1.2860382564771475.$$

Note that the mean values of $P(t)$ and $2^{\mathbb{E}[P(t)]}$ represent means for uniformly distributed t; they do not correspond to means over integers n with fixed $\lfloor \log_2 n \rfloor$, as $\log_2 n$ does not have a uniformly distributed fractional part over integers n with fixed $\lfloor \log_2 n \rfloor$.

7 Discussion

We have solved the problem of finding an exact expression for the asymptotic growth of a_n, the minimal Colijn-Plazzotta rank among unlabeled binary rooted trees with n leaves. We find that a_n has periodically varying exponential growth, with exponential order depending on $\{\log_2 n\}$ (Section 3). Its value lies in $[1.246020832983662, 1.3342827071604892)$, where the lower bound is achieved if $\{\log_2 n\} = 0$ and the upper bound is approached as $\{\log_2 n\} \to 0^+$ (Sections 5 and 4). We have obtained the tight upper bound $a_n < 2c_2^n$ for all $n \geq 1$, where $c \approx 1.3342827071604892$ (Corollary 11), improving upon an earlier bound.

The growth of a_n is slowest when n is a power of two and fastest when n is slightly larger than a power of two. This result captures the "jumps" that occur in CP rank near powers of two. For example, in [9, Figure 1] the ratio $a_n/(2\alpha^n)$ for $\alpha \approx 1.24602$ is near 1 if n is a power of two but sharply increases when n is one larger than a power of two. The jumps are visible numerically: $a_{32} = 2279$ yet $a_{33} = 20369$, and $a_{64} = 2598061$ yet $a_{65} = 207440176$. The dependence of the exponential growth of a_n on $\{\log_2 n\}$ reflects these discontinuities.

Maranca & Rosenberg [7] studied an extension of CP rank to strictly and at-most-k-furcating trees, $k \geq 2$, where each internal node of a strictly k-furcating tree has exactly k children, and each internal node of an at-most-k-furcating tree has at least two and at most k children. For such trees, the same questions about the minimal and maximal rank among trees with n leaves can be posed. The work of [4] contains theorems that can potentially be used for asymptotics in these more general cases, whose analyses we defer to future work.

The Colijn-Plazzotta rank has been suggested for use in tree balance indices [3, 9]. Our results characterize the minimal value of the rank across trees with a fixed number of nodes, so that a statistic such as $[f(t) - a_n]/(b_n - a_n)$ or $[\log f(t) - \log a_n]/(\log b_n - \log a_n)$ normalized to lie in $[0, 1]$ can be used as a measure of the balance of a tree.

References

1 A. V. Aho and N. J. A. Sloane. Some doubly exponential sequences. *Fibonacci Q.*, 11:429–437, 1973.

2 C. Colijn and G. Plazzotta. A metric on phylogenetic tree shapes. *Syst. Biol.*, 67:113–126, 2018.

3 M. Fischer, L. Herbst, S. Kersting, L. Kühn, and K. Wicke. *Tree balance indices: a comprehensive survey.* Springer, Berlin, 2023.

4 H.-K. Hwang, S. Janson, and T.-H. Tsai. Exact and asymptotic solutions of a divide-and-conquer recurrence dividing at half: theory and applications. *ACM Trans. Alg.*, 13:47, 2017.

5 H.-K. Hwang, S. Janson, and T.-H. Tsai. Identities and periodic oscillations of divide-and-conquer recurrences splitting at half. *Adv. Appl. Math*, 155:102653, 2024.
6 H.-K. Hwang and J.-M. Steyaert. On the number of heaps and the cost of heap construction. In B. Chauvin, P. Flajolet, D. Gardy, and A. Mokkadem, editors, *Mathematics and Computer Science II. Trends in Mathematics*, pages 295–310, Basel, 2002. Birkhäuser.
7 A.R.P Maranca and N.A. Rosenberg. Bijections between the multifurcating unlabeled rooted trees and the positive integers. *Adv. Appl. Math.*, 153:102612, 2024.
8 S.-Y. Oh and J. C. Kieffer. Fractal compression rate curves in lossless compression of balanced trees. In *2010 IEEE International Symposium on Information Theory (ISIT) Proceedings*, pages 106–110, 2010.
9 N. A. Rosenberg. On the Colijn-Plazzotta numbering scheme for unlabaled binary rooted trees. *Discr. Appl. Math.*, 291:88–98, 2021.

Appendix

A Upper bound on $x|\mu_x|$

We provide an upper bound for $x|\mu_x|$. This bound is utilized in Sections 4 and 5, as well as for a result in Section 3. First, we need an upper bound on $|\log_2 x|$ for all positive reals x.

▶ **Lemma 13.** *For all $x > 0$, $|\log_2 x| \leq \frac{1}{2}(\log_2 e)\,|x - \frac{1}{x}|$.*

Proof. For $x \geq 1$, we show $\log x \leq \frac{1}{2}(x - \frac{1}{x})$. Writing a function $f(x) = \log(x) - \frac{x}{2} + \frac{1}{2x}$, we have $f(1) = 0$ and $f'(x) = -(1-x)^2/(2x^2) \leq 0$ for $x \geq 1$. As a function that begins at 0 for $x = 1$ and is nonincreasing for $x \geq 1$, $f(x) \leq 0$ for $x \geq 1$. For $0 < x < 1$, we show $\log x \geq \frac{1}{2}(x - \frac{1}{x})$. This statement follows by noting $\frac{1}{x} > 1$ and applying the case of $x \geq 1$. ◀

Next, we need a uniform lower bound on the expression $1 + g_{\lfloor n/2\rfloor}/g^2_{\lceil n/2\rceil} - 1/g_{\lceil n/2\rceil} + 11/(16g^2_{\lceil n/2\rceil})$ for $n \geq 66$; the use of $n \geq 66$ follows Proposition 1.

▶ **Lemma 14.** *For all integers $n \geq 66$, $1 + g_{\lfloor n/2\rfloor}/g^2_{\lceil n/2\rceil} - 1/g_{\lceil n/2\rceil} + 11/(16g^2_{\lceil n/2\rceil}) > \frac{1}{2}$.*

Proof. Because $g_n = (2a_n + 1)/4$ is positive, we have that $1 + g_{\lfloor n/2\rfloor}/g^2_{\lceil n/2\rceil} - 1/g_{\lceil n/2\rceil} + 11/(16g^2_{\lceil n/2\rceil}) > 1 - 1/g_{\lceil n/2\rceil} \geq 1 - 1/g_{33}$, where the last inequality follows from the monotonicity of g_n. We have $1 - 1/g_{33} \approx 0.9999 > \frac{1}{2}$, as desired. ◀

We are now ready for the main result of this appendix.

▶ **Lemma 15.** *For all integers $x \geq 245$, $x|\mu_x| < (2\lambda_1 + \mu_2)/2$.*

Proof. First note that $1 + g_{\lfloor x/2\rfloor}/g^2_{\lceil x/2\rceil} - 1/g_{\lceil x/2\rceil} + 11/(16g^2_{\lceil x/2\rceil}) > 0$ for $x \geq 245$ by the stronger result in Lemma 14. Using Lemma 13,

$$\begin{aligned}
x|\mu_x| &= x\left|\log_2\left(1 + \frac{g_{\lfloor x/2\rfloor}}{g^2_{\lceil x/2\rceil}} - \frac{1}{g_{\lceil x/2\rceil}} + \frac{11}{16g^2_{\lceil x/2\rceil}}\right)\right| \\
&\leq \frac{x(\log_2 e)}{2}\left|\frac{\frac{g_{\lfloor x/2\rfloor}}{g^2_{\lceil x/2\rceil}} - \frac{1}{g_{\lceil x/2\rceil}} + \frac{11}{16g^2_{\lceil x/2\rceil}}}{1 + \frac{g_{\lfloor x/2\rfloor}}{g^2_{\lceil x/2\rceil}} - \frac{1}{g_{\lceil x/2\rceil}} + \frac{11}{16g^2_{\lceil x/2\rceil}}} + \frac{g_{\lfloor x/2\rfloor}}{g^2_{\lceil x/2\rceil}} - \frac{1}{g_{\lceil x/2\rceil}} + \frac{11}{16g^2_{\lceil x/2\rceil}}\right| \\
&\leq \frac{x(\log_2 e)}{2}\left[\left|\frac{\frac{g_{\lfloor x/2\rfloor}}{g^2_{\lceil x/2\rceil}} - \frac{1}{g_{\lceil x/2\rceil}} + \frac{11}{16g^2_{\lceil x/2\rceil}}}{1 + \frac{g_{\lfloor x/2\rfloor}}{g^2_{\lceil x/2\rceil}} - \frac{1}{g_{\lceil x/2\rceil}} + \frac{11}{16g^2_{\lceil x/2\rceil}}}\right| + \left|\frac{g_{\lfloor x/2\rfloor}}{g^2_{\lceil x/2\rceil}} - \frac{1}{g_{\lceil x/2\rceil}} + \frac{11}{16g^2_{\lceil x/2\rceil}}\right|\right] \\
&\leq \frac{x(\log_2 e)}{2}\left[\frac{\frac{g_{\lfloor x/2\rfloor}}{g^2_{\lceil x/2\rceil}} + \frac{1}{g_{\lceil x/2\rceil}} + \frac{11}{16g^2_{\lceil x/2\rceil}}}{1 + \frac{g_{\lfloor x/2\rfloor}}{g^2_{\lceil x/2\rceil}} - \frac{1}{g_{\lceil x/2\rceil}} + \frac{11}{16g^2_{\lceil x/2\rceil}}} + \frac{g_{\lfloor x/2\rfloor}}{g^2_{\lceil x/2\rceil}} + \frac{1}{g_{\lceil x/2\rceil}} + \frac{11}{16g^2_{\lceil x/2\rceil}}\right].
\end{aligned}$$

Using Proposition 1 with the fact that $g_n = (2a_n + 1)/4$, we have that for $n \geq 66$,

$$\frac{2}{(1.2)^n} > \frac{1}{g_n} = \frac{4}{2a_n + 1} > \frac{4}{2(1.34)^n + 1} \tag{8}$$

$$\frac{(1.2)^n}{2} + \frac{1}{4} < \frac{a_n}{2} + \frac{1}{4} = g_n < \frac{(1.34)^n}{2} + \frac{1}{4}. \tag{9}$$

The bounds in eqs. 8 and 9 and Lemma 14 then yield

$$\begin{aligned} x|\mu_x| &< \frac{x(\log_2 e)}{2}\Bigg[\Bigg[\frac{\frac{(1.34)^{\lfloor x/2\rfloor}}{2}+\frac14}{\left(\frac{(1.2)^{\lceil x/2\rceil}}{2}\right)^2} + \frac{2}{(1.2)^{\lceil x/2\rceil}} + \frac{11}{16}\left(\frac{2}{(1.2)^{\lceil x/2\rceil}}\right)^2\Bigg]\Big/\left(\frac12\right) \\ &\quad + \frac{\frac{(1.34)^{\lfloor x/2\rfloor}}{2}+\frac14}{\left(\frac{(1.2)^{\lceil x/2\rceil}}{2}\right)^2} + \frac{2}{(1.2)^{\lceil x/2\rceil}} + \frac{11}{16}\left(\frac{2}{(1.2)^{\lceil x/2\rceil}}\right)^2\Bigg] \\ &= \frac{3x(\log_2 e)}{2}\left[\left(\frac{(1.34)^{\lfloor x/2\rfloor}}{2}+\frac14\right)\left(\frac{2}{(1.2)^{\lceil x/2\rceil}}\right)^2 + \frac{2}{(1.2)^{\lceil x/2\rceil}} + \frac{11}{16}\left(\frac{2}{(1.2)^{\lceil x/2\rceil}}\right)^2\right] \\ &\leq \frac{3x(\log_2 e)}{2}\left[\left(\frac{(1.34)^{x/2}}{2}+\frac14\right)\left(\frac{2}{(1.2)^{x/2}}\right)^2 + \frac{2}{(1.2)^{x/2}} + \frac{11}{16}\left(\frac{2}{(1.2)^{x/2}}\right)^2\right] \\ &= (3\log_2 e)\left[x\left(\frac{\sqrt{1.34}}{1.2}\right)^x + x\left(\frac{1}{\sqrt{1.2}}\right)^x + \frac{15}{8}x\left(\frac{1}{1.2}\right)^x\right]. \end{aligned} \tag{10}$$

Eq. 10 sums constant multiples of three terms of the form xa^x, where $a < 1$. For $a < 1$, function $f(x) = xa^x$ attains its maximum at $x_{\max} = -1/\log a$ and is decreasing for $x > x_{\max}$. With $\sqrt{1.34}/1.2, 1/\sqrt{1.2}$ and $1/1.2$ in the role of a, $x_{\max}$ evaluates to approximately 27.7880, 10.9696, and 5.4848, respectively. The sum of three decreasing functions is also decreasing. It follows that for $x \geq -1/\log(\sqrt{1.34}/1.2) \approx 27.7880$, the quantity in eq. 10 is decreasing.

To show that the quantity in eq. 10 is less than $(2\lambda_1 + \mu_2)/2$ for $x \geq 245$, it suffices to show that if $x = 245$ is inserted into eq. 10, the result is bounded above by $(2\lambda_1 + \mu_2)/2$; indeed, with $x = 245$, we get 0.15718 in eq. 10, while $(2\lambda_1+\mu_2)/2 \approx 0.1609640474436812$. ◀

B Properties of $Q(n)$

We give two results about $Q(n)$: Lemma 16 for Section 3, and Lemma 17 for Section 6.

▶ **Lemma 16.** $\lim_{n\to\infty} Q(n) = 0$.

Proof. We apply Lemma 15. For all $n \geq 245$, noting $(2\lambda_1 + \mu_2)/2 < \mu_2/2$ because $\lambda_1 < 0$,

$$Q(n) = \sum_{k=1}^{\infty} 2^{-k}\mu_{2^k n} \leq \sum_{k=1}^{\infty} 2^{-k}|\mu_{2^k n}| < \sum_{k=1}^{\infty} 2^{-k}\frac{\mu_2}{2(2^k n)} \leq \sum_{k=1}^{\infty} 2^{-k}\frac{\mu_2}{2(2^k n)}.$$

For $n \geq 245$, we have $Q(n) < [\mu_2/(2n)]\sum_{k=1}^{\infty} 4^{-k} = \mu_2/(6n)$, so $Q(n) \to 0$ as $n \to \infty$. ◀

▶ **Lemma 17.** *For all positive integers n, $Q(n) > 0$.*

Proof. By definition, $\mu_n = \log_2[1 + (g_{\lfloor n/2\rfloor} - g_{\lceil n/2\rceil} + \frac{11}{16})/g^2_{\lceil n/2\rceil}]$. For any integer $n \geq 1$, $\mu_{2n} = \log_2[1 + (g_n - g_n + \frac{11}{16})/g_n^2] = \log_2[1 + 11/(16g_n^2)] > 0$, noting $g_n > 0$ because $a_n > 0$.

By definition of $Q(n)$, for any positive integer $n \geq 1$, $Q(n) = \sum_{k=1}^{\infty} 2^{-k}\mu_{2^k n} = \sum_{k=1}^{\infty} 2^{-k}\mu_{2^k n} > 0$, where the last inequality follows because $\mu_{2^k n} > 0$ for each $k \geq 1$. ◀

C Proofs of Lemmas 4 and 5 for Section 4

This appendix proves Lemmas 4 and 5, used in the proof of Theorem 6. First, we prove two additional lemmas, Lemmas 18 and 19, needed for the proof of Lemma 4.

▶ **Lemma 18.** $\mu_{2^k+1} + 2\lambda_1 + \mu_2 > [2^k/(2^k+1)](\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2)$ *for all integers* $k \geq 11$ *and all* $t \in \big[\log_2(2^k+1), 1\big)$.

Proof. Because we assume $k \geq 11$ and $t > 0$, $\lceil 2^{k+t}\rceil \geq 245$ and $2^k + 1 \geq 245$. We apply Lemma 15 twice, to $\lceil 2^{k+t}\rceil$ and then to $2^k + 1$. From the first application, we obtain $\lceil 2^{k+t}\rceil |\mu_{\lceil 2^{k+t}\rceil}| < (2\lambda_1 + \mu_2)/2$, from which $2^k\,|\mu_{\lceil 2^{k+t}\rceil}| < (2\lambda_1 + \mu_2)/2$.

From the second application, we obtain $(2^k+1)\,|\mu_{2^k+1}| < (2\lambda_1 + \mu_2)/2$. We then have

$$\frac{2^k}{2^k+1}\mu_{\lceil 2^{k+t}\rceil} - \mu_{2^k+1} \leq \left|\frac{2^k}{2^k+1}\mu_{\lceil 2^{k+t}\rceil}\right| + |\mu_{2^k+1}|$$
$$< \frac{1}{2^k+1}\frac{2\lambda_1+\mu_2}{2} + \frac{1}{2^k+1}\frac{2\lambda_1+\mu_2}{2} = \frac{1}{2^k+1}(2\lambda_1+\mu_2).$$

Adding $\mu_{2^k+1} + [2^k/(2^k+1)](2\lambda_1+\mu_2)$ to both sides, we obtain the result. ◀

▶ **Lemma 19.** $\mu_x + 2\lambda_1 + \mu_2 > 0$ *for all integers* $x \geq 1$.

Proof. For $1 \leq x \leq 244$, we verify the finite number of cases computationally. For $x \geq 245$, we can use Lemma 15 to obtain $x|\mu_x| < (2\lambda_1 + \mu_2)/2$. Noting that $\lambda_1 < 0$, we have $|\mu_x| < \mu_2/(2x)$, from which $\mu_x > -\mu_2/(2x)$ because $\mu_2 > 0$. We then have $\mu_x + 2\lambda_1 + \mu_2 > -\mu_2/(2x) + 2\lambda_1 + \mu_2 \geq -\mu_2/(2\cdot 245) + 2\lambda_1 + \mu_2 \approx 0.31957706816604603 \geq 0$. ◀

We are now ready to provide a lower bound on $\mu_{2^k+1} - 2^{-t}\mu_{\lceil 2^{k+t}\rceil}$, applicable for all $k \geq 1$ and all $t \in (0,1)$, and independent of k. In particular, we prove Lemma 4.

Proof of Lemma 4. The desired inequality is equivalent to

$$\mu_{2^k+1} + 2\lambda_1 + \mu_2 > 2^{-t}(\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2). \tag{11}$$

By Lemma 19, $\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2 > 0$ for all positive integer values of $\lceil 2^{k+t}\rceil$, and specifically for $k \geq 11$ and $t \in (0,1)$. Therefore, the right-hand side of eq. 11 is strictly decreasing in t other than at discontinuities: values of t at which $\lceil 2^{k+t}\rceil$ increments by 1. For fixed k, the discontinuities are precisely those values of t at which 2^{k+t} is one of the integers $2^k, 2^k+1, \ldots, 2^{k+1}-1$, the values $t = \log_2(2^k+n) - k$ for integers n, $0 \leq n \leq 2^k - 1$.

To verify inequality 11 for all $t \in (0,1)$, it suffices to check points at which t approaches a discontinuity from the right. For $t \to 0^+$, inequality 11 becomes $\mu_{2^k+1} + 2\lambda_1 + \mu_2 > 2^{-t}(\mu_{2^k+1} + 2\lambda_1 + \mu_2)$, which holds from the positivity of $\mu_{2^k+1} + 2\lambda_1 + \mu_2 > 0$ by Lemma 19.

At $t = \log(2^k+n) - k$ for integers $1 \leq n \leq 2^k - 1$, because the discontinuity is approached from the right, $t > \log_2(2^k+1) - k$, so that $2^{-t} < 2^{-[\log_2(2^k+1)-k]} = 2^k/(2^k+1)$. Lemma 18 gives $\mu_{2^k+1} + 2\lambda_1 + \mu_2 > [2^k/(2^k+1)](\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2) > 2^{-t}(\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2)$. ◀

Proof of Lemma 5. Moving terms with t to one side, we must prove, for all $t \in (0,1)$,

$$\sum_{k=1}^{10} 2^{-k}(\mu_{2^k+1} + 2\lambda_1 + \mu_2) > 2^{-t}\sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2). \tag{12}$$

Because $\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2 > 0$ by Lemma 19 for all k, $1 \leq k \leq 10$, and $t \in (0,1)$, the right-hand side of eq. 12 is decreasing except at values of t where $\lceil 2^{k+t}\rceil$ increments by one: set $S = \{\log_2(2^k+n) - k : 1 \leq k \leq 10, 0 \leq n \leq 2^k - 1\}$. Hence, to verify eq. 12 for all $t \in (0,1)$, it suffices to examine only the limits as t approaches points in S from the right.

First, for $0 \in S$, as $t \to 0^+$, inequality 12 approaches $\sum_{k=1}^{10} 2^{-k}(\mu_{2^k+1} + 2\lambda_1 + \mu_2) > 2^{-t} \sum_{k=1}^{10} 2^{-k}(\mu_{2^k+1} + 2\lambda_1 + \mu_2)$, which holds by Lemma 19, noting $t > 0$.

Next, denote the points in finite set $S' = \{\log_2(2^k + n) - k : 1 \le k \le 10,\ 1 \le n \le 2^k - 1\}$ by $t_1 < t_2 < \ldots < t_K$, where $K = |S'|$. Notice that if $t \in (t_i, t_{i+1}]$ for some i, then the right-hand side of inequality 12 is maximized as $t \to t_i^+$. Furthermore, for all $t \in (t_i, t_{i+1}]$:

$$
\begin{aligned}
2^{-t} \sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t} \rceil} + 2\lambda_1 + \mu_2) &< \lim_{t \to t_i^+} 2^{-t} \sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t} \rceil} + 2\lambda_1 + \mu_2) \\
&< 2^{-t_i} \sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t_i+\epsilon} \rceil} + 2\lambda_1 + \mu_2),
\end{aligned}
$$

where $\epsilon > 0$ satisfies $t_i + \epsilon < t_{i+1}$ for all i, $1 \le i \le K-1$; we can take $\epsilon = \frac{1}{2} \min_{1 \le i \le K-1}(t_{i+1} - t_i)$. Hence, to prove inequality 12, it suffices to prove the stronger inequality

$$
\sum_{k=1}^{10} 2^{-k}(\mu_{2^k+1} + 2\lambda_1 + \mu_2) > 2^{-t_i} \sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t_i+\epsilon} \rceil} + 2\lambda_1 + \mu_2). \tag{13}
$$

for each i, $1 \le i \le K$. The advantage of inequality 13 over 12 is that it can be computationally verified by testing a finite number of points. In particular, we consider each $t_i \in S'$, choose an appropriate ϵ ($\epsilon = 10^{-16}$ suffices), and verify inequality 13 with that t_i and ϵ. ◀

D Proof of Lemmas 7 and 8 for Section 5

This appendix proves two lemmas used for Theorem 9. First, Lemma 20 gives a refined upper bound for $|\mu_x|$ that improves upon Lemma 15. This lemma is needed for Lemma 7.

▶ **Lemma 20.** *For all integers $x \ge 267$, $2^{\lceil \log_2 x \rceil}|\mu_x| < (2\lambda_1 + \mu_2)/2$.*

Proof. From eq. 10, $|\mu_x| \le 3(\log_2 e)\,[(\sqrt{1.34}/1.2)^x + (1/\sqrt{1.2})^x + 15/\big(8(1.2)^x\big)]$. Because $2^{\lfloor \log_2 x \rfloor} \le 2^{\log_2 x} = x$, it follows that $2^{\lceil \log_2 x \rceil} \le 2(2^{\lfloor \log_2 x \rfloor}) \le 2x$. Hence, it suffices to prove that $2x|\mu_x| \le (2\lambda_1 + \mu_2)/2$ for integers $x \ge 267$. That is, we can show the stronger inequality

$$
2x(3 \log_2 e)\left[\left(\frac{\sqrt{1.34}}{1.2}\right)^x + \frac{1}{(\sqrt{1.2})^x} + \frac{15}{8(1.2)^x}\right] < \frac{2\lambda_1 + \mu_2}{2}.
$$

As in A, the left-hand side is the sum of three terms of the form xa^x, where $a < 1$. By the same argument, this sum is decreasing for $x \ge -1/\log(\sqrt{1.34}/1.2) \approx 27.7880$. Hence, it suffices to show that the sum is less than $(2\lambda_1 + \mu_2)/2$ at $x = 267$, which can be accomplished computationally. Therefore, $2^{\lceil \log_2 x \rceil}|\mu_x| < (2\lambda_1 + \mu_2)/2$ for $x \ge 267$. ◀

Proof of Lemma 7. Moving all the terms involving t to one side, we must prove

$$
2^{-t}(\mu_{\lceil 2^{k+t} \rceil} + 2\lambda_1 + \mu_2) > 2^{-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2). \tag{14}
$$

Lemma 19 finds that $\mu_{\lceil 2^{k+t} \rceil} + 2\lambda_1 + \mu_2 > 0$ for all $k \ge 11$ and $t \in (0, 1)$. Therefore, except at discontinuities where 2^{k+t} is an integer and $\lceil 2^{k+t} \rceil$ increments by 1, the left-hand side strictly decreases as we increase t. It suffices to check inequality 14 at those discontinuities and as $t \to 1^-$. First, as $t \to 1^-$, inequality 14 becomes $2^{-t}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2) > 2^{-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2)$, a result that follows because $\mu_{2^{k+1}} + 2\lambda_1 + \mu_2 > 0$ and $0 < t < 1$.

Next, because $t \in (0,1)$, 2^{k+t} is an integer if and only if $t = \log_2(2^k+n) - k$ for an integer $1 \le n \le 2^k - 1$. Plugging $t = \log_2(2^k+n) - k$ into inequality 14 yields $[2^k/(2^k+n)](\mu_{2^k+n} + 2\lambda_1 + \mu_2) > 2^{-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2)$. Because $\mu_{2^k+n} + 2\lambda_1 + \mu_2 > 0$, it suffices to prove the stronger inequality $[2^k/(2^{k+1}-1)](\mu_{2^k+n} + 2\lambda_1 + \mu_2) > 2^{-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2)$, or equivalently, $[2^k/(2^{k+1}-1)]\mu_{2^k+n} - \frac{1}{2}\mu_{2^{k+1}} > -\left[\frac{1}{2}/(2^{k+1}-1)\right](2\lambda_1 + \mu_2)$.

By Lemma 20, which applies for $k \ge 11$ because $2^k + n \ge 267$, we have $2^{\lceil \log_2(2^k+n) \rceil}|\mu_{2^k+n}| < (2\lambda_1 + \mu_2)/2$ and $2^{k+1}|\mu_{2^k+n}| < (2\lambda_1 + \mu_2)/2$. Furthermore, by Lemma 15, we have that $2^{k+1}|\mu_{2^{k+1}}| < (2\lambda_1 + \mu_2)/2$, from which $(2^{k+1}-1)|\mu_{2^{k+1}}| < (2\lambda_1 + \mu_2)/2$. We then obtain

$$\begin{aligned}\frac{2^k}{2^{k+1}-1}\mu_{2^k+n} - \frac{1}{2}\mu_{2^{k+1}} &\ge -\left|\frac{2^k}{2^{k+1}-1}\mu_{2^k+n}\right| - \left|\frac{1}{2}\mu_{2^{k+1}}\right| \\ &\ge -\frac{1}{2}\frac{1}{2(2^{k+1}-1)}(2\lambda_1+\mu_2) - \frac{1}{2}\frac{1}{2(2^{k+1}-1)}(2\lambda_1+\mu_2) = -\frac{1}{2(2^{k+1}-1)}(2\lambda_1+\mu_2). \blacktriangleleft\end{aligned}$$

Proof of Lemma 8. Moving all the terms involving t to one side, we must prove

$$2^{-t}\sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t} \rceil} + 2\lambda_1 + \mu_2) > \sum_{k=1}^{10} 2^{-k-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2). \tag{15}$$

The terms in both summands are positive (Lemma 19). Therefore, except at the discontinuities in the left-hand side, the left-hand side strictly decreases with t. Hence, it suffices to check inequality 15 precisely at the discontinuities of the left-hand side and as $t \to 1^-$.

As $t \to 1^-$, the left-hand side of inequality 15 decreases to $\lim_{t\to 1^{-1}} 2^{-t}\sum_{k=1}^{10} 2^{-k}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2) = \sum_{k=1}^{10} 2^{-k-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2)$, verifying inequality 15 at $t \to 1^-$.

Hence, it remains to check inequality 15 at the discontinuities of the left-hand side, the points in $S' = \{\log_2(2^k+n) - k : 1 \le k \le 10, 1 \le n \le 2^k - 1\}$ used in C. We can computationally verify that inequality 15 holds for the finitely many points in S'. ◀