Journal of Integer Sequences, Vol. 28 (2025),
Article 25.5.5

The Comma Sequence is Finite in Other Bases

Robert Dougherty-Bliss
Department of Mathematics
Dartmouth College
27 N. Main St.
Hanover, NH 03755
USA
rdbliss@dartmouth.edu

Natalya Ter-Saakov
Department of Mathematics
Rutgers University
110 Frelinghuysen Rd.
Piscataway, NJ 08854
USA
nt399@rutgers.edu

Abstract

The comma sequence (1, 12, 35, 94, ...) is the lexicographically earliest sequence such that the difference of consecutive terms equals the concatenation of the digits on either side of the comma separating them. The behavior of a "generalized comma sequence" depends on the base in which the numbers are written, as well as the sequence's initial values. We give a computational proof that all comma sequences in bases 3 through 633 are finite.

Angelini et al. conjectured a generating function formula related to the comma sequence and, from this, predicted that the final element of a comma sequence in base

b should be roughly $\exp(O(b))$. We give a combinatorial proof of Angelini's conjecture, but also numerical evidence that the prediction about the final element is wrong. We provide a new random model for the comma sequence that predicts a final term of $\exp(O(b \log b))$, which aligns with our simulations.

Dedicated to Éric Angelini, a great friend of the OEIS.

1 Introduction

In 2006, Éric Angelini submitted A121805, the "comma sequence," to the On-Line Encyclopedia of Integer Sequences (OEIS) [7]. It is defined by a peculiar rule, best illustrated by listing the initial terms and their differences:

$$1 \underbrace{\ ,\ }_{11} 12 \underbrace{\ ,\ }_{23} 35 \underbrace{\ ,\ }_{59} 94 \underbrace{\ ,\ }_{41} 135 \underbrace{\ ,\ }_{51} 186 \underbrace{\ ,\ }_{62} 248 \underbrace{\ ,\ }_{83} 331 \underbrace{\ ,\ }_{13} 344, \ldots$$

The comma sequence is the lexicographically earliest sequence of positive integers such that the difference of consecutive terms equals the concatenation of the digits on either side of the comma separating them. For example, the sequence contains $12, 35$, which has a difference of $35 - 12 = 23$, the concatenation of 2 (the digit to the left of the comma) and 3 (the digit to the right of the comma). The amazing fact about the comma sequence is that it contains exactly 2,137,453 terms. It starts at 1, reaches 99999945, and then terminates. Naturally, the question is why the sequence persists so long, why it terminated at that number in particular, and what happens if we slightly change the setup.

A (generalized) comma sequence in base b is the lexicographically earliest sequence which satisfies the comma rule with a particular fixed initial value. Angelini et al. [1] studied generalized comma sequences and obtained a number of results that we have extended. First, they proved that all comma sequences in base 3 are finite and conjectured that this is true for all bases. We have extended this to more bases.

Theorem 1. *Comma sequences with arbitrary positive initial values are finite in bases 3 through 633.*

The proof of this theorem includes reducing the problem to a finite set of cases, then checking them on a large computing cluster with a program written in the Go language. We give the mathematical details of this proof in Section 3, and the computational details in Section 6.

While thinking about the duration of comma sequences in different bases, Angelini et al. conjectured the following, which we have proved [1, Conjecture 7.1].

Theorem 2. *The number $D(b)$ of comma sequences in base b with initial values in $[b^m - b^2, b^m)$ (with $m \geq 2$) that do not reach b^m is independent of m and has ordinary generating function*

$$\sum_{b=2}^{\infty} D(b) t^b = (1-t)^{-1} \left(\sum_{b=1}^{\infty} \frac{t^{b(b+3)/2}}{(1-t^b)} - t^2 \right) = t^3 + 2t^4 + 4t^5 + 5t^6 + \cdots \tag{1}$$

Angelini et al. (with help from Václav Kotěšovec) relied on this conjecture to reason that the final element of a base-b comma sequence should be roughly e^{2b}. More precisely, they estimated that a comma sequence ought to survive about $2b/\log(2b)$ "danger intervals," which we define in Section 2. This estimate is accurate for small bases, but becomes worse for larger ones. In Section 5, we introduce a model that more accurately matches empirical simulations for large bases, but depends on a probabilistic conjecture that we cannot prove. It predicts that comma sequences should survive about $b/2+1$ intervals, which translates to an estimated length of $b^{b/2+1} = e^{O(b \log b)}$.

We have been unable to prove that all comma sequences in bases other than 2 are finite. We could push our computations further than base 629 by finding small optimizations or using more computing power, but we do not see how this leads to a better theoretical understanding. The main problem remains open.

2 Background and examples

In this section we summarize some relevant background on comma sequences and give a number of examples. We begin with a more explicit definition of generalized comma sequences.

Definition 3. The *generalized comma sequence* in base-b with initial condition v is the sequence $(a(n))_{n\geq 1}$ defined as follows: $a(1)=v$; for $n>1$, if x is the least significant base-b digit of $a(n-1)$, then

$$a(n) = a(n-1) + bx + y, \tag{2}$$

where y is the most significant digit of $a(n)$ and is the smallest such y, if one exists. If no such y exists, then the sequence terminates.

Every positive integer has at most two positive integers which satisfy the comma-concatenation rule. For example, starting with the decimal number 14, we could write either "14, 59" or "14, 60." The definition says that we always choose the smallest next integer.

The comma sequence corresponds to $v=1$ and $b=10$. The first ten terms of the base-10 comma sequences with initial values $v=1,2,\ldots,10$ are as follows:

$$1, 12, 35, 94, 135, 186, 248, 331, 344, 387\ldots$$
$$2, 24, 71, 89, 180, 181, 192, 214, 256, 319\ldots$$
$$3, 36 \text{ (terminates)}$$
$$4, 48, 129, 221, 233, 265, 318, 402, 426, 490\ldots$$
$$5, 61, 78, 159, 251, 263, 295, 348, 432, 456\ldots$$
$$6, 73, 104, 145, 196, 258, 341, 354, 397, 471\ldots$$
$$7, 85, 136, 197, 269, 362, 385, 439, 534, 579\ldots$$

$$8, 97, 168, 250, 252, 274, 317, 390, 393, 427 \ldots$$
$$9, 100, 101, 112, 133, 164, 206, 268, 351, 364 \ldots$$
$$10, 11, 23, 58, 139, 231, 243, 275, 328, 412 \ldots$$

These sequences are all finite. Their lengths are recorded in A330128.

The first ten terms of the base-3 comma sequences with initial values $v = 1, 2, 3$ are as follows (given in base 3):

$$(1)_3, (12)_3, (110)_3, (111)_3, (122)_3, (221)_3, (1002)_3, (1100)_3, (1101)_3, (1112)_3 \ldots$$
$$(2)_3, (100)_3, (101)_3, (112)_3, (211)_3 \text{ (terminates)}$$
$$(10)_3, (11)_3 \text{ (terminates)}$$

(We use the notation $(d_k, d_{k-1}, \ldots, d_0)_b$ to represent the integer $d_k b^k + d_{k-1} b^{k-1} + \cdots + d_0$ in base-b where the d_i are base-b digits.) These sequences are also all finite; the first lasts for 17 terms.

The numbers which do not have successors, meaning integers at which comma sequences terminate, were termed *landmines* by Angelini et al. [1]. They are precisely the integers in base-b of the form

$$(b-1, b-1, \ldots, b-1, x, y)_b$$

where x and y are nonzero digits which sum to $b-1$ [1, Theorem 5.2]. For example, in base-10, the integers 45, 99972, and 9999918 are all landmines. We will call these *terminal* points or values.

Terminal points live in the following small intervals just to the left of powers of b.

Definition 4. The base-b *danger intervals* are the intervals $[b^k - b^2, b^k)$ for $k \geq 2$.

Some properties about comma sequences are best stated in terms of graphs. Let C_b be the directed graph on the positive integers which contains the edge $i \to j$ if and only if "i, j" satisfies the comma-concatenation rule in base-b. (This does not mean that the comma sequence starting at i *would* have j as its next term, just that j *could* be the next term, provided that it is the smallest option.)

Every vertex in C_b has out-degree at most two and in-degree at most 1. The only vertices with in-degree 0 are of the form

$$(d, 0, 0, \ldots, 0)_b,$$

where $d \in \{2, 3, \ldots, b-1\}$ [1, Theorems 5.4, 5.5].

Angelini et al. also showed that C_b is the disjoint union of finitely many directed trees. As a corollary, C_b contains an infinite path, meaning that if comma sequences were allowed to pick either of two possible successors, then there would be at least one infinite comma sequence in every base [1, Theorem 5.6].

3 How to prove finiteness

Neil Sloane and Giovanni Resta communicated to us the rough idea behind their proof that all comma sequences in base-3 are finite. In this section we describe their idea and its generalization to arbitrary bases. This section contains primarily mathematical details, while Section 6 contains more details about the computation and its implementation.

3.1 Overview

Angelini et al. proved that the behavior of comma sequences is determined by their values within relatively small intervals. Specifically, if you know the last time a comma sequence is in the interval

$$[d \cdot b^k - b^2, d \cdot b^k),$$

then you can predict the last time the sequence is in the "next" interval

$$[(d+1) \cdot b^k - b^2, (d+1) \cdot b^k)$$

without actually computing the intermediate terms.

The idea rests on a simple observation. Here are some terms from the original comma sequence:

$$997, 1068, 1149, 1240, 1241, 1252, 1273, 1304, 1345, 1396, 1457, 1528, \ldots$$

Their differences are

$$71, 81, 91, 01, 11, 21, 31, 41, 51, 61, 71, \ldots$$

The observation is that, while the leading digit remains 1, the last digit of the difference remains 1, and the first digit follows a simple arithmetic progression. Angelini et al. gave a description of this progression which allows us to predict that the last term in $[2000 - 100, 2000)$ is 1988, given only the information that the last term in $[1000 - 100, 1000)$ is 997. The formula for the arithmetic progression is given in equation (3).

Definition 5. Given a base-b digit d, an integer $0 \leq u < b^2$, and a positive integer k, we write (d, u, k) to represent the integer $d \cdot b^k - u$.

The tuples (d, u, k) represent points in the intervals $[d \cdot b^k - u, d \cdot b^k)$. The observation of Angelini et al. amounts to saying that we can compute a map

$$(d, u, k) \to (d', u', k')$$

which tells us how comma sequences move from one interval to the next. For example, the jump from 997 to 1988 is represented as $(1, 3, 2) \to (2, 12, 3)$.

Our contribution is that the map $(d, u, k) \to (d', u', k')$ is determined by only finitely many k. Specifically, for each base b there exists a finite graph G'_b which contains a cycle if and only if there are infinite base-b comma sequences. Figure 2 is a drawing of G'_3, which has no cycles. Figure 3 is a condensed drawing of G'_4, which also does not contain a cycle. Therefore, all comma sequences in bases 3 and 4 are finite. The only exception we know of is $b = 2$; see Figure 1.

3.2 Infinite graph construction

Here we construct an infinite directed graph that represents the comma map

$$(d, u, k) \to (d', u', k').$$

We first need a technical lemma which says that we do not need to consider *all* (d, u, k).

Let us first illustrate this with an example. If a base-10 comma sequence reaches the value $(1, 99, 3)$, then it proceeds as follows:

$$(1, 99, 3) \to (1, 80, 3) \to (1, 71, 3).$$

The last point is the last value the sequence takes in the interval $[900, 1000)$. It turns out that comma sequences behave identically on *all* danger interval. For example, if a base-10 comma sequence reaches $(1, 99, 7)$, then it proceeds as follows:

$$(1, 99, 7) \to (1, 80, 7) \to (1, 71, 7).$$

All that has changed is k, but the behavior of the u's are the same. The below lemma states that this is true in general.

Lemma 6 (Minimal u's)**.** *For any base b, any integer $k > 2$, and any nonzero base-b digit d, there exists a smallest finite set $U(b, d)$ that satisfies the following properties:*

- *(**a**) If a base-b comma sequence has a value in $[d \cdot b^k - b^2, d \cdot b^k)$, then its final value in this interval is $d \cdot b^k - u$ for some $u \in U(b, d)$.*
- *(**b**) $U(b, d)$ is independent of k.*
- *(**c**) If $d \neq 1$, then*

$$U(b, d) = \{(r, s)_b \mid r + s < b,\ 0 < s < b\} \cup \{(r, s)_b \mid r + s = b,\ s < d\}.$$

- *(**d**) If $d = 1$, then*

$$U(b, 1) = \{(r, s)_b \mid r + s \le b,\ 0 < s < b\}.$$

Proof. Begin with $d > 1$. The final value of a base-b comma sequence in $[d \cdot b^k - b^2, d \cdot b^k)$ has the form $(d - 1, b - 1, \dots, b - 1, x, y)_b$ for some $0 \le x, y < b$. Because this is the last term in the interval, the next term is

$$(d - 1, b - 1, \dots, b - 1, x, y) + (y, d)_b,$$

The relationship between (x, y) and (r, s) is

$$(x, y)_b + (r, s)_b = b^2.$$

There are two cases: First, $x + y \geq b$. Since $y > 0$, we have $(r, s) = (b - 1 - x, b - y)$ and the inequality $x + y \geq b$ is equivalent to $r + s \leq b - 1$ and $s > 0$.

Second, $x + y = b - 1$ and $y + d \geq b$. Note that

$$(d - 1, b - 1, \ldots, b - 1, x, y) + (y, d - 1)_b$$

is also no longer in the interval, so $y + d - 1 \geq b$. This again implies $y > 0$, and so $(r, s) = (b - 1 - x, b - y)$, and the conditions

$$\begin{aligned} x + y &= b - 1 \\ y + d &\geq b + 1 \end{aligned}$$

are equivalent to

$$\begin{aligned} r + s &= b \\ s &< d. \end{aligned}$$

The reasoning for $d = 1$ is similar. The only difference is in the second case, when the inequality $y + d - 1 \geq b$ becomes $y + b - 1 \geq b$. The rest of the argument is identical. $\square$

As an example of the lemma, here are the sets $U(3, d)$:

$$\begin{aligned} U(3, 1) &= \{1, 2, 4, 5, 7\} \\ U(3, 2) &= \{1, 2, 4, 7\}. \end{aligned}$$

Definition 7. The digraph G_b consists of vertices labeled (d, u, k) with base-b digit d, a minimal $u \in U(b, d)$, and an integer $k \geq 0$. The edge $(d, u, k) \to (d', u', k')$ exists if the latter is the immediate image of the former under the comma map.

An infinite comma sequence in base b corresponds to an infinite path in G_b. Therefore, our main goal is to determine whether such paths exist.

3.3 Finiteness of the comma map

We now study the comma map $(d, u, k) \to (d', u', k')$ in more detail.

Given (d, u, k) which is not a terminal point, we want to compute the next value (d', u', k') as a function of (d, u, k). For example, if we have a base-10 sequence at the value $(6, 8, 3)$ (also known as 5992), where does it go next? Both d' and k' are easy:

$$d' = \begin{cases} d + 1 & \text{if } d < b - 1; \\ 1 & \text{otherwise,} \end{cases}$$

$$k' = \begin{cases} k & \text{if } d < b - 1; \\ k + 1 & \text{otherwise.} \end{cases}$$

So a partial answer is $(6, 8, 3) \to (7, u', 3)$. The main difficulty is determining u'.

Proposition 8. *For each fixed base b, there exists an integer $L(b)$ such that the comma map $(d, u, k) \to (d', u', k')$ can be computed for sufficiently large k knowing only d, u, and $k \bmod L(b)$. The point where k becomes "sufficiently large" depends only on b and can be computed explicitly.*

Proof. Define the list

$$S_b(d, u) = \langle ((md - u) \bmod b)b + d \mid 0 \le m < b/\gcd(b, d)\rangle. \tag{3}$$

Angelini et al. proved that the differences between consecutive terms of a comma sequence cycle through the arithmetic progression $S_b(d, u)$ until they reach (d', u', k') (see [1, sec. 6] and A121805). To compute u', we start at $d \cdot b^k - u$ and add the elements of S_b in order, repeating until we get as close to $(d + 1)b^k$ as possible without going over. The distance between $(d+1)b^k$ and $db^k - u$ is $b^k + u$. Using entire copies of $S_b(d, u)$, we can get as close as

$$(b^k + u) \bmod \sum S_b(d, u) \tag{4}$$

where $\sum S_b(d, u)$ is the sum of the entries of the cycle. After this, we manually add elements of $S_b(d, u)$, starting from the beginning.

In particular, u' depends only on the cycle, u, and (4), but not k directly. It is well-known that $b^k \bmod m$, for integers b and $m \ge 1$, is eventually periodic in k[1]. It follows that for each cycle $S_b(d, u)$, there exists a period $l(d, u)$ and a nonnegative integer $k_0(d, u)$ such that

$$b^k \equiv b^{k+l(d,u)} \pmod{\sum S_b(d, u)}$$

for all $k \ge k_0(d, u)$. If we set

$$L(b) = \operatorname{lcm}\{l(d, u) \mid 1 \le d < b, \quad u \in U(b, d) \cup \{0\}\}, \tag{5}$$

then (4) is eventually periodic with period $L(b)$ for all d and u. $\square$

The technical point of forcing k to be "sufficiently large" is not important. If there were a truly infinite comma sequence, then its values would pass by arbitrarily large b^k's, meaning that the periodicity of (4) would kick in at some point.

[1] Any C-finite sequence taken with a fixed modulus is eventually periodic. Famous examples include exponential sequences and the Fibonacci numbers [4, 5].

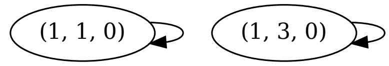

Figure 1: Drawing of G_2'. The loop from $(1, 1, 0)$ to itself demonstrates that a binary comma sequence which reaches $2^k - 1$ advances to $2^{k+1} - 1$, and that this continues forever. The loop from $(1, 3, 0)$ to itself shows that the same thing happens from $2^k - 3$ to $2^{k+1} - 3$.

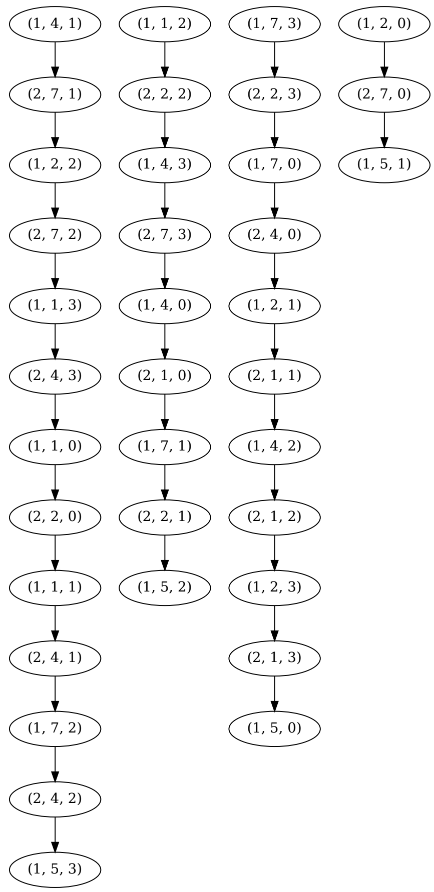

Figure 2: Drawing of G_3'. To prove that all base-3 comma sequences are finite, it suffices to check that all vertices eventually lead to a vertex with out-degree 0.

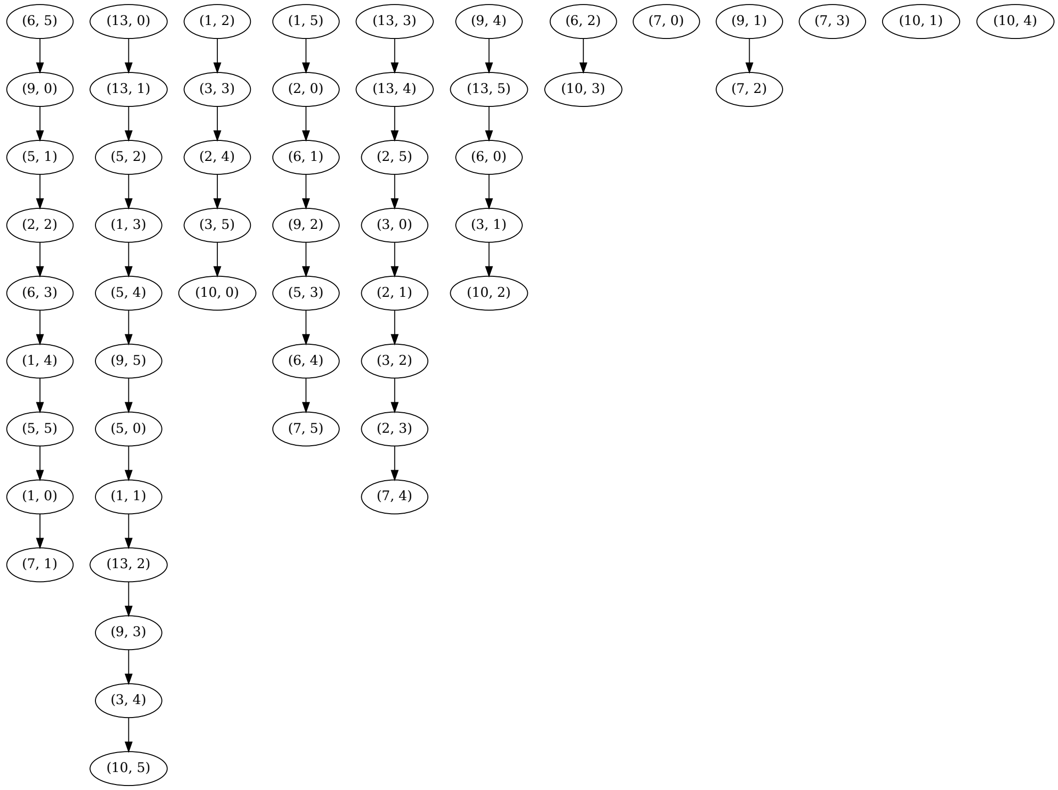

Figure 3: Modified drawing of G'_4 which contains only vertices of the form $(1, u, k)$, with all intermediate vertices replaced by one edge. To prove that all base-4 comma sequences are finite, it suffices to check that all vertices eventually lead to a vertex with out-degree 0. This modification contains 54 vertices, while G'_4 contains 144.

3.4 Finite graph construction

Armed with the results of the previous section about the comma map, we now show how to reduce the infinite graph G_b to a finite graph G'_b.

Definition 9. The digraph G'_b consists of vertices (d, u, κ) where d is a base-b digit, u is an element of $U(b, d)$, and $[\kappa]$ is an equivalence class mod $L(b)$. The edge $(d, u, [\kappa]) \to (d', u', [\kappa'])$ exists provided that the edge $(d, u, k) \to (d', u', k')$ exists in G_b for sufficiently large integers $k \equiv \kappa \pmod{L(b)}$ and $k' \equiv \kappa' \pmod{L(b)}$.

Lemma 10. *G'_b is well-defined and has both in- and out-degree at most* 1*.*

Proof. The periodicity of (4) for sufficiently large k shows that the computation of $(d, u, k) \to (d', u', k')$ in G_b depends only on $k \bmod L(b)$, so G'_b is well-defined. It further shows that the in- and out-degree of $(d, u, [\kappa])$ in G'_b are the same as (d, u, k) in G_b for sufficiently large $k \equiv \kappa \pmod{L(b)}$. Vertices in G_b have out-degree at most 1 (by the lexicographically earliest construct) and in-degree at most 1 (by Lemma 5.3 in [1]). $\square$

Proposition 11. *Comma sequences with arbitrary initial conditions in base-b terminate if and only if G'_b contains no cycles.*

Proof. The existence of infinite comma sequences is equivalent to the existence of infinite walks on the vertices of G_b, which is in turn equivalent to the existence of a cycle in G'_b. $\square$

This proposition implies a relatively simple, direct way to prove that all base b comma sequences are finite: Check that G'_b contains no cycles. The runtime of this depends directly on $L(b)$, because Lemma 10 implies that G'_b is the disjoint union of independent paths and cycles. Our initial calculations, limited to bases no higher than 23, used this idea. However, there is a far better method.

Proposition 12. *Comma sequences with arbitrary initial conditions in base-b terminate if none of the points $(d, u, 0)$ lie on a cycle.*

Because edges in G'_b are of the form

$$(d, u, \kappa) \to (d', u', \kappa + 1),$$

a cycle in G'_b contains every value of κ mod $L(b)$. In particular, to check that no cycle exists, it suffices to check that no point $(d, u, 0)$ leads to a cycle. We carried out this computation for $b \in \{3, 4, \ldots, 633\}$ and observed that there were no cycles in G'_b, thereby proving that all comma sequences in these bases are finite.

4 The geometric model

Definition 13. Let $I(b)$ be the set of initial values in $[b^k - b^2, b^k)$ (with $b, k \geq 2$) for which the comma sequence in base-b does not reach b^k. Let $D(b) = |I(b)|$ be the size of $I(b)$. (As mentioned in the previous section, this quantity is independent of k.)

Angelini et al.'s probabilistic model makes the simplifying assumption that comma sequences enter the danger intervals $[b^k - b^2, b^k)$ uniformly at random, and that each interval should be considered independent. Under this model we would expect to hit a terminal point after $b^2/D(b)$ intervals. In this section, we prove Theorem 2 and as a corollary the asymptotic estimate

$$\frac{b^2}{D(b)} \sim \frac{2b}{\log 2b}.$$

While this result is correct, the underlying model is slightly off. See Section 5 for better estimates.

To prove Theorem 2, we provide an injection from $I(b-1)$ to $I(b)$ and characterize the elements not covered. We then demonstrate a bijection with representations of b as the difference of two positive triangular numbers, A136107.

For brevity, we refer to the base-b integers of the form $(b-1, b-1, \ldots, b-1, x, y)_b$ as $\overline{(x, y)_b}$.

Lemma 14. *For $y > 0$, $\overline{(x, y)}_{b-1}$ is the parent of $\overline{(r, s)}_{b-1}$ in a base-$(b-1)$ comma sequence if and only if $\overline{(x+1, y)}_b$ is the parent of $\overline{(r+1, s)}_b$ in base-b.*

For example, $921 = (9, 2, 1)_{10}$ leads to $940 = (9, 4, 0)_{10}$. The lemma says that this is equivalent to the observation that $(10, 3, 1)_{11} = 1244$ leads to $(10, 5, 0)_{11} = 1265$ in base-11.

Proof. By the comma rule, we have

$$\overline{(r, s)}_{b-1} = \overline{(x, y)}_{b-1} + (y, b-2)_{b-1}.$$

There must be a carry in the units digit because y is positive and s is less than $b-1$, so $y + (b-2) = s + b - 1$. Because there is no carry in the tens digit this gives $x + y + 1 = r$. Together these imply $y + (b-1) = s + b$ and $(x+1) + y + 1 = (r+1)$, so $\overline{(x+1, y)}_b$ is followed by $\overline{(r+1, s)}_b$ in base-b. (Note that $r+1$ and $x+1$ are digits base-b since r and x are digits in base-$(b-1)$.) The other direction follows the same logic. $\square$

Lemma 15. *The function $f\colon I(b-1) \to I(b)$ defined by*

$$f(\overline{(r, s)}_{b-1}) = \overline{(r+1, s)}_b$$

is an injection that misses only $\overline{(1, b-2)}_b$ and the elements of the form $\overline{(0, y)}_b$ (with $0 < y < b-1$).

This map obviously takes terminal points to terminal points, but it is more general. For example, the base-10 value $31 = (3, 1)_{10}$ leads to the terminal point $45 = (4, 5)_{10}$. If we apply this map to both elements the path is preserved, and we get $(4, 1)_{11}$, which leads to $(5, 6)_{11}$.

Proof. First, we must show that $\overline{(r+1, s)}_b$ is actually an element of $I(b)$. If $\overline{(r, s)}_{b-1}$ is a terminal point in base-$(b-1)$, then $\overline{(r+1, s)}_b$ is a terminal point in base-b. By Lemma 14, if $x, y > 0$ and $\overline{(x, y)}_{b-1}$ leads to the terminal point $\overline{(r, s)}_{b-1} \neq \overline{(x, y)}_{b-1}$ in base-$(b-1)$, then $\overline{(x+1, y)}_b$ leads to $\overline{(r+1, s)}_b$ in base b. Note that all elements of the form $\overline{(x, 0)}_b$ escape. This means that the only elements of $I(b)$ not covered are the terminal point $\overline{(1, b-2)}_b$ and the points of the form $\overline{(0, y)}_b$ (with $0 < y < b-1$). $\square$

Let $T(n) = \frac{n(n+1)}{2}$ be the nth triangular number.

Lemma 16. *There exists a bijection between the uncovered elements of $I(b)$ and the ways to represent b as the difference of two positive triangular numbers.*

Proof. For every b, there exists a trivial representation $T(b) - T(b-1)$ which we map to $\overline{(1, b-2)}_b$. For the rest, we claim the element $\overline{(0, y)}_b$ fails to escape the interval if and only if $b = T(y+1) - T(a)$ for some $y > a > 0$.

If we continue the comma sequence beginning at $\overline{(0, y)}_b$, then as long as we are in the interval we obtain the sequence of values $\overline{(x(n), y(n))}_b$ which satisfy

$$\overline{(x(0), y(0))}_b = \overline{(0, y)}_b$$

and the recurrences

$$\begin{aligned} y(n+1) &= y(n) - 1 \\ x(n+1) &= x(n) + y(n) + 1. \end{aligned}$$

From these recurrences it is easy to prove $y(n) = y - n$ and

$$x(n) + y(n) = (n+1)y - \frac{n(n-1)}{2}.$$

If $\overline{(0, y)}_b$ hits a mine in n steps then $x(n) + y(n) = b - 1$. Then,

$$b = (n+1)y - \frac{n(n-1)}{2} = T(y+1) - T(y-n).$$

Note that $\overline{(x(n), y(n))}_b$ is a mine only if $y(n) > 0$. Therefore, $T(y-n)$ is also positive.

Now suppose that $b = T(y+1) - T(a)$ for some $y > a > 0$. Then, $y - a$ steps after $\overline{(0, y)}_b$, we have $x(n) + y(n) = b - 1$ (by the same calculation as above). We know that $x(y-a) > 0$ since $y - a > 0$ and $y(y-a) = a > 0$ so this must be a terminal point. Thus $\overline{(0, y)}_b$ fails to escape. $\square$

Proof of Theorem 2. From the injection in Lemma 15 and the bijection in Lemma 16, we know that, for $b \geq 3$, $D(b) - D(b-1) = r(b)$, the number of ways to represent b as the difference of two positive triangular numbers (recorded in A136107). This implies

$$D(b) = \sum_{k=1}^{b} r(b) - 1$$

for $b \geq 2$. On the other hand, the ordinary generating function for the sequence $r(b)$ was found by Václav Kotěšovec [7, seq. A136107] to be

$$\sum_{b=1}^{\infty} r(b) t^b \;=\; \sum_{n \geq 1} \frac{t^{n(n+3)/2}}{1 - t^n}.$$

Using the formula for the generating function of partial sums [8], we see that the ordinary generating function of $D(b)$ is

$$\sum_{b=1}^{\infty} D(b) t^b \;=\; \frac{1}{1-t} \sum_{n \geq 1} \frac{t^{n(n+3)/2}}{(1 - t^n)} - \frac{t^2}{1-t},$$

which matches (1). $\square$

5 A more accurate model

In this section we give an estimate of the length of a base-b comma sequence without the independence estimate used by Angelini et al.

In Section 3, we constructed a finite graph G'_b which contains no cycles if and only if all comma sequences in base b are finite. By randomly sampling points from this graph, we obtain a different probabilistic model of comma sequence behavior. This model predicts the lengths of comma sequences more accurately than the geometric model in [1], provided that there truly are no cycles.

The expected number of danger intervals which a comma sequence survives can reasonably be computed as

$$|V(G'_b)|^{-1} \sum_{v \in V(G'_b)} h(v),$$

where $h(v)$ is the number of edges between $v \in V(G'_b)$ and the end of the path in which it resides, and $V(G'_b)$ is the set of all vertices in G'_b. A simple calculation shows that this equals

$$|V(G'_b)|^{-1} \sum_{P} \binom{|P|}{2},$$

where the sum is over all paths P in G'_b and $|P|$ is number of vertices in a path. By adjusting the denominator, we can rephrase this as an expectation in terms of a uniformly randomly chosen path P:

$$|V(G'_b)|^{-1} \sum_{P} \binom{|P|}{2} = \frac{\#\text{ of paths}}{|V(G'_b)|} \mathbb{E}\left[\binom{|P|}{2}\right].$$

Recall that all paths in G'_b begin at a point of the form $(d, 0, \kappa)$ with κ taken mod $L(b)$ (as in (5)), so there are $(b-2)L(b)$ paths in total.

Empirical testing led us to the following conjecture.

Conjecture 17. As $b \to \infty$, the path length $|P|$ is approximately exponentially distributed with rate close to $(b/2+1)^{-1}$. In particular, $\mathbb{E}[|P|] \approx b/2 + 1$ and $\mathrm{Var}(|P|) \approx (b/2+1)^2$.

With this conjecture and the preceding remarks, we may very roughly estimate the number of danger intervals survived as

$$\frac{b}{2} + 1 + O(b^{-1}). \tag{6}$$

The geometric model of [1] predicts that roughly $2b/\log(2b)$ intervals should be survived. Figure 4 gives evidence that $b/2+1$ is a better estimate. It also suggests that $b/2+1$ may also be slightly smaller than the true mean.

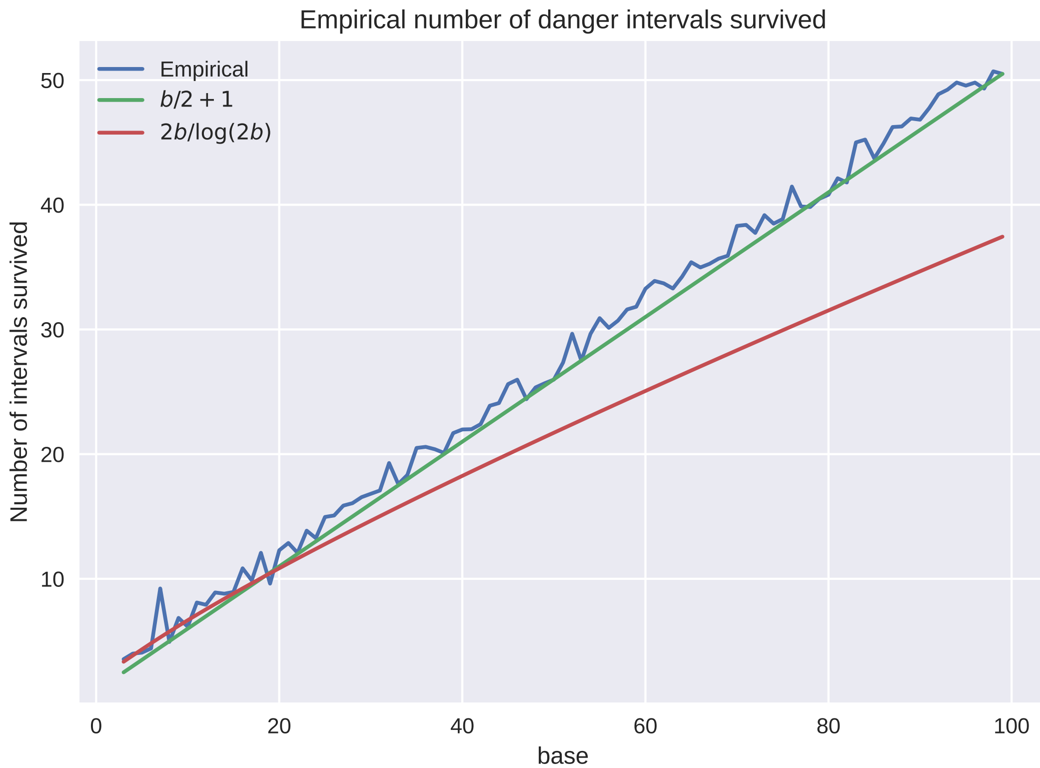

Figure 4: Average number of danger intervals survived for comma sequences in base b, estimated empirically using the initial values $1, 2, \ldots, b^2-1$. Note that the empirical curve begins close to $2b/\log(2b)$, but pulls away for later bases.

6 Computational details

The considerations of Section 3 lead to Algorithm 1. Our parallel implementation is in the Go programming language [2, 3].

The algorithm starts at each point $(d, u, 0)$ and applies the comma map until it hits a loop or a terminal point. In the case that a terminal point is always reached, the runtime depends on how far each vertex is from a terminal point. In the previous section we conjectured that paths have an average length of $b/2+1$, and we are picking b^3 different points, so we expect the average runtime to be around $O(b^4)$. We can obtain constant time improvements by noticing that the points $(d, u, 0)$ are processed independently, so the procedure `IsFinite` is trivially parallelized.

More importantly, we can save a lot of time by noticing that `AdvancePoint` duplicates effort over repeated calls. The computations in that procedure depend only on d, u, and

Algorithm 1 Check whether base-b comma sequences are finite.

```
procedure IsFinite(b)
    for 1 ≤ d < b, u ∈ U(b, d) do
        P, P_0 ← (d, u, 0)
        repeat
            P ← AdvancePoint(b, P)
            if P = P_0 then
                return infinite
            end if
        until isMine(P)
    end for
    return finite
end procedure
```

```
procedure AdvancePoint(b, d, u, k)
    if d = b − 1 then
        d' ← 1
        k' ← (k + 1) mod L(b)
    else
        d' ← d + 1
        k' ← k
    end if
    S ← Cycle(d, u)                                  ▷ see (3)
    (k_0, l) ← Order(b, ∑ S)                         ▷ b^k is periodic with period l for k ≥ k_0
    gap ← (b^(k_0 + (k − k_0) mod l) + u) mod ∑ S
    if gap = 0 then
        gap ← LastElem(S)                            ▷ the shortcut does not apply if the gap is 0
    else
        i ← 0
        while gap > S[i] do
            if gap < b^2 and d = b − 1 and IsMine(b, (d', gap, k')) then
                break
            end if
            gap ← gap − S[i]
            i ← i + 1
        end while
    end if
    u' ← gap
    return (d', u', k')
end procedure
```

$k \bmod l(d,u)$, where $l(d,u)$ is the period appearing in (5). So, while `IsFinite` must consider $k \bmod L(b)$, we can cache the results of `AdvancePoint(d, u, k mod l(d, u))`. This cache requires a memory footprint of roughly

$$O\left((\log b)^3 \sum_{d,u} l(d,u) \log l(d,u)\right).$$

This has not been a bottleneck thus far. (Base $b = 629$ required approximately 2GB of memory.)

As a minor remark, it proved difficult to find a good reference algorithm for determining precisely when and with what period $b^k \bmod m$ repeats for *arbitrary* integers b and $m \geq 1$. For completeness, this is how it is done:

- If $\gcd(b, m) = 1$, then the period equals the smallest positive k such that $b^k \equiv 1 \pmod{m}$, and the cycle begins immediately [6, ch. 6].
- If $\gcd(b, m) = g > 1$, then let k_0 be the smallest positive integer such that $\gcd(b^k, m)$ is constant for $k \geq k_0$. Then $b^k \equiv b^{k+l} \pmod{m}$ for $k \geq k_0$ if and only if $b^l \equiv 1 \pmod{m/g}$. Now apply the previous case to b and m/g.

7 Acknowledgments

We thank Neil Sloane and Giovanni Resta for bringing this problem and several key ideas to our attention, and Luke Pebody for his suggestion of a far superior way to search for cycles in G_b'. We also thank the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey for providing access to the Amarel cluster and associated research computing resources (https://oarc.rutgers.edu), as well as the office of Research Computing at Dartmouth College for access to the Discovery Cluster (https://rc.dartmouth.edu/).

References

[1] Éric Angelini, Michael S. Branicky, Giovanni Resta, N. J. A. Sloane, and David W. Wilson, The comma sequence: A simple sequence with bizarre properties, *Fibonacci Quart.* **62** (2024), 215–232.

[2] Alan Donovan and Brian Kernighan, *The Go Programming Language*, Addison-Wesley Professional, 2015.

[3] Robert Dougherty-Bliss, Github repository “rdbliss/comma”, 2024. https://github.com/rdbliss/comma.

[4] Paul Erdős, Carl Pomerance, and Eric Schmutz, Carmichael's lambda function, *Acta Arith.* **58** (1991), 363–385.

[5] John D. Fulton and William L. Morris, On arithmetical functions related to the Fibonacci numbers, *Acta Arith.* **2** (1969), 105–110.

[6] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, 1975. Fourth edition.

[7] OEIS Foundation Inc., The On-Line Encyclopedia of Integer Sequences, 2025. Available at `https://oeis.org`.

[8] Herbert Wilf, *generatingfunctionology*, CRC Press, 2005.

2020 *Mathematics Subject Classification*: Primary 11B25; Secondary 05A15.
Keywords: comma sequence, generating function, computer proof, cycle detection.

(Concerned with sequences A121805, A136107, and A330128.)

Received January 6 2025; revised versions received January 7 2025; August 23 2025. Published in *Journal of Integer Sequences*, September 19 2025.

Return to Journal of Integer Sequences home page.