#A74 INTEGERS 23 (2023)

A DOMBI COUNTEREXAMPLE WITH POSITIVE LOWER DENSITY

Jeffrey Shallit[1]
School of Computer Science, University of Waterloo, Waterloo, ON, Canada
shallit@uwaterloo.ca

Received: 4/24/23, Revised: 7/24/23, Accepted: 9/19/23, Published: 10/2/23

Abstract

Let $r(k, A, n)$ denote the number of representations of n as a sum of k elements of a set $A \subseteq \mathbb{N}$. In 2002, Dombi conjectured that if A is co-infinite, then the sequence $(r(k, A, n))_{n \geq 0}$ cannot be strictly increasing. Using tools from automata theory and logic, we give an explicit counterexample where $\mathbb{N} \setminus A$ has positive lower density.

1. Introduction

Let $\mathbb{N} = \{0, 1, \ldots\}$ be the natural numbers, and let $A \subseteq \mathbb{N}$. Define $r(k, A, n)$ to be the number of k-tuples of elements of A that sum to n. Dombi [7] conjectured that there is no infinite set F such that $r(3, \mathbb{N} \setminus F, n)$ is strictly increasing. Recently Bell et al. [2] found a counterexample to this conjecture; also see the recent paper of Kiss, Sándor, and Yang [9]. However, the F of their example is quite sparse; it has upper density 0. In this note we give a simple explicit example of an F such that $r(3, \mathbb{N} \setminus F, n)$ is strictly increasing and F has positive lower density. The novelty in our approach is the use of tools from automata theory and logic.

2. Brief Introduction to Automata

A *deterministic finite automaton* (DFA) is a simple model of a computer that takes strings over a finite alphabet Σ as input, and either accepts or rejects them. The set of all accepted strings is called the *language recognized by* the automaton. A DFA consists of a finite number of states, Q, and labeled transitions between them, specified by a transition function $\delta : Q \times \Sigma \to Q$. Starting in the distinguished start state q_0, the automaton processes each symbol of the input in turn, moving from state to state according to δ. If it ends up in one of the distinguished accepting

DOI: 10.5281/zenodo.8399647

[1]Research supported in part by NSERC grant 2018-04118.

states, specified by a set F, the automaton accepts; otherwise it rejects. For more about automata theory, consult any basic textbook in the area, such as [8].

A set $S \subseteq \mathbb{N}$ is said to be *k-automatic* if there is a DFA recognizing the language of base-k representations of members of S. It is known that if S is k-automatic, then so is its complement $\mathbb{N} \setminus S$. For more about automatic sequences, see, for example, [1].

A *deterministic finite automaton with output* (DFAO) is a small variation on this model. Here the notion of acceptance/rejection for states is replaced by an output chosen from an output set, Δ. The output on input x is the output associated with the last state reached. A sequence $(a(n))_{n\geq 0}$ is said to be k-automatic if there is a DFAO that, after processing the input n represented in base k, reaches a state with output $a(n)$.

As an example, let $F = \{3, 12, 13, 14, 15, 48, 49, 50, \ldots\}$ be the set of natural numbers whose base-2 expansion (ignoring leading zeros) is of even length and begins with 11. This is a 2-automatic set, and the automaton is depicted in Figure 1. Here 0 is the initial state, and 3 is the only accepting state, denoted by the double circle. The input is a binary representation of n, starting with the most significant digit.

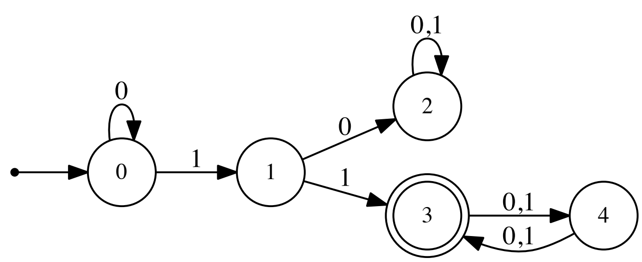

Figure 1: Automaton for F.

The language recognized by this automaton is

$$\{11, 011, 0011, 1100, 1101, 1110, 1111, \ldots\}.$$

For example, 14 in base 2 is 1110. On input 1110, the automaton visits states $0, 1, 3, 4$, and ends in state 3, which is an accepting state. Hence 14 is accepted, as it should be. On the other hand, 9 in base 2 is 1001. On input 1001, the automaton visits states $0, 1, 2, 2$, and ends in state 2, which not an accepting state. Hence 9 is rejected, and it is not in F.

A very important result about automatic sequences, originally due to Büchi, is the following:

Theorem 1. *The first-order logical theory of automatic sequences is decidable.*

Roughly speaking, this means that the truth of any well-formed first-order logic formula about automatic sequences (using a set of basic operations such as indexing, equality, and addition) can be determined with an algorithm. As an example of the kinds of claims that can be so expressed, consider the assertion that every natural number is a sum of two elements of $\mathbb{N} - F$. It can be expressed as

$$\forall n \ \exists i, j \notin F \text{ such that } n = i + j,$$

and hence is decidable with the algorithm. (It is true, by the way.)

For more about this celebrated result, see [5, 4, 11].

3. Linear Representations

Another tool we will need is linear representations. A sequence $(a(n))_{n\geq 0}$ has a *k-linear representation* if there exist an integer $t \geq 1$ and

- a $1 \times t$ row vector v;
- a $t \times t$-matrix-valued morphism μ with domain $\{0, 1, \ldots, k-1\}$; and
- a $t \times 1$ column vector w

such that $a(n) = v\mu(x)w$ for all strings x that represent n in base k (whether or not they have leading zeros). The integer t is called the *rank* of a linear representation.

Given linear representations for one or more sequences, we can, using a simple construction involving block matrices, effectively compute linear representations for any linear combination of them.

Linear representations are connected to automatic sequences in the following way [6]:

Theorem 2. *Suppose φ is a first-order logical formula about a k-automatic sequence, with free variables $n, x_1, x_2, \ldots x_j$. Then the number of j-tuples $(x_1, \ldots, x_j)$ such that $\varphi(n, x_1, \ldots, x_j)$ evaluates to true has a k-linear representation that can be effectively computed.*

We say two linear representations are *equivalent* if they represent the same sequence. Another extremely useful result is the following [3, Chap. 2]:

Theorem 3. *There is an algorithm for minimizing linear representations; that is, finding an equivalent linear representation of minimum rank.*

Finally, we will also use the following result [11, §4.11]:

Theorem 4. *It is decidable if an integer sequence $(a(n))_{n\geq 0}$ given by a k-linear representation takes only finitely many values. If so, then it is k-automatic, and the automaton is computably deducible from the representation.*

For more about linear representations, see [3, 11].

4. Our Example Set F

Let F be as defined in Section 2.

For a set $X \subseteq \mathbb{N}$, define $D_X(n) = \frac{1}{n}|\{X \cap \{0,1,\ldots,n-1\}\}|$. Recall that the *lower density* of X is defined to be $\liminf_{n\to\infty} D_X(n)$ and the *upper density* is $\limsup_{n\to\infty} D_X(n)$.

Proposition 1. *The lower density of F is $1/9$ and the upper density is $1/3$.*

Proof. The characteristic sequence of F is

$$000\,1\,\overbrace{\{0\cdots0\}}^{8}\overbrace{\{1\cdots1\}}^{4}\overbrace{\{0\cdots0\}}^{32}\overbrace{\{1\cdots1\}}^{16}\cdots\overbrace{\{0\cdots0\}}^{2\cdot4^n}\overbrace{\{1\cdots1\}}^{4^n}\cdots.$$

So the lower density of F is

$$\liminf_{n\to\infty} D_X(3\cdot4^n) = \frac{1+4+16+\cdots+4^{n-1}}{3\cdot4^n} = \frac{\frac{4^n-1}{3}}{3\cdot4^n} = \frac{1}{9},$$

and the upper density is

$$\limsup_{n\to\infty} D_X(4^n) = \frac{1+4+16+\cdots+4^{n-1}}{4^n} = \frac{\frac{4^n-1}{3}}{4^n} = \frac{1}{3}.$$

□

Theorem 5. *The sequence $r(3,\mathbb{N}\setminus F,n)$ is strictly increasing.*

Proof. Here is an outline of the proof. Define $A := \mathbb{N}\setminus F$ and $d(n) := r(3,A,n) - r(3,A,n-1)$. We will show that $d(n) > 0$ for all n. To do this, we show

$$d(n) \geq 4d(\lfloor n/4\rfloor) - 18. \tag{1}$$

and then use an easy induction.

To prove the bound in Equation (1), we show that $f(n) := d(n) - 4d(\lfloor n/4\rfloor)$ is a 2-automatic sequence, and we explicitly determine the automaton. Once we have the automaton for f, we can determine the range of f simply by examining the (finitely many) outputs associated with the states.

To find the DFAO for f, we first observe that $A := \mathbb{N}\setminus F$ is a 2-automatic set since F is. We can now use Theorem 1 to conclude that

$$G := \{(n,i,j,k) \,:\, n = i+j+k \text{ for } i,j,k \in A\}$$

is also 2-automatic. The automaton for G can be computed explicitly by free software called `Walnut` [10, 11], with the following commands:

```
morphism x "0->01 1->23 2->22 3->44 4->33":
morphism y "0->0 1->0 2->0 3->1 4->0":
promote X x:
image FF y X:
def g "FF[i]=@0 & FF[j]=@0 & FF[k]=@0 & n=i+j+k":
```

The resulting automaton has 143 states.

Now, from Theorem 2, we know that the number $r(3, A, n)$ of triples (i, j, k) corresponding to a particular n has a linear representation. Furthermore, this linear representation is computable from the automaton for G by the `Walnut` command

```
def r3an n "$g(i,j,k,n)":
```

In the same way we can compute linear representations for

$$r(3, A, n-1), r(3, A, \lfloor n/4 \rfloor), \text{ and } r(3, A, \lfloor n/4 \rfloor - 1)$$

using the following commands:

```
def r3anm1 n "$g(i,j,k,n-1)":
def r3an4 n "$g(i,j,k,n/4)":
def r3an4m1 n "$g(i,j,k,n/4-1)":
```

These have rank $143, 446, 446$ respectively. From these four linear representations we can compute a linear representation for the linear combination

$$\begin{aligned} f(n) &:= d(n) - 4d(\lfloor n/4 \rfloor) \\ &= r(3, A, n) - r(3, A, n-1) - 4(r(3, A, \lfloor n/4 \rfloor) - r(3, A, \lfloor n/4 \rfloor - 1)). \end{aligned}$$

The resulting linear representation has rank 1178.

Now we can use Theorem 3 to minimize this linear representation, resulting in a linear representation (v', γ', w') for f of rank 16. We give it explicitly below:

$$v'^T = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad w' = \begin{bmatrix} -3 \\ -2 \\ -1 \\ -3 \\ -6 \\ -5 \\ -1 \\ -3 \\ -6 \\ -6 \\ -3 \\ -3 \\ 3 \\ 1 \\ -1 \\ 3 \end{bmatrix}$$

$$\gamma'(0) =$$

$$\frac{1}{276}\begin{bmatrix}
276 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 276 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 276 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 276 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 276 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 276 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 276 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 276 & 0 \\
0 & 0 & 0 & -616 & -348 & 344 & 156 & -202 & 156 & 414 & 208 & -680 & -8 & 692 & -572 & -634 \\
0 & 0 & 0 & 244 & -36 & -288 & -566 & 405 & -566 & 161 & 12 & 1408 & -280 & -252 & 542 & 971 \\
0 & 0 & 0 & 276 & 552 & 92 & 46 & 23 & 46 & -253 & -368 & 92 & 276 & -460 & 138 & 253 \\
0 & 0 & 0 & -380 & -324 & 168 & 334 & -219 & 334 & 161 & 384 & -668 & 56 & 492 & -274 & -553 \\
0 & 0 & 0 & 448 & 504 & 168 & 472 & 126 & 196 & -322 & -168 & -392 & 56 & -336 & 140 & -70 \\
0 & 0 & 0 & -344 & -180 & 216 & 712 & -252 & 712 & 92 & 60 & -1148 & 164 & 396 & -556 & -964 \\
0 & 0 & 0 & -132 & -252 & -84 & -6 & 75 & -6 & 207 & 84 & 12 & -120 & 168 & -162 & -195 \\
0 & 0 & 0 & 364 & 444 & 56 & -226 & 203 & -226 & -161 & -332 & 360 & 80 & -388 & 62 & 429
\end{bmatrix}$$

$$\gamma'(1) =$$

$$\frac{1}{552}\begin{bmatrix}
0 & 552 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 552 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 552 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 552 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 552 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 552 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 552 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 552 \\
0 & 0 & 0 & -104 & -600 & -16 & -724 & 218 & -724 & 322 & 568 & 1632 & -496 & 32 & 692 & 942 \\
0 & 0 & 0 & -444 & -396 & 420 & 306 & -99 & 306 & 345 & 408 & -1164 & 48 & 816 & -294 & -957 \\
0 & 0 & 0 & -300 & 180 & 612 & 714 & -231 & 714 & 69 & -336 & -1428 & 480 & 432 & -870 & -945 \\
0 & 0 & 0 & 616 & 72 & -528 & -1076 & 846 & -1076 & 230 & -24 & 2704 & -544 & -600 & 1124 & 1922 \\
0 & 0 & 0 & 324 & 468 & 156 & 642 & 393 & 642 & -483 & 120 & -732 & 144 & -312 & -54 & -249 \\
0 & 0 & 0 & 376 & 216 & -480 & -284 & 606 & -284 & 230 & -72 & 1120 & -160 & -696 & 428 & 890 \\
0 & 0 & 0 & 772 & 420 & -596 & -726 & 565 & -726 & -207 & -232 & 1820 & -352 & -1016 & 914 & 1291 \\
0 & 0 & 0 & -980 & -516 & 748 & 1302 & -1049 & 1302 & -69 & 632 & -2788 & 464 & 1264 & -1186 & -2351
\end{bmatrix}$$

Using Theorem 4, we can check that the range of this linear representation is finite and explicitly deduce a DFAO M for it. By inspection of M, we see that the range of f is

$$\{-18, -15, -14, -12, -11, -10, -9, -8, -7, -6, -5, -4, -3, -2, -1, \\ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 18\}.$$

Thus we have shown $f(n) = d(n) - 4d(\lfloor n/4 \rfloor) \geq -18$.

We now verify by induction on n that $d(n) > n/5 + 7$ for $n \geq 87$. The base case is $87 \leq n < 348$, and is easily checked.

Now assume $n \geq 348$ and that $d(n') > n'/5 + 7$ for $87 \leq n' < n$. Then by induction

$$d(n) \geq 4d(\lfloor n/4 \rfloor) - 18 > 4(\lfloor n/4 \rfloor/5 + 7) - 18 \geq 4((n/4 - 1)/5 + 7) - 18 > n/5 + 7,$$

as desired.

After checking that $d(n) > 0$ for $0 \leq n < 87$, it follows that $d(n) > 0$ for all n. $\square$

References

[1] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations.* Cambridge University Press, 2003.

[2] J. P. Bell and J. Shallit, Counterexamples to a conjecture of Dombi in additive number theory, *Acta Math. Hung.* **169** (2023), 562–565.

[3] J. Berstel and C. Reutenauer, *Noncommutative Rational Series With Applications*, Vol. 137 of *Encyclopedia of Mathematics and Its Applications*, Cambridge University Press, 2011.

[4] V. Bruyère, G. Hansel, C. Michaux, and R. Villemaire, Logic and p-recognizable sets of integers, *Bull. Belgian Math. Soc.* **1** (1994), 191–238. Corrigendum, *Bull. Belg. Math. Soc.* **1** (1994), 577.

[5] J. R. Büchi, Weak secord-order arithmetic and finite automata, *Z. Math. Logik Grundlagen Math.* **6** (1960), 66–92. Reprinted in S. Mac Lane and D. Siefkes, eds., *The Collected Works of J. Richard Büchi*, Springer-Verlag, 1990, pp. 398–424.

[6] É. Charlier, N. Rampersad, and J. Shallit, Enumeration and decidable properties of automatic sequences, *Internat. J. Found. Comp. Sci.* **23** (2012), 1035–1066.

[7] G. Dombi, Additive properties of certain sets, *Acta Arith.* **103** (2002), 137–146.

[8] J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley, 1979.

[9] S. Z. Kiss, C. Sándor, and Q.-H. Yang, On monotone increasing representation functions, *Bull. Aust. Math. Soc.* (2023), https://doi.org/10.1017/S0004972723000679.

[10] H. Mousavi, Automatic theorem proving in `Walnut`, preprint, arXiv:1603.06017 [cs.FL], available at http://arxiv.org/abs/1603.06017, 2016.

[11] J. Shallit, *The Logical Approach to Automatic Sequences: Exploring Combinatorics on Words with* `Walnut`, Vol. 482 of *London Math. Soc. Lecture Notes Series*, Cambridge University Press, 2022.