On Integer Sequences Whose First Iterates are Linear

Jean-Paul Allouche
CNRS, LRI
Bâtiment 490
F-91405 Orsay Cedex
France
allouche@lri.fr

Narad Rampersad and Jeffrey Shallit
School of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada
nrampersad@math.uwaterloo.ca
shallit@graceland.uwaterloo.ca

Abstract

In this paper we discuss the functional equation $a(a(n)) = dn$, where $(a(n))_{n\geq 0}$ is a monotone sequence of non-negative integers. Mallows observed this equation has a unique solution for $d = 2$, and Propp observed the same thing for $d = 3$. We show that the equation has uncountably many solutions for $d \geq 4$. Further, we give a complete description for the lexicographically least such sequence, showing that the first difference sequence can be generated by a finite automaton.

AMS 2000 Classification: Primary 39B05; Secondary 11B85, 11A63.

1 Introduction

C. L. Mallows observed that there is a unique monotone sequence $(a(n))_{n\geq 0}$ of non-negative integers such that

$$a(a(n)) = 2n \tag{1}$$

for $n \neq 1$; see sequence A007378 in Sloane's on-line Encyclopedia of Integer Sequences [6]. (It is not difficult to see that no monotone sequence can satisfy Eq. (1) for *all* $n \geq 0$, including $n = 1$.) In 1979 J. Propp [5] introduced the sequence $(s(n))_{n\geq 0}$, defined to be the unique monotone sequence such that $s(s(n)) = 3n$. This is sequence A003605 in Sloane's Encyclopedia [6].

The first and third authors recently showed [2] that the sequence $(a(n))_{n\geq 0}$ is 2-regular in the sense of [1] and that the sequence $(s(n))_{n\geq 0}$ is 3-regular. A natural question to ask is,

> For $d \geq 4$ do there exist monotone sequences $\mathbf{a} = (a(n))_{n\geq 0}$ that satisfy $a(a(n)) = dn$ for all $n \geq 0$?

In this paper we prove that there are actually *uncountably* many such sequences for $d \geq 4$. We also study the lexicographically least monotone sequence satisfying $a(a(n)) = dn$ for all $n \geq 0$ and prove it is d-regular.

2 Uncountably many sequences

Although there are unique monotone sequences satisfying $a(a(n)) = dn$ for $d = 2, 3$, for $d \geq 4$ it turns out that there are uncountably many such sequences. We give a complete proof in the case $d = 4$; the general case is similar and we omit the details.

Theorem 1 *Let $b_1, \ldots, b_n$ be positive integers. For all choices*

$$\begin{aligned} b_1 &\in \{5, 6\} \\ b_2 &\in \{29, 30\} \\ &\vdots \\ b_i &\in \{2 \cdot 4^i - 3, 2 \cdot 4^i - 2\} \end{aligned}$$

where $i = 1, 2, \ldots$, there exists a strictly increasing sequence $\mathbf{a} = (a(n))_{n \geq 0}$ of non-negative integers such that $a(a(n)) = 4n$ and $a(4^i - 1) = b_i$.

Proof. We first define a *partial sequence.*

Definition 2 Let k be a positive integer and let $s : \{0, 1, 2, \ldots, 4^k - 1\} \to \mathbb{N}$. Then s is called a *partial sequence of length* 4^k if

1. s is strictly increasing;
2. $s(s(n)) = 4n$ for all $n < 2 \cdot 4^{k-1}$;
3. for $i = 1, 2, \ldots, k$, we have $s(4^i - 1) \in \{2 \cdot 4^i - 3, 2 \cdot 4^i - 2\}$.

We will show that for any $k \geq 0$ and any partial sequence a of length 4^k, we can always extend the domain of a to $\{0, 1, 2, \ldots, 4^{k+1} - 1\}$ such that, for both choices of $a(4^{k+1} - 1) \in \{2 \cdot 4^k - 3, 2 \cdot 4^k - 2\}$, a is a partial sequence of length 4^{k+1}.

Certainly this is true for $k = 0$ since the partial sequence of length 1

$$a(0) = 0$$

can be extended to the partial sequence of length 4 given by

$$a(0) = 0, \quad a(1) = 2, \quad a(2) = 4, \quad a(3) = 5$$

or the partial sequence of length 4 given by

$$a(0) = 0, \quad a(1) = 2, \quad a(2) = 4, \quad a(3) = 6$$

Now assume that for some $k \geq 0$, we have a partial sequence a of length 4^k. We will show that the domain of a can be extended such that a becomes a partial sequence of length 4^{k+1}.

We first note that a must have

$$a(0) = 0, \quad a(1) = 2, \quad a(2) = 4$$

since any other choices would violate either the rule $a(a(n)) = 4n$ or the requirement that the sequence be strictly increasing. By induction then, for $i = 0, 1, \ldots, k$, we have $a(2\cdot 4^i) = 4^{i+1}$.

Note that for each choice for the value of $a(4^{k-i} - 1)$, where $i = 0, \ldots, k-1$, the requirement that $a(a(n)) = 4n$ fixes exactly two values of the sequence in the range $4^k \leq n < 4^{k+1}$. Specifically, an easy induction shows that these values are:

$$a(2 \cdot 4^k - 4^i \cdot 3) = 4^{k+1} - 4^{i+1} \tag{2}$$

$$a(4^{k+1} - 4^{i+1}) = 2 \cdot 4^{k+1} - 4^{i+1} \cdot 3 \tag{3}$$

if $a(4^{k-i} - 1) = 2 \cdot 4^{k-i} - 3$, and

$$a(2 \cdot 4^k - 4^i \cdot 2) = 4^{k+1} - 4^{i+1} \tag{4}$$

$$a(4^{k+1} - 4^{i+1}) = 2 \cdot 4^{k+1} - 4^{i+1} \cdot 2 \tag{5}$$

if $a(4^{k-i} - 1) = 2 \cdot 4^{k-i} - 2$.

Our first task is to show that each of the terms defined by equations (1)-(4) is distinct. Specifically, we need to show that for all distinct integers i and j, such that $0 \leq i, j < k$, the following inequalities hold:

$$2 \cdot 4^k - 4^i \cdot 3 \neq 2 \cdot 4^k - 4^j \cdot 2 \tag{6}$$

and for all integers i and j, such that $0 \leq i, j < k$:

$$2 \cdot 4^k - 4^i \cdot 3 \neq 4^{k+1} - 4^{j+1} \tag{7}$$

and

$$2 \cdot 4^k - 4^i \cdot 2 \neq 4^{k+1} - 4^{j+1} \tag{8}$$

(5) is clearly true, since $4^i \cdot 3 \neq 4^j \cdot 2$ for all i, j. (6) and (7) are also true, since $2 \cdot 4^k - 4^i \cdot 3 < 2 \cdot 4^k$, $2 \cdot 4^k - 4^i \cdot 3 < 2 \cdot 4^k$, and $4^{k+1} - 4^{i+1} > 2 \cdot 4^k$, for all integers i such that $0 \leq i < k$.

We have already seen that the rule $a(a(n)) = 4n$ determines the values of the terms in the sequence for certain integers n such that $4^k \leq n < 4^{k+1}$. Our next task is to show that for those terms that have not been fixed by this rule, we are able to choose values that preserve the required properties of the sequence.

We give a procedure to assign values to $a(n)$, with $4^k \leq n < 4^{k+1}$, such that at the end of the procedure a is a partial sequence of length 4^{k+1}.

EXTEND(a)

1. Either define $a(4^{k+1}-1) = 2\cdot 4^{k+1}-3$ or define $a(4^{k+1}-1) = 2\cdot 4^{k+1}-2$.
2. Repeat (3)-(5) until for all $h = 0, 1, \ldots, 4^{k+1}-1$ the value of $a(h)$ is defined.
3. For $h = 0$ to $2\cdot 4^{k-1}-1$ repeat the following:
   If $a(h)$ is defined and $a(a(h))$ is not defined, set $a(a(h)) = 4h$.
4. Let l be the smallest integer such that there exists an integer i, with $2 \le i \le 4$, satisfying the following:
   (a) $a(l)$ and $a(l+i)$ are defined, and
   (b) $a(l+1), a(l+2), \ldots, a(l+i-1)$ are not defined.
5. Let j be the integer such that $a(j) = l$ and $a(j+1) = l+i$.
   If $i = 4$, define
   $$a(l+1) = 4j+1, \quad a(l+2) = 4j+2, \quad a(l+3) = 4j+3.$$
   If $i = 3$, choose
   $$(a(l+1), a(l+2)) \in \{(4j+1, 4j+2), (4j+1, 4j+3), (4j+2, 4j+3)\}.$$
   If $i = 2$, choose
   $$a(l+1) \in \{4j+1, 4j+2, 4j+3\}.$$

To see that steps 4 and 5 are correct, note that since for all $n \le l$ the values of $a(n)$ have already been defined, and since $a(l)$ and $a(l+i)$ have already been determined by the rule $a(a(n)) = 4n$ (steps 3 and 4 guarantee this), it must be the case that there exists a number j such that $a(j) = l$ and $a(j+1) = l+i$. Hence, we see that $a(a(j)) = a(l) = 4j$ and $a(a(j+1)) = a(l+i) = 4j+4$. Additionally, since $a(l+i) - a(l) = 4$, we also see that i is at most 4 (otherwise, a would not be strictly increasing). Therefore, if $i = 4$ we must choose values for $a(l+1), a(l+2), a(l+3)$ in the following way:

$$a(l+1) = 4j+1, \quad a(l+2) = 4j+2, \quad a(l+3) = 4j+3$$

Similarly, if $i = 3$ we can choose values for $a(l+1)$ and $a(l+2)$ in three different ways among $\{4j+1, 4j+2, 4j+3\}$, and if $i = 2$ we can choose $a(l+1)$ to be one of $\{4j+1, 4j+2, 4j+3\}$. In this way we ensure that a is strictly increasing.

The last thing that remains to be shown is that choosing values in this way does not affect our freedom to choose the value of $a(4^{k+1}-1)$ in step 1. To do this, we first note that the only terms of a that can possibly influence the choice of $a(4^{k+1}-1)$ are $a(2\cdot 4^k-3)$, $a(2\cdot 4^k-2)$, and $a(2\cdot 4^k-1)$, as will be shown below. (Note that we have already shown that $a(2\cdot 4^k)$ must have the value 4^{k+1}).

Since $a(4^k-1) \in \{2\cdot 4^k-3, 2\cdot 4^k-2\}$, we must have either $a(2\cdot 4^k-3) = 4^{k+1}-4$ or $a(2\cdot 4^k-2) = 4^{k+1}-4$. This implies that the value of $a(4^{k+1}-4)$ is always either $2\cdot 4^{k+1}-3\cdot 4$

or $2 \cdot 4^{k+1} - 2 \cdot 4$. Therefore, the only terms of a that can affect the choice of $a(4^{k+1} - 1)$ are those that affect the values of $a(4^{k+1} - 4)$, $a(4^{k+1} - 3)$, $a(4^{k+1} - 2)$, and $a(4^{k+1} - 1)$. However, since $n = 2 \cdot 4^k - 3$ is the smallest possible value of n such that $a(n) = 4^{k+1} - 4$, and since $a(2 \cdot 4^k) = 4^{k+1}$, we see that the only terms of a that can affect the values of $a(4^{k+1} - 4)$, $a(4^{k+1} - 3)$, $a(4^{k+1} - 2)$, and $a(4^{k+1} - 1)$ are $a(2 \cdot 4^k - 3)$, $a(2 \cdot 4^k - 2)$, and $a(2 \cdot 4^k - 1)$. We will now see how to ensure that suitable values are assigned to $a(2 \cdot 4^k - 3)$, $a(2 \cdot 4^k - 2)$, and $a(2 \cdot 4^k - 1)$.

Suppose that we had chosen $a(4^k - 1) = 2 \cdot 4^k - 3$. Then, since $a(2 \cdot 4^k) = 4^{k+1}$, and since we must be left free to choose the value of $a(4^{k+1} - 1)$, we must have the following:

$$a(2 \cdot 4^k - 3) = 4^{k+1} - 4$$

$$a(2 \cdot 4^k - 2) = 4^{k+1} - 3$$

$$a(2 \cdot 4^k - 1) = 4^{k+1} - 2$$

The iterative rule $a(a(n)) = 4n$ yields in turn:

$$a(4^{k+1} - 4) = 2 \cdot 4^{k+1} - 3 \cdot 4$$

$$a(4^{k+1} - 3) = 2 \cdot 4^{k+1} - 2 \cdot 4$$

$$a(4^{k+1} - 2) = 2 \cdot 4^{k+1} - 1 \cdot 4$$

leaving us free to choose the value of $a(4^{k+1} - 1)$ as we like.

On the other hand, suppose that we had chosen $a(4^k - 1) = 2 \cdot 4^k - 2$. Then, we must choose the value of $a(2 \cdot 4^k - 1)$ to be either $4^{k+1} - 3$ or $4^{k+1} - 2$. However, if we choose $a(2 \cdot 4^k - 1) = 4^{k+1} - 3$, we would have the following:

$$a(2 \cdot 4^k - 2) = 4^{k+1} - 4$$

$$a(2 \cdot 4^k - 1) = 4^{k+1} - 3$$

and the iterative rule $a(a(n)) = 4n$ would yield:

$$a(4^{k+1} - 4) = 2 \cdot 4^{k+1} - 2 \cdot 4$$

$$a(4^{k+1} - 3) = 2 \cdot 4^{k+1} - 1 \cdot 4$$

thus forcing the following:

$$a(4^{k+1} - 2) = 2 \cdot 4^{k+1} - 3$$

$$a(4^{k+1} - 1) = 2 \cdot 4^{k+1} - 2$$

and we have lost our freedom to choose the value of $a(4^{k+1} - 1)$. Therefore, we must choose $a(2 \cdot 4^k - 1) = 4^{k+1} - 2$, which gives the following:

$$a(2 \cdot 4^k - 2) = 4^{k+1} - 4$$

$$a(2 \cdot 4^k - 1) = 4^{k+1} - 2$$

and the iterative rule $a(a(n)) = 4n$ yields:

$$a(4^{k+1} - 4) = 2 \cdot 4^{k+1} - 2 \cdot 4$$

$$a(4^{k+1} - 2) = 2 \cdot 4^{k+1} - 1 \cdot 4$$

Here we can choose the value of $a(4^{k+1} - 3)$ to be one of $\{2 \cdot 4^{k+1} - 7, 2 \cdot 4^{k+1} - 6, 2 \cdot 4^{k+1} - 5\}$, and our freedom to choose $a(4^{k+1} - 1)$ is preserved.

Hence, we have shown that given a partial sequence a of length 4^k, we can extend the domain of a such that, for both choices of $a(4^{k+1} - 1) \in \{2 \cdot 4^k - 3, 2 \cdot 4^k - 2\}$, a is a partial sequence of length 4^{k+1}. The result now follows by induction.

3 Regularity of the sequences

We now turn to a simple description of the lexicographically least monotone sequence satisfying $a(a(n)) = dn$ for all $n \geq 0$. This description is in terms of the so-called regular sequences of Allouche and Shallit [1, 2]. Recall that we say that a sequence $\mathbf{a} = (a(n))_{n\geq 0}$ is *d-regular* if the set of subsequences

$$\{(a(d^i n + b))_{n\geq 0} \ : \ i \geq 0, \ 0 \leq b < d^i\}$$

generates a finitely-generated $\mathbb{Z}$-module.

We will show that the sequence $(a(n))_{n\geq 0}$ is d-regular. As a consequence it follows that $a(n)$ can be described in terms of the base-d expansion of n. More precisely, there exists an integer r, a matrix-valued morphism $\mu : \{0, 1, \ldots, d-1\} \to \mathbb{Z}^{r\times r}$ and two vectors $v, w \in \mathbb{Z}^r$ such that if the base-d expansion of n is $n = \sum_{0\leq i\leq t} a_i d^i$, then $a(n) = v^T \mu(a_0 a_1 \cdots a_t) w$. We do not give these maps explicitly here, although they can be deduced from what follows.

Since, as noted previously, the cases $d = 2, 3$ have been studied by Mallows and Propp, respectively, we restrict ourselves here to $d \geq 4$.

Theorem 3 *Let d be an integer ≥ 4. Define the sequence $\mathbf{a} = (a(n))_{n\geq 0}$ by*

$$a(0) = 0, \quad a(1) = 2, \quad a(d-1) = 2d - 3$$

and for all $k \geq 0$, $i \in [0, d^k - 1]$, *we have*

$$\begin{aligned}
a(xd^k + i) &= (x + d - 2)d^k + i, && \text{for all } x \in [2, d-2],\\
a(xd^k + i) &= (x - d + 2)d^{k+1} + di, && \text{for all } x \in [d, 2d-4],\\
a((2d-3)d^k + i) &= (d-1)d^{k+1} + i, &&\\
a((2d-2)d^k + i) &= (d^2 - d + 1)d^k + i, &&\\
a((2d-1)d^k + i) &= (d-1)d^{k+1} + a(d^k + i), &&\\
a(d(d-1)d^k + i) &= (2d-3)d^{k+1} + di, &&\\
a((d^2 - d + 1)d^k + i) &= (2d-2)d^{k+1} + di, &&\\
a(xd^k + i) &= (d^2 + x - 2)d^k + i, && \text{for all } x \in [d^2 - d + 2, d^2 - 2],\\
a((d^2 - 1)d^k + i) &= 2(d-1)d^{k+1} + a((d-1)d^k + i). &&
\end{aligned}$$

Then the sequence $\mathbf{a}$ *is increasing and it satisfies* $a(a(n)) = dn$, for all $n \geq 0$. *This sequence* $\mathbf{a}$ *is the lexicographically least increasing sequence* $\mathbf{u} = (u(n))_{n \geq 0}$ *satisfying* $u(u(n)) = dn$, for all $n \geq 0$. *Furthermore the sequence* $\mathbf{a}$ *is* d*-regular.*

The proof of this theorem takes many steps and is broken up into pieces in the next few sections.

First, we note that the relations in Theorem 3 above do indeed define a sequence $\mathbf{a}$. If $n \in \{0, 1, d-1\}$, this is clear. If $n \in [2, d-2]$ use the first relation with $k = 0$. If $n \geq d$, then there exists a unique $k > 0$ such that $d^k \leq n \leq d^{k+1} - 1$. Writing $n = xd^k + y$ with $y \in [0, d-1]$, we see that $x \in [1, d-1]$. If x belongs to $[2, d-2]$, use the first relation. If $x \in \{1, d-1\}$, then write $y = zd^{k-1} + t$, with $t \in [0, d^{k-1} - 1]$. Hence $n = (xd + z)d^{k-1} + t$. Then use the remaining relations. Note that the value of $a(d-1)$ and the relation $a(d(d-1)d^k + i) = 2d^{k+2} - 3d^{k+1} + di$ together show that for all $k \geq 0$ we have

$$a((d-1)d^k) = 2d^{k+1} - 3d^k. \tag{9}$$

Theorem 4 *Define the sequence* $\mathbf{b} = (b(n))_{n \geq 0}$ *by*

$$\text{for all } n \geq 0,\ b(n) := a(n+1) - a(n).$$

The relations defining the sequence $\mathbf{a}$ *and the remark* (9) *above imply:*

$$b(0) = 2, \quad b(1) = d - 2, \quad b(d-1) = 3,$$

and for all $i \in [0, d^k - 1]$

$$\begin{aligned}
b(xd^k + i) &= 1, \quad \text{for all } x \in [2, d-2], \\
b(xd^k + i) &= d, \quad \text{for all } x \in [d, 2d-4], \\
b((2d-3)d^k + i) &= 1, \\
b((2d-2)d^k + i) &= 1, \\
b((2d-1)d^k + i) &= b(d^k + i), \\
b(d(d-1)d^k + i) &= d, \\
b((d^2 - d + 1)d^k + i) &= d, \\
b(xd^k + i) &= 1, \quad \text{for all } x \in [d^2 - d + 2, d^2 - 2], \\
b((d^2 - 1)d^k + i) &= b((d-1)d^k + i),
\end{aligned}$$

Corollary 5 *These relations show that the first difference sequence* **b** *takes on only the (not necessarily distinct) values* $1, 2, 3, d-2, d$. *Hence the sequence* **a** *is strictly increasing.*

4 The sequence a satisfies $a(a(n)) = dn$

We prove in this section that for all $n \geq 0$ we have $a(a(n)) = dn$. First, we have

$$a(a(0)) = 0, \ \ a(a(1)) = a(2) = a(2d^0 + 0) = d,$$

and

$$a(a(d-1)) = a(2d-3) = a(dd^0 + d - 3) = 2d + d(d-3) = d^2 - d = d(d-1).$$

Next, we note that the result is true if the most significant digit of n in base d belongs to $[2, d-2]$. For if k is the largest exponent of d in the base d expansion of such an n, we have $n = xd^k + i$, where $i \in [0, d^k - 1]$ and $x \in [2, d-2]$. Hence $x + d - 2 \in [d, 2d-4]$, and

$$a(a(n)) = a((x + d - 2)d^k + i) = xd^{k+1} + di = dn.$$

We now prove, by induction on $k \geq 0$, that for all n belonging to $[d^k, d^{k+1} - 1]$ and having its most significant digit either equal to 1 or to $(d-1)$, we have $a(a(n)) = dn$. We see from what precedes that this holds for $k = 0$. Suppose this is true up to $k - 1$ for some $k \geq 1$ and let us prove the result for k.

(1) Let $n \in [d^k, d^{k+1} - 1]$ have its most significant digit equal to 1. Then $n = d^k + \lambda d^{k-1} + j$, with $\lambda \in [0, d-1]$ and $j \in [0, d^{k-1} - 1]$.

(1a) If $\lambda \in [0, d-4]$, then $\lambda + d \in [d, 2d-4]$ and $\lambda + 2 \in [2, d-2]$. Hence, noting that $dj \leq d^k - 1$,

$$a(a(n)) = a(a((d+\lambda)d^{k-1} + j)) = a((\lambda+2)d^k + dj) = (\lambda + d)d^k + dj = dn.$$

(1b) If $\lambda = d-3$, then, remembering that $j \leq d^{k-1} - 1$,

$$\begin{aligned} a(a(n)) &= a(a((2d-3)d^{k-1} + j)) = a((d-1)d^k + j) \\ &= a(d(d-1)d^{k-1} + j) = (2d-3)d^k + dj = dn. \end{aligned}$$

(1c) If $\lambda = d-2$, then

$$a(a(n)) = a(a((2d-2)d^{k-1} + j)) = a((d^2 - d + 1)d^{k-1} + j) = (2d-2)d^k + dj = dn.$$

(1d) If $\lambda = d-1$, we first note that the result is clear for $k = 1$. For we then have $j = 0$, and hence

$$a(a(n)) = a(a((2d-1)) = a(d(d-1) + a(1)) = a(d^2 - d + 2) = 2d^2 - d = dn.$$

We can now suppose that $k \geq 2$.

We note that, since **a** is increasing, we have

$$2d^{k-1} = a(d^{k-1}) \leq a(d^{k-1} + j) < a(2d^{k-1}) = d^k.$$

Hence there exist $s \in [2, d-1]$ and $t \in [0, d^{k-1} - 1]$ such that

$$a(d^{k-1} + j) = sd^{k-1} + t.$$

We thus have

$$\begin{aligned} a(a(n)) &= a(a((2d-1)d^{k-1} + j)) = a((d-1)d^k + a(d^{k-1} + j)) \\ &= a((d-1)d^k + sd^{k-1} + t) = a((d(d-1) + s)d^{k-1} + t). \end{aligned}$$

(1d.i) If $2 \leq s \leq d-2$, then $(d(d-1) + s) \in [d(d-1) + 2, d^2 - 2]$, hence

$$\begin{aligned} a(a(n)) &= a((d(d-1) + s)d^{k-1} + t) = (2d^2 - d + s - 2)d^{k-1} + t \\ &= (2d-1)d^k - 2d^{k-1} + a(d^{k-1} + j) \end{aligned}$$

To prove that $a(a(n)) = dn$, we have to show that under the assumptions above

$$a(d^{k-1} + j) = 2d^{k-1} + dj.$$

We write

$$\begin{aligned} a(d^{k-1} + j) &= sd^{k-1} + t \leq (d-2)d^{k-1} + d^{k-1} - 1 = (d-1)d^{k-1} - 1 \\ &< (d-1)d^{k-1} = a((2d-3)d^{k-2}). \end{aligned}$$

Since **a** is increasing, this implies

$$d^{k-1} + j < (2d-3)d^{k-2}, \text{ hence } j < (d-3)d^{k-2}.$$

This implies the existence of $\alpha \in [0, d-4]$ and of $\beta \in [0, d^{k-2} - 1]$ such that $j = \alpha d^{k-2} + \beta$. Since $d + \alpha \in [d, 2d-4]$ we then have

$$a(d^{k-1} + j) = a((d+\alpha)d^{k-2} + \beta) = (\alpha + 2)d^{k-1} + \beta d = dj + 2d^{k-1}.$$

(1d.ii) If $s = d-1$, then $a(d^{k-1}+j) = (d-1)d^{k-1}+t$. We thus have

$$\begin{aligned} a(a(n)) &= a((d^2-1)d^{k-1}+t) = 2(d-1)d^k + a(((d-1)d^{k-1}+t) \\ &= 2(d-1)d^k + a(a(d^{k-1}+j)) = 2(d-1)d^k + d(d^{k-1}+j) = dn, \end{aligned}$$

where we used the induction hypothesis in the penultimate equality.

(2) Let $n \in [d^k, d^{k+1}-1]$ with most significant digit equal to $(d-1)$. Then $n = (d-1)d^k + \lambda d^{k-1} + j$, with $\lambda \in [0, d-1]$ and $j \in [0, d^{k-1}-1]$.

(2a) If $\lambda = 0$, then $n = (d-1)d^k + j = d(d-1)d^{k-1} + j$, with $j \in [0, d^{k-1}-1]$. Hence, since $dj \in [0, d^k - 1]$,

$$a(a(n)) = a(a(d(d-1)d^{k-1}+j)) = a((2d-3)d^k + dj) = (d-1)d^{k+1} + dj = dn.$$

(2b) If $\lambda = 1$, then $n = (d-1)d^k + d^{k-1} + j = (d^2-d+1)d^{k-1}+j$, hence, noting that $dj \in [0, d^k-1]$,

$$a(a(n)) = a(a((d^2-d+1)d^{k-1}+j)) = a((2d-2)d^k+dj) = (d^2-d+1)d^k+dj = dn.$$

(2c) If $\lambda \in [2, d-2]$. Hence $d(d-1)+\lambda \in [d^2-d+2, d^2-2]$. We have, since $(\lambda-2)d^{k-1}+j \in [0, d^{k-1}-1]$,

$$\begin{aligned} a(a(n)) &= a(a((d(d-1)+\lambda)d^{k-1}+j)) = a((2d^2-d+\lambda-2)d^{k-1}+j) \\ &= a((2d-1)d^k + (\lambda-2)d^{k-1}+j) \\ &= (d-1)d^{k+1} + a(d^k + (\lambda-2)d^{k-1}+j) \\ &= (d-1)d^{k+1} + a((d+\lambda-2)d^{k-1}+j) \\ &= (d-1)d^{k+1} + \lambda d^k + dj = dn. \end{aligned}$$

(2d) If $\lambda = d-1$, we note that

$$2d^k - 3d^{k-1} = a((d-1)d^{k-1}) \le a((d-1)d^{k-1}+j) < a(d^k) = 2d^k.$$

Hence there exists $z \in [(d-1)d^{k-1}, d^k-1]$ such that $a((d-1)d^{k-1}+j) = d^k + z$. We then write

$$\begin{aligned} a(a(n)) &= a(a((d^2-1)d^{k-1}+j)) = a(2(d-1)d^k + a((d-1)d^{k-1}+j)) \\ &= a(2(d-1)d^k + d^k + z) \\ &= a((2d-1)d^k + z) = (d-1)d^{k+1} + a(d^k+z) \\ &= (d-1)d^{k+1} + a(a((d-1)d^{k-1}+j)) \\ &= (d-1)d^{k+1} + (d-1)d^k + j = dn, \end{aligned}$$

where we used the induction hypothesis in the penultimate equality.

5 The sequence a is minimal

We now prove that if an increasing sequence $\mathbf{u} = (u(n))_{n\geq 0}$ satisfies $u(u(n)) = dn$, for all $n \geq 0$, then $\mathbf{u} \geq \mathbf{a}$ (for the lexicographical order).

A first simple remark is that, since $\mathbf{a}$ is increasing and $a(1) = 2$, then we must have, for all $n \geq 1$, the inequality $a(n) \geq n + 1$.

A second useful remark is that $(u \circ u)(n) = u(u(n)) = dn$ for all $n \geq 0$ implies

$$\text{for all } n \geq 0,\ u(dn) = du(n).$$

Namely for all $n \geq 0$ we can write

$$du(n) = (u \circ u)(u(n)) = u(u(u(n))) = u((u \circ u)(n)) = u(dn).$$

This implies in particular that $u(0) = 0 = a(0)$.

(1) We first study "small" values of n, i.e., $n \leq d$.

We have $u(1) \neq 1$, otherwise $d = u(u(1)) = u(1) = 1$. Hence either $u(1) > 2 = a(1)$, and $\mathbf{u} > \mathbf{a}$, or $u(1) = 2$. In the latter case, we thus have

$$u(2) = u(u(1)) = d.$$

Then, since $\mathbf{u}$ is increasing, we have, for all $j \in [2, d-1]$,

$$u(j) \geq u(2) + j - 2 = d + j - 2 = a(j).$$

An easy finite induction on $j \in [2, d-1]$ shows that either $\mathbf{u} > \mathbf{a}$, or $u(n) = a(n)$ for all $n \in [0, d-1]$.

Then we have $u(d) = du(1) = da(1) = a(d)$.

(2) We now prove that if $u(n) = a(n)$ for all $n \in [0, N-1]$ for some $N \geq d+1$, then either $\mathbf{u} > \mathbf{a}$ or $u(N) = a(N)$. Let t be the unique integer such that $a(t) \leq N < a(t+1)$.

If $N = a(t)$, then $t < a(t) = N$, hence $u(t) = a(t) = N$. Thus

$$u(N) = u(u(t)) = dt = a(a(t)) = a(N).$$

We hence can suppose that $a(t) < N < a(t+1)$. We know that the first difference $a(t+1) - a(t) = b(t)$ belongs to $\{1, 2, 3, d-2, d\}$. Since the interval $(a(t), a(t+1))$ is not empty (it contains N) we cannot have $b(t) = 1$. **Let us suppose that** $d \geq 6$, so that $a(t+1) - a(t) = b(t)$ belongs to the set of exactly four different values $\{2, 3, d-2, d\}$. (The cases $d = 4, 5$ introduce some minor technical difficulties arising from the fact that $2 = d-2$ for $d = 4$, and $3 = d-2$ for $d = 5$; these complicate the proof somewhat, and we leave these details to the reader.) We note that $b(t) = 2$ can happen only if $t = 0$, hence is excluded ($N \geq d+1$). Hence $a(t+1) - a(t)$ belongs to $\{3, d-2, d\}$.

(2a) $a(t+1)-a(t)=3$.

Using the relations satisfied by the sequence $\mathbf{b}$, we see that this case occurs if and only if there exists $\ell \geq 1$ such that $t = d^\ell - 1$. Thus

$$a(t) = a(d^\ell - 1) = 2d^\ell - 3.$$

(Check this relation for $\ell = 1, 2$, then prove it by induction using that

$$\begin{aligned} a(d^{\ell+1}-1) &= a((d^2-1)d^{\ell-2} + d^{\ell-2} - 1) \\ &= 2(d-1)d^{\ell-1} + a((d-1)d^{\ell-2} + d^{\ell-2} - 1) \\ &= 2(d-1)d^{\ell-1} + a(d^{\ell-1} - 1).) \end{aligned}$$

Since $a(t+1) = a(d^\ell) = 2d^\ell$, we have $2d^\ell - 3 < N < 2d^\ell$. In particular $u(2d^\ell - 3) = a(2d^\ell - 3)$ from the induction hypothesis. But

$$\begin{aligned} a(2d^\ell - 3) &= a(a(d^\ell - 1) = d^{\ell+1} - d \\ a(2d^\ell - 2) &= a((2d-1)d^{\ell-1} + d^{\ell-1} - 2) \\ &= (d-1)d^\ell + a(2d^{\ell-1} - 2) = \ldots = d^{\ell+1} - d + 1 \\ a((2d^\ell - 1) &= a((2d-1)d^{\ell-1} + d^{\ell-1} - 1) \\ &= (d-1)d^\ell + a(2d^{\ell-1} - 1) = \ldots = d^{\ell+1} - d + 2. \end{aligned}$$

Since $\mathbf{u}$ is increasing, we have for $j = 0, 1, 2$

$$u(2d^\ell - 3 + j) \geq u(2d^\ell - 3) + j = a(2d^\ell - 3) + j = a(2d^\ell - 3 + j).$$

Hence either $\mathbf{u} > \mathbf{a}$ or $u(n) = a(n)$ for all $n \in [a(t), a(t+1))$.

(2b) $a(t+1) - a(t) = d - 2$.

We have $a(t) \leq N - 1$ hence $u(a(t)) = a(a(t)) = dt$. Using the relations satisfied by the sequence $\mathbf{b}$ we see that this happens if and only if there exists $\ell \geq 0$ such that $t = 2d^\ell - 1$. We have just seen that $a(2d^\ell - 1) = d^{\ell+1} - d + 2$ and we have $a(2d^\ell) = d^{\ell+1}$. Hence, noting that $a(d^{\ell+1} - d + 2) = a(a(2d^\ell - 1)) = 2d^{\ell+1} - d$, and $a(d^{\ell+1} - 1) = 2d^{\ell+1} - 3$, we have

$$d^{\ell+1} - d + 2 < N \leq d^{\ell+1} - 1$$

and

$$2d^{\ell+1} - d < a(N) \leq 2d^{\ell+1} - 3.$$

Since $\mathbf{a}$ is increasing, and since the intervals $(d^{\ell+1} - d + 2, d^{\ell+1} - 1]$ and $(2d^{\ell+1} - d, 2d^{\ell+1} - 3]$ have the same cardinality $d - 3$, we must have

$$\text{for all } j \in [0, d-3],\ a(d^{\ell+1} - d + 2 + j) = a(d^{\ell+1} - d + 2) + j.$$

Since $\mathbf{u}$ is increasing, we thus have for all $j \in [0, d-3]$,

$$u(d^{\ell+1} - d + 2 + j) \geq u(d^{\ell+1} - d + 2) + j = a(d^{\ell+1} - d + 2) + j = a(d^{\ell+1} - d + 2 + j).$$

Again this gives either $\mathbf{u} > \mathbf{a}$, or $u(n) = a(n)$ for all $n \in [a(t), a(t+1))$.

(2c) $a(t+1) - a(t) = d$.

Again we have $a(t) \le N-1$. Hence $u(a(t)) = a(a(t)) = dt$. We have

$$a(t) < N < a(t+1)$$

and

$$dt < a(N) < dt + d.$$

Since **a** is increasing, and since the intervals $(a(t), (a(t+1))$ and $(dt, dt+d)$ have the same cardinality $d-1$, we must have for all $j \in [0, d)$,

$$a((a(t)+j) = a(a(t)) + j = dt + j.$$

Since **u** is increasing we conclude just as above that either $\mathbf{u} > \mathbf{a}$ or $u(n) = a(n)$ for all $n \in [a(t), a(t+1))$.

6 Small and large values of the sequence a

Let d be an integer ≥ 3, and let $\mathbf{a} = (a(n))_{n \ge 0}$ be the lexicographically least sequence satisfying $a(a(n)) = dn$. A referee asked about the existence of $\lim_{n \to \infty} a(n)/n$. The following theorem addresses this question.

Theorem 6 *We have* $a(n) = \frac{2d-3}{d-1} n$ *for* $n = (d-1)d^n$, $n \ge 0$, *while* $a(n) = \frac{d(d-1)}{2d-3} n$ *for* $n = (2d-3)d^n$, $n \ge 0$.

Proof. Follows immediately from Theorem 3.

Since $\frac{2d-3}{d-1} < \frac{d(d-1)}{2d-3}$ for $d \ge 3$, it therefore follows that $\lim_{n \to \infty} a(n)/n$ does not exist.

7 The sequence a is d-regular

To prove that the sequence **a** is d-regular it suffices to prove that its first difference sequence **b** is d-automatic. There are many different, equivalent definitions of what it means for a sequence to be automatic [3]; for our purposes the following definition suffices.

Definition 7 Let Σ, Δ be finite alphabets, and let Σ^* (resp., Δ^*) denote the set of all finite strings over Σ (resp., Δ). A *morphism* is a map $h : \Sigma^* \to \Delta^*$ such that $h(xy) = h(x)h(y)$ for all $x, y \in \Sigma^*$. A morphism is said to be *d-uniform* if $|h(a)| = d$ for all $a \in \Sigma$. A morphism is said to be a *coding* if it is 1-uniform. An infinite word **w** is said to be a fixed point of h if $h(\mathbf{w}) = \mathbf{w}$. A sequence $(r(n))_{n \ge 0}$ is *d-automatic* if it can be written as the image (under a coding) of a fixed point of a d-uniform morphism.

Theorem 8

(a) If $d \geq 6$, then the sequence $\mathbf{b}$ *is the fixed point beginning with* 2 *of the d-uniform morphism*

$$\begin{array}{rcl} 2 & \to & 2\ \delta\ \overbrace{1\ 1\ \cdots\ 1}^{d-3}\ 3 \\ \delta & \to & \overbrace{d\ d\ \cdots\ d}^{d-3}\ 1\ 1\ \delta \\ 1 & \to & \overbrace{1\ 1\ \cdots\ 1}^{d} \\ d & \to & \overbrace{d\ d\ \cdots\ d}^{d} \\ 3 & \to & d\ d\ \overbrace{1\ 1\ \cdots\ 1}^{d-3}\ 3, \end{array}$$

where $\delta = d - 2$.

(b) If $d = 4$, then the sequence $\mathbf{b}$ *is the image by the coding φ of the fixed point beginning with c of the* 4*-uniform morphism*

$$\begin{array}{rcl} c & \to & c\ 2\ 1\ 3 \\ 2 & \to & 4\ 1\ 1\ 2 \\ 1 & \to & 1\ 1\ 1\ 1 \\ 4 & \to & 4\ 4\ 4\ 4 \\ 3 & \to & 4\ 4\ 1\ 3 \end{array}$$

where $\varphi(c) = 2$ and $\varphi(i) = i$ for $1 \leq i \leq 4$.

(c) If $d = 5$, then the sequence $\mathbf{b}$ *is the image by the coding ψ of the fixed point beginning with* 2 *of the* 5*-uniform morphism*

$$\begin{array}{rcl} 2 & \to & 2\ 3\ 1\ 1\ 3 \\ 3 & \to & 5\ 5\ 1\ 1\ 3 \\ 1 & \to & 1\ 1\ 1\ 1\ 1 \\ 5 & \to & 5\ 5\ 5\ 5\ 5 \end{array}$$

where $\psi(i) = i$ for $i \in \{1, 2, 3, 5\}$.

Proof.
This follows immediately from Theorem 4.

Corollary 9 *The sequence* $\mathbf{a}$ *is d-regular.*

Proof. By Theorem 2.3 of [1], it follows that since $\mathbf{b}$ is d-automatic, it is also d-regular. By the remark on p. 173 of [1], it follows that $\mathbf{a}$ is d-regular.

8 Open questions

Among open questions coming from our study, we mention the following.

1. If we choose $u(1) > 2$ and if there is a monotone sequence $\mathbf{u}$ such that $u(u(n)) = dn$, for all ≥ 0, is the lexicographically least such sequence a d-regular sequence?

2. What happens if the function $n \to dn$ is replaced by an increasing function? More precisely, let $\mathbf{f} = (f(n))_{n\geq 0}$ be an increasing sequence of integers. Is there an increasing sequence $\mathbf{g} = (g(n))_{n\geq 0}$ such that $g(g(n)) = f(n)$ for all $n \geq 0$? If it is the case and if the sequence $\mathbf{f}$ is d-regular for some integer $d \geq 2$, is it true that the lexicographically least such $\mathbf{g}$ is also d-regular?

3. One can also study the continuous case (see, for example, [4, pp. 53–61; 117–126]) of these functional equations. Is there any connection between the discrete and continuous versions?

References

[1] J.-P. Allouche and J. O. Shallit, The ring of k-regular sequences, *Theoret. Comput. Sci.* **98** (1992), 163–197.

[2] J.-P. Allouche and J. O. Shallit, The ring of k-regular sequences, II, *Theoret. Comput. Sci.* **307** (2003), 3–29.

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

[4] A. Delachet, *Calcul Différentiel et Intégral*, Que sais-je ?, Presses Universitaires de France, Paris, 1968.

[5] J. Propp, Problem proposal 474, *Crux Math.* **5** (1979), 229. Solution by G. Patruno, **6** (1980), 198.

[6] N. J. A. Sloane, *On-line Encyclopedia of Integer Sequences*, published electronically at `http://www.research.att.com/~njas/sequences/`