# Avoiding fractional powers over the natural numbers (extended abstract)

Lara Pudwell* and Eric Rowland†

June 22, 2014

**Abstract**

We study the structure of the lexicographically least infinite $\frac{a}{b}$-power-free word on the alphabet $\mathbb{Z}_{\geq 0}$, showing that for many rationals $\frac{a}{b}$ this word is a fixed point of a uniform morphism.

## 1 Introduction

Beginning with work of Thue [8, 9, 4], researchers have been interested in the avoidability of repetitions in infinite words. For example, it is easy to see that nonempty squares (words of the form $ww$ where $w$ is a nonempty word) are unavoidable in sufficiently long words on a binary alphabet, but Thue exhibited an infinite square-free word on a ternary alphabet.

Here we are interested in avoiding fractional powers. Let $a$ and $b$ be relatively prime positive integers. If $v = v_1 v_2 \cdots v_l$ is a nonempty word whose length $l$ is divisible by $b$, define

$$v^{a/b} := v^{\lfloor a/b \rfloor} v_1 v_2 \cdots v_{l \cdot \mathrm{frac}(a/b)},$$

where $\mathrm{frac}(\frac{a}{b}) = \frac{a}{b} - \lfloor \frac{a}{b} \rfloor$ is the fractional part of $\frac{a}{b}$. We say that $v^{a/b}$ is an $\frac{a}{b}$-*power*. Note that $|v^{a/b}| = \frac{a}{b}|v|$. If $\frac{a}{b} > 1$, then a word $w$ is an $\frac{a}{b}$-power if and only if $w$ can be written $v^e x$ where $e$ is a non-negative integer, $x$ is a prefix of $v$, and $|w|/|v| = a/b$. For example, $011101 = (0111)^{3/2}$ is a $\frac{3}{2}$-power. We say that a word is $\frac{a}{b}$-*power-free* if none of its nonempty factors are $\frac{a}{b}$-powers. Avoiding $\frac{3}{2}$-powers, for example, means avoiding factors $xyx$ where $|x| = |y| \geq 1$. Avoiding $\frac{5}{4}$-powers means avoiding factors $xyzux$ where $|x| = |y| = |z| = |u| \geq 1$.

If one does not know whether $\frac{a}{b}$-powers are avoidable on a given alphabet $\Sigma$, it is common to gain intuition by choosing an order for $\Sigma$ and attempting to construct a long finite $\frac{a}{b}$-power-free word by using the standard backtracking algorithm. If an infinite $\frac{a}{b}$-power-free word on $\Sigma$ does not exist, then the backtracking algorithm will identify the length of the longest $\frac{a}{b}$-power-free words. If an infinite $\frac{a}{b}$-power-free word on $\Sigma$ does exist, then the backtracking algorithm eventually computes prefixes of the lexicographically least such word. The lexicographically least $\frac{a}{b}$-power-free word is a canonical representative of the set of $\frac{a}{b}$-power-free words, so its structure is of interest.

On (ordered) finite alphabets, there has not been much success in identifying the structure of the lexicographically least infinite $\frac{a}{b}$-power-free word. Even characterizing the lexicographically least square-free word on $\{0, 1, 2\}$ is an open problem [3, §1.10].

On an infinite alphabet, however, the problem seems to be more tractable.

*Valparaiso University, USA

†University of Liege, Belgium, supported in part by a Marie Curie Actions COFUND fellowship

**Notation.** Let $a$ and $b$ be relatively prime positive integers such that $\frac{a}{b} > 1$. We define $\mathbf{w}_{a/b}$ to be the lexicographically least infinite word on $\mathbb{Z}_{\geq 0}$ avoiding $\frac{a}{b}$-powers.

Guay-Paquet and Shallit [6] showed that the lexicographically least square-free word on $\mathbb{Z}_{\geq 0}$ is

$$\mathbf{w}_2 = 010201030102010401020103010201 05 \cdots .$$

More generally, for an integer $a \geq 2$ we have $\mathbf{w}_a = \varphi^\omega(0)$, where $\varphi : \mathbb{Z}_{\geq 0}^* \to \mathbb{Z}_{\geq 0}^*$ is the morphism defined by $\varphi(n) = 0^{a-1}(n{+}1)$. The letters of $\mathbf{w}_a$ satisfy the recurrence $\mathbf{w}_a(ai{+}(a{-}1)) = \mathbf{w}_a(i){+}1$ for all $i \geq 0$, where we index letters in a word beginning at 0.

Shallit and the second author [7] gave a recurrence for the letters of

$$\mathbf{w}_{3/2} = 001102\,100112\,001103\,100113\,001102\,100114\,001103\,100112 \cdots .$$

The word $\mathbf{w}_{3/2}$ is 6-regular in the sense of Allouche and Shallit [1, 2]; informally, this means that the $i$th letter can be computed directly from the base-6 digits of $i$. Part of the motivation of the present study is to put this '6' into context by studying $\mathbf{w}_{a/b}$ systematically.

In this extended abstract, we show that for many rational numbers $\frac{a}{b}$, the word $\mathbf{w}_{a/b}$ is the fixed point of a $k$-uniform morphism for some integer $k$. (Recall that a morphism $\varphi$ on an alphabet $\Sigma$ is *$k$-uniform* if $|\varphi(n)| = k$ for all $n \in \Sigma$.)

## 2 Morphisms

It turns out that for $\frac{a}{b} \geq 2$ the word $\mathbf{w}_{a/b}$ is easy to describe. For example, for $\frac{a}{b} = \frac{5}{2}$ one computes

$$\mathbf{w}_{5/2} = 00001\,00001\,00001\,00001\,00002\,00001\,00001\,00001\,00001\,00002 \cdots$$

and observes that $\mathbf{w}_{5/2}$ agrees with $\mathbf{w}_5$ on a long prefix. In fact these two words are the same. More generally, for $\frac{a}{b} \geq 2$, the lexicographically least $\frac{a}{b}$-power-free word is a word we have already seen.

**Theorem 1.** *Let $a, b$ be relatively prime positive integers such that $\frac{a}{b} \geq 2$. Then $\mathbf{w}_{a/b} = \mathbf{w}_a$.*

Therefore it suffices to study $\mathbf{w}_{a/b}$ for rationals satisfying $1 < \frac{a}{b} < 2$. As a first example, let's consider

$$\mathbf{w}_{5/3} = 0000101\,0000101\,0000101\,0000101\,0000102\,0000101\,0000102 \cdots .$$

By examining a prefix of $\mathbf{w}_{5/3}$, one guesses the following theorem, which establishes the structure of $\mathbf{w}_{5/3}$.

**Theorem 2.** *Let $\varphi$ be the 7-uniform morphism defined by*

$$\varphi(n) = 000010(n+1)$$

*for all $n \in \mathbb{Z}_{\geq 0}$. Then $\mathbf{w}_{5/3} = \varphi^\omega(0)$.*

Similarly, by examining a prefix of $\mathbf{w}_{9/5}$, one guesses the following.

**Theorem 3.** *Let $\varphi$ be the 13-uniform morphism defined by*

$$\varphi(n) = 000000001000(n+1)$$

*for all $n \in \mathbb{Z}_{\geq 0}$. Then $\mathbf{w}_{9/5} = \varphi^\omega(0)$.*

At first it is not clear why 7 is the correct length for $\mathbf{w}_{5/3}$ and why 13 is the correct length for $\mathbf{w}_{9/5}$. However, the two morphisms in these theorems are quite similar; they only differ in their run lengths. In fact they belong to an infinite family of morphisms that generate words $\mathbf{w}_{a/b}$ for certain rationals, and we can generalize Theorems 2 and 3 as follows.

**Theorem 4.** *Let $a, b$ be relatively prime positive integers such that $\frac{5}{3} \leq \frac{a}{b} < 2$ and $\gcd(b, 2) = 1$. Let $\varphi$ be the $(2a - b)$-uniform morphism defined by*

$$\varphi(n) = 0^{a-1}\, 1\, 0^{a-b-1}\, (n+1)$$

*for all $n \in \mathbb{Z}_{\geq 0}$. Then $\mathbf{w}_{a/b} = \varphi^\omega(0)$.*

On the other hand, there also seem to be many "sporadic" words $\mathbf{w}_{a/b}$ that are fixed points of uniform morphisms but do not belong to general families. For example, the length $k = |\varphi(n)| = 733$ for the morphism in the following theorem is somewhat mysterious in that it has no obvious relationship to 8/5.

**Theorem 5.** *There is a 733-uniform morphism*

$$\varphi(n) = 000000010010000001001000000010011 \cdots 1001000000010100(n+2)$$

*such that $\mathbf{w}_{8/5} = \varphi^\omega(0)$.*

There are two steps in the proof of each of these theorems. The first step is to verify that the morphism $\varphi$ is $\frac{a}{b}$*-power-free* (that is, $\varphi(w)$ is $\frac{a}{b}$-power-free whenever $w$ is $\frac{a}{b}$-power-free). The second step is to verify that $\varphi$ is *lexicographically least* with respect to $\frac{a}{b}$ (that is, if $w$ is $\frac{a}{b}$-power-free and decrementing any letter introduces an $\frac{a}{b}$-power, then decrementing any letter in $\varphi(w)$ introduces an $\frac{a}{b}$-power ending at that position). Since the word 0 is $\frac{a}{b}$-power-free and lexicographically least of its length, if $\varphi$ is an $\frac{a}{b}$-power-free, lexicographically least morphism then $\mathbf{w}_{a/b} = \varphi^\omega(0)$.

## 3 Automatic verification that $\varphi$ is $\frac{a}{b}$-power-free

Theorems 4 and 5 represent a small sample of the potential results regarding $\mathbf{w}_{a/b}$; there are several symbolic families of morphisms analogous to the family in Theorem 4 and many more sporadic words as well. (Additionally, as we discuss in Section 4, some words are not fixed points of morphisms but are nonetheless related to fixed points of morphisms.) Of course, one would ideally prove a single theorem that captures all these cases. However, the structures are diverse enough that it is not clear how to unify them. The next best thing, then, is to identify a general proof scheme so that each individual proof may be carried out automatically. Here we briefly describe how to automatically verify that a morphism is $\frac{a}{b}$-power-free.

The basic idea is that we use the special form of the morphisms that arise to reduce the statement that $\varphi$ is $\frac{a}{b}$-power-free to a finite case analysis and then use software to carry out the case analysis. In the case of an individual rational number (as in Theorem 5) this is more or less straightforward using the results below. However, for parameterized morphisms that are symbolic in $a$ and $b$ (as in Theorem 4), a substantial amount of symbolic computation is required.

We use the following concept.

**Definition.** Let $k \geq 2$ and $\ell \geq 1$. Let $\varphi$ be $k$-uniform morphism on $\Sigma$. We say that $\varphi$ *locates words of length* $\ell$ if there exists an integer $j$ such that, for all words $w, x \in \Sigma^*$ with $|x| = \ell$, every occurrence of the factor $x$ in $\varphi(w)$ begins at a position congruent to $j$ modulo $k$. That is, the position of each $x$ is uniquely determined modulo $k$.

The notion of locating words of length $\ell$ is related to the concept of a *synchronization point* introduced by Cassaigne [5]. If $\varphi$ locates words of length $\ell$, then it locates words of length $\ell+1$, since if $|x| = \ell+1$ then the position of the length-$\ell$ prefix of $x$ is determined modulo $k$.

For example, the morphism $\varphi(n) = 000010(n+1)$, for which $\mathbf{w}_{5/3} = \varphi^\omega(0)$, locates words of length 4. Lemma 6 and Proposition 7 use the locating length to establish an upper bound on the length of factors of $\varphi(w)$ that need to be checked to see if they are $\frac{a}{b}$-powers.

**Lemma 6.** *Let $a, b$ be relatively prime positive integers such that $1 < \frac{a}{b} < 2$. Let $k \geq 2$ such that* $\gcd(b,k) = 1$, *and let $\ell \geq 1$. Let $\varphi$ be a $k$-uniform morphism on a (finite or infinite) alphabet $\Sigma$ such that*

- *$\varphi$ locates words of length $\ell$, and*
- *for all $n, n' \in \Sigma$, the words $\varphi(n)$ and $\varphi(n')$ differ in at most one position.*

*Then $w$ contains an $\frac{a}{b}$-power whenever $\varphi(w)$ contains an $\frac{a}{b}$-power $(xy)^{a/b} = xyx$ with $|x| \geq \ell$.*

**Proposition 7.** *Assume the hypotheses of Lemma 6. Let $I_{\min} \in \mathbb{Q}$ such that $1 < I_{\min} < \frac{a}{b}$, and let $c \geq d \geq 0$ be integers such that $\ell = ca - db$. Suppose additionally that for every $\frac{a}{b}$-power-free word $w \in \Sigma^*$ the word $\varphi(w)$ contains no $\frac{a}{b}$-power $(xy)^{a/b} = xyx$ of length $am$ for $m \leq \lfloor \frac{c \cdot I_{\min} - d}{I_{\min} - 1} \rfloor$. Then $\varphi$ is $\frac{a}{b}$-power-free.*

Given particular values of $\ell$ and $\frac{a}{b}$, we may have several choices of $c, d$, and of course there are many choices for $I_{\min}$. But we apply Proposition 7 to families of morphisms for which $I_{\min}, c, d$ are fixed.

In light of Proposition 7, it remains to automate the verification that, for any sequence $n_0, n_1, \ldots$ and for finitely many values of $m$, no word of length $am$ in $\varphi(n_0)\varphi(n_1)\cdots$ is an $\frac{a}{b}$-power. For each $m$ in range we slide a window of length $am$ through the word $\varphi(n_0)\varphi(n_1)\cdots$ and verify that no factor is an $\frac{a}{b}$-power. For a symbolic morphism parameterized by $a$ and $b$, this word is necessarily given in its run-length encoding, with symbolic block lengths in $a$ and $b$. Therefore we must be able to decide whether certain linear inequalities in $a$ and $b$ are true or false for $\frac{a}{b}$ restricted to a given interval. If an inequality is true for some $\frac{a}{b}$ in the interval and false for other $\frac{a}{b}$ in the interval, we restart the test with the interval broken into two subintervals at the point where equality holds.

# 4 Scope

The statements of Theorems 2–5, and many others that we have omitted here, were discovered by computing prefixes of $\mathbf{w}_{a/b}$ for 910 different rationals $1 < \frac{a}{b} < 2$. We have identified morphism-related structure in $\mathbf{w}_{a/b}$ for 520 of these 910 rationals. Not all of these words are fixed points of uniform morphisms, but they satisfy some recurrence

$$\mathbf{w}_{a/b}(ki + t) = \mathbf{w}_{a/b}(i) + d \tag{1}$$

for all $i \geq 0$, where there is a transient length $t$ before the self-similar behavior, governed by a $k$-uniform morphism, begins. The corresponding morphisms are $\frac{a}{b}$-power-free but not necessarily lexicographically least. The following is one of the more complex examples we have found.

**Theorem 8.** *Let $a, b$ be relatively prime positive integers such that $\frac{9}{7} < \frac{a}{b} < \frac{4}{3}$ and $\gcd(b, 24) = 1$. Then the $(24a - 15b)$-uniform morphism*

$$\begin{aligned}\varphi(n) = {} & 0^{a-b-1}\,1\,0^{2a-2b-1}\,1\,0^{-a+2b-1}\,1\,0^{2a-2b-1}\,1\,0^{a-b-1}\,1\,0^{-2a+3b-1}\,1\,0^{4a-5b-1}\,1 \\ & 0^{-a+2b-1}\,1\,0^{2a-2b-1}\,1\,0^{a-b-1}\,1\,0^{-2a+3b-1}\,1\,0^{-2a+3b-1}\,1\,0^{5a-6b-1}\,1 \\ & 0^{-2a+3b-1}\,1\,0^{4a-5b-1}\,1\,0^{a-b-1}\,1\,0^{-2a+3b-1}\,1\,0^{3a-3b-1}\,1\,0^{-2a+3b-1}\,1 \\ & 0^{a-b-1}\,1\,0^{-3a+4b-1}\,1\,0^{5a-6b-1}\,1\,0^{2a-2b-1}\,1\,0^{a-b-1}\,1\,0^{-2a+3b-1}\,1 \\ & 0^{3a-3b-1}\,1\,0^{-2a+3b-1}\,1\,0^{4a-5b-1}\,1\,0^{a-b-1}\,1\,0^{-2a+3b-1}\,1\,0^{2a-2b-1}\,2 \\ & 0^{a-b-1}\,1\,0^{-2a+3b-1}\,1\,0^{3a-3b-1}\,1\,0^{-2a+3b-1}\,1\,0^{a-b-1}\,1\,0^{a-b-1}\,(n+2)\end{aligned}$$

*locates words of length $5a - 4b$ and is $\frac{a}{b}$-power-free.*

Words of the form $\varphi^{\omega}(0)$ are $k$-regular when $\varphi$ is $k$-uniform, and the words $\mathbf{w}_{a/b}$ which aren't fixed points of morphisms but which satisfy Equation (1) are also $k$-regular, so a natural question is the following.

**Open question.** For which $1 < \frac{a}{b} < 2$ does there exist an integer $k$ such that $\mathbf{w}_{a/b}$ is $k$-regular?

## References


[1] Jean-Paul Allouche and Jeffrey Shallit, The ring of $k$-regular sequences, *Theoretical Computer Science* **98** (1992) 163–197.

[2] Jean-Paul Allouche and Jeffrey Shallit, The ring of $k$-regular sequences II, *Theoretical Computer Science* **307** (2003) 3–29.

[3] Jean-Paul Allouche and Jeffrey Shallit. *Automatic Sequences: Theory, Applications, Generalizations.* Cambridge University Press, Cambridge, 2003.

[4] Jean Berstel, Axel Thue's Papers on Repetitions in Words: a Translation, Number 20 in Publications du Laboratoire de Combinatoire et d'Informatique Math'ematique, Université du Québec à Montréal, February 1995.

[5] Julien Cassaigne, An algorithm to test if a given circular HD0L-language avoids a pattern, Information processing '94, Vol. I (Hamburg, 1994), 459–464, IFIP Trans. A Comput. Sci. Tech., A-51, North-Holland, Amsterdam, 1994.

[6] Mathieu Guay-Paquet and Jeffrey Shallit, Avoiding squares and overlaps over the natural numbers, *Discrete Mathematics* **309** (2009) 6245–6254.

[7] Eric Rowland and Jeffrey Shallit, Avoiding 3/2-powers over the natural numbers, *Discrete Mathematics* **312** (2012) 1282–1288.

[8] Axel Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana* **7** (1906) 1–22. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 139–158.

[9] Axel Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana* **1** (1912) 1–67. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 413–478.