ITERATES OF FUNCTIONS DEFINED IN TERMS OF DIGITAL REPRESENTATIONS OF THE INTEGERS

BY

JOHANN A. THIEL

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Bruce Berndt, Chair
Professor A. J. Hildebrand, Director of Research
Professor Bruce Reznick
Professor Emeritus Kenneth Stolarsky

Abstract

For a fixed base, John H. Conway's RATS sequences are generated by iterating the following procedure on an initial integer: Reverse the digits of the integer, Add the reversal to the original, Then Sort the resulting digits in increasing order. For example, 334+433=767, which gets sorted into 677. In base 10, Conway discovered the curious sequence: 12333334444, 55666667777, 123333334444, 556666667777, ... Although the sequence is not periodic, it does display some periodic-like behavior which we refer to as "quasiperiodic." Conway conjectured that all RATS sequences in base 10 are either eventually periodic, or they eventually lead to the previously mentioned quasiperiodic sequence. In this thesis, we study RATS sequences in various bases. In particular, we prove an Erdős-Kac type result for the periods of RATS sequences in base 3; we establish a connection between RATS sequences in general bases and Lyndon words; and we construct infinite families of bases for which there exist RATS sequences having certain prescribed periodicity properties, e.g., we show that there are infinitely many bases for which we can construct quasiperiodic RATS sequences all of a similar type. In the final chapter, we consider a similar iteration process, the reverse-add process. We present data and heuristic arguments on a problem of D.H. Lehmer asking whether every sequence obtained by this process contains a palindrome.

To Amber

Acknowledgments

First and foremost, I would like to thank my advisor, A. J. Hildebrand. Over the years he has spent countless hours helping me become a better mathematician. In the process, I have learned a great deal about good writing, problem solving, teaching, and Perl. Without a doubt, his great advice has helped me during my time as a graduate student. I would also like to thank Bruce Berndt, Bruce Reznick, and Kenneth Stolarsky for all of their helpful advice and suggestions while on my thesis committee.

Many thanks go to the friends I met while in graduate school. Bart Snapp, Jesse Beder, Will Green, Tim LeSaulnier and the entirety of office 178 in Altgeld Hall have helped keep me in good spirits, even during the most stressful times.

I would also like to acknowledge the support from National Science Foundation grant DMS 0838434 "EMSW21MCTP: Research Experience for Graduate Students."

Lastly I would like to thank my family. I could not have finished what I started without the endless support of my parents, my sister, and my loving wife.

Table of Contents

Chapter 1

Introduction

> Most problems involving the digits of numbers are recreative rather than deep. In fact it is hard to imagine a problem in Digitology which could not be solved in a finite number of steps or trials. [...] It is always dangerous to comment upon the difficulty of an unsolved problem but I believe that it stands apart from the common digital problems and to such a degree that it is a challenge not only to Digitologists but to any mathematician.
>
> –*D.H. Lehmer*

In 1949, D.R. Kaprekar [9] created a simple process with an unusual outcome. Begin with a positive integer written in base 10 containing four digits (not all the same), compute the difference between the largest and the smallest integers that can be formed with those four digits, and repeat. In iterating this process, he found that all such four digit numbers eventually gave the same result, 6174, a fixed point of the process.

Example We apply Kaprekar's process beginning with 4785.

$$\begin{array}{rr} 8754 & 7641 \\ -\ 4578 & -\ 1467 \\ \hline 4176 & 6174 \end{array}$$

In this particular case, it only takes two iterations of the Kaprekar process to arrive at 6174.

In generalizing the Kaprekar process to longer digits and other bases, other numbers have appeared with somewhat similar properties to 6174. For example, every positive integer written in base 10 containing three digits (not all the same) always leads to 495 under enough iterations of the Kaprekar process. However, in general, we do not have that all m digit long numbers converge to some fixed point of the Kaprekar process, some numbers, like 53955, give rise to periodic behavior.

***

John H. Conway has invented many mathematical games, including the well known Game of Life. In 1990, Conway [1] described a new game inspired by the Kaprekar process that he dubbed Reverse-Add-Then-Sort (RATS). The rules are simple: begin with any positive integer n written in base 10, add to it the number obtained by reversing the order of the digits of n, then sort the result in order of increasing digits from left to right (discarding any zeros), and repeat. The game stops if the sequence generated by iterating this process ever repeats a term, and therefore becomes periodic.

Example We play the RATS game beginning with $n = 9$.

$$\begin{array}{r} 9 \\ +\ 9 \\ \hline 18 \end{array} \quad \begin{array}{r} 18 \\ +\ 81 \\ \hline 99 \end{array} \quad \begin{array}{r} 99 \\ +\ 99 \\ \hline 198 \end{array} \quad \begin{array}{r} 189 \\ +\ 981 \\ \hline 1170 \end{array} \quad \begin{array}{r} 117 \\ +\ 711 \\ \hline 828 \end{array} \quad \begin{array}{r} 288 \\ +\ 882 \\ \hline 1170 \end{array}$$

In this particular case, the game ends after six iterations and yields the sequence

$$9 \to 18 \to 99 \to 189 \to 117 \to 288 \to 117 \to \cdots, \tag{1.1}$$

with the terms 288 and 117 forming a cycle with period 2.

Conway discovered that the RATS sequence generated by starting with 1, referred to as Conway's sequence, eventually leads to the following monotonically increasing sequence

$$\cdots \to 12333333334444 \to 55666666667777 \to 123333333334444 \to 556666666667777 \to \cdots, \tag{1.2}$$

and conjectured that this sequence is unique in the following sense:

Conjecture (Conway's Conjecture, [1]). *Every RATS sequence is either eventually periodic or eventually part of Conway's sequence* (1.2).

Curtis McMullen [7] has verified Conway's conjecture up to 100,000,000. In the process, he has also discovered infinite families of numbers that generate periodic RATS sequences with periods 1–24.

Despite the fact that Conway's sequence is not eventually periodic, it does seem to exhibit a pattern that almost repeats every other term. The only difference between every other term in the sequence is that a fixed digit increases in appearance by one. We will refer to this type of behavior as quasiperiodic. (See Definition 2.5 for a precise definition.)

Base 10 is a natural starting point for the RATS game, as it coincides well with the arithmetic most individuals are used to; however, there is nothing particularly special about base 10 that prevents the game from being played in other bases. In this thesis, we will investigate the behavior of RATS sequences for general bases. Questions we are interested in include:

- Can we find periodic RATS sequences in base 10 for any desired period?
- Are there other quasiperiodic RATS sequences in base 10 besides (1.2)? Conway's conjecture implies there are none.
- Are there other bases where we can generate quasiperiodic sequences? If so, what kind of quasiperiods can we find?
- Given a period p, are there infinitely many bases for which there exist periodic RATS sequences with period p?
- Are there bases for which there exist periodic RATS sequences for all sufficiently large periods?
- Do any RATS sequences exhibit behavior which is neither eventually periodic nor quasiperiodic? Conway's conjecture implies that the answer is no for base 10.

We will not provide answers to all of these questions, as some remain unsolved. They are simply meant as examples to motivate our work.

Some of these questions have been investigated in the literature. (See [8], Section F32, for a list of results and related problems.) The current state of the art is summarized by the following results:

- In base 10, there are periodic RATS sequences for any desired period ≥ 2. (Cooper–Kennedy, [2])
- In base 3, all RATS sequences are eventually periodic. (Gentges, [6])
- Quasiperiodic RATS sequences exist in all bases b satisfying $b \equiv 1 \bmod 9$. (Cooper–Shattuck, [13], Gentges, [6])
- If q is a prime or a Fermat pseudoprime, then there exists a quasiperiodic RATS sequence with quasiperiod q with respect to the base $b = (2^q - 1)^2 + 1$. (Cooper–Shattuck, [13])

Using the above questions as a starting point, and expanding on the above results, we obtain the following new results:

- **Periodic RATS sequences with period 1 can only occur in base 2. (Theorem 3.2)**

- **There are no quasiperiodic RATS sequences with period 1 in any base. (Theorem 6.13)** Previously, bases 2, 3, and 10 were the only known examples for which this was known.
- **In base 3, the period of a "random" integer is approximately $\frac{1}{2}\log_2 v$, where v denotes the number of times the digits 2 appears in the base 3 representation. (Corollary 4.14, Theorem 4.16)**
- **There is an infinite family of bases for which there exist periodic RATS sequences for all arbitrarily large periods. (Theorem 6.1)** Previously, bases 3 and 10 were the only known examples of bases for which one could exhibit periodic RATS sequences for an infinite number of distinct periods. We extend this result to an infinite family of bases.
- **For every $q > 2$, there is an infinite family of bases for which there exist quasiperiodic RATS sequences with quasiperiod q. (Theorem 6.4)** Previously, the existence of quasiperiodic RATS sequences with a given quasiperiod q had been only known for prime and pseudoprime values of q. Also, no single value $q \geq 7$ had been known for which there exist quasiperiodic RATS sequences with respect to multiple bases.
- **There are no quasiperiodic RATS sequences for any base of the form $b = 2^t + 1$ for $t \geq 0$. (Corollary 6.19)** This is the first such nonexistence result for bases > 10. We will prove this, and some related results, by establishing a connection between quasiperiodic RATS sequences and so-called Lyndon words, a concept from the theory of combinatorics on words. (Section 6.4)

The list above is not all inclusive. It is meant as a brief summary of the kinds of results that we have been able to obtain. Our results suggest that the following vast generalization of Conway's conjecture must be true:

Conjecture. *In any base $b \geq 2$, every RATS sequence is either eventually periodic or eventually quasiperiodic.*

***

Fifty-two years prior to Conway's RATS game, D.H. Lehmer [11] wrote about a very similar topic. He considered the following process: begin with any positive integer n written in base 10, add to it the number obtained by reversing the order of the digits of n, and repeat. This is the RATS process minus the "then-sort" portion.

Example Lehmer's "reverse-add" process for $n = 9$.

$$
\begin{array}{rrrrrr}
9 & 18 & 99 & 198 & 1089 & 10890 \\
+\ 9 & +\ 81 & +\ 99 & +\ 891 & +\ 9801 & +\ 09801 \\
\hline
18 & 99 & 198 & 1089 & 10890 & 20691
\end{array}
$$

In this particular case, the first six iterations yield the sequence

$$9 \to 18 \to 99 \to 198 \to 1089 \to 10890 \to 20691 \to \cdots . \tag{1.3}$$

Lehmer was interested in panlindromic numbers (numbers whose digital representations are the same forwards and backwards) generated by this process. In the example above, we see a palindrome in the first six iterations of the process. Lehmer noted that all one and two digit numbers always lead to some palindrome after enough iterations. This led him to ask the following question: Does every integer eventually lead to a palindrome upon iterating the reverse-add process enough times?

Lehmer's computations suggest that this is the case, with one notable exception. After computing the first 73 iterates of 196, Lehmer noticed that no iterate up to that point was a palindrome. Later computations by Yamashita [17] expanded on Lehmer's work and showed that the first 1000 iterates of 196 were not palindromes. Further computational efforts by a Walker [16] and VanLandingham [15] showed that the first $725,000,000$ iterates of 196 yield no palindromes. In particular, VanLandingham showed that if some iterate of 196 is palindromic, the palindrome formed will be at least $300,000,000$ digits long!

In Chapter 7 we will formalize this problem. We will also give more computational evidence and establish conjectures that suggest that the answer to Lehmer's question seems to be a resounding "no."

Chapter 2

The RATS process in base 10

2.1 Notation and terminology

In order to study the RATS game, we are motivated to formalize the rules mathematically. Using the definition below, we will translate the RATS game into the language of discrete dynamical systems.

Definition 2.1. *For any positive integer n written in base 10, let $\overline{n}$ be the digit formed by reversing the digits of n, and let n' be the digit formed by ordering the nonzero digits of n in increasing order from left to right. Define the function $R\colon \mathbb{N} \to \mathbb{N}$ by $R(n) = (n + \overline{n})'$.*

The function R represents a single iteration of the RATS process. We are interested in the behavior of the sequence obtained by repeatedly applying R to whatever output it generates. To denote iterates of R, we adopt superscript notation. For any nonnegative integer t, let $R^t(n) = R(R^{t-1}(n))$ if $t \geq 1$ and $R^0(n) = n$.

Definition 2.2. *We call $\{R^i(n)\}_{i=0}^{\infty}$ the **RATS sequence generated by** n.*

Example

$$\begin{aligned} R^2(845363) &= R((845363 + 363548)') \\ &= R((1208911)') \\ &= R(111289) \\ &= (111289 + 98211)' \\ &= (109500)' \\ &= 159 \end{aligned}$$

Definition 2.3. *If $R^p(n) = n$ for some $p > 0$, we say that n and the RATS sequence generated by n are **periodic**. If p is the least integer with this property, we say that the **period** is p. If $R^t(n)$ is periodic for some $t \geq 0$, we say that n and the RATS sequence generated by n are **preperiodic**[1].*

[1] In [7], periodic sequences are referred to as "cycles" and preperiodic sequences are "tributaries" to a cycle.

Example The sequence (1.1) shows that 117 and 288 are periodic with period 2 and 9 is preperiodic with period 2.

From the definitions above, we see that all periodic sequences are preperiodic[2], but not vice versa. Also note that in the definition of periodic elements n, we only require that $R^p(n) = n$ for *some* p. This is because if $R^p(n) = n$, then $R^{t+p}(n) = R^t(R^p(n)) = R^t(n)$ for all $t \geq 0$; that is, the sequence $\{R^t(n)\}_{t=0}^{\infty}$ is periodic, in the usual sense, with period p.

Suppose that a subsequence of the RATS sequence generated by n is bounded by some fixed constant. Then, by the pigeonhole principle, there must be some pair $i, j \geq 0$, with $i < j$, such that $R^i(n) = R^j(n)$. It follows that the RATS sequence generated by n is eventually periodic, i.e., preperiodic. Hence we obtain that RATS sequences can behave in one of two ways.

Proposition 2.4. *The RATS sequence generated by n, $\{R^i(n)\}_{i=0}^{\infty}$, is preperiodic, or $R^i(n) \to \infty$ as $i \to \infty$.*

For convenience of notation, we will adopt the exponential notation used by Guy [7] when necessary. In exponential notation, the base denotes the digit and the exponent denotes the number of times that digit appears; e.g., 118110005 will be denoted as $1^3 8^2 1^2 0^3 5$.

With this notation we can rewrite Conway's sequence (1.2) as

$$\cdots \to 123^8 4^4 \to 5^2 6^8 7^4 \to 123^9 4^4 \to 5^2 6^9 7^4 \to \cdots . \tag{2.1}$$

We mentioned in the previous chapter that this sequence displays a kind of periodic behavior we referred to as quasiperiodic. The following definition clarifies what what we mean by this.

Definition 2.5. *If $n = 1^{m_1} \cdots (k-1)^{m_{k-1}} k^{m_k} (k+1)^{m_{k+1}} \cdots 9^{m_9}$ and $R^{qt}(n) = 1^{m_1} \cdots (k-1)^{m_{k-1}} k^{m_k+t} (k+1)^{m_{k+1}} \cdots 9^{m_9}$, with $m_i \geq 0$ for $1 \leq i \leq 9$, for some $q > 0$ and all $t \geq 0$, we say that n and the RATS sequence generated by n are **quasiperiodic**. If q is the least integer with this property, we say that the **quasiperiod** is q. The digit k which increases in count after q iterations is called the **growing digit**. If $R^t(n)$ is quasiperiodic for some $t \geq 0$, we say that n and the RATS sequence generated by n are **prequasiperiodic**[3].*

The particular form of the definition of "quasiperiodic" may seem arbitrary. This definition is motivated by the fact that Conway's sequence and all other known nonpreperiodic sequences are of this form.

Example From (2.1) we find that, for $m \geq 3$, $123^m 4^4$ and $5^2 6^m 7^4$ are quasiperiodic elements with quasiperiod 2 and growing digits 3 and 6, respectively. Also, the full Conway sequence (2.1), starting at 1, is prequasiperiodic.

[2]This goes against convention, but is designed to avoid having to say "periodic and/or preperiodic" repeatedly.

[3]In [13], quasiperiodic sequences are called "divergent" and the quasiperiod is called the "length."

Conway's conjecture asserts that every RATS sequence is either preperiodic or eventually part of (2.1). It may seem that, in light of this conjecture, the above definitions are rather useless. If one believes that Conway is correct, then there is essentially only one quasiperiodic RATS sequence (with quasiperiod 2). The definitions for (pre)quasiperiodic sequences, however, will be useful when considering the RATS game in more general bases.

2.2 Periodic RATS sequences

In this section, we describe the periods that arise from the RATS game.

Definition 2.6. *Let $\mathcal{P}$, called the **period set**, be the set of all p for which there are periodic elements with period p.*

From the first example in Chapter 1, the sequence $117 \to 288 \to 117 \to \cdots$ shows that 117 is periodic with period 2, so $2 \in \mathcal{P}$. Cooper and Kennedy were able to completely characterize the period set $\mathcal{P}$. For completeness, we give a proof of this below.

Theorem 2.7 (Cooper–Kennedy, [2]). *We have $\mathcal{P} = \{p : p \geq 2\}$. That is, for every $p \geq 2$, there exists a periodic RATS sequence with period p, and there does not exist a periodic RATS sequence with period 1.*

Proof. We begin by showing that $1 \notin \mathcal{P}$. Suppose, by contradiction, that there is some periodic element n with period 1. Then $R(n) = (n + \overline{n})' = n$. Note that the digits of $(n + \overline{n})'$, and hence those of n, are in increasing order. Hence n is of the form $s* \cdots *t$, where s and t are the smallest and largest digits of n, respectively, and the asterisks denote unknown or unspecified digits. If $s \geq 2$, we must have that $s + t \leq 9$, for otherwise a carry occurs when computing $n + \overline{n}$, causing the leftmost digit of $R(n) = n$ to be a 1, a contradiction to the assumption that $s \geq 2$. On the other hand, if $s + t \leq 9$, then the rightmost digit of $R(n) = n$ is equal to $s + t$, a contradiction to the definition of t. Therefore, we must have that $s = 1$. If $t \leq 8$, then $1 + t$ is the largest digit of $R(n) = n$, which again contradicts the definition of t. Therefore, we must have that $t = 9$.

Thus, n is of the form

$$n = 1^{m_1} 2^{m_2} 3^{m_3} 4^{m_4} 5^{m_5} 6^{m_6} 7^{m_7} 8^{m_8} 9^{m_9}$$

with $m_1, m_9 \geq 1$. Since $m_1, m_9 \geq 1$, n must be of the form $1* \ldots *9$. Then $n + \overline{n}$ can be calculated as

$$\begin{array}{r} 1* \ldots *9 \\ +9* \ldots *1 \\ \hline 1x* \ldots *0 \end{array}$$

with $x \in \{0, 1\}$. If $x = 0$, then $R(n)$ has fewer digits than n, a contradiction to the assumption that $R(n) = n$. (Recall that zero digits are discarded when computing $R(n)$.) Therefore, $x = 1$ and $m_1 \geq 2$. With this additional information, the sum $n + \overline{n}$ takes the form

$$\begin{array}{r} 11* \ldots *y9 \\ +\ 9y* \ldots *11 \\ \hline 11** \ldots *z0 \end{array}$$

with $y \in \{8, 9\}$. If $y = 8$, then $z = 0$, which also gives that $R(n)$ has fewer digits than n, a contradiction. Therefore, $y = 9$ and $m_9 \geq 2$.

By induction, it follows that $m_1, m_9 \geq k$ for any positive integer k, a contradiction. Therefore, $1 \notin \mathcal{P}$.

That every integer $p \geq 2$ occurs as a period follows from the following results, which can be verified by direct calculations (see Cooper–Kennedy, [2]).

- The integers 117, $1^6 7^{15}$, and $1^4 2^6 5^2$ are periodic with periods 2, 4, and 6, respectively.
- For any odd integer $p \geq 3$, the integer $1^{6 \cdot 2^{p-3}} 3^{2 \cdot 2^{p-3} - 1}$ is periodic with period p.
- For any even integer $p \geq 8$, the integer $1^{216 \cdot 2^{p-8}} 3^{40 \cdot 2^{p-8} - 1}$ is periodic with period p.

The desired result follows immediately. □

2.3 Quasiperiodic RATS sequences

In the same spirit as Section 2.2, we look to describe the quasiperiods that arise from the RATS game.

Definition 2.8. *Let $\mathcal{Q}$, called the* ***quasiperiod set****, be the set of all q for which there are quasiperiodic elements with quasiperiod q.*

Conway's sequence (see (2.1)) gives that $2 \in \mathcal{Q}$. Unlike the case for $\mathcal{P}$, we do not have a complete description of $\mathcal{Q}$. However, if Conway's conjecture is true, then Conway's sequence is, up to finitely many terms, the only quasiperiodic sequence, so $\mathcal{Q} = \{2\}$. In its full strength, this remains open, but we will show that $\mathcal{Q}$ can only contain even numbers (see Proposition 2.11).

Suppose that n is quasiperiodic with quasiperiod q and growing digit k. Then, given $\epsilon > 0$, there is a $t_0 := t_0(\epsilon)$ such that the ratio of the sum of all exponents to the exponent of k in $R^t(n)$ is smaller than $1+\epsilon$ if $t \geq t_0$ and $t \equiv 0 \bmod q$. This is due to the fact that for any positive integer $t \equiv 0 \bmod q$, $R^t(n)$ has fixed exponents for all digits different from k, while the exponent of k goes to infinity as $t \to \infty$.

It follows that the most popular digit of $R^t(n)$, with $t \equiv 0 \bmod q$ and t large enough, is k. In fact, choosing ϵ small enough, we can arrange that, in addition, the most popular digit of $R^{t+1}(n)$ is $2k$ if $k < 5$, and $2k-9$ if $k \geq 5$. Since all of the exponents of $R^t(n)$, save that of k, are fixed, it follows that, for t large enough, $R^{t+1}(n)$ is quasiperiodic with quasiperiod q and growing term $2k$ or $2k-9$.

Lemma 2.9. *Let $A = \{1, 2, 4, 5, 7, 8\}$. If there exists a quasiperiodic integer with growing digit k for some $k \in A$, then there exists a quasiperiodic integer with growing digit k for any $k \in A$. Moreover, the quasiperiod of any integer with growing digit in the set A is a multiple of* 6.

Proof. If there exists a quasiperiodic integer n with growing digit 1 and quasiperiod q, then by the discussion preceding the statement of the lemma, for $t \equiv 1 \bmod q$ with t large enough, $R^t(n)$ is quasiperiodic with growing digit 2. Repeating this argument five more times yields the existence of quasiperiodic elements with growing digits 4, 8, $16-9=7$, and $14-9=5$. Hence if 1 occurs as a growing digit of a quasiperiodic integer, then so do 2, 4, 5, 7, and 8. The same argument can be made with any value $k \in A$. Furthermore, the above argument shows that it takes six iterations of the RATS process to return to the starting growing digit. Hence the quasiperiod is a multiple of 6. ∎

Lemma 2.10. *Suppose that n is quasiperiodic with growing digit k. Then $k \neq 9$.*

Proof. Let q denote the quasiperiod of n. Suppose that $k = 9$ and that $n = 1^{m_1}2^{m_2}\ldots 8^{m_8}9^{m_9}$ with $m_9 > 0$. First assume that $m_1 > 0$. Without loss of generality, we can assume that $m_9 > M := \sum_{i=1}^{8} m_i > 0$. Then there exists a positive integer t such that $R^t(n) = 1^{m_1}2^{m_2}\ldots 8^{m_8}9^{2^q m_9}$. Then $R^t(n) + \overline{R^t(n)}$ can be calculated as

$$\begin{array}{r} 1\ldots12\ldots23\ldots34\ldots45\ldots56\ldots67\ldots78\ldots89\ldots99\ldots99\ldots99\ldots99\ldots99\ldots99\ldots99\ldots99\ldots99 \\ +\ 9\ldots99\ldots99\ldots99\ldots99\ldots99\ldots99\ldots99\ldots99\ldots98\ldots87\ldots76\ldots65\ldots54\ldots43\ldots32\ldots21\ldots11 \\ \hline 11\ldots12\ldots23\ldots34\ldots45\ldots56\ldots67\ldots78\ldots89\ldots98\ldots87\ldots76\ldots65\ldots54\ldots43\ldots32\ldots21\ldots10 \end{array}$$

which shows that $R^{t+1}(n) = 1^{2m_1}2^{2m_2}\ldots 8^{2m_8}9^{2^q m_9 - M}$. Continuing in this way shows that $R^{t+q}(n) = 1^{2^q m_1}2^{2^q m_2}\ldots 8^{2^q m_8}9^{2^q m_9-(2^q-1)M}$, a contradiction to the assumption that n is quasiperiodic with quasiperiod q.

In the case where $m_1 = 0$, notice that the above sum would give that $R(n)$ has at least one 1 and has growing digit 9. Arguing the same way with $R(n)$ in place of n yields the same contradiction to the assumption that n is quasiperiodic with quasiperiod q. So $k \neq 9$. □

Proposition 2.11. *If $q \in \mathcal{Q}$, then q is even.*

Proof. If n is quasiperiodic with growing digit k and quasiperiod q, then, by Lemma 2.10, we have that $k \neq 9$. If $k \in \{1, 2, 4, 5, 7, 8\}$, from Lemma 2.9 it follows that $6|q$. Similarly, if $k \in \{3, 6\}$, then $2|q$. □

One way to make further progress towards Conway's conjecture would be to show that, indeed, $\mathcal{Q} = \{2\}$. This is, however, not enough to prove the conjecture, since the possibility exists that there are elements that are neither preperiodic nor prequasiperiodic, or that there are multiple disjoint sequences with quasiperiod 2. The difficulty in showing that certain quasiperiods are not in $\mathcal{Q}$ is that they require dealing with a tremendous number of cases. Even showing that there are no quasiperiodic RATS sequences, with quasiperiod 2, that do not intersect with Conway's sequence seems difficult.

2.4 Extension to general bases

In this section we begin by generalizing the RATS game, and related definitions, to general integer bases.

Definition 2.12. *Fix an integer $b \geq 2$. For any positive integer n written in base b, let $\overline{n} := \overline{n_b}$ be the digit formed by reversing the digits of n, and let $n' := n'_b$ be the digit formed by ordering the digits of n in increasing order from left to right. Define the function $R_b \colon \mathbb{N} \to \mathbb{N}$ by $R_b(n) = (n + \overline{n})'$.*

Example

$$\begin{aligned} R_3(112) &= (112 + 211)' \\ &= 1100 \\ &= 11 \end{aligned}$$

Definition 2.13. *We call $\{R_b^i(n)\}_{i=0}^{\infty}$ the **RATS sequence generated by n in base b**.*

As before, we are interested in the behavior of RATS sequences in base b. We extend the definitions of (pre)periodic, (pre)quasiperiodic, the period set, and the quasiperiod set (the latter two denoted as $\mathcal{P}_b$ and $\mathcal{Q}_b$, respectively) to RATS sequences in base b, in the obvious way.

Not every current result for base 10 RATS sequences can be extended to general bases. There are examples of bases where $|\mathcal{P}_b| < \infty$, as well as examples where $\mathcal{Q}_p = \emptyset$, and even cases where $|\mathcal{Q}_p| > 1$. In the chapters that follow, we will study RATS sequences for general bases using the questions and results raised in this chapter, and the previous, as motivation.

In subsequent chapters, the subscript b denoting the choice of base will be suppressed whenever its value is clear from the context.

Chapter 3

The RATS process in base 2

3.1 Periodicity

When studying RATS sequences for general bases, it seems natural to start with base 2. Due to the low number of digits available to represent positive integers, one could reasonably expect that it should be easier to determine the behavior of any RATS sequence. Indeed, this is the case, as the following theorem shows.

Theorem 3.1. *(Guy, [7]) Every RATS sequence in base* 2 *is preperiodic and* $\mathcal{P}_2 = \{1\}$.

Proof. Let n be a positive integer written in base 2. Since the RATS process eliminates zeros after sorting, we have that $R(n) = 1^m$ for some $m > 0$. This is true for any n, so to understand RATS sequences in base 2, it is enough to consider sequences generated by integers of the form 1^m. A direct calculation gives

$$\begin{array}{r} 11\ldots11 \\ +\ 11\ldots11 \\ \hline 111\ldots10 \end{array}$$

As the sum above shows, $R(1^m) = 1^m$ for any $m > 0$. So it follows that, in base 2, every RATS sequence is preperiodic with period 1. □

What is interesting about Theorem 3.1 is that one can show that periodic RATS sequences of period 1 can only occur in base 2.

Theorem 3.2. *We have that* $1 \in \mathcal{P}_b$ *if and only if* $b = 2$.

Proof. (⇐) This direction follows from Theorem 3.1.

(⇒) If $b > 2$, then one can show that $1 \notin \mathcal{P}_2$ using the same argument as in Theorem 2.7. We omit the details. □

3.2 Inverse images

With a complete characterization of the behavior of RATS sequences in base 2, we turn our attention to inverse images of the map R. It is clear that, for any $m > 1$, the set $R^{-1}(1^m)$ is infinite. With this in mind, we seek to count the number of integers in $R^{-1}(1^m)$ containing a fixed number of digits in their base 2 representation.

Definition 3.3. *For any positive integers k and m with $k \geq m$, let*

$$I(k,m) = \#\{R^{-1}(1^m) \cap [10^{k-1}, 1^k]\}.$$

In other words, $I(k,m)$ counts the number of positive integers with k digits in their base 2 representation that get mapped to 1^m after one iteration of the RATS process. (Note: In this definition, 10^{k-1} is to be interpreted as the integer with base 2 representation $1\underbrace{0\ldots0}_{k-1}$.)

Proposition 3.4. *Let $k > 5$ be an integer.*

(i) $I(k,2) = \begin{cases} 2 & \text{if } k \text{ is odd,} \\ 2^{(k-2)/2} + 2 & \text{if } k \text{ is even.} \end{cases}$

(ii) $I(k,3) = \begin{cases} 2^{(k-3)/2} + 5 & \text{if } k \text{ is odd,} \\ 2^{(k-4)/2} + 3 & \text{if } k \text{ is even.} \end{cases}$

(iii) $I(k,k-1) = \begin{cases} 2^{(k-3)/2} + 2 & \text{if } k \text{ is odd,} \\ 1 & \text{if } k \text{ is even.} \end{cases}$

(iv) $I(k,k) = \begin{cases} 1 & \text{if } k \text{ is odd,} \\ 2^{(k-2)/2} + 1 & \text{if } k \text{ is even.} \end{cases}$

Proof. The results can be proved by an exhaustive examination of cases. We will give the details for (i). The cases (ii)–(iv) can be proved similarly.

We use the convention that the jth digit, from left to right, of the base 2 representation of an integer a will be denoted as a_j. That is, if a has k digits in its base 2 representation, we write $a = 1a_2a_3\ldots a_k$ with $a_i \in \{0,1\}$.

Case 1. **k is odd.** Let a be an integer with k digits in base 2 such that $R(a) = 1^2$. Let $k_0 = (k-1)/2$ and $y = a_{k_0+1}$. Thus, y is the "middle" digit of a.

Case 1a. $a_k = 0$ **and** $a + \overline{a}$ **is** k **digits long.** Then a has the form $a = 1* \ldots *y* \ldots *0$. Since k is odd, the digit y will line up against itself in the sum $a + \overline{a}$. We then compute $a + \overline{a}$ as follows.

$$\begin{array}{rl} 1* \ldots *y* \ldots *0 & \longleftarrow a \\ +\ 0* \ldots *y* \ldots *1 & \longleftarrow \overline{a} \\ \hline 1* \ldots *** \ldots *1 & \longleftarrow a + \overline{a} \end{array}$$

The above sum shows that $(a + \overline{a})_k = 1$. Since $R(a) = 1^2$, the digit 1 can only appear twice in the base 2 representation of $a + \overline{a}$. Thus, we have that

$$(a + \overline{a})_i = 0 \tag{3.1}$$

for $1 < i < k$; that is, the above sum must be of the form

$$\begin{array}{rl} 1* \ldots *y* \ldots *0 & \longleftarrow a \\ +\ 0* \ldots *y* \ldots *1 & \longleftarrow \overline{a} \\ \hline 10 \ldots 000 \ldots 01 & \longleftarrow a + \overline{a} \end{array}$$

We will show that we necessarily have $a_i = 0$ for $1 < i < k$. Suppose that a is not of this form. Let $1 < k_1 < k$ be the smallest integer such that either $a_{k_1} = 1$ or $a_{k-k_1+1} = 1$. That is, k_1 is the second column in the above sum, from both the right and the left, to have any 1's appearing in either summand. This leads to three possibilities.

Case $a_{k_1} \neq a_{k-k_1+1}$. As the two sums below show, in this case we have $(a+\overline{a})_{k-k_1+1} = 1$, contradicting (3.1) for $i = k - k_1 + 1$.

$$\begin{array}{r} 10 \ldots 01* \ldots *y* \ldots *00 \ldots 00 \\ +\ 00 \ldots 00* \ldots *y* \ldots *10 \ldots 01 \\ \hline 1* \ldots *** \ldots *** \ldots *\mathbf{1}0 \ldots 01 \end{array} \qquad \begin{array}{r} 10 \ldots 00* \ldots *y* \ldots *10 \ldots 00 \\ +\ 00 \ldots 01* \ldots *y* \ldots *00 \ldots 01 \\ \hline 1* \ldots *** \ldots *** \ldots *\mathbf{1}0 \ldots 01 \end{array}$$

Therefore, this case cannot happen.

Case $a_{k_1} = a_{k-k_1+1} = 1$ *and* $k_1 = 2$. As the sum below shows, the carry from the 2nd column of the sum causes $a+\overline{a}$ to be $k+1$ digits long, a contradiction to our initial assumption that $a+\overline{a}$ is k digits long.

$$\begin{array}{rl} 1 & \leftarrow \text{carries} \\ 11*\ldots*y*\ldots*10 & \\ +\ 01*\ldots*y*\ldots*11 & \\ \hline \mathbf{1}0**\ldots***\ldots*01 & \end{array}$$

Therefore, this case is also impossible.

Case $a_{k_1} = a_{k-k_1+1} = 1$ *and* $k_1 > 2$. As the sum below shows, the carry from the k_1th column of the sum gives that $(a+\overline{a})_{k_1-1} = 1$, a contradiction of (3.1) for $i = k_1 - 1$.

$$\begin{array}{r} 1 \\ 10\ldots01*\ldots*y*\ldots*10\ldots00 \\ +\ 00\ldots01*\ldots*y*\ldots*10\ldots01 \\ \hline 1*\ldots\mathbf{1}**\ldots***\ldots**0\ldots01 \end{array}$$

Therefore, this case cannot happen.

Having excluded all alternatives, we conclude that $a_i = 0$ for $1 < i < k$.

Case 1b. $a_k = 0$ **and** $a+\overline{a}$ **is** $k+1$ **digits long.** Again, a has the form $a = 1*\ldots*y*\ldots*0$. We compute the sum $a+\overline{a}$ as

$$\begin{array}{r} 1*\ldots*y*\ldots*0 \\ +\ 0*\ldots*y*\ldots*1 \\ \hline 10*\ldots***\ldots*1 \end{array}$$

where (3.1) holds for $1 < i < k$ by the same reasoning as in Case 1a.

Suppose that $a_2 \neq a_{k-1}$. Then, as the sums below show, $(a+\overline{a})_k = 1$, a contradiction of (3.1) for $i = k$, so $a_2 = a_{k-1}$. This leads to two possible cases.

$$\begin{array}{r} 11\ldots*y*\ldots00 \\ +\ 00\ldots*y*\ldots11 \\ \hline 10*\ldots***\ldots\mathbf{1}1 \end{array} \qquad \begin{array}{r} 10\ldots*y*\ldots10 \\ +\ 01\ldots*y*\ldots01 \\ \hline 10*\ldots***\ldots\mathbf{1}1 \end{array}$$

Case $a_2 = a_{k-1} = 0$. Then, as the sum below shows, the lack of a carry from the 2nd column implies that $a+\overline{a}$ cannot be $k+1$ digits long, a contradiction to our initial assumption that $a+\overline{a}$ is $k+1$ digits long.

$$\begin{array}{r} 10* \ldots *y* \ldots *00 \\ +\ 00* \ldots *y* \ldots *01 \\ \hline 1** \ldots *** \ldots *01 \end{array}$$

Therefore, this case is impossible.

Case $a_2 = a_{k-1} = 1$. In this situation, the $(k-1)$th column in the sum produces a carry, as the sum below shows.

$$\begin{array}{r} 1 \\ 11* \ldots *y* \ldots *10 \\ +\ 01* \ldots *y* \ldots *11 \\ \hline 10** \ldots *** \ldots *01 \end{array}$$

In order to avoid a contradiction to (3.1) for $i = k-1$, we must have that $a_3 \neq a_{k-2}$. That is, we have one of two sums below.

$$\begin{array}{r} 1 \\ 110* \ldots *y* \ldots *110 \\ +\ 011* \ldots *y* \ldots *011 \\ \hline 10*** \ldots *** \ldots *001 \end{array} \qquad \begin{array}{r} 1 \\ 110* \ldots *y* \ldots *010 \\ +\ 011* \ldots *y* \ldots *111 \\ \hline 10*** \ldots *** \ldots *001 \end{array}$$

Continuing with this line of reasoning, we see that, in order to avoid a contradiction to (3.1) for $2 < i < k_0 + 1$, we must have that $a_i \neq a_{k-i+1}$ for $2 < i < k_0 + 1$. In particular, the sum will introduce a carry in the $(k_0 + 1)$th column, the column associated with the middle digit y. This gives that $(a + \overline{a})_{k_0+1} \equiv 1 + y + y \bmod 2 \equiv 1 \bmod 2$, contradicting (3.1) for $i = k_0 + 1$. Therefore, Case 1b cannot occur.

Case 1c. $a_k = 1$**.** Then a has the form $a = 1* \ldots *y* \ldots *1$. We compute the sum $a + \overline{a}$ as

$$\begin{array}{r} 1 \\ 1* \ldots *y* \ldots *1 \\ +\ 1* \ldots *y* \ldots *1 \\ \hline 1v* \ldots *z* \ldots *0 \end{array}$$

where $v, z \in \{0, 1\}$. Since the digit 1 can appear only twice in the base 2 representation of $a + \overline{a}$, to avoid a contradiction to the assumption that $R(a) = 1^2$, we have that at least one of the two variables v or z must be 0.

If $v = 1$, then from the above comment, we immediately find that $z = 0$. Furthermore, arguing as in Case 1b, it follows that this is impossible. Therefore, we must have $v = 0$.

We will show that we necessarily have that $a_2 = a_{k-1} = 0$. Suppose that this is not the case, that is, either $a_2 = a_{k-1} = 1$ or $a_2 \neq a_{k-1}$.

Case $a_2 = a_{k-1} = 1$. As the sum below shows, in this case, we find that $v = 1$, a contradiction to our initial assumption that $v = 0$.

$$\begin{array}{r} 1 \qquad\qquad\quad 1 \quad \\ 11*\ldots*y*\ldots*11 \\ +\ 11*\ldots*y*\ldots*11 \\ \hline 1\mathbf{1}**\ldots*z*\ldots*10 \end{array}$$

Therefore, this case is impossible.

Case $a_2 \neq a_{k-1}$. Then, as the sums below show, $(a + \overline{a})_3 = 1$. If this were not the case, we would get a carry into the 1st column of the sum, contradicting the assumption that $v = 0$. This forces $z = 0$. Again, arguing as in Case 1b, we will arrive at a contradiction.

$$\begin{array}{r} 1 \qquad\qquad\quad \\ 10*\ldots*y*\ldots*11 \\ +\ 11*\ldots*y*\ldots*01 \\ \hline 101*\ldots*z*\ldots*00 \end{array} \qquad \begin{array}{r} 1 \quad \\ 11*\ldots*y*\ldots*01 \\ +\ 10*\ldots*y*\ldots*11 \\ \hline 101*\ldots*z*\ldots*00 \end{array}$$

Therefore, this case cannot happen.

Having excluded all alternatives, we have that that $a_2 = a_{k-1} = 0$. Again, computing $a + \overline{a}$ with this new information gives the following sum.

$$\begin{array}{r} 1 \quad \\ 10*\ldots*y*\ldots*01 \\ +\ 10*\ldots*y*\ldots*01 \\ \hline 10**\ldots*z*\ldots*10 \end{array}$$

Arguing similarly to Case 1a, it follows that $a_i = 0$ for $1 < i < k$.

Since Cases $1a$–$1c$ exhaust all possibilities, we now count the number of possible a's from each case. Case $1a$ and Case $1c$ each yield one element, while Case $1b$ yields none. This gives that $I(k, 2) = 2$ for k odd.

Case 2. k is even. Let a be an integer such that $R(a) = 1^2$.

Case 2a. $a_k = 0$ and $a + \overline{a}$ is k digits long. Using a similar argument to Case $1a$, it follows that $a_i = 0$ for $1 < i < k$.

Case 2b. $a_k = 0$ and $a + \overline{a}$ is $k + 1$ digits long. Using a similar argument to Case $1b$, it follows that this case cannot occur.

Case 2c. $a_k = 1$. Then a is of the form $a = 1 * \ldots * 1$. We compute the sum $a + \overline{a}$ as

$$\begin{array}{r} 1 \\ 1* \ldots *1 \\ +\ 1* \ldots *1 \\ \hline 1v* \ldots *0 \end{array}$$

where $v \in \{0, 1\}$.

If $v = 0$, we can use an argument similar to Cases $1b$ and $1c$ to find that $a_i = 0$ for $1 < i < k$.

If $v = 1$, we show that we necessarily have $a_i \neq a_{k-i+1}$ for $1 < i < k$ by induction. Suppose that $a_2 = a_{k-1}$. This leads to two cases, $a_2 = a_{k-1} = 0$ and $a_2 = a_{k-1} = 1$.

Case $a_2 = a_{k-1} = 0$. As the sum below shows, the lack of a carry from the 2nd column implies that $v = 0$, a contradiction to our initial assumption that $v = 1$.

$$\begin{array}{r} 1 \\ 10* \ldots *01 \\ +\ 10* \ldots *01 \\ \hline 1\mathbf{0}** \ldots *10 \end{array}$$

Therefore, this case is impossible.

Case $a_2 = a_{k-1} = 1$. As the sum below shows, $(a+\overline{a})_1 = (a+\overline{a})_2 = 1$. In order to avoid a contradiction to (3.1) for $i = k - 1$, we must have that $a_3 \neq a_{k-2}$.

$$\begin{array}{r} 1 \quad\quad 11 \\ 11{*}\ldots{*}11 \\ +\ 11{*}\ldots{*}11 \\ \hline 11{*}{*}\ldots{*}00 \end{array}$$

In fact, arguing as in Case 1b, we find that $a_i \neq a_{k-i+1}$ for $2 < i < k-2$ to avoid a contradiction to (3.1) for $2 < i < k-2$. However, this cannot occur, as an example sum below shows, since the carries introduced by having $a_i \neq a_{k-i+1}$ for $2 < i < k-2$ lead to $(a + \overline{a})_3 = 1$, a contradiction of (3.1) for $i = 3$.

$$\begin{array}{r} 1111\ldots111 \\ 111{*}\ldots{*}011 \\ +\ 110{*}\ldots{*}111 \\ \hline 11\mathbf{1}00\ldots0000 \end{array}$$

Thus, we have that $a_2 \neq a_{k-1}$. This, in turn, creates another carry which must be counteracted by having $a_3 \neq a_{k-2}$. This carry process continues and yields that $a_i \neq a_{k-i+1}$ for $1 < i < k$.

Again, having exhausted all possibilities with Cases 2a–2c, we now count the number of possible a's from each case. Case 2a yields one element, Case 2b yields none, and Case 2c yields $2^{(k-2)/2} + 1$. This gives that $I(k, 2) = 2^{(k-2)/2} + 2$ for k even. □

At this time, there is no known general formula for $I(k, m)$. For small k and $3 < m < k - 1$, no obvious pattern arises from computational experiments (see Figure 3.1), and the method of proof of Proposition 3.4 becomes cumbersome, even in the case $I(k, 4)$.

However, we can give a heuristic argument that leads to a plausible conjecture. Suppose that k is fixed and that we choose an integer a with k digits in its base 2 representation. If a is taken at random, then we can reasonably expect that, most of the time, $a + \overline{a}$ has roughly an equal number of 0's and 1's. After sorting, this would give that $R(a) = 1^{k'}$ with $k' \approx k/2$. This suggests that $I(k, m)$ should be maximal, as a function of m, when $m \approx k/2$. The data in Figure 3.1 seems to support this heuristic.

A closer examination of such data suggests the following conjecture.

Conjecture. *For any integer $k > 1$, if $k \equiv a \bmod 4$ with $a \in \{-1, 0, 1, 2\}$, then* $\max_m I(k, m) = I(k, \frac{k+|a|}{2})$.

k \ m	2	3	4	5	6	7	8	9	10
2	**2**								
3	**3**	1							
4	**4**	1	3						
5	2	**9**	4	1					
6	6	5	**15**	1	5				
7	2	9	**32**	14	6	1			
8	10	7	**45**	24	32	1	9		
9	2	13	40	**109**	60	21	10	1	
10	18	11	87	93	**171**	57	67	1	17

Figure 3.1: $I(k, m)$ values. Maximal values in each row are in boldface.

Chapter 4

The RATS process in base 3

4.1 Periodicity

By Theorem 3.2, it follows that base 3 is the first base with RATS sequences that may have periods > 1 or quasiperiods. In [6], Gentges completely characterized $\mathcal{P}_3$ and $\mathcal{Q}_3$ (see Theorem 4.1). We will give an independent proof of this result that yields additional information, including a complete characterization of all periodic elements.

Theorem 4.1 (Gentges, [6])**.** *Every RATS sequence in base* 3 *is preperiodic and* $\mathcal{P}_3 = \{p : p \geq 3\}$.

This result seems to be a hybrid between what is known about RATS sequences in bases 2 and 10. As in the case of base 2, we have $\mathcal{Q}_3 = \emptyset$, and as in the case base 10, $\mathcal{P}_3$ contains all sufficiently large periods.

Before proving Theorem 4.1, we introduce some helpful notation and prove some necessary, auxiliary lemmas.

Definition 4.2. *We define a one-to-one correspondence between sorted integers in base* 3 *and* $X := \mathbb{Z}^2_{\geq 0} \setminus \{(0,0)\}$ *by* $1^u 2^v \leftrightarrow (u, v)$. *With this correspondence, we use* $R(u, v)$ *to mean* $R(1^u 2^v)$ *and we extend any terminology from Chapter 2 to* (u, v).

Since our aim is to show that every RATS sequence in base 3 is preperiodic, we can assume, without loss of generality, that the starting integer for the sequence is already sorted. The following lemma characterizes the output of R for any given, sorted, input.

Lemma 4.3. *Given any element* (u, v) *representing a sorted integer, we have that:*

(i) *if* $v = 0$, $R(u, v) = (0, u)$;

(ii) *if* $u > v > 0$, $R(u, v) = (2v, 0)$;

(iii) *if* $0 < u \leq v$, $R(u, v) = (2u, v - u)$;

(iv) *if* $u = 0$, $R(u, v) = (2, v - 1)$.

Proof. If $u = 0$, we have that

$$\begin{array}{r} 22\ldots22 \\ +\ 22\ldots22 \\ \hline 122\ldots21 \end{array}$$

So $R(0, v) = (2, v - 1)$, proving (iv).

The assertions (i)–(iii) can all be proved in a similar fashion. □

Lemma 4.4. *Given any* (u, v)*, there is a* $t > 0$ *such that* $R^t(u, v) = (0, v')$ *for some* v'*.*

Proof. **Case 1.** $v = 0$**.** Then $R(u, 0) = (0, u)$, by Lemma 4.3(i), so the statement holds with $t = 1$.

Case 2. $u > v > 0$**.** Then $R^2(u, v) = R(2v, 0) = (0, 2v)$ by using Lemma 4.3 (ii) and (i). So the statement holds with $t = 2$.

Case 3. $0 < u \leq v$**.** Then $R(u, v) = (2u, v - u)$, again by Lemma 4.3. If $2u \leq v - u$, the output $(2u, v - u)$ falls into the same case, while increasing the first coordinate and decreasing the second. Hence, repeated applications of R can only produce an output of the form of Case 3 a finite number of times before reaching an element of the form (u', v') with $u' > v' \geq 0$, i.e., an element that falls into Case 1 or Case 2.

Case 4. $u = 0$**.** Then $R(0, v) = (2, v - 1)$, which leads to Case 3 if $v \geq 3$, to Case 2 if $v = 2$, and to Case 1 if $v = 1$. □

The above argument can be illustrated by the directed graph in Figure 4.1. The vertices are labeled according to the cases in the lemma, and the directed edges show how the RATS process may evolve when iterating R.

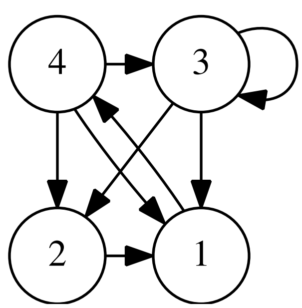


Figure 4.1: Directed graph of cases in Lemma 4.4.

It follows from Lemma 4.4 that for any (u, v), there is a minimal $t > 0$ such that $R^t(u, v) = (0, v')$ for some v'. The following lemma gives an explicit formula for t and v' in the special case when $u = 0$ and $v \neq 2^k - 1$ for $k \geq 1$.

Lemma 4.5. *If $2^k - 1 < v < 2^{k+1} - 1$ for some $k \geq 1$ and $t > 0$ is minimal such that $R^t(0, v) = (0, v')$ for some v', then $t = k + 2$ and*

$$v' = 2v - 2^{k+1} + 2. \tag{4.1}$$

In particular,

$$v' \leq v, \tag{4.2}$$

since $v \leq 2^{k+1} - 2$.

Proof. By repeated applications of Cases (iv) and (iii) from Lemma 4.3, it follows that

$$R^s(0, v) = (2^s, v - 2^s + 1) \tag{4.3}$$

for $0 < s \leq k$. (Note that (4.3) holds even in the case $v = 2^k - 1$.)

If $v = 2^k$, then, by (4.3) and two further applications of Cases (ii) and (i) from Lemma 4.3, $R^{k+2}(0, v) = R^2(2^k, 1) = R(2, 0) = (0, 2v - 2^{k+1} + 2) = (0, 2)$. Since $v = 2^k \geq 2$, the assertion holds in this case.

If $2^k < v < 2^{k+1} - 1$, then, by (4.3) and Lemma 4.3,

$$R^{k+2}(0, v) = R^2(2^k, v - 2^k + 1) = R(2v - 2^{k+1} + 2, 0) = (0, 2v - 2^{k+1} + 2). \tag{4.4}$$

The upper bound $v < 2^{k+1} - 1$ is equivalent to the inequality $2v - 2^{k+1} + 2 \leq v$, so the assertion also holds in this case. □

Proposition 4.6. *In base 3, for any integer $p \geq 3$, there exists a RATS sequence that is periodic with period p. In particular, $(0, v)$ is periodic with period p if and only if $v = 2^{p-1} - 2$ for some $p \geq 3$.*

Proof. ($\Leftarrow$) If $v = 2^{p-1} - 2$ for some $p \geq 3$, then, by (4.3) and Lemma 4.3, the RATS sequence generated by $(0, 2^{p-1} - 2)$ after p iterations is

$$(0, 2^{p-1} - 2) \overset{\text{(iv)}}{\to} (2, 2^{p-1} - 3) \overset{\text{(iii)}}{\to} \underbrace{\cdots}_{p-3 \text{ times}} \overset{\text{(iii)}}{\to} (2^{p-2}, 2^{p-2} - 1) \overset{\text{(ii)}}{\to} (2^{p-1} - 2, 0) \overset{\text{(i)}}{\to} (0, 2^{p-1} - 2). \tag{4.5}$$

Therefore, $(0, v)$ is periodic with period p.

($\Rightarrow$) Let $(0, v)$ be periodic with period p. Now if $2^k - 1 \leq v < 2^{k+1} - 1$ for some integer $k > 1$, then $R^k(0, v) = (2^k, v - 2^k + 1)$ by (4.3).

If $v = 2^k - 1$, then $R^{k+1}(0, v) = R(2^k, 0) = (0, 2^k)$. Since, by assumption, $(0, v)$ is periodic, this implies that $(0, 2^k)$ is periodic. However, the following calculation shows that $(0, 2^k)$ leads to the periodic cycle $(0, 2) \overset{\text{(iv)}}{\to} (2, 1) \overset{\text{(ii)}}{\to} (2, 0) \overset{\text{(i)}}{\to} (0, 2)$ and hence is itself not periodic if $k > 1$:

$$(0, 2^k) \overset{\text{(iv)}}{\to} (2, 2^k - 1) \overset{\text{(iii)}}{\to} \cdots \overset{\text{(iii)}}{\to} (2^k, 1) \overset{\text{(ii)}}{\to} (2, 0). \tag{4.6}$$

This shows that $v \notin \{2^k - 1, 2^k\}$.

Hence we must have that $2^k < v < 2^{k+1} - 1$. Suppose that p is not minimal such that $R^p(0, v) = (0, v')$ for some v'. Then we can write $p = p_1 + p_2 + \cdots + p_N$ for some $N > 1$ such that $0 < p_1 < p$ is minimal such that $R^{p_1}(0, v) = (0, v_1)$ for some v_1, $0 < p_2 < p - p_1$ is minimal such that $R^{p_2}(0, v_1) = (0, v_2)$ for some v_2, etc. Notice that, for $0 < i < N$, $2^{k_i} < v_i < 2^{k_i+1} - 1$ for some $k_i > 1$, since, if $v_{i_0} = 2^{k_{i_0}} - 1$ or $v_{i_0} = 2^{k_{i_0}}$ for some k_{i_0}, the assumption that $(0, v)$ is periodic implies that, using a similar argument as above, $v = 2$, a contradiction to the assumption that $2^k < v < 2^{k+1} - 1$ for some $k > 1$. By Lemma 4.5, it follows that $v = v_N \le v_{N-1} \le \cdots \le v_1 \le v$. Therefore, $v_1 = v$. So $R^{p_1}(0, v) = (0, v)$, contradicting the assertion that p is the period of $(0, v)$. So p must be minimal such that $R^p(0, v) = (0, v')$ for some v'. By (4.4), we find that $v = 2v - 2^{p-1} + 2$, as desired.

The case $1 \le v < 3$ can be handled by inspection. This completes the proof of the proposition. □

We now completely characterize all periodic elements (u, v) with the following key lemma.

Corollary 4.7. *The element (u, v) is periodic with period p if and only if there is some t such that $R^t(0, 2^{p-1} - 2) = (u, v)$. In particular, if this condition holds, (u, v) appears in the sequence given by (4.5).*

Proof. (⇐) By Proposition 4.6, $(0, 2^{p-1} - 2)$ is periodic with period p. Hence if there is a t such that $R^t(0, 2^{p-1} - 2) = (u, v)$, then (u, v) is also periodic with period p.

(⇒) By Lemma 4.4, there is a t' such that $R^{t'}(u, v) = (0, v')$ for some v'. If (u, v) is periodic with period p, then this implies that $(0, v')$ is also periodic with period p. By Proposition 4.6, this gives that $v' = 2^{p-1} - 2$. In particular, $(u, v) = R^p(u, v) = R^{p-t'}(R^{t'}(0, v')) = R^{p-t'}(0, 2^{p-1-2})$. So the assertion holds with $t = p - t'$.

The final assertion of the corollary follows from the fact that (4.5) is the periodic sequence generated by $(0, 2^{p-1} - 2)$. □

Lemma 4.8. *Given any (u, v), there is a t such that $R^t(u, v) = (0, 2^{p-1} - 2)$ for some $p \ge 3$.*

Proof. By Lemma 4.4, there is a minimal $t > 0$ such that $R^t(u, v) = (0, v')$ for some v'. So, it is enough to prove the lemma for $u = 0$.

Suppose that $2^k - 1 \leq v < 2^{k+1} - 1$ for some positive integer k. Then Lemma 4.5 shows that either $v = 2^k - 1$ or $v' \leq v$.

Case 1. If $v = 2^k - 1$, then there is a t such that $R^t(0, v) = (0, 2)$ (see (4.6)).

Case 2. If $v = v'$, $(0, v)$ is periodic, so the assertion holds by Proposition 4.6.

Case 3. If $v' < v$, we apply Lemma 4.4 to $(0, v')$ in place of $(0, v)$. This third case can only occur a finite number of times before a repeated application of Lemma 4.4 leads to Case 1 or 2. □

Proof of Theorem 4.1. Lemma 4.8 establishes that all RATS sequences in base 3 are preperiodic, Proposition 4.6 shows that $\{p : p \geq 3\} \subseteq \mathcal{P}_3$, and Theorem 3.2 shows that $1 \notin \mathcal{P}_3$. It remains to show that $2 \notin \mathcal{P}_3$.

Suppose that (u, v) is periodic with period 2. Let $(u', v') = R(u, v)$. By studying Figure 4.1, if follows that the only way to have (u, v) periodic with period 2 would be to have $u = 0$ and $v' = 0$ (Case 4 followed by Case 1 of Lemma 4.1), or, alternatively, one would need to have $0 < u \leq v$ and $0 < u' \leq v'$ (both Case 3 of Lemma 4.1).

If $u = 0$, then $(u', v') = (2, v-1)$, so $v - 1 = v' = 0$. Since (u, v) has period 2, $(0, v) = R(u', v') = R(2, 0)$, which implies that $v = 2$, a contradiction. Now if both (u, v) and (u', v') are such that $0 < u \leq v$ and $0 < u' \leq v'$, then $(u, v) = R^2(u, v) = (4u, v - 3u)$, which implies that $u = 0$, a contradiction to the fact that u must be positive. Therefore, $2 \notin \mathcal{P}_3$. □

4.2 Distribution of periods

In the previous section we showed that every RATS sequence in base 3 is preperiodic and we constructed elements $(u, v) \in X$ that are periodic with any desired period $p \geq 3$. In this section, we seek to answer the following question: Given $p \geq 3$ and a set $Y \subset X$, how many elements $(u, v) \in Y$ are preperiodic with period p? In other words, given some collection of sorted integers $1^u 2^v$ in base 3, how are the periods of the RATS sequences they generate distributed? We will answer this question completely for sets of the form $\{(0, v) : 0 \leq v < 2^k\}$, and obtain partial results for other sets.

We begin with a definition followed by a key result that will be used repeatedly in this section. The key result is special in that it allows us to identify the exact period of $(0, v)$ in terms of the binary representation of v with no additional information.

v	base 2	period of $(0,v)$
32	100000	3
33	100001	3
34	100010	4
35	100011	3
36	100100	4
37	100101	4
38	100110	5
39	100111	3
40	101000	4
41	101001	4
42	101010	5

v	base 2	period of $(0,v)$
43	101011	4
44	101100	5
45	101101	5
46	101110	6
47	101111	3
48	110000	4
49	110001	4
50	110010	5
51	110011	4
52	110100	5
53	110101	5

Figure 4.2: Base 2 representation of v and period of $(0,v)$.

Definition 4.9. *For a positive integer $p \geq 3$, let*

$$B(p) := \{v \in \mathbb{Z}_{>0} : v = 1\underbrace{*\dots*0}_{p-3 \text{ 1's}}1\dots1 \text{ in binary}\},$$

where we allow for the possibility that the bracketed block or the rightmost block of consecutive 1's is empty.

Example From Figure 4.2, we see that $45, 50 \in B(5)$, since $45 = 1\underbrace{0110}_{2 \text{ 1's}}1$ and $50 = 1\underbrace{10010}_{2 \text{ 1's}}$. We also have that $7, 35 \in B(3)$, since $7 = 111$ and $35 = 1\underbrace{000}_{0 \text{ 1's}}11$.

The membership of a positive integer v in $B(p)$ is based on the following procedure. Take the integer v in binary, remove the leftmost 1 and all of the 1's beyond the rightmost 0, count the number of remaining 1's in the string of digits left and call the count p'. Then $v \in B(p'+3)$. In particular, this is a well-defined procedure and every v belongs to $B(p)$ for some unique p. A careful examination of Figure 4.2 suggests that if $v \in B(p)$, then $(0,v)$ is preperiodic with period p. Indeed this is the case, as we will show in Proposition 4.11.

As noted above, the sets $B(p)$ are pairwise disjoint and thus form a partition of the positive integers. The key lemma of interest is that, in some sense, the RATS process does not disturb this partition.

Lemma 4.10. *If $v \in B(q)$ for some $q \geq 3$ and there is a $t > 0$ such that $R^t(0,v) = (0,v')$, then $v' \in B(q)$.*

Proof. Suppose that we are given $(0,v)$ with $v \in B(q)$ for some $q \geq 3$. Furthermore, suppose that $2^{k-1} \leq v < 2^k$; i.e., that v is k digits long.

Case 1. If $v = 2^{k-1}$, then $v \in B(3)$ since $v = 10\dots0$. By Lemma 4.5, $R^{k+2}(0,v) = R^2(2^{k-1},1) = R(2,0) = (0,2)$. Note that $2 \in B(3)$.

Case 2. If $v = 2^k - 1$, then $v \in B(3)$ since $v = 1 \dots 1$. By (4.3), $R^{k+1}(0, v) = R(2^k, 0) = (0, 2^k)$. Note that $2^k \in B(3)$.

Case 3. If $2^{k-1} < v < 2^k - 1$, then, by (4.4), $R^{k+2}(0, v) = (0, 2v - 2^k + 2)$. As in the previous two cases, we aim to show that if $v \in B(q)$ for some $q \geq 3$, then $2v - 2^k + 2 \in B(q)$.

Subtracting 2^{k-1} from v simply plucks the leftmost zero from the binary representation of v. If $v \in B(q)$ for some $q > 3$, then

$$\begin{aligned} v - 2^{k-1} &= 1\underbrace{* \dots * 0}_{p-3 \text{ 1's}} 1 \dots 1 - 10 \dots 0 \\ &= 1\underbrace{* \dots * 0}_{p-4 \text{ 1's}} 1 \dots 1, \end{aligned}$$

so $v - 2^{k-1} \in B(q-1)$. Adding 1 to the result gives

$$\begin{aligned} v - 2^{k-1} + 1 &= 1\underbrace{* \dots * 0}_{p-4 \text{ 1's}} 1 \dots 1 + 1 \\ &= 1\underbrace{* \dots * 1}_{p-3 \text{ 1's}} 0 \dots 0, \end{aligned} \tag{4.7}$$

so $v - 2^{k-1} + 1 \in B(q)$. If $v \in B(3)$, then

$$\begin{aligned} v - 2^{k-1} + 1 &= 10 \dots 01 \dots 1 - 10 \dots 0 + 1 \\ &= 1 \dots 1 + 1 \\ &= 10 \dots 0, \end{aligned} \tag{4.8}$$

so $v - 2^{k-1} + 1 \in B(3)$. Therefore, for any v in Case 3, if $v \in B(q)$ for some $q \geq 3$, then $v - 2^{k-1} + 1 \in B(q)$.

Multiplication by 2 in binary is equivalent to simply tacking on an extra 0 at the end of $v - 2^{k-1} + 1$. From (4.7) and (4.8), we see that if $v - 2^{k-1} + 1 \in B(q)$ for some $q \geq 3$, then $2v - 2^k + 2 \in B(q)$.

What we have shown is that, by appealing to Lemma 4.5, if $v \in B(q)$ for some $q \geq 3$ and $t > 0$ is minimal such that $R^t(0, v) = (0, v')$, then $v' \in B(q)$. We now remove the word "minimal" from the previous statement with the following argument. If $t > 0$ is such that $R^t(0, v) = (0, v')$ and t is not minimal, then there is a partition of $t = t_1 + t_2 + \cdots + t_N$ for some $N > 1$ such that $0 < t_1 < t$ is minimal such that $R^{t_1}(0, v) = (0, v_1)$ for some v_1, $0 < t_2 < t - t_1$ is minimal such that $R^{t_2}(0, v_1) = (0, v_2)$ for some v_2, etc. It follows, from the above case analysis, that $v_1, v_2, \dots, v_N \in B(q)$. Since $v' = v_N$, the assertion of the lemma holds. □

Proposition 4.11. *For any positive integer v, $(0, v)$ is preperiodic with period $p \geq 3$ if and only if $v \in B(p)$.*

Proof. ($\Rightarrow$) If $(0, v)$ is preperiodic with period $p \geq 3$, then, by Proposition 4.6 and Lemma 4.8, there is a t such that $R^t(0, v) = (0, 2^{p-1} - 2)$. By Lemma 4.10, if $v \in B(q)$ for some $q \geq 3$, then $2^{p-1} - 2 \in B(q)$. Now $2^{p-1} - 2 = 1\underbrace{1\ldots1}_{p-3\text{ 1's}}0$, so $2^{p-1} - 2 \in B(p)$. Since the sets $B(q)$ and $B(p)$ are disjoint if $p \neq q$, we must have that $p = q$. Therefore, we have $v \in B(p)$.

($\Leftarrow$) By Lemma 4.8, there is a t such that $R^t(0, v) = (0, 2^{q-1} - 2)$ for some $q \geq 3$. If $v \in B(p)$, then so $2^{q-1} - 2 \in B(p)$. In particular, reasoning as above, $p = q$, so $(0, v)$ is preperiodic with period p by Proposition 4.6. □

The importance of Proposition 4.11 is due to Lemma 4.4. Since, for any $(u, v) \in X$, there is a minimal t such that $R^t(u, v) = (0, v')$ for some v', to know the period of (u, v) we simply need to know the form of the binary representation of v'.

Definition 4.12. *For any set $Y \subset X$ and any integer $p \geq 3$, let*

$$N(Y, p) = \#\{(u, v) \in Y : (u, v) \text{ is preperiodic with period } p\}.$$

We now compute $N(Y, p)$ for special cases of Y.

Proposition 4.13. *For $p \geq 3$ and $k \geq p - 1$, let $Y_k := \{(0, v) : 1 \leq v < 2^k\}$. Then*

$$N(Y_k, p) = \begin{cases} \binom{k}{p-1} & \text{if } p > 3, \\ \binom{k}{2} + \binom{k}{1} & \text{if } p = 3. \end{cases}$$

Proof. By Proposition 4.11, $(0, v)$ is preperiodic with period p if and only if $v \in B(p)$.

Suppose first that $p > 3$. If $2^{j-1} \leq v < 2^j$, then v has j digits in its binary representation. We now count the number of ways in which v can have j digits in its binary representation and be of the form $v = 1\underbrace{*\ldots*}_{p-3\text{ 1's}}0\,1\ldots1$. Suppose that the rightmost block of all 1's of v is of length i, where $0 \leq i \leq j + 1 - p$. Then there are $j - i - 2$ unspecified digits of v of which $p - 3$ are 1's and the rest are 0's. The number of ways this can happen is $\binom{j-i-2}{p-3}$. So the total number of integers v with j digits in their binary representation of the necessary form is

$$\binom{j-2}{p-3} + \binom{j-3}{p-3} + \cdots + \binom{p-3}{p-3} = \binom{j-1}{p-2}.$$

Summing over all possible $j \leq k$, we find that

$$N(Y_k, p) = \sum_{j=p-1}^{k} \binom{j-1}{p-2} = \binom{k}{p-1}.$$

Now suppose that $p = 3$. If $2^{j-1} \leq v < 2^j$, then v has j digits in its binary representation. We now count the number of ways in which v can have j digits in its binary representation and be of the form $v = 10\ldots01\ldots1$. Clearly there are only j numbers of this form. Summing over all possible $j \leq k$, we find that

$$N(Y_k, 3) = \sum_{j=1}^{k} j = \frac{k(k+1)}{2} = \binom{k+1}{2} = \binom{k}{2} + \binom{k}{1}.$$

□

Corollary 4.14. *Let $P(v)$ denote the period of $(0, v)$. Then, for any fixed M,*

$$\lim_{k\to\infty} \left(\frac{1}{2^k} \# \left\{ v \leq 2^k : \frac{P(v) - k/2}{\sqrt{k/2}} \leq M \right\} \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{M} e^{-t^2/2}\, dt. \tag{4.9}$$

and

$$\lim_{k\to\infty} \left(\frac{1}{2^k} \# \left\{ v \leq 2^k : \frac{P(v) - \frac{1}{2}\log_2 v}{\sqrt{\frac{1}{2}\log_2 v}} \leq M \right\} \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{M} e^{-t^2/2}\, dt \tag{4.10}$$

In other words, asymptotically, the periods of $(0, v)$ for $1 \leq v \leq 2^k$ have a Gaussian distribution with mean $\frac{1}{2}\log_2 v$ and standard deviation $\sqrt{\frac{1}{2}\log_2 v}$.

Proof. First note that, by Theorem 4.1, $P(v) \geq 3$ for any v. Also note that $P(2^k) = 3$ by Proposition 4.11. By Proposition 4.13, for k suffiently large,

$$\begin{aligned}
\# \left\{ v \leq 2^k : \frac{P(v) - k/2}{\sqrt{k/2}} \leq M \right\} &= \# \left\{ v \leq 2^k : P(v) \leq k/2 + M\sqrt{k/2} \right\} \\
&= 1 + N(Y_k, 3) + \sum_{3<p\leq k/2+M\sqrt{k/2}} N(Y_k, p) \\
&= \binom{k}{0} + \binom{k}{1} + \binom{k}{2} + \sum_{3<p\leq k/2+M\sqrt{k/2}} \binom{k}{p-1} \\
&= \sum_{1\leq p\leq k/2+M\sqrt{k/2}} \binom{k}{p-1}.
\end{aligned}$$

An application of the de Moivre–Laplace limit theorem (see [5], p. 186) gives (4.9).

To prove (4.10), notice first that, for $v \le 2^k$,

$$\frac{P(v) - \frac{1}{2}\log_2 v}{\sqrt{\frac{1}{2}\log_2 v}} \ge \frac{P(v) - k/2}{\sqrt{k/2}},$$

so

$$\begin{aligned}\limsup_{k\to\infty}\left(\frac{1}{2^k}\#\left\{v \le 2^k : \frac{P(v) - \frac{1}{2}\log_2 v}{\sqrt{\frac{1}{2}\log_2 v}} \le M\right\}\right) &\le \lim_{k\to\infty}\left(\frac{1}{2^k}\#\left\{v \le 2^k : \frac{P(v) - k/2}{\sqrt{k/2}} \le M\right\}\right)\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{M} e^{-t^2/2}\,dt\end{aligned}$$

by (4.9). This proves the upper bound in (4.10).

It remains to prove the lower bound in (4.10). Observe that for any v with $2^{k-k^{1/3}} \le v \le 2^k$, we have $k - k^{1/3} \le \log_2 v \le k$. In particular, for such v,

$$\begin{aligned}\frac{P(v) - \frac{1}{2}\log_2 v}{\sqrt{\frac{1}{2}\log_2 v}} &\le \frac{P(v) - (k - k^{1/3})/2}{\sqrt{(k - k^{1/3})/2}}\\ &= \frac{P(v) - k/2}{\sqrt{(k - k^{1/3})/2)}} + \frac{k^{1/3}/2}{\sqrt{(k - k^{1/3})/2}}\\ &= \frac{1}{\sqrt{1 - k^{-2/3}}}\left(\frac{P(v) - k/2}{\sqrt{k/2}} + \frac{1}{\sqrt{2}k^{1/6}}\right)\\ &\le \left(1 + \frac{2}{k^{2/3}}\right)\left(\frac{P(v) - k/2}{\sqrt{k/2}} + \frac{1}{\sqrt{2}k^{1/6}}\right),\end{aligned} \tag{4.11}$$

provided k is sufficiently large. So, for any $\epsilon > 0$, by (4.11),

$$\begin{aligned}\#\left\{v \le 2^k : \frac{P(v) - k/2}{\sqrt{k/2}} \le M - \epsilon\right\} &= \#\left\{2^{k-k^{1/3}} \le v \le 2^k : \frac{P(v) - k/2}{\sqrt{k/2}} \le M - \epsilon\right\} + O\big(2^{k-k^{1/3}}\big)\\ &\le \#\left\{2^{k-k^{1/3}} \le v \le 2^k : \frac{P(v) - \frac{1}{2}\log_2 v}{\sqrt{\frac{1}{2}\log_2 v}} \le M - \epsilon + \right.\\ &\qquad \left. + \frac{2(M-\epsilon)}{k^{2/3}} + \frac{1}{\sqrt{2}k^{1/6}} + \frac{\sqrt{2}}{k^{5/6}}\right\} + O\big(2^{k-k^{1/3}}\big)\\ &\le \#\left\{2^{k-k^{1/3}} \le v \le 2^k : \frac{P(v) - \frac{1}{2}\log_2 v}{\sqrt{\frac{1}{2}\log_2 v}} \le M - \epsilon + \frac{2M}{k^{2/3}} + \frac{3}{\sqrt{2}k^{1/6}}\right\}\\ &\qquad + O\big(2^{k-k^{1/3}}\big)\end{aligned}$$

$$\leq \#\left\{ v \leq 2^k : \frac{P(v) - \frac{1}{2}\log_2 v}{\sqrt{\frac{1}{2}\log_2 v}} \leq M - \epsilon + \frac{2M}{k^{2/3}} + \frac{3}{\sqrt{2}k^{1/6}} \right\}$$
$$+ O\big(2^{k-k^{1/3}}\big).$$

In particular, for all k large enough such that

$$\frac{2M}{k^{2/3}} + \frac{3}{\sqrt{2}k^{1/6}} < \epsilon,$$

we have

$$\#\left\{ v \leq 2^k : \frac{P(v) - k/2}{\sqrt{k/2}} \leq M - \epsilon \right\} \leq \#\left\{ v \leq 2^k : \frac{P(v) - \frac{1}{2}\log_2 v}{\sqrt{\frac{1}{2}\log_2 v}} \leq M \right\} + O\big(2^{k-k^{1/3}}\big).$$

Therefore, taking the limit as $k \to \infty$, we find that

$$\liminf_{k\to\infty} \left(\frac{1}{2^k} \#\left\{ v \leq 2^k : \frac{P(v) - \frac{1}{2}\log_2 v}{\sqrt{\frac{1}{2}\log_2 v}} \leq M \right\} \right) \geq \lim_{k\to\infty} \left(\frac{1}{2^k} \#\left\{ v \leq 2^k : \frac{P(v) - k/2}{\sqrt{k/2}} \leq M - \epsilon \right\} \right)$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{M-\epsilon} e^{-t^2/2}\, dt$$

for all $\epsilon > 0$. This proves the lower bound in (4.10). □

Proposition 4.15. *For $p > 3$, let $Y(j,i) = \{(u,v) : 2^j \leq u < 2^{j+1}, 2^i \leq v < 2^{i+1}\}$. Then,*

$$N(Y(j,i),p) = \begin{cases} 2^j \binom{i}{p-3} & \text{if } i < j, \\ 2^i \left[\binom{i}{p-2} + \binom{i-1}{p-3} \right] + \binom{i}{p-3} - \sum_{k=p-2}^{i} 2^{k-1} \left[\binom{k-1}{p-3} + \binom{k-2}{p-4} \right] & \text{if } i = j. \end{cases}$$

Proof. Suppose first that $i < j$. Then $0 < v < 2^{i+1} \leq 2^j \leq u$, so, by Lemma 4.3, $R^2(u,v) = R(2v,0) = (0,2v)$. This means that the period of (u,v) depends only on v. Letting $P(u,v)$ denote the period of (u,v),

$$N(Y(j,i),p) = \sum_{\substack{2^j \leq u < 2^{j+1} \\ 2^i \leq v < 2^{i+1} \\ P(u,v)=p}} 1 \tag{4.12}$$
$$= 2^j \sum_{\substack{2^i \leq v < 2^{i+1} \\ P(0,2v)=p}} 1$$
$$= 2^j \sum_{\substack{2^{i+1} \leq v < 2^{i+2} \\ P(0,v)=p,\ 2|v}} 1.$$

The last sum in (4.12) is counting integers v that have $i+2$ digits that are of the form $v = 1\underbrace{*\ldots*0}_{p-3 \text{ 1's}}$. Note that the lack of trailing 1's is due to the fact that v is even. Thus, v has i unspecified digits of which $p-3$ must be 1's and the rest all 0's. Clearly, there are $\binom{i}{p-3}$ such v. Therefore, the proposition holds for $i < j$.

Now suppose that $i = j$. Then,

$$N(Y(i,i),p) = \sum_{\substack{2^i \le v,u < 2^{i+1} \\ P(u,v)=p}} 1 = \sum_{\substack{2^i \le v \le u < 2^{i+1} \\ P(u,v)=p}} 1 + \sum_{\substack{2^i \le u < v < 2^{i+1} \\ P(u,v)=p}} 1. \tag{4.13}$$

Of the last two sums in (4.13), the first is

$$\begin{aligned} \sum_{\substack{2^i \le v \le u < 2^{i+1} \\ P(u,v)=p}} 1 &= \sum_{\substack{2^i \le v < 2^{i+1} \\ P(0,2v)=p}} \sum_{v \le u < 2^{i+1}} 1 \\ &= \sum_{\substack{2^i \le v < 2^{i+1} \\ P(0,2v)=p}} (2^{i+1} - v) \\ &= \sum_{\substack{2^{i+1} \le v < 2^{i+2} \\ P(0,v)=p,\ 2|v}} \left(2^{i+1} - \frac{v}{2}\right) \\ &= 2^{i+1}\binom{i}{p-3} - \frac{1}{2} \sum_{\substack{2^{i+1} \le v < 2^{i+2} \\ P(0,v)=p,\ 2|v}} v, \end{aligned} \tag{4.14}$$

where the last equality holds due to (4.12). We now evaluate the last sum in (4.14). Notice that the last sum is summing all numbers of the form $v = 2^{i+1} + a_i 2^i + \cdots + a_1 2$ with exactly $p-3$ a_m's equal to 1 and the rest all 0. The total number of v's in the last sum is $\binom{i}{p-3}$ by the argument above. Moreover, for any fixed m with $1 \le m \le i$, there are $\binom{i-1}{p-4}$ integers in the sum with $a_m = 1$. Hence the last sum in (4.14) is

$$\begin{aligned} \sum_{\substack{2^{i+1} \le v < 2^{i+2} \\ P(0,v)=p,\ 2|v}} v &= 2^{i+1}\binom{i}{p-3} + \binom{i-1}{p-4} \sum_{m=1}^{i} 2^m \\ &= 2^{i+1}\binom{i}{p-3} + 2\binom{i-1}{p-4}(2^i - 1). \end{aligned} \tag{4.15}$$

Combining (4.14) and (4.15), we find that

$$\sum_{\substack{2^i \le v \le u < 2^{i+1} \\ P(u,v)=p}} 1 = 2^i \left[\binom{i}{p-3} - \binom{i-1}{p-4} \right] + \binom{i-1}{p-4}. \tag{4.16}$$

For the second sum in (4.13), we find that

$$\begin{aligned}
\sum_{\substack{2^i \le u < v < 2^{i+1} \\ P(u,v)=p}} 1 &= \sum_{\substack{2^i \le u < v < 2^{i+1} \\ P(0,2(v-u))=p}} 1 \\
&= \sum_{\substack{1 \le v < 2^i \\ P(0,2v)=p}} (2^i - v) \\
&= \sum_{\substack{1 < v < 2^{i+1} \\ P(0,v)=p,\ 2|v}} \left(2^i - \frac{v}{2}\right),
\end{aligned}$$

where the first equality follows from the fact that, by Lemma 4.3, $R^3(u,v) = (0, 2(v-u))$, since $u < v \le 3u$. So,

$$\begin{aligned}
\sum_{\substack{1 \le v < 2^{i+1} \\ P(0,v)=p,\ 2|v}} \left(2^i - \frac{v}{2}\right) &= \sum_{k=0}^{i} \sum_{\substack{2^k \le v < 2^{k+1} \\ P(0,v)=p,\ 2|v}} \left(2^i - \frac{v}{2}\right) \\
&= \sum_{k=p-2}^{i} \left[2^i \binom{k-1}{p-3} - \frac{1}{2} \sum_{\substack{2^k \le v < 2^{k+1} \\ P(0,v)=p,\ 2|v}} v \right] \\
&= 2^i \binom{i}{p-2} - \sum_{k=p-2}^{i} \left[2^{k-1} \binom{k-1}{p-3} + \binom{k-2}{p-4} (2^{k-1} - 1) \right] \\
&= 2^i \binom{i}{p-2} - \sum_{k=p-2}^{i} 2^{k-1} \left[\binom{k-1}{p-3} + \binom{k-2}{p-4} \right] + \sum_{k=p-2}^{i} \binom{k-2}{p-4} \\
&= 2^i \binom{i}{p-2} + \binom{i-1}{p-3} - \sum_{k=p-2}^{i} 2^{k-1} \left[\binom{k-1}{p-3} + \binom{k-2}{p-4} \right],
\end{aligned} \tag{4.17}$$

using similar reasoning as was used for the first sum in (4.13).

Combining (4.16) and (4.17) gives

$$\begin{aligned}
N(Y(i,i),p) &= 2^i\left[\binom{i}{p-3}-\binom{i-1}{p-4}\right]+\binom{i-1}{p-4}\\
&\quad +2^i\binom{i}{p-2}+\binom{i-1}{p-3}-\sum_{k=p-2}^{i}2^{k-1}\left[\binom{k-1}{p-3}+\binom{k-2}{p-4}\right]\\
&= 2^i\left[\binom{i}{p-2}+\binom{i}{p-3}-\binom{i-1}{p-4}\right]+\binom{i-1}{p-3}+\binom{i-1}{p-4}\\
&\qquad -\sum_{k=p-2}^{i}2^{k-1}\left[\binom{k-1}{p-3}+\binom{k-2}{p-4}\right]\\
&= 2^i\left[\binom{i}{p-2}+\binom{i-1}{p-3}\right]+\binom{i}{p-3}-\sum_{k=p-2}^{i}2^{k-1}\left[\binom{k-1}{p-3}+\binom{k-2}{p-4}\right].
\end{aligned}$$

The last equality follows after two applications of the binomial coefficient identity $\binom{n}{k}=\binom{n-1}{k}+\binom{n-1}{k-1}$. □

The exact formula for $N(Y(j,i),p)$ when $j<i$ is significantly harder to compute. This is because, in this case, we are given elements (u,v) with $u<v$. If k is the smallest integer such that $v\le(2^{k+1}-1)u$, then $R^{k+2}(u,v)=R^2(2^k,v-u(2^k-1))=(0,2v-2u(2^k-1))$. So the period of (u,v) depends on the base 2 representation of $2v-2u(2^k-1)$, which is difficult to predict. We expect, however, that over large subsets of X, periods are distributed nicely, as in Corollary 4.14.

At this time, we cannot prove an Erdős-Kac type result for sets of the form $\{(u,v):0\le u,v\le V\}$ due to the difficulty encountered in the discussion above. Instead, we prove a weaker Hardy-Ramanujan type result for these sets.

Theorem 4.16. *Let $Y(V)=\{(u,v):0\le u,v<V\}$ and let $P(u,v)$ denote the period of (u,v). Then, for every $\epsilon>0$,*

$$\lim_{V\to\infty}\frac{1}{V^2}\#\left\{(u,v)\in Y(V):\left|P(u,v)-\frac{1}{2}\log_2(V)\right|>\frac{\epsilon}{2}\log_2(V)\right\}=0. \tag{4.18}$$

Proof. Let $\epsilon>0$ be given. Fix a positive integer $M>0$. Then

$$\left\{(u,v)\in Y(V):\left|P(u,v)-\frac{1}{2}\log_2(V)\right|>\frac{\epsilon}{2}\log_2(V)\right\}=A_1(V)\cup A_2(V)\cup A_3(V)\cup A_4(V),$$

where

$$A_1(V) = \Big\{(u,v) \in Y(V) : \Big|P(u,v) - \frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V),\ (2^M-1)u \le v\Big\},$$
$$A_2(V) = \bigcup_{i=1}^{M-1} A_{2,i}(V),$$
$$A_3(V) = \bigcup_{i=1}^{M-1} \Big\{(u,v) \in Y(V) : \Big|P(u,v) - \frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V),\ (2^i-1)u = v\Big\},$$
$$A_4(V) = \Big\{(u,v) \in Y(V) : \Big|P(u,v) - \frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V),\ v < u\Big\},$$

with

$$A_{2,i}(V) = \Big\{(u,v) \in Y(V) : \Big|P(u,v) - \frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V),\ (2^i-1)u < v < (2^{i+1}-1)u\Big\} \tag{4.19}$$

for $1 \le i \le M-1$.

We first estimate $\#A_1(V)$ by

$$\begin{aligned} \#A_1(V) &= \#\Big\{(u,v) \in Y(V) : \Big|P(u,v) - \frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V),\ (2^M-1)u \le v\Big\} \\ &\le \#\{(u,v) \in Y(V) : (2^M-1)u \le v\} \\ &\le \frac{V}{2^M-1} \cdot V + O(V). \end{aligned} \tag{4.20}$$

For $\#A_3(V)$, we estimate this quantity crudely by

$$\begin{aligned} \#A_3(V) &\le \bigcup_{i=1}^{M-1} \#\Big\{(u,v) \in Y(V) : \Big|P(u,v) - \frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V),\ (2^i-1)u = v\Big\} \\ &\le \bigcup_{i=1}^{M-1} \#\{(u,v) \in Y(V) : (2^i-1)u = v\} \\ &\le 2MV. \end{aligned} \tag{4.21}$$

Recall that, by Lemma 4.3, for $0 < v < u$, we have $R^2(u,v) = (0, 2v)$ and hence $P(u,v) = P(0, 2v)$. So, for $\#A_4(V)$, we have that

$$\begin{aligned}\#A_4(V) &= \#\Big\{(u,v)\in Y(V) : \Big|P(u,v)-\frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V),\ v<u\Big\} \\ &\le V\cdot\#\Big\{v\le V : \Big|P(0,2v)-\frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V)\Big\} \\ &\le V\cdot\#\Big\{v'\le 2V : \Big|P(0,v')-\frac{1}{2}\log_2(2V)\Big| > \frac{\epsilon}{2}\log_2(2V)-\frac{1+\epsilon}{2}\Big\} \\ &= o(V^2)\end{aligned} \tag{4.22}$$

as $V\to\infty$, where the little–oh bound follows from Corollary 4.14. (Note that, in the notation in Corollary 4.14, $P(0,2v)=P(2v)$.)

For $\#A_2(V)$, we first estimate $\#A_{2,i}(V)$ for $1\le i\le M-1$. Note that, by (4.4), if $(2^i-1)u<v<(2^{i+1}-1)u$, then $R^{i+2}(u,v)=(0,2(v-(2^i-1)u))$ and hence $P(u,v)=P(0,2(v-(2^i-1)u))$. So, we have that

$$\begin{aligned}\#A_{2,i}(V) &= \#\Big\{(u,v)\in Y(V) : \Big|P(u,v)-\frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V),\ (2^i-1)u<v<(2^{i+1}-1)u\Big\} \\ &= \#\Big\{(u,v)\in Y(V) : \Big|P(0,2(v-(2^i-1)u)-\frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V), \\ &\qquad 0<2(v-(2^i-1)u)<2^{i+1}u\Big\} \\ &\le V\cdot\#\Big\{v\le V : \Big|P(0,2v)-\frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V)\Big\} \\ &= o(V^2)\end{aligned} \tag{4.23}$$

by (4.22). It follows that

$$\begin{aligned}\#A_2(V) &\le \sum_{i=1}^{M-1}\#A_{2,i}(V) \\ &= o(V^2)\end{aligned} \tag{4.24}$$

as $V\to\infty$, for any fixed M.

Combining all four estimates (4.20), (4.21), (4.22), and (4.24), we find that

$$\begin{aligned}&\limsup_{V\to\infty}\frac{1}{V^2}\#\Big\{(u,v)\in Y(V) : \Big|P(u,v)-\frac{1}{2}\log_2(V)\Big| > \frac{\epsilon}{2}\log_2(V)\Big\} \\ &\le \limsup_{V\to\infty}\frac{1}{V^2}\sum_{i=1}^{4}\#A_i(V) \\ &\le \frac{1}{2^M-1}.\end{aligned} \tag{4.25}$$

Since (4.25) holds for all fixed M, letting $M \to \infty$, we get (4.18). $\square$

Theorem 4.16 roughly states that most base 3 sorted integers $1^u 2^v$, with $0 \leq u, v < V$, have their period close to $\frac{1}{2} \log_2 V$.

Chapter 5

The RATS process in bases 4 and 5

After some extensive computations [7], Curtis McMullen conjectured that RATS sequences in bases 2, 3, 4, 5 , 6, 7, 8, and 9 are all preperiodic. In particular, this suggests that Conway's conjecture is also special in the sense that Conway's sequence (see (1.2)) is the first instance of a quasiperiodic sequence. Currently, McMullen's conjecture has been proved only for bases 2 and 3.

Conjecture (McMullen, [7]). *For bases* $4, 5, 6, 7, 8$, *and* 9, *every RATS sequence is preperiodic.*

In the case of base 4 and 5, computational evidence suggests an even stronger statement for the behavior of RATS sequences can be made. For base 4, we have verified the following conjecture for all integers $\leq 4^{1,000,000}$.

Conjecture. *For bases* 4 *and* 5, *every RATS sequence is preperiodic and* $\mathcal{P}_4 = \mathcal{P}_5 = \{2\}$. *That is, the only period available in either case is* 2.

A case by case analysis of RATS sequences in base 4 and 5 is significantly more challenging that in base 3. In the case of base 4, theoretical methods have whittled down the problem to a single case. However, simple combinatorial arguments have not been sufficient to prove McMullen's conjecture. For illustrative purposes, we include a characterization of the output of R_4 given any sorted integer input. (For base 4, we use an analogous one-to-one correspondence between sorted integers $1^u 2^v 3^w$ and vectors (u, v, w) as was used in Definition 4.2.)

Lemma 5.1. *Given any element* (u, v, w) *representing a sorted integer* $1^u 2^v 3^w$, *we have that:*

(i) *If* $v = w = 0$, $R_4(u, v, w) = (0, u, 0)$;

(ii) *If* $u = w = 0$, $R_4(u, v, w) = (v, 0, 0)$;

(iii) *If* $u = v = 0$, $R_4(u, v, w) = (1, 1, w - 1)$;

(iv) *If* $u = v$ *and* $w = 0$, $R_4(u, v, w) = (0, 0, 2u)$;

(v) *If* $u > v$ *and* $w = 0$, $R_4(u, v, w) = (0, u - v, 2v)$;

(vi) *If* $u < v$ *and* $w = 0$, $R_4(u, v, w) = (v - u, 0, u)$*;*

(vii) *If* $u = w$ *and* $v = 0$, $R_4(u, v, w) = (2u, 0, 0)$*;*

(viii) *If* $u > w$ *and* $v = 0$, $R_4(u, v, w) = (2w - 1, u - w - 1, 1)$*;*

(ix) *If* $u < w$ *and* $v = 0$, $R_4(u, v, w) = (2u, 0, w - u)$*;*

(x) *If* $u > v + w$, $R_4(u, v, w) = (2w - 1, u - (v + w) - 1, v + 1)$*;*

(xi) *If* $u = v + w$, $R_4(u, v, w) = (2w, 0, 0)$*;*

(xii) *If* $v + w > u > w$, $R_4(u, v, w) = (3w + v - u, 0, 0)$*;*

(xiii) *If* $u = w$, $R_4(u, v, w) = (u + v + w, 0, 0)$*;*

(xiv) *If* $v = w$ *and* $u = 0$, $R_4(u, v, w) = (2, w - 1, 0)$*;*

(xv) *If* $v > w$ *and* $u = 0$, $R_4(u, v, w) = (v - w + 2, 2w - 1, 0)$*;*

(xvi) *If* $v < w$ *and* $u = 0$, $R_4 = (2, 2v - 1, w - v)$.

Proof. The method for proving (i)–(xvi) is similar to the method used in Lemma 4.3. We omit the details. □

In the case of base 5, we will show in the following chapter that $\mathcal{Q}_5 = \emptyset$, i.e., that there are no quasiperiodic elements. This does not prove McMullen's conjecture in this case, as it is unknown if there are elements in base 5 that are neither preperiodic nor periodic. For base 4, the possibility of quasiperiodic elements remains, but seem exceedingly unlikely given the strong numerical evidence mentioned above.

Chapter 6

The RATS process in general bases

In the previous chapters, we studied preperiodic and quasiperiodic RATS sequences in very specific bases. In this chapter, we explore the existence of infinite families of bases that have a common property.

6.1 Existence of periodic sequences

We saw in Chapter 2 and 4 that there are bases b such that $|\mathcal{P}_b| = \infty$, namely $b = 3$ and $b = 10$. (Recall that $\mathcal{P}_b$ is the period set in base b.) In this section, we show that there is an infinite family of bases for which the same result is true.

For ease of notation, we introduce the following two conventions. First, to avoid ambiguities, we use parentheses with exponential notation to denote digits larger than 9. For example, in base 22, $1(11)^3(14)$ is the five digit number with digits 1, 11, 11, 11, 14. Second, by a sum written in the form

$$\begin{array}{ccccc} & 2^2 & 2^3 & 3^3 & 4^2 \\ + & 4^2 & 3^3 & 2^3 & 2^2 \\ \hline \end{array}$$

we mean the following sum:

$$\begin{array}{r} 2222233344 \\ +\ 4433322222 \\ \hline \end{array}$$

This will help visualize what the "reverse-add" portion of the RATS process produces. For example, if the above sums were in base 10, they would give the sum $2^5 3^3 4^2 + \overline{2^5 3^3 4^2} = 5^2 6^6 5^2$.

Theorem 6.1. *For $m > 1$ and $b = 3 \cdot 2^m - 2$, we have $|\mathcal{P}_b| = \infty$. More precisely, we have:*

(i) If m is odd, then $\mathcal{P}_b$ contains all sufficiently large even integers.

(ii) If m is even, then $\mathcal{P}_b$ contains all sufficiently large integers.

Proof. We will first show that, for any even $k > 0$, $1^{2^k(2^m-1)}(2^m-1)^{2^k-1}$ is periodic with period $m+1+k$. This proves that when m is odd, $\mathcal{P}_b$ contains all sufficiently large even numbers, and when m is even, $\mathcal{P}_b$ contains all sufficiently large odd numbers. In particular, $|\mathcal{P}_b| = \infty$.

We begin by iterating the RATS process $m+1$ times with the starting element $1^{2^k(2^m-1)}(2^m-1)^{2^k-1}$. Since $2^k(2^m-1) = 2^k - 1 + 2^k(2^m-2) + 1$, the reverse-add portion of the RATS process produces

$$\begin{array}{cccc} & 1^{2^k-1} & 1^{2^k(2^m-2)+1} & (2^m-1)^{2^k-1} \\ + & (2^m-1)^{2^k-1} & 1^{2^k(2^m-2)+1} & 1^{2^k-1} \\ \hline & (2^m)^{2^k-1} & 2^{2^k(2^m-2)+1} & (2^m)^{2^k-1} \end{array}$$

(Note that no carries occur since the base $b = 3 \cdot 2^m - 2$ satisfies $b > 2^m$.) After the sorting step, we obtain

$$R_b(1^{2^k(2^m-1)}(2^m-1)^{2^k-1}) = 2^{2^k(2^m-2)+1}(2^m)^{2(2^k-1)}.$$

Since $2^k(2^m-2)+1 = 2(2^k-1) + 2^k(2^m-4) + 3$, we get

$$\begin{array}{cccc} & 2^{2(2^k-1)} & 2^{2^k(2^m-4)+3} & (2^m)^{2(2^k-1)} \\ + & (2^m)^{2(2^k-1)} & 2^{2^k(2^m-4)+3} & 2^{2(2^k-1)} \\ \hline & (2^m+2)^{2(2^k-1)} & 4^{2^k(2^m-4)+3} & (2^m+2)^{2(2^k-1)} \end{array}$$

Therefore,

$$R_b(2^{2^k(2^m-2)+1}(2^m)^{2(2^k-1)}) = 4^{2^k(2^m-4)+3}(2^m+2)^{4(2^k-1)}.$$

Again, since $2^k(2^m-4)+3 = 4(2^k-1) + 2^k(2^m-8) + 7$, we get

$$\begin{array}{cccc} & 4^{4(2^k-1)} & 4^{2^k(2^m-8)+7} & (2^m+2)^{4(2^k-1)} \\ + & (2^m+2)^{4(2^k-1)} & 4^{2^k(2^m-8)+7} & 4^{4(2^k-1)} \\ \hline & (2^m+6)^{4(2^k-1)} & 8^{2^k(2^m-8)+7} & (2^m+6)^{4(2^k-1)} \end{array}$$

Therefore,

$$R_b(4^{2^k(2^m-4)+3}(2^m+2)^{4(2^k-1)}) = 8^{2^k(2^m-8)+7}(2^m+6)^{8(2^k-1)}.$$

By continuing this process, we see that if $0 < j \le m$, then the jth term in the RATS sequence generated by $1^{2^k(2^m-1)}(2^m-1)^{2^k-1}$ is given by

$$(2^j)^{2^k(2^m-2^j)+2^j-1}(2^m+2^{j+1}-2)^{2^j(2^k-1)}. \tag{6.1}$$

In particular, by (6.1), the mth term in the RATS sequence is $(2^m)^{2^m-1}(2^{m+1}-2)^{2^m(2^k-1)}$.

The mth term is the first instance in which the larger digit has the higher exponent. So now the reverse-add portion of the RATS process has the following form:

$$\begin{array}{ccccc} & (2^m)^{2^m-1} & (2^{m+1}-2)^{2^m(2^k-2)+1} & (2^{m+1}-2)^{2^m-1} \\ + & (2^{m+1}-2)^{2^m-1} & (2^{m+1}-2)^{2^m(2^k-2)+1} & (2^m)^{2^m-1} \\ \hline & 1^{2^m} & (2^m-1)^{2^m(2^k-2)+1} & 1^{2^m-2}\,0 \end{array}$$

(Note that carries occur at every step in the sum.) Therefore,

$$R_b((2^m)^{2^m-1}(2^{m+1}-2)^{2^m(2^k-1)}) = 1^{2^{m+1}-2}(2^m-1)^{2^m(2^k-2)+1}.$$

Continuing this process in a similar fashion, since $2^m(2^k-2)+1 = 2^{m+1}-2+2^m(2^k-4)+3$, we get

$$\begin{array}{ccccc} & 1^{2^{m+1}-2} & (2^m-1)^{2^m(2^k-4)+3} & (2^m-1)^{2^{m+1}-2} \\ + & (2^m-1)^{2^{m+1}-2} & (2^m-1)^{2^m(2^k-4)+3} & 1^{2^{m+1}-2} \\ \hline & (2^m)^{2^{m+1}-2} & (2^{m+1}-2)^{2^m(2^k-4)+3} & (2^m)^{2^{m+1}-2} \end{array}$$

Therefore,

$$R_b(1^{2^{m+1}-2}(2^m-1)^{2^m(2^k-2)+1}) = (2^m)^{2^{m+2}-4}(2^{m+1}-2)^{2^m(2^k-4)+3}.$$

Since $2^m(2^k-4)+3 = 2^{m+2}-4+2^m(2^k-8)+7$, we get

$$\begin{array}{ccccc} & (2^m)^{2^{m+2}-4} & (2^{m+1}-2)^{2^m(2^k-8)+7} & (2^{m+1}-2)^{2^{m+2}-4} \\ + & (2^{m+1}-2)^{2^{m+2}-4} & (2^{m+1}-2)^{2^m(2^k-8)+7} & (2^m)^{2^{m+2}-4} \\ \hline & 1^{2^{m+2}-3} & (2^m-1)^{2^m(2^k-8)+7} & 1^{2^{m+2}-5}\,0 \end{array}$$

Therefore,

$$R_b((2^m)^{2^{m+2}-4}(2^{m+1}-2)^{2^m(2^k-4)+3}) = 1^{2^{m+3}-8}(2^m-1)^{2^m(2^k-8)+7}.$$

Continuing this process, we see that if $m+1 < j \leq k$, then the jth term in the RATS sequence generated by $1^{2^k(2^m-1)}(2^m-1)^{2^k-1}$ is given by

$$(A)^{2^j(2^m-1)}(B)^{2^m(2^k-2^j)+2^j-1}, \tag{6.2}$$

where $A = 1$ and $B = 2^m - 1$ if j is even, and $A = 2^m$ and $B = 2^{m+1}-2$ if j is odd. Taking $j = k$ in

(6.2), we find that, since k is even, the $(m+1+k)$th term in the RATS sequence is $1^{2^k(2^m-1)}(2^m-1)^{2^k-1}$, our starting element. Thus this sequence is periodic with period $m+1+k$, as claimed. This proves that $|\mathcal{P}_b| = \infty$ and part (i) of the theorem.

To obtain part (ii), it is enough to show that for all even $k > 0$, $1^{2^k(2^{m+1}-2)(2^{m+1}+1)}(2^m-1)^{2^k(2^{m+1}-1)}$ is periodic with period $2(m+1)+k$. This can be done using the same method as above. We omit the details. □

Notice that since $10 = 3 \cdot 2^2 - 2$, the theorem applies to base $b = 10$ with $m = 2$, and gives that $\mathcal{P}_{10}$ contains all sufficiently large integers. In this case, as a slightly stronger result, Cooper and Kennedy [2] showed that the period set is as large as possible, that is, $\mathcal{P}_{10} = \{p : p \geq 2\}$ (see Theorem 2.7). At this time it is not known if there are infinitely many bases such that their period sets contain all periods $p \geq p_0$ for some fixed $p_0 > 1$.

Another point of interest is that base 3 is not in the family described by Theorem 6.1, yet it too has the property that $|\mathcal{P}_3| = \infty$. Clearly we have not identified all such bases for which the period set is infinite. In fact, the family $\{3 \cdot 2^m - 2 : m > 1\}$, covered by Theorem 6.1, has asymptotic density zero. A reasonable follow-up to Theorem 6.1 would be to examine the possibility of the existence of a family of bases, with nonzero asymptotic density, having infinite period sets.

6.2 Existence of quasiperiodic sequences

Conway's sequence (1.2) shows that $\mathcal{Q}_{10}$, the quasiperiod set in base 10, contains the element $q = 2$. In particular, the set $\mathcal{Q}_{10}$ is not empty. In this section we will construct infinite families of bases b for which $\mathcal{Q}_b \neq \emptyset$. We begin with a theorem of Cooper, Gentges, and Shattuck.

Theorem 6.2 (Cooper–Shattuck, [13], Gentges, [6]). *For any base b satisfying $b \equiv 1 \bmod 9$, with $b \geq 10$, we have $2 \in \mathcal{Q}_b$, i.e., there exists a quasiperiodic RATS sequence of period 2.*

Sketch of Proof. If $b \equiv 1 \bmod 9$, then either $b \equiv 1 \bmod 18$ or $b \equiv 10 \bmod 18$.

Case $b = 18k+1$. The element

$$123^3 4^4 5^{12} 6^{16} 7^{48} 8^{64} \cdots (6k)^m (6k+1)^{(23 \cdot 64^n - 32)/36}$$

is quasiperiodic with quasiperiod 2 if m is sufficiently large.

Case $b = 18k + 10$**.** The element

$$123^3 4^4 5^{12} 6^{16} 7^{48} 8^{64} \cdots (6k+1)^{(3\cdot 64^n)/4} (6k+2)^{64^n} (6k+3)^m (6k+4)^{(44\cdot 64^n - 8)/9}$$

is quasiperiodic with quasiperiod 2 if m is sufficiently large. □

Theorem 6.2 shows that bases containing at least one quasiperiodic sequence make up a positive proportion of all integers. The following corollary makes this statement explicit.

Corollary 6.3. *Let*

$$\delta = \liminf_{N\to\infty} \frac{\#\{b \leq N : \mathcal{Q}_b \neq \emptyset\}}{N}. \tag{6.3}$$

Then $\delta \geq 1/9$*.*

Along the lines of Theorem 6.2, Cooper and Shattuck [13] were also able to conclude that for bases $b = (2^q - 1)^2 + 1$, with q a prime or Fermat pseudoprime, we have $q \in \mathcal{Q}_b$. In the next theorem we expand on their result in two directions. First, we will remove the primality constraint on the quasiperiod q and second, we will exhibit an infinite family of bases b for which q is a quasiperiod.

Theorem 6.4. *Let* $q > 2$*. If* $b \equiv 1 \bmod (2^q - 1)^2$ *with* $b > 1$*, then* $q \in \mathcal{Q}_b$*.*

Proof. Suppose that $b = (2^q - 1)^2 k + 1$, where $k \geq 1$. Let $v = (2^q - 1)k$. We will construct a quasiperiodic integer in base b with quasiperiod q and growing digit v.

Consider an integer of the form

$$1^{m_1} 2^{m_2} \ldots (v-1)^{m_{v-1}} \boldsymbol{v}^{M+m_v} (v+1)^{\sum_{i=1}^{v} m_i}, \tag{6.4}$$

where $M > 0$, $m_i \geq 0$ for $i = 1, \ldots, v$, and where the digit in boldface denotes the growing digit. We will derive a system of equations and conditions that the exponents m_i must satisfy in order for the integer (6.4) to have quasiperiod q. We then will show that this system has a solution provided M is large enough.

We see that the reverse-add portion of the RATS process produces

	1^{m_1}	2^{m_2}	$\ldots$	$(v-1)^{m_{v-1}}$	v^{m_v}	$\boldsymbol{v}^M$	$(v+1)^{m_v}$	$\ldots$	$(v+1)^{m_1}$
$+$	$(v+1)^{m_1}$	$(v+1)^{m_2}$	$\ldots$	$(v+1)^{m_{v-1}}$	$(v+1)^{m_v}$	$\boldsymbol{v}^M$	v^{m_v}	$\ldots$	1^{m_1}
	$(v+2)^{m_1}$	$(v+3)^{m_2}$	$\ldots$	$(2v)^{m_{v-1}}$	$(2v+1)^{m_v}$	$(\boldsymbol{2v})^M$	$(2v+1)^{m_v}$	$\ldots$	$(v+2)^{m_1}$

(Note that no carries occur in the sum and every digit of the sum is nonzero.) Sorting the result gives the second term in the RATS sequence,

$$(v+2)^{2m_1}(v+3)^{2m_2}\dots(2v-1)^{2m_{v-2}}(\boldsymbol{2v})^{M+2m_{v-1}}(2v+1)^{2m_v}. \tag{6.5}$$

In order for the growing digit $\boldsymbol{2v}$ of (6.5) to (approximately) line up against itself in the next sum, we need $2\sum_{i=1}^{v-2} m_i \approx 2m_v$. Let M_1 be such that

$$2m_v = M_1 + 2\sum_{i=1}^{v-2} m_i \tag{6.6}$$

and assume $M_1 \geq 0$ for now. We repeat the above process and get

$$\begin{array}{cccccccc} & (v+2)^{2m_1} & (v+3)^{2m_2} & \dots & (2v)^{M_1} & (\boldsymbol{2v})^{M+2m_{v-1}-M_1} & (2v+1)^{M_1} & \dots & (2v+1)^{2m_1} \\ + & (2v+1)^{2m_1} & (2v+1)^{2m_2} & \dots & (2v+1)^{M_1} & (\boldsymbol{2v})^{M+2m_{v-1}-M_1} & (2v)^{M_1} & \dots & (v+2)^{2m_1} \\ \hline & (3v+3)^{2m_1} & (3v+4)^{2m_2} & \dots & (4v+1)^{M_1} & (\boldsymbol{4v})^{M+2m_{v-1}-M_1} & (4v+1)^{M_1} & \dots & (3v+3)^{2m_1} \end{array}$$

Sorting the result gives the third term in the RATS sequence,

$$(3v+3)^{4m_1}(3v+4)^{4m_2}\dots(4v-1)^{4m_{v-3}}(\boldsymbol{4v})^{M+2m_{v-1}+4m_{v-2}-M_1}(4v+1)^{2M_1}. \tag{6.7}$$

In order for the growing digit $\boldsymbol{4v}$ of (6.7) to line up against itself in the next sum, we need $4\sum_{i=1}^{v-3} m_i \approx 2M_1$. Let M_2 be such that $2M_1 = M_2 + 4\sum_{i=1}^{v-3} m_i$ and assume $M_2 \geq 0$ for now. We repeat the reverse-add portion of the RATS process and get

$$\begin{array}{cccccccc} & (3v+3)^{4m_1} & \dots & (4v)^{M_2} & (\boldsymbol{4v})^{M+2m_{v-1}+4m_{v-2}-M_1-M_2} & (4v+1)^{M_2} & \dots & (4v+1)^{4m_1} \\ + & (4v+1)^{4m_1} & \dots & (4v+1)^{M_2} & (\boldsymbol{4v})^{M+2m_{v-1}+4m_{v-2}-M_1-M_2} & (4v)^{M_2} & \dots & (3v+3)^{4m_1} \\ \hline & (7v+4)^{4m_1} & \dots & (8v+1)^{M_2} & (\boldsymbol{8v})^{M+2m_{v-1}+4m_{v-2}-M_1-M_2} & (8v+1)^{M_2} & \dots & (7v+4)^{4m_1} \end{array}$$

Sorting the result gives the fourth term in the RATS sequence,

$$(7v+4)^{8m_1}(7v+5)^{8m_2}\dots(8v-1)^{8m_{v-4}}(\boldsymbol{8v})^{M+2m_{v-1}+4m_{v-2}+8m_{v-3}-M_1-M_2}(8v+1)^{2M_2}. \tag{6.8}$$

Repeating this process a total of $q-1$ times gives that the qth term in the RATS sequences is

$$\begin{aligned}((2^{q-1}-1)v+q)^{2^{q-1}m_1}((2^{q-1}-1)v+q+1)^{2^{q-1}m_2}\dots \\ \dots(2^{q-1}v-1)^{2^{q-1}m_{q-1}}(\boldsymbol{2^{q-1}v})^{M+\sum_{i=1}^{q-1}2^i m_{v-i}-\sum_{i=1}^{q-2}M_i}(2^{q-1}v+1)^{2M_{q-2}},\end{aligned} \tag{6.9}$$

where the M_j's are defined to be such that

$$2M_j = M_{j+1} + 2^{j+1} \sum_{i=1}^{v-(j+2)} m_j \tag{6.10}$$

for $j = 1, \ldots, q-2$ and are assumed to be nonnegative.

We repeat the reverse-add portion of the RATS process one more time and get

$$\begin{array}{cccccc} & ((2^{q-1}-1)v+q)^{2^{q-1}m_1} & \ldots & (\mathbf{2^{q-1}v})^{M+\sum_{i=1}^{q-1} 2^i m_{v-i} - \sum_{i=1}^{q-1} M_i} & \ldots & (2^{q-1}v+1)^{2^{q-1}m_1} \\ + & (2^{q-1}v+1)^{2^{q-1}m_1} & \ldots & (\mathbf{2^{q-1}v})^{M+\sum_{i=1}^{q-1} 2^i m_{v-i} - \sum_{i=1}^{q-1} M_i} & \ldots & ((2^{q-1}-1)v+q)^{2^{q-1}m_1} \\ \hline & 1\,(q+1)^{2^{q-1}m_1} & \ldots & \boldsymbol{v}^{M+\sum_{i=1}^{q-1} 2^i m_{v-i} - \sum_{i=1}^{q-1} M_i} & \ldots & (q+1)^{2^{q-1}m_1-1}\,q \end{array}$$

(This is the first instance where carries occur, they occur at every column and every digit of the sum is nonzero.) Sorting the result gives the $(q+1)$th term in the RATS sequence,

$$1q(q+1)^{2^q m_1 - 1}(q+2)^{2^q m_2} \ldots \boldsymbol{v}^{M+2m_{v-1}+\cdots+2^q m_{v-q} - (M_1+\cdots+M_{q-1})}(v+1)^{2M_{q-1}}. \tag{6.11}$$

We seek to choose the m_i's such that $(q+1)$th term is identical to the first term of the RATS sequence, except that the exponent of v should be one larger than the exponent of v in the starting element. Comparing exponents in (6.4) and (6.11) for all digits except v leads to the following set of equations:

$$\begin{aligned} m_1 &= 1, \\ m_2 &= 0, \\ m_3 &= 0, \\ &\vdots \\ m_{q-1} &= 0, \\ m_q &= 1, \\ m_{q+1} &= 2^q m_1 - 1 = 2^q - 1, \\ m_{q+2} &= 2^q m_2 = 0, \\ m_{q+3} &= 2^q m_3 = 0, \\ &\vdots \\ m_{2q-1} &= 2^q m_{q-1} = 0, \\ m_{2q} &= 2^q m_q = 2^q, \\ m_i &= 2^q m_{i-q} \text{ for } 2q < i \le v-1, \end{aligned} \tag{6.12}$$

and

$$2M_{q-1} = \sum_{i=1}^{v} m_i. \tag{6.13}$$

Note that these conditions automatically imply that the exponent of v in (6.11) must be one larger than that in (6.4). This is due to the fact that the total number of digits is preserved by the RATS process in every step leading from (6.4) to (6.11) except for the final step, where carries result in one additional digit. We see that (6.12) immediately gives nonnegative integer values of m_i for $1 \le i \le v-1$. What remains is to show that m_v and the quantities M_j defined by (6.10) and (6.6) are nonnegative integers. Furthermore, we need to ensure that the exponents of the growing digit in (6.4)–(6.9) and (6.11) are all nonnegative, in order for these expressions to make sense.

For m_v, combining (6.13), (6.6) and (6.10) for $j = 1, \ldots, q-2$, we find that

$$\begin{aligned}
2m_v &= M_1 + 2\sum_{i=1}^{v-2} m_i, \\
4m_v &= 2M_1 + 4\sum_{i=1}^{v-2} m_i, \\
4m_v &= M_2 + 4\sum_{i=1}^{v-3} m_i + 4\sum_{i=1}^{v-2} m_i, \\
8m_v &= 2M_2 + 8\sum_{i=1}^{v-3} m_i + 8\sum_{i=1}^{v-2} m_i, \\
&\vdots \\
2^q m_v &= 2M_{q-1} + 2^q\sum_{i=1}^{v-q} m_i + \cdots + 2^q\sum_{i=1}^{v-2} m_i + 2^q\sum_{i=1}^{v-1} m_i, \\
2^q m_v &= \sum_{i=1}^{v} m_i + 2^q\sum_{i=1}^{v-q} m_i + \cdots + 2^q\sum_{i=1}^{v-2} m_i + 2^q\sum_{i=1}^{v-1} m_i.
\end{aligned}$$

In particular, we find that $m_v = \frac{1}{2^q-1} L(q,v)$, where

$$L(q,v) = \left[2^q\left(\sum_{i=1}^{v-q} m_i + \cdots + \sum_{i=1}^{v-2} m_i\right) + \sum_{i=1}^{v-1} m_i\right]. \tag{6.14}$$

Assume that m_v is an integer. (We will establish below that this is indeed the case.) It immediately follows from (6.6) that M_1 is also an integer and that

$$\begin{aligned} M_1 &= 2m_v - 2\sum_{i=1}^{v-2} m_i \\ &= \frac{2}{2^q-1}\left[2^q\left(\sum_{i=1}^{v-q} m_i + \cdots + \sum_{i=1}^{v-2} m_i\right) + \sum_{i=1}^{v-1} m_i\right] - 2\sum_{i=1}^{v-2} m_i \\ &= \left(\frac{2^{q+1}}{2^q-1} - 2\right)\sum_{i=1}^{v-2} m_i + \frac{2}{2^q-1}\left[2^q\left(\sum_{i=1}^{v-q} m_i + \cdots + \sum_{i=1}^{v-3} m_i\right) + \sum_{i=1}^{v-1} m_i\right] \\ &\geq 0, \end{aligned}$$

where the last inequality holds since $\frac{2^{q+1}}{2^q-1} - 2 \geq 0$ and $m_i \geq 0$ for all $1 \leq i \leq v-1$.

Similarly, using (6.10) with $j = 1$, since M_1 is an integer, it follows that M_2 is also an integer and that

$$\begin{aligned} M_2 &= 2M_1 - 4\sum_{i=1}^{v-3} m_i \\ &= \left(\frac{2^{q+2}}{2^q-1} - 4\right)\sum_{i=1}^{v-2} m_i + \frac{4}{2^q-1}\left[2^q\left(\sum_{i=1}^{v-q} m_i + \cdots + \sum_{i=1}^{v-3} m_i\right) + \sum_{i=1}^{v-1} m_i\right] - 4\sum_{i=1}^{v-3} m_i \\ &= \left(\frac{2^{q+2}}{2^q-1} - 4\right)\left[\sum_{i=1}^{v-2} m_i + \sum_{i=1}^{v-3} m_i\right] + \frac{4}{2^q-1}\left[2^q\left(\sum_{i=1}^{v-q} m_i + \cdots + \sum_{i=1}^{v-4} m_i\right) + \sum_{i=1}^{v-1} m_i\right] \\ &\geq 0, \end{aligned}$$

where the last inequality holds by a similar argument as above.

Continuing in this way, we see that the numbers M_j defined by (6.10) and (6.6) are indeed nonnegative integers. Taking $M > 0$ sufficiently large, which we are free choose, we find that every growing digit exponent in (6.4)–(6.9) and (6.11) is a nonnegative integer. Therefore, (6.4) is a quasiperiodic element with quasiperiod q.

To complete the proof, we will show that $m_v = \frac{1}{2^q-1}L(q,v)$ is an integer, or equivalently, that

$$L(q,v) \equiv 0 \bmod 2^q - 1. \tag{6.15}$$

The proof is broken down into four cases, depending on the residue of v modulo q.

Case 1. Suppose that $v \equiv 0 \bmod q$. Then $v = sq$ for some $s \geq 1$. By (6.14),

$$L(q,v) = \left[2^q\left(\sum_{i=1}^{v-q} m_i + \cdots + \sum_{i=1}^{v-2} m_i\right) + \sum_{i=1}^{v-1} m_i\right] \tag{6.16}$$
$$= \left[2^q\left(\sum_{i=1}^{(s-1)q} m_i + \cdots + \sum_{i=1}^{sq-2} m_i\right) + \sum_{i=1}^{sq-1} m_i\right].$$

Using (6.12) and (6.13), we see that

$$\sum_{i=1}^{sq-1} m_i = 1 + 1 + (2^q - 1) + 2^q + 2^q(2^q-1) + 2^{2q} + \cdots + 2^{(s-2)q}(2^q - 1) + 2^{(s-1)q}$$
$$= 1 + 2^q + 2^{2q} + \cdots + 2^{(s-1)q}$$
$$= \sum_{j=0}^{s-1} 2^{qj}.$$

For the remaining sums in the definition of $L(q,v)$, notice that $m_i = 0$ for $v - q + 1 < i < v - 2$, so the sums inside the parenthesis of (6.16) are identical, with the exception of the first sum, $\sum_{i=1}^{(s-1)q} m_i$, which is missing the term $(2^q - 1)2^{(s-1)q}$. Thus, we find that

$$L(q,v) = 2^q\left((q-1)\sum_{j=0}^{s-1} 2^{qj} - (2^q-1)2^{(s-2)q}\right) + \sum_{j=0}^{s-1} 2^{qj} \tag{6.17}$$
$$= (2^q(q-1)+1)\sum_{j=0}^{s-1} 2^{qj} - (2^q-1)2^{q(s-1)}$$
$$\equiv (1(q-1)+1)\sum_{j=0}^{s-1} 1 \bmod 2^q - 1$$
$$\equiv sq \bmod 2^q - 1$$
$$\equiv v \bmod 2^q - 1$$
$$\equiv (2^q-1)k \bmod 2^q - 1$$
$$\equiv 0 \bmod 2^q - 1.$$

Therefore, (6.15) holds, as desired.

Case 2. Suppose that $v \equiv 1 \bmod q$. Then $v = sq + 1$ for some $s \geq 1$. By a similar computation as in

(6.17) and (6.14),

$$
\begin{aligned}
L(q,v) &= (2^q(q-1)+1)\sum_{j=0}^{s-1} 2^{qj} + 2^{sq} \\
&\equiv sq+1 \bmod 2^q-1 \\
&\equiv 0 \bmod 2^q-1.
\end{aligned}
$$

Therefore, (6.15) holds, as desired.

Case 3. Suppose that $v \equiv 2 \bmod q$. Then $v = sq+2$ for some $s \geq 1$. By a similar computation as in (6.17) and (6.14),

$$
\begin{aligned}
L(q,v) &= 2^q\left((q-1)\sum_{j=0}^{s-1} 2^{qj} + 2^{sq}\right) + \sum_{j=0}^{s-1} 2^{qj} + 2^{q(s+1)} \\
&= (2^q(q-1)+1)\sum_{j=0}^{s-1} 2^{qj} + 2^{q(s+1)+1} \\
&\equiv sq+2 \bmod 2^q-1 \\
&\equiv 0 \bmod 2^q-1.
\end{aligned}
$$

Therefore, (6.15) holds, as desired.

Case 4. Suppose that $v \equiv c \bmod q$, where $2 < c < q$. Then $v = sq+c$ for some $s \geq 1$. By a similar computation as in (6.17) and (6.14),

$$
\begin{aligned}
L(q,v) &= (2^q(q-1)+1)\sum_{j=0}^{s-1} 2^{qj} + (2^q(c-1)+1)2^{q(s-1)} + (2^q(c-2)+1)(2^q-1)2^{q(s-1)} \\
&\equiv sq+c \bmod 2^q-1 \\
&\equiv 0 \bmod 2^q-1.
\end{aligned}
$$

Therefore, (6.15) holds, as desired.

So by exhausting all possibilities, we see that m_v is always a nonnegative integer. □

Theorem 6.4 shows that the asymptotic density of bases with nonempty quasiperiod sets is higher than the lower bound given by Corollary 6.3.

Corollary 6.5. *With δ as in* (6.3)*, we have* $\delta \geq \frac{1}{9} + \frac{1}{49}\left(1 - \frac{1}{9}\right)$.

Proof. By Theorem 6.2, any base $b \equiv 1 \bmod 9$ is such that $2 \in \mathcal{Q}_b$. By Theorem 6.4 for $q = 3$, any base $b \equiv 1 \bmod 49$ is such that $3 \in \mathcal{Q}_b$. The set of bases that are either congruent to 1 mod 9 or 1 mod 49 has asymptotic density $(1/9) + (1/49)(1 - 1/9)$ by the Chinese remainder theorem. It follows immediately that $\delta \geq (1/9) + (1/49)(1 - 1/9)$. □

Theorems 6.2 and 6.4 also imply that there exist bases b such that $\mathcal{Q}_b$ contains any prescribed finite set of distinct positive integers > 1.

Corollary 6.6. *For any set $A = \{q_1, q_2, \ldots, q_k\}$ of distinct positive integers > 1, there exists a base b such that $A \subseteq \mathcal{Q}_b$.*

Proof. Let $b = 18(2^{q_1} - 1)^2(2^{q_2} - 1)^2 \cdots (2^{q_k} - 1)^2 + 1$. Then $b \equiv 1 \bmod 9$ and $b \equiv 1 \bmod (2^{q_i} - 1)^2$ for each i. The first condition implies that $2 \in \mathcal{Q}_b$ by Theorem 6.2 and the second condition implies that $q_i \in \mathcal{Q}_b$ for any $q_i > 2$, by Theorem 6.4. □

Corollary 6.6 shows that there are bases with arbitrarily large quasiperiod sets. However, it is still unknown if there are bases with distinct quasiperiodic RATS sequences with the same quasiperiod. It is conjectured that this is never the case.

Conjecture (Cooper–Shattuck, [13])**.** *No base has two or more distinct quasiperiodic RATS sequences with the same quasiperiod.*

The definition for quasiperiodic sequences can sometimes seem rather narrow. However, all computational evidence compiled so far has always yielded that the behavior of RATS sequences always fit the descriptions of preperiodic or prequasiperiodic sequences. We conjecture that this is always the case.

Conjecture. *In any base, every RATS sequence is preperiodic or prequasiperiodic.*

6.3 Behavior of growing digits

In this section we introduce a finite map, f_b, that describes the behavior of growing digits in a quasiperiodic sequence in base b. We establish properties of this map and use these results to prove nonexistence results for quasiperiodic sequences.

We begin with Conway's sequence

$$\cdots \to 123^m4^4 \to 5^26^m7^4 \to 123^{m+1}4^4 \to 5^26^{m+1}7^4 \to \cdots . \tag{6.18}$$

as a motivating example.

By examining this sequence, it is clear that 3 is the growing digit associated with the quasiperiodic element 123^m4^4. When performing the reverse-add portion of the RATS process, we see that the digit 3 primarily lines up against itself in the sum. This is the reason that $6 = 3+3$ is the growing digit of the next term in the sequence, $5^26^m7^2$. Performing the reverse-add portion of the RATS process, again, we see that the digit 6 primarily lines up against itself in the sum. Due to a propagating carry, the most popular digit of the next iterate is again 3, since $3 \equiv 6+6+1 \mod 10$.

In general, if k is the growing digit of some quasiperiodic element in base b, and $k < b/2$, then $2k$ is the growing digit of the next term in the RATS sequence. (Some care must be taken in the case where $k = (b-1)/2$.) If $k \geq b/2$, then a propagating carry makes $2k+1-b$ the growing digit of the next term in the RATS sequence. Motivated by this idea, we make the following definition.

Definition 6.7. *Given a base b, let*

$$f_b(k) = \begin{cases} 2k & \text{if } 0 < k < b/2, \\ 2k+1-b & \text{if } b/2 \leq k < b. \end{cases}$$

Proposition 6.8. *Suppose that b is a base with a quasiperiodic element n. If k is the growing digit of n, then $f_b(k)$ is the growing digit of $R_b(n)$. In particular, if n is quasiperiodic with quasiperiod q, then $f_b^q(k) = k$.*

Proof. The discussion above Definition 6.7 shows that given an element n, if the digit $k \neq (b-1)/2$ appears in n with a vastly higher count than all other digits combined, then $f_b(k)$ gives the digit that will be the most popular in $R_b(n)$. Applying this reasoning to the growing digit of a quasiperiodic element, we obtain the desired result.

The case where $k = (b-1)/2$ is the the most popular digit if n requires some additional care. In this case, the predominant digit of $R_b(n)$ could be either $b-1$ or 0, depending on the existence of carries to the right of the block of digits k. However, since the RATS process discards zero digits, this would result in $R_b(n)$ having fewer digits than n (for n large enough), a contradiction to the definition of a quasiperiodic element. Thus, if n is a quasiperiodic element with growing digit $k = (b-1)/2$, then the growing digit of $R_b(n)$ must be $2k = b-1 = f_b(k)$.

From the above work, it follows that $f_b^q(k) = k$ by repeating the above reasoning q times. □

Example From (6.18), the growing digit of 123^m4^4 is 3 and $f_{10}^2(3) = f_{10}(6) = 3$.

Definition 6.9. *Given a base b and a digit $0 < k < b$, we say that k is a* ***fixed point*** *of f_b if there is a $t > 0$ such that $f_b^t(k) = k$. If t is the least integer with this property, we say that the* ***order*** *of k is t.*

Corollary 6.10. *The quasiperiod of a quasiperiodic element is a multiple of the order of the corresponding growing digit with respect to the map* f_b*.*

Proof. Suppose that k is the growing digit of a quasiperiodic element with quasiperiod q. By Proposition 6.8, $f_b^q(k) = k$; that is, k is a fixed point of f_b. Suppose that the order of k is t. By the Euclidean algorithm, there are nonnegative integers s and r such that $q = st + r$ with $0 \leq r < t$. Then $k = f_b^q(k) = f_b^{st+r}(k) = f_b^r(f_b^{st}(k)) = f_b^r(k)$. If $r > 0$, then we have a contradiction to the assumed order of k. Hence $r = 0$ and $t|q$. Therefore, the quasiperiod of a quasiperiodic element is a multiple of the order of the corresponding digit. □

The above results show that a growing digit of a quasiperiodic element is a fixed point of the corresponding map f_b. However, the converse is not true in general. The element $b-1$ is always a fixed point (of order 1), but is never a growing digit of a quasiperiodic element, as the following results show.

Lemma 6.11. *For any base* b*,* k *is a fixed point of order* 1 *if and only if* $k = b-1$*.*

Proof. ($\Rightarrow$) If k is a fixed point of order 1, then either $k = f_b(k) = 2k$ or $k = f_b(k) = 2k+1-b$. The first case gives that $k = 0$, which is a contradiction, and the second case gives that $k = b-1$.

($\Leftarrow$) A simple calculation shows that, since $b > 1$, $b-1 \geq b/2$, so $f_b(b-1) = 2(b-1)+1-b = b-1$. □

Lemma 6.12. *For any base* b*, if a quasiperiodic element has growing digit* k*, then* $k \neq b-1$*.*

Proof. For the base $b = 10$, this was proved in Lemma 2.10. The proof for general bases $b > 2$ follows by the same argument, with $b-1$ in place of the digit 9. For $b = 2$, Theorem 3.1 shows that the quasiperiod set $\mathcal{Q}_2$ is empty, so the statement holds trivially in this case. □

Theorem 6.13. *For any base* b*, we have* $1 \notin \mathcal{Q}_b$*. In other words, in any base, there are no quasiperiodic elements with quasiperiod* 1*.*

Proof. By Theorem 3.1, we have $\mathcal{Q}_2 = \emptyset$, so the assertion holds for $b = 2$. Suppose $b > 2$ and that there is a quasiperiodic element n with quasiperiod 1 and growing digit k. By Corollary 6.10, k must be fixed point of f_b of order 1. By Lemma 6.11, k must be $b-1$, a contradiction to Lemma 6.12. □

Theorem 6.13 serves as an analogue of Theorem 3.2. The result is slightly different in that base 2 does not play a special role. Continuing in this same vein, we investigate the existence (or nonexistence) of fixed points of f_b of a given order for general bases.

Lemma 6.14. *For any base* $b > 2$*, there is a fixed point of* f_b *of order* 2 *if and only if* $b \equiv 1 \bmod 3$*.*

Proof. (⇒). Suppose that k is a fixed point of f_b of order 2. We consider four cases.

Case 1. $0 < k, f_b(k) < b/2$. Then $k = f_b^2(k) = 4k$, which implies that $k = 0$, a contradiction, since $k > 0$. So this case cannot happen.

Case 2. $0 < k < b/2 \leq f_b(k) < b$. Then $k = f_b^2(k) = 2(2k) + 1 - b = 4k + 1 - b$. This equation is equivalent to $3k = b - 1$, which implies that $b \equiv 1 \bmod 3$.

Case 3. $0 < f_b(k) < b/2 \leq k < b$. Then $k = f_b^2(k) = 2(2k + 1 - b) = 4k + 2 - 2b$. This equation is equivalent to $3k = 2(b - 1)$, which implies that $b \equiv 1 \bmod 3$.

Case 4. $b/2 \leq k, f_b(k) < b$. Then $k = f_b^2(k) = 2(2k + 1 - b) + 1 - b = 4k + 3 - 3b$. This equation is equivalent to $k = b - 1$, a contradiction, since, by Lemma 6.12, $k < b - 1$. So this case cannot happen.

Having exhausted all possibilities, we see that we must have that $b \equiv 1 \bmod 3$.

(⇐) Suppose that $b \equiv 1 \bmod 3$. Let $k = (b - 1)/3$. Then, since $0 < k < b/2$, $f_b(k) = 2k = 2(b - 1)/3$. For $b \geq 4$ we have $2(b - 1)/3 > b/2$, so $f_b^2(k) = 4(b - 1)/3 + 1 - b = (b - 1)/3 = k$. Therefore, k is a fixed point of f_b of order 2. □

Corollary 6.15. *If $2 \in \mathcal{Q}_b$, for some base $b > 2$, then $b \equiv 1 \bmod 3$.*

Proof. The result follows from Lemma 6.14 and Proposition 6.8. □

Corollary 6.15 shows that the density, if it exists, of bases containing a quasiperiodic sequence with quasiperiod 2 is at most 1/3. On the other hand, Theorem 6.2 shows that quasiperiodic sequences with quasiperiod 2 exist in any base b satisfying $b \equiv 1 \bmod 9$. Hence this density is at least 1/9.

The following lemma shows that, for any given order t, there is a b such that f_b has a fixed point of order t. It should not be surprising that this is the case, as Theorems 6.2 and 6.4 show that quasiperiodic sequences can be constructed for any desired quasiperiod. What is interesting about the lemma is that the bases in the statement do not match the form of those in Theorems 6.2 and 6.4.

Lemma 6.16. *If $b = 2^t$, then 1 is a fixed point of order t.*

Proof. The lemma follows quickly from noticing that, for $b = 2^t$, we have $f_b^s(1) = 2^s$ for $1 \leq s < t$, and $f_b^t(1) = 2^t + 1 - b = 1$. □

Results like Corollary 6.15 and Lemma 6.16 can be used to narrow down bases that allow the existence of quasiperiodic RATS sequences of a specified quasiperiod. So if we are interested in finding quasiperiodic RATS sequences, by Proposition 6.8, it seems natural to search for bases b that have many fixed points of f_b.

Definition 6.17. *For a base b, let $S_b = \{k : 0 < k < b, f_b^t(k) = k \text{ for some } t\}$. In other words, S_b is the set of all fixed points of f_b.*

The following result gives an exact description of bases b for which f_b has only one fixed point. Notice that by Lemma 6.11, $b-1$ is always a fixed point of f_b, so we always have $|S_b| \geq 1$.

Theorem 6.18. *For any base $b > 2$, we have $|S_b| = 1$ if and only if $b = 2^t + 1$, where $t \in \mathbb{N}$.*

Proof. ($\Leftarrow$) Assume that $b = 2^t + 1$ for some t. If $t = 1$, then $b = 2^1 + 1 = 3$, and direct verification shows that in this case, $k = 2$ is the only fixed point of f_b. Hence, $|S_b| = 1$. Suppose now that $t \geq 2$. Let $\{1, 2, \ldots, b-1\} = \bigcup_{i=0}^{\infty} A_i$, where $A_i = \{k : 0 < k < b, 2^i \mid\mid k\}$. In other words, we partition the set $\{1, 2, \ldots, b-1\}$ according to the highest power of 2 dividing each digit. Notice that all but finitely many A_i's are empty; in particular, since $b - 1 = 2^t$, we have $\{1, 2, \ldots, b-1\} = \bigcup_{i=0}^{t} A_i$.

The set A_0 represents all odd numbers in $\{1, \ldots, 2^t\}$. Let $m \in A_0$. If $m < b/2$, then $f_b(m) = 2m$. If $m \geq b/2$, then $f_b(m) = 2m - b + 1 = 2m - 2^t$. Since $2 \mid f_b(m)$ and $4 \nmid f_b(m)$ in either case, we find that $f_b(A_0) \subseteq A_1$. Now let $m' \in A_1$. Then $m' = 2m''$ with $2 \nmid m''$. In particular, $m'' \in A_0$. Since $m'' = m'/2 < b/2$, $f_b(m'') = 2m'' = m'$. So we find that $A_1 \subseteq f_b(A_0)$. Therefore, we have $f_b(A_0) = A_1$.

Continuing in this fashion, we find that $f_b(A_i) = A_{i+1}$ for $0 \leq i < t$. Notice that $A_t = \{2^t\}$, so for any starting digit k, there is a t' such that $f_b^{t'}(k) = 2^t$. Since, by Lemma 6.11, $b - 1 = 2^t$ is a fixed point of order one, there cannot be any other fixed points. This is equivalent to the statement that $|S_b| = 1$.

($\Rightarrow$) For the converse direction, we first note that if m is such that $0 < m < b$, then, by Definition 6.7,

$$f_b^{-1}(m) = \begin{cases} \left\{\frac{m}{2}, \frac{m}{2} + \frac{b-1}{2}\right\} & \text{if } m \text{ is even and } b \text{ is odd,} \\ \emptyset & \text{if } m \text{ is odd and } b \text{ is odd,} \\ \left\{\frac{m}{2}\right\} & \text{if } m \text{ is even and } b \text{ is even,} \\ \left\{\frac{m}{2} + \frac{b-1}{2}\right\} & \text{if } m \text{ is odd and } b \text{ is even.} \end{cases} \tag{6.19}$$

We will use this fact extensively in this portion of the proof.

Assume now that $|S_b| = 1$. By Lemma 6.11, we know that $b - 1$ is a fixed point of f_b, so $S_b = \{b - 1\}$. By the pigeonhole principle, for every $0 < k < b - 1$ there is a $t > 0$ such that $f_b^t(k)$ is a fixed point of f_b. By assumption, $b - 1$ is the only fixed point, so $f_b^t(k) = b - 1$. In particular, $|f_b^{-1}(b-1)| > 1$, which, by

(6.19), implies $b-1$ is even and

$$f_b^{-1}(b-1) = \left\{\frac{b-1}{2}, b-1\right\} = \left\{\frac{c(b-1)}{2} : 1 \le c \le 2\right\},$$

where c runs through integers in the given range. Thus b must be of the form $b-1 = 2^{i_0}b_0$ with $i_0 > 0$ and $2 \nmid b_0$. If $i_0 > 1$, then using (6.19) again, we get

$$\begin{aligned} f_b^{-2}(b-1) &= \left\{\frac{b-1}{4}, \frac{b-1}{4} + \frac{b-1}{2}\right\} \cup f_b^{-1}(b-1) \\ &= \left\{\frac{b-1}{4}, \frac{3(b-1)}{4}\right\} \cup \left\{\frac{b-1}{2}, b-1\right\} \\ &= \left\{\frac{b-1}{4}, \frac{b-1}{2}, \frac{3(b-1)}{4}, b-1\right\} \\ &= \left\{\frac{c(b-1)}{4} : 1 \le c \le 4\right\}. \end{aligned}$$

If $i_0 > 2$, then

$$\begin{aligned} f_b^{-3}(b-1) &= \left\{\frac{b-1}{8}, \frac{b-1}{8} + \frac{b-1}{2}, \frac{b-1}{8} + \frac{b-1}{4}, \frac{b-1}{8} + \frac{b-1}{4} + \frac{b-1}{2}\right\} \cup f_b^{-2}(b-1) \\ &= \left\{\frac{b-1}{8}, \frac{5(b-1)}{8}, \frac{3(b-1)}{8}, \frac{7(b-1)}{8}\right\} \cup \left\{\frac{b-1}{4}, \frac{b-1}{2}, \frac{3(b-1)}{4}, b-1\right\} \\ &= \left\{\frac{c(b-1)}{8} : 1 \le c \le 8\right\}. \end{aligned}$$

Continuing in this way, we see that, for $1 \le i \le i_0$,

$$f_b^{-i}(b-1) = \left\{\frac{c(b-1)}{2^i} : 1 \le c \le 2^i\right\}.$$

In particular, for $1 \le i \le i_0$,

$$|f_b^{-i}(b-1)| = 2^i. \tag{6.20}$$

Notice that for $i < i_0$, the elements in $f_b^{-i}(b-1)$ are all even, but in $f_b^{-i_0}(b-1)$, half of the elements are odd. This implies, by (6.19), that

$$f_b^{-(i_0+1)}(b-1) = f_b^{-i_0}(b-1)$$

and hence $f_b^{-(i_0+t)}(b-1) = f_b^{-i_0}(b-1)$ for all $t > 0$. Hence $\{0 < k < b : k \text{ odd}\} \subseteq f_b^{-i_0}(b-1)$. Since, by (6.20), $f_b^{-i_0}(b-1)$ contains 2^{i_0} elements, and, by the above observation, half of these elements are odd, this implies that $\#\{0 < k < b : k \text{ odd}\} \leq 2^{i_0-1}$ and hence $(b-1)/2 = 2^{i_0-1}b_0 \leq 2^{i_0-1}$. Thus $b_0 = 1$ and $b = 2^{i_0}b_0 + 1 = 2^{i_0} + 1$, so b is of the desired form.

In the case $i_0 = 1$, a similar argument shows that $b = 2^1 + 1 = 3$, which is also of the desired form. □

Corollary 6.19. *For any base $b = 2^t + 1$, where $t \geq 0$, we have $\mathcal{Q}_b = \emptyset$. In other words, there are no quasiperiodic RATS sequences in base $b = 2^t + 1$.*

Proof. Suppose that there is a quasiperiodic element n in base $b = 2^t + 1$. By Proposition 6.8, the growing digit of n must be fixed point of f_b. Theorem 6.18 shows that if $b = 2^t + 1$, then the only fixed point of f_b is $b - 1$. It follows that $b - 1$ must be the growing digit of n, a contradiction to Lemma 6.12. Therefore, there are no quasiperiodic RATS sequences in base b. □

Corollary 6.19 shows that there is an infinite family of bases for which no quasiperiodic RATS sequences exist. Unfortunately, this is not enough to conclude that all sequences in such bases are preperiodic. For example, McMullen's conjecture (Chapter 5) is still open in the case $b = 5 = 2^2 + 1$. However, this corollary gives an infinite family for which the quasiperiod set must be empty. There is no analogous result for period sets that excludes certain periods from occurring in some infinite family.

6.4 Lyndon words connection

We now introduce the concept of Lyndon words, a combinatorial object on words. In order to properly introduce the subject, we will need the following basic definitions.

Definition 6.20. *Let Σ be a set, called an **alphabet**. A **word** in Σ of **length** t is a string of length t that only uses characters from the alphabet Σ.*

For our purposes, we are only interested in working with the alphabet $\Sigma = \{A, B\}$. From this point on, unless stated otherwise, this will be our assumed alphabet.

Example Let $w = ABAAB$. Then w is a word of length 5.

Definition 6.21. *By a **cyclic rotation** of a word w, we mean a word generated by removing a (possibly empty) block of characters at the right side of w and tacking it on to the left side of w.*

Example If $w = ABAAB$, then the cyclic rotations of w are $ABAAB$, $BABAA$, $ABABA$, $AABAB$, and $BAABA$.

Definition 6.22. *We say that a word w of length t is a* ***Lyndon word*** *of length t if w differs from all of its nontrivial cyclic rotations*[1]*.*

Example For example, $w = AAAAB$ is a Lyndon word of length 5 since its nontrivial cyclic rotations, $BAAAA$, $ABAAA$, $AABAA$ and $AAABA$ all differ from w. On the other hand, $w = AABAAB$ is not a Lyndon word of length 6 since its nontrivial cyclic rotations, $BAABAA$, $ABAABA$, $AABAAB$, $BAABAA$, and $ABAABA$ are not all distinct from w.

Lyndon words play a role akin to primes in terms of cyclic rotations. By the Chen–Fox–Lyndon theorem (see [12], p. 67), if a given word is not a Lyndon word, then some cyclic rotation of it can be expressed as a concatenation of multiple copies of a shorter word. This process of decomposition of a word into smaller and smaller words is analogous to decomposing an integer into its prime factors. Lyndon words play the role of the "indecomposable factors" for cyclic rotations.

Example Let $w = AABAAB$. As noted above, w is not a Lyndon word, but it is the concatenation of two copies of the Lyndon word AAB.

To make the connection to primes even stronger, notice that all words, with at least two distinct characters, of length p, where p is a prime, are Lyndon words.

We will show a connection between Lyndon words and fixed points of f_b. Given a base b and a digit $0 < k < b$, let

$$A_b(k) = 2k \text{ and } B_b(k) = 2k + 1 - b.$$

Then, Definition 6.7 can be written as

$$f_b(k) = \begin{cases} A_b(k) & \text{if } 0 < k < b/2, \\ B_b(k) & \text{if } b/2 \le k < b. \end{cases} \tag{6.21}$$

Definition 6.23. *Given a base b, we associate with every word $w = w_1 w_2 \cdots w_t$, where $w_i \in \{A, B\}$, the map $\phi_b(w) = w_1 \circ w_2 \circ \cdots \circ w_t$, where $\circ$ denotes function composition and the w_i's are to be interpreted as the maps $A(k) := A_b(k)$ and $B(k) := B_b(k)$ defined above.*

Example If $w = AABA$, then $\phi_b(w) = A \circ A \circ B \circ A$. Thus $\phi_b(w)(k) = A \circ A \circ B \circ A(k) = 2(2(2(2k) - (b-1))) = 2^4 k - 2^2(b-1)$.

[1]Our use of the term Lyndon word is slightly different from its customary use, which also requires the word w be lexicographically first among the set of all cyclic rotations of w. This interpretation is more convenient for our purposes.

Lemma 6.24. *Let w be a word of length t. If $w = w_1 w_2 \dots w_t$, then*

$$\phi_b(w)(k) = 2^t k - (b-1)a, \tag{6.22}$$

where

$$a = \sum_{i=1}^{t} a_i 2^{i-1}$$

and

$$a_i = \begin{cases} 0 & \textit{if } w_i = A, \\ 1 & \textit{if } w_i = B \end{cases}$$

for $1 \le i \le t$.

Proof. The lemma is trivially true in the case $t = 1$. So now suppose that the lemma holds for all words of length a fixed length $(t-1) > 0$. Let w be a word of length t; that is, $w = w_1 w_2 \dots w_t$. Then $\phi_b(w) = w_1 \circ \phi_b(w')$, where $w' = w_2 \dots w_t$. Since w' is a word of length $t-1$, the assertion of the lemma holds. In particular,

$$\phi_b(w')(k) = 2^{t-1} k - (b-1)a',$$

where

$$a' = \sum_{i=2}^{t} a_i 2^{i-2}.$$

Case $w_1 = A$. Then $\phi_b(w) = w_1 \circ \phi_b(w') = 2\phi_b(w') = 2^t k - (b-1)2a'$. From the formula for a' above we find that $2a' = \sum_{i=2}^{t} a_i 2^{i-1} = \sum_{i=1}^{t} a_i 2^{i-1}$ with $a_1 = 0$. Hence the assertion of the lemma holds.

Case $w_1 = B$. Then $\phi_b(w) = w_1 \circ \phi_b(w') = 2\phi_b(w') - (b-1) = 2^t k - (b-1)(2a'+1)$. From the formula for a' above we find that $2a' + 1 = 1 + \sum_{i=2}^{t} a_i 2^{i-1} = \sum_{i=1}^{t} a_i 2^{i-1}$ with $a_1 = 1$. Hence the assertion of the lemma holds.

Having exhausted all cases, the lemma follows. □

It should be noted that the constant a appearing in Lemma 6.24 is distinct for distinct words of the same length. This follows from the fact that a is an integer whose binary representation is $a = a_t \dots a_2 a_1$, with $a_i \in \{0, 1\}$, and hence is in one-to-one correspondence with words $w = w_1 w_2 \dots w_t$ of length t over the alphabet $\{A, B\}$.

Lemma 6.25. *If w and w' are words of length t such that $\phi_b(w)(k) = \phi_b(w')(k)$ for some k, then $w = w'$.*

Proof. By Lemma 6.24, there are integers a and a' such that $\phi_b(w)(k) = 2^t k - (b-1)a$ and $\phi_b(w')(k) = 2^t k - (b-1)a'$. By assumption, $2^t k - (b-1)a = \phi_b(w)(k) = \phi_b(w')(k) = 2^t k - (b-1)a'$, which is equivalent to $a = a'$. By the comments following Lemma 6.24, this implies that $w = w'$. □

Lemma 6.26. *Suppose that $w = w_1 w_2 \dots w_t$ is a word and $0 < k < b$. If $\phi_b(w)(k) = k$, then $w_i \circ \cdots \circ w_t(k) = f_b^{t-i+1}(k)$ for $1 \le i \le t$. In particular, k is a fixed point of f_b.*

Proof. Suppose that the conclusion of the lemma is false. That is, there is some largest $1 \le i_0 \le t$ such that $w_{i_0} \circ \cdots \circ w_t(k) \ne f_b^{t-i_0+1}(k)$. Then $0 < w_{i_0+1} \circ \cdots \circ w_t(k) = f_b^{t-i_0}(k) < b$.

Case $f_b^{t-i_0}(k) < b/2$. By Definition 6.7, we must have that $f_b^{t-i_0+1}(k) = 2f_b^{t-i_0}(k)$. Since $w_{i_0} \circ \cdots \circ w_t(k) \ne f_b^{t-i_0+1}(k)$, we must have that $w_{i_0} \circ f_b^{t-i_0}(k) = 2f_b^{t-i_0}(k) - (b-1) < 0$. This cannot occur, as it will give that, by Lemma 6.24, $k = \phi_b(w)(k) = w_1 \circ \cdots \circ w_{i_0-1}(2f_b^{t-i_0}(k) - (b-1)) = 2^{i_0-1}(2f_b^{t-i_0}(k) - (b-1)) - (b-1)a < 0$, since $a \ge 0$, a contradiction to the assumption that $0 < k < b$. Therefore, this case is impossible.

Case $b/2 \le f_b^{t-i_0}(k) < b$. By Definition 6.7, we must have that $f_b^{t-i_0+1}(k) = 2f_b^{t-i_0}(k) - (b-1)$. Since $w_{i_0} \circ \cdots \circ w_t(k) \ne f_b^{t-i_0+1}(k)$, we must have that $w_{i_0} \circ f_b^{t-i_0}(k) = 2f_b^{t-i_0}(k) > b-1$. This cannot occur, as it will give that, by Lemma 6.24, $k = \phi_b(w)(k) = w_1 \circ \cdots \circ w_{i_0-1}(2f_b^{t-i_0}(k)) = 2^{i_0-1}(2f_b^{t-i_0}(k)) - (b-1)a > b-1$, since $a \le 2^{i_0-1} - 1$, a contradiction to the assumption that $0 < k < b$. Therefore, this case is impossible.

Having exhausted all cases, we see that we contradict the assumption that the lemma was false. Therefore, the assertion holds. □

Definition 6.27. *Suppose we are given a base b. We associate with every fixed point k of f_b the word $w = \psi_b(k)$ such that $\phi_b(w)(k) = k$ and w is minimal in length.*

It is not immediately obvious that the map defined in Definition 6.27 is well–defined, i.e., that such a word w exists and is unique. The existence of a word w satisfying $\phi_b(w)(k) = k$ follows from the definition of f_b in (6.21) and the assumption that k is a fixed point of f_b. Lemma 6.25 shows that for each given length, there is at most one word w with $\phi_b(w)(k) = k$. So among all words w satisfying $\phi_b(w)(k) = k$, there is a unique one of shortest length. Therefore, the map in Definition 6.27 is well–defined.

Example For $b = 10$, $k = 3$ is a fixed point of order 2, namely $3 = f_{10}^2(3) = B \circ A(3)$. Hence $\psi_{10}(3) = BA$.

Lemma 6.28. *Given a base b, if k and k' are distinct fixed points of f_b of order t, then their associated words, $w = \psi_b(k)$ and $w' = \psi_b(k')$, are also distinct.*

Proof. By Lemma 6.24, we see that if $w = w'$, then k and k' are distinct solutions to a polynomial equation of degree 1, a contradiction. Therefore, we have $w \ne w'$. □

Lemma 6.29. *Given a base b, if k is a fixed point of f_b of order $t > 0$, then the words $w = \psi_b(k)$ and $w' = \psi_b(f_b(k))$ differ by a cyclic rotation.*

Proof. Let $k' = f_b(k)$. Suppose first that $k' = k$. Then $w = w'$, so the assertion holds in this case.

Now suppose that $k' \neq k$. Since k is a fixed point of order t, so is $k' = f_b(k)$. Hence, by Lemma 6.28, w and w' are distinct. We have that

$$k' = f_b(k) = f_b(f_b^t(k)) = f_b(f_b^{t-1}(k')). \tag{6.23}$$

The first and last equalities in (6.23) give that the rightmost character of w is the same as the leftmost character of w'. We also have

$$f_b^2(k) = f_b^2(f_b^{t-1}(k')) = f_b(k'). \tag{6.24}$$

Therefore, by (6.24), the second rightmost character of w is the same as the rightmost character of w'. Continuing on in this way, we see that w' is obtained by removing the rightmost character of w and tacking it on to the left side of w. By Definition 6.21, w' is a cyclic rotation of w. □

Corollary 6.30. *Given a base b, if k is a fixed point of f_b of order t, then the set of words associated to the fixed points $f_b^i(k)$, for $0 \leq i \leq t$, is the set of cyclic rotations of $\psi_b(k)$.*

Proof. The result follows immediately from Lemma 6.29. □

Corollary 6.31. *Given a base b, if k and k' are fixed points of f_b of order t such that the associated words $w = \psi_b(k)$ and $w' = \psi_b(k')$ are the same up to a cyclic rotation, then there is some t' such that $f_b^{t'}(k) = k'$.*

Proof. By Corollary 6.30, there are nonnegative integers $0 \leq i, j < t$ such that $f_b^i(k)$ and $f_b^j(k')$ are associated with the same word. By Lemma 6.28, this implies that $f_b^i(k) = f_b^j(k')$. Furthermore, we find that $f_b^{i+s}(k) = f_b^{j+s}(k')$ for all $s \geq 0$. Choosing $s = t - j$, we find that $k' = f_b^t(k') = f_b^{t-j+j}(k') = f_b^{t-j+i}(k)$. Therefore, the corollary follows with $t' = t - j + i$. □

Corollary 6.30 yields a natural partition of the fixed points of f_b into different classes according to their associated words. That is, two fixed points are in the same class if their associated words are the same up to a cyclic rotation. The next proposition shows that the only words that occur as words associated with a fixed point are Lyndon words.

Proposition 6.32. *Given a base b, if k is a fixed point of f_b of order $t > 1$, then its associated word $w = \psi_b(k)$ is a Lyndon word.*

Proof. Suppose that w is not a Lyndon word. Then, by the Chen–Fox–Lyndon theorem ([12], p. 67), some cyclic rotation of w is a word that is formed by concatenating a shorter word w', of length $t' < t$ with $t'|t$, multiple times. Without loss of generality, we assume that w can be decomposed this way. This gives that

$$\phi_b(w) = \phi_b(w') \circ \cdots \circ \phi_b(w'). \tag{6.25}$$

By Corollary 6.30 and (6.25), it follows that $\psi_b(f_b^{t'}(k)) = w$ by performing t' cyclic rotations of w. By Lemma 6.28, we must have that $f_b^{t'}(k) = k$, which contradicts the assumption on the order of k. Therefore, w is a Lyndon word. □

Proposition 6.32 shows that words associated to fixed points of f_b, for some base b, are always Lyndon words. The next theorem gives necessary and sufficient conditions on b in order to guarantee that there is a bijection between Lyndon words of length t and fixed points of f_b of order t.

Theorem 6.33. *Given a base b, every Lyndon word of length t is associated to some fixed point of f_b of order t if and only if $b \equiv 1 \bmod (2^t - 1)$. In particular, given a Lyndon word of length t, say w, there exists a base b and a fixed point k of f_b of order t such that its associated word is w.*

Proof. ($\Rightarrow$) By assumption, there is a digit k in base b such that $\psi_b(k) = BA\ldots A$. (Note that $BA\ldots A$ is always a Lyndon word for any length.) In particular, we have, by Lemma 6.24, $k = \phi_b(BA\ldots A)(k) = 2^t k - (b-1)$. This is equivalent to the statement $b \equiv 1 \bmod (2^t - 1)$.

($\Leftarrow$) Let w be a Lyndon word of length t. By Lemma 6.24, we have $\phi_b(w)(x) = 2^t x - (b-1)a$ for some integer $0 \le a \le 2^t - 1$. Since $b \equiv 1 \bmod (2^t - 1)$, we see that

$$k_w = \frac{(b-1)a}{2^t - 1}$$

is a positive integer such that $k_w = \phi_b(w)(k_w)$. In particular, $0 < k_w < b$, so by Lemma 6.26, k_w is a fixed point of f_b.

Let $w' = \psi_b(k_w)$ be the word associated with k_w according to Definition 6.27; that is, w' is the word of minimal length satisfying $k_w = \phi_b(w')(k_w)$. Let t' be the length of w'. By the minimality of w', we have $t' \le t$. We seek to show that $w = w'$. If $t' = t$, then, by Lemma 6.25, $w = w'$. So suppose that $t' < t$. Then there are nonnegative integers s and r such that $t = st' + r$ with $0 \le r < t'$. Let $w = w_1 w_2 \ldots w_{st'+r}$ and $w'' = w'w'\ldots w'$, where w'' is of length st'. By Lemma 6.26, $w_{r+1} \circ \cdots \circ w_{st'+r}(k_w) = f_b^{st'}(k_w) = \phi_b(w'')(k_w) = k_w$. This implies that $k_w = w_1 \circ \cdots \circ w_r(k_w)$. Therefore, by Definition 6.27, $r = 0$ and $t'|t$. So w and w'' are words of the same length satisfying $k_w = \phi_b(w)(k_w) = \phi_b(w'')(k_w)$. By Lemma 6.25, we have $w = w''$. This implies that w is generated by concatenating multiple copies of w', a contradiction to the assumption that w is a Lyndon word. Therefore, for any given Lyndon word w, there is a fixed point of f_b whose associated word is w. □

This section shows that, if we are interested in searching for quasiperiodic elements for a base b, then a good place to start is by searching for fixed points of f_b. The connection to Lyndon words shows that it is not always worth looking for a specific quasiperiod in every base. Theorem 6.33 suggests bases that are more likely to yield fixed points of a particular order.

Looking back to Theorems 6.2 and 6.4, we see that the growing digit of any quasiperiodic RATS sequence currently known always has the same type of Lyndon word associated with it, namely, some cyclic rotation of $BA \dots A$ (in fact, every base in Theorem 6.4 is also of the form needed in Theorem 6.33). It is natural to ask if there are examples of bases for which quasiperiodic RATS sequences exist with growing digits associated to a different Lyndon word. Results like Corollary 6.19 and Theorem 6.33 do not give definitive answers, but they greatly aid in redirecting the search.

Chapter 7

Lehmer's Reverse-Add process

7.1 Notation and terminology

We wish to study Lehmer's question on the existence of palindromes in sequences generated by the reverse-add process. As in Chapter 2, we begin with some definitions to formalize the question.

Definition 7.1. *For any positive integer n written in base 10, let $\overline{n}$ be the digit formed by reversing the digits of n. A **palindrome** is any integer n such that $n = \overline{n}$. We also define the function $T\colon \mathbb{N} \to \mathbb{N}$ by $T(n) = n + \overline{n}$.*

The above definition makes the similarity between Lehmer's reverse-add process and the RATS game even more obvious. As was the case in Chapter 2, we will use superscript notation on T to denote multiple iterations of the process.

Definition 7.2. *We call $\{T^i(n)\}_{i=0}^{\infty}$ the **reverse-add sequence generated by** n.*

Example

$$\begin{aligned} T^2(845363) &= T(845363 + 363548) \\ &= T(1208911) \\ &= 1208911 + 1198021 \\ &= 2406932 \end{aligned}$$

Definition 7.3. *Define the function $\tau\colon \mathbb{N} \to \mathbb{N}$ by $\tau(n) = t$ if and only if t is the smallest nonnegative integer such that $T^t(n)$ is a palindrome. If no such t exists, define $\tau(n) = \infty$. We call $\tau(n)$ the **stopping time** of n.*

Example From (1.3), we see that $\tau(9) = 2$ since $T^2(9) = 99$ is a palindrome and $T(9) = 18$ is not.

We can now reformulate Lehmer's question as: Given any integer n, is $\tau(n) < \infty$?

So far, no theoretical methods have been able to answer Lehmer's question. However, the overwhelming computational evidence collected by Doucette [3] and VanLandingham [15] suggest that the answer, in general, is no. In fact, Doucette's work suggests that for most integers n, $\tau(n)$ is infinite.

Conjecture (Problem 196)**.** *The integer* 196 *has infinite stopping time, i.e.,* $\tau(196) = \infty$.

Even though Lehmer never found a stopping time for 196, he did note that

$$T^{56}(196) = 93\ \ 4217310162393261013712 4\ \ 28. \tag{7.1}$$

This 27 digit number is *almost* a palindrome in that all but the first pair and last pair of digits match up. So despite the fact that the reverse-add sequence generated by 196 has not generated any palindromes, the extent by which each iterate "misses" can, in some sense, be measured.

Definition 7.4. *Let n_i represent the ith digit of n. If n has a k digit long base* 10 *representation, then the* ***palindromicity quotient*** *of n in base b is given by*

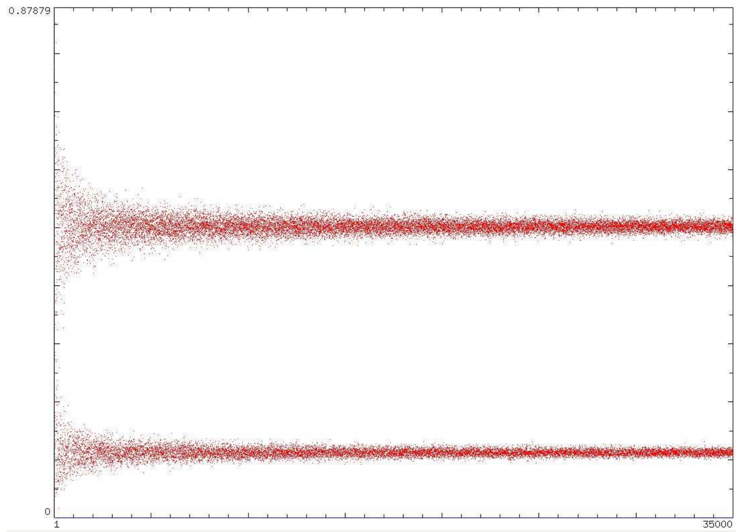
$$\text{palq}(n) = \frac{\#\{(n_i, n_{k-i}) : n_i = n_{k-i}, 0 \leq i \leq k)\}}{k}.$$

Example If $n = \underline{11}3\underline{8}7\underline{11}$, then $\text{palq}(n) = 5/7$. For (7.1), the palindromicity quotient of $T^{56}(196)$ is $23/27 \approx 0.85$.

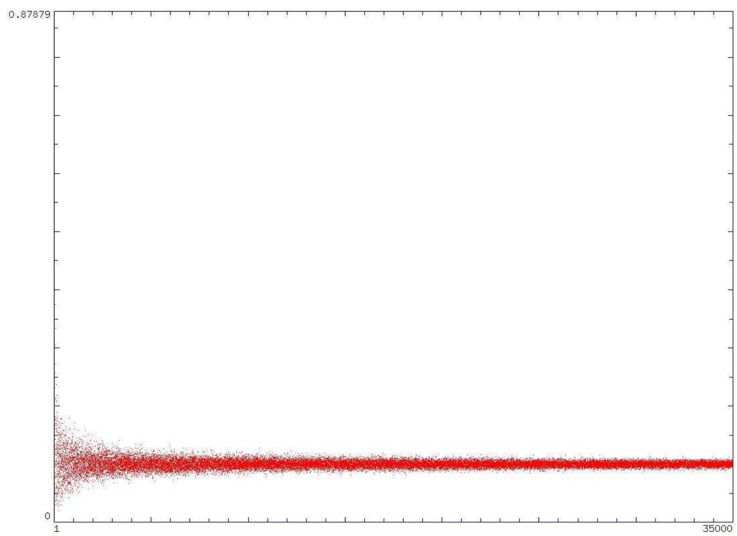
The palindromicity quotient measures how close an integer is to being a palindrome by giving the proportion of digits that match accordingly. Figure 7.1 shows that iterates of the reverse-add process display more palindrome-like behavior than randomly generated integers of equal length. From Figure 7.1a, it appears that the generated sequence of palindromicity quotients has two accumulation points at roughly 1/2 and 1/10. Upon close inspection of the data, the palindromicity quotient is close to 1/10 whenever the particular iterate is one digit longer thant the previous, and close to 1/2 otherwise. This seems to suggest two things:

- Digits that appear side by side in numbers generated by the reverse-add process are "independent" from each other in the sense that they appear to be distributed in the same way that they would be in a randomly generated number (see Figure 7.1b).
- The reverse-add process produces more palindrome-like numbers than a random number generator when the iterate is of the same length as the previous term.

Another thing to notice about Figure 7.1a is that after many iterations, the palindromicity quotient seems to be bounded away from 1. The trend of the graph suggests that the palindromicity quotient will never be



(a) Reverse-add process iterates



(b) Randomly generated integers

Figure 7.1: Palindromicity quotients for 35000 trials with 10000 digit long integers.

higher than 0.88 after 35000 reverse-add iterations. In particular, this would imply that there are at most a finite number of times for which the reverse-add sequence generated a palindrome. This behavior is not uncommon in random trials like the one in Figure 7.1, and thus leads to the following conjecture.

Conjecture. *Given an integer n, $T^i(n)$ is a palindrome for only a finite number of i.*

7.2 Extension to general bases

As was the case for the RATS process, we can extend Definitions 7.1, 7.3, and 7.4 to other bases in the natural way. In doing so, we can ask the same questions as we did before, and in some cases answers them.

Proposition 7.5. *(Duncan, [4], Trigg [14], Seals, [15]) There are numbers in base* 2, 4, 8, 11, 16, 17, 20, 26 *and* 32 *that never lead to palindromes under the reverse-add process.*

Proposition 7.5 immediately answers Lehmer's original question in a few cases, but the problem is still open for general bases. Numerical evidence suggests that the last conjecture in Section 7.1 should be extended to general bases as well.

There are very few concrete results in the literature related to Lehmer's reverse-add process. The lack of structure in the iterates, unlike the RATS process, makes it difficult to predict patterns or establish results. For example, if $T_b^i(n)$ is the ith reverse-add iterate of n in base b, then $\log_b\left(T_b^i(n)\right)$ should roughly equal the total number of digits in the base b representation of $T_b^i(n)$. Furthermore, the limit as $i \to \infty$ of

$$\frac{\log_b\left(T_b^i(n)\right)}{i}, \tag{7.2}$$

if it exists, gives, roughly, the average chance that the next iterate in the reverse-add sequence will be a digit longer (see Figure 7.2). Initial computations like those in Figure 7.2 suggested that, as $b \to \infty$, (7.2) tends to $1/3$. However, this initial guess does not appear to hold true if the base is a power of 2. In other words, the growth of the length of terms in reverse-add generated sequences appears to be vary depending on the base.

b	$\frac{\log_b\left(T_b^i(n)\right)}{i}$
2	0.5013
3	0.5996
8	0.3947
10	0.4153
25	0.3479
100	0.3550
500	0.3490
729	0.3350
999	0.3430
1000	0.3630
1001	0.3420
1024	0.3970

Figure 7.2: Average digit growth rate of the reverse-add sequence generated by $n = 1$ after $i = 50000$ iterates in various bases.

Many questions regarding reverse-add sequences remain open, but very few methods of attack exist to address them. As a first approximation, perhaps a probabilistic model could shed some light on this problem. As motivation, we look to the Lagarias–Weiss [10] probabilistic model used to study the Collatz conjecture. Such techniques would not necessarily solve the problem, but they could aid us in understanding some of the complexities that arise from such an easily stated problem.

References

[1] John H. Conway. Play it again..and again... *Quantum*, 63:30–31, November/December 1990.

[2] Curtis Cooper and Robert E. Kennedy. Base 10 RATS cycles and arbitrarily long base 10 RATS cycles. In *Applications of Fibonacci numbers, Vol. 8 (Rochester, NY, 1998)*, pages 83–93. Kluwer Acad. Publ., Dordrecht, 1999.

[3] Jason Doucette. `http://www.jasondoucette.com/worldrecords.html`, 1999.

[4] D. C. Duncan. Sujets d'étude. *Sphinx (Bruxelles)*, 9:91–92, 1939.

[5] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[6] Karen Gentges. On Conway's RATS. Master's thesis, Central Missouri State University, 1998.

[7] Richard K. Guy. Conway's RATS and other reversals. *The American Mathematical Monthly*, 96:425–428, 1989.

[8] Richard K. Guy. *Unsolved problems in number theory.* Problem Books in Mathematics. Springer-Verlag, New York, third edition, 2004.

[9] D. R. Kaprekar. Another solitaire game. *Scripta Mathematica*, 15:244–245, 1949.

[10] J. C. Lagarias and A. Weiss. The $3x+1$ problem: two stochastic models. *Ann. Appl. Probab.*, 2(1):229–261, 1992.

[11] D. H. Lehmer. Sujets d'étude. *Sphinx (Bruxelles)*, 8:12–13, 1938.

[12] M. Lothaire. *Combinatorics on words.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997. With a foreword by Roger Lyndon and a preface by Dominique Perrin, Corrected reprint of the 1983 original, with a new preface by Perrin.

[13] Steven Shattuck and Curtis Cooper. Divergent RATS sequences. *Fibonacci Quart.*, 39(2):101–106, 2001.

[14] Charles W. Trigg. Versum sequences in the binary system. *Pacific J. Math.*, 47:263–275; corrections, ibid. 49 (1973), 619, 1973.

[15] Wade VanLandingham. `http://www.p196.org`, 2006.

[16] John Walker. `http://www.fourmilab.ch/documents/threeyears/threeyears.html`, 1990.

[17] Koji Yamashita. A problem on palindromic numbers. *Math. Japon.*, 45(1):159–163, 1997.